Excuse-Driven Present Bias*

Luca Flóra Drucker and Marc Kaufmann
(Central European University)

June 24, 2022
(*Preliminary draft*)
Click here for latest version

Abstract

We test whether people behave in a more present-biased way when they can excuse such behavior. We run two experiments, one on the Amazon Mechanical Turk and one with students in Luxembourg, to elicit subjects' willingness to work (WTW) today and at a future date. We elicit this WTW against an alternative that provides no excuses and one that provides an excuse: while the no-excuse alternative always requires to work, the excuse alternative adds a 10% chance of not having to do extra work. In the first experiment, we find that the WTW today drops by $0.11 more than the WTW in two days when we move from the no-excuse to the excuse alternative, as if the excuse alternative was worth more when it allows avoiding working today. This result cannot be explained by risk and time preferences that do not depend on other alternatives present. In the second experiment, we test another potential excuse besides risk: a different type of task. The results do not support that a different task would act as an excuse for postponing work. For the chance of not having to do work in the future, we get non-significant results that nevertheless point in the same direction as the MTurk results. We discuss both experiments and describe a planned follow-up study with the goal of replicating our finding with excuses 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 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*

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

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.[1]

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 an excuse, we should find that the former work less, because they have a reason to work less. The same reason should lead to the same change in willingness to work. In two experiments we look at whether instead willingness to work is different when there is a behavior – avoiding present work – for which people can use the reason as an excuse. We provide two types of alternatives for work: *no-excuse* and *excuse* alternatives. In the first experiment, the no-excuse alternative requires subjects to do extra future work for sure, while the 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 [2016], who finds that people use risk in charitable giving as an excuse to donate less to charity. We find that the willingness to work today drops by \$0.11 ($N = 147$, p-value 0.011) more than the willingness to work at a future date, even though the reason to work less is the same in the two cases. We interpret this phenomenon as excuse-driven present bias.

In the second experiment, besides the chance of no future work, we test another reason: a different type of task. Here, the no-excuse alternative is doing the same task in the future people are already working on. The excuse alternative, however, involves a different task in the future. The idea is that people can explain to themselves postponing work to the future by just choosing to do the other task, therefore, hiding their *implicit preference*, present bias,

[1]Another reason why we refer to present bias is that we focus on short-term time preferences with future decisions at most several days in the future. As others have noted, it is calibrationally implausible to have measureable discount factors at such short time frames that are due to exponential discounting. For example, a weekly discount of 0.99 (which is virtually indistinguishable from 1.00) leads to a yearly discount factor of $0.99^{52} \approx 0.59$. Thus our claim is that whatever short-term impatience we find is likely to be present bias or some other present-focused behavior, rather than impatience of the exponential discounting variety.

behind a choice between tasks (similarly as in the framework described in Cunningham and de Quidt [2016]). This reason does not seem to work as an excuse: the willingness to work today is not significantly different from the willingness to work in the future when moving from the no-excuse to the excuse alternative. In this experiment, the chance of no future work as an excuse shows similar – however, non-significant – results to the first experiment.

We describe our within-subjects design to identify excuse-making based on change in willingness to work in Section 2. We then present the details of the first experiment and the implementation on Amazon Mechanical Turk (MTurk) in Section 3. In Section 4 we show the results of the first experiment in more detail. Section 5 discusses the drawbacks of the implementation on MTurk and describes the second experiment which we ran online with students in the LISER-LAB in Luxembourg. The aim was to replicate the first results and to strengthen our identification on two fronts. The first issue was to complement risk as an excuse with other, cleaner excuses, to rule out all state-dependent preferences with risk and time interactions. 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 would 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).

Section 6 presents the results from the second experiment and discusses why we might have failed to replicate the MTurk results. In the first experiment, to minimize MTurk workers substituting our tasks with other MTurk tasks, we chose to have the same amount of work in all choices. Some tasks were, however, *hard* while others *easy*, and participants decided about the number of hard tasks in their choices. When moving to the lab, we chose to have a certain amount of required tasks in each session, and the choices were about *extra* tasks on top of the required amount instead of hard vs. easy tasks. We suspect that part of our failure to replicate the results was that students might still have substituted our tasks with other activities. The other part was probably lack of power to detect an effect size of similar magnitude to that in the MTurk study. Our lab study lasted four weeks and we had significant attrition, potentially leading to low power. In this section we also describe how we want to test if choosing extra tasks or hard vs. easy tasks really makes a difference, and to corroborate our findings from the first experiment. Finally, in Section 7 we discuss our results, other possible mechanisms to explain them and how we rule these out, and the potential of excuse-making in present bias.

If 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 [2021]). At the same time, it leaves room for learning in situations with clear feedback, such as in Le Yaouanq and Schwardmann [2019].[2] 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.

2 Identification of excuse-driven present bias

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

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

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 *no-excuse alternative* and an *excuse alternative* that allow for small and large choice distortions respectively.

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

No excuse:

$$\begin{aligned} \text{work in the future} + \$X &\quad\sim\quad \text{no-excuse alternative} \\ \text{work in the present} + \$X' &\quad\sim\quad \text{no-excuse alternative} \end{aligned}$$

Excuse:

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

Then $\Delta_C = X' - X$ is the difference between willingness to work in the future and in the present when there are no possibilities for making excuses (control), while $\Delta_T = Y' - Y$ is the same difference when there are possibilities for making excuses (treatment). We present the identification using the main no-excuse and excuse alternatives we provide in our experiments: 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 [2016], who finds that people distort risky choices in charitable giving in a way that is beneficial to them. However, the identification is similar for any type of excuse we can think of.

The main concern of the design is to take care of the difference in utility between the no-excuse and the 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 no-excuse condition, and by β_E the present bias parameter in the 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 (no-excuse or 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 excuse setting.

No excuse (control):

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

Excuse (treatment):

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

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:

$$\begin{aligned} \Delta_C &= d(x_1)(1-\beta) \\ \Delta_T &= d(x_1)(1-\beta) + 0.9d(x_2)(\beta - \beta_E) \end{aligned}$$

We finally obtain the difference of these differences,

$$\Delta\Delta = \Delta_T - \Delta_C = 0.9d(x_2)(\beta - \beta_E)$$

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.[3] If, however, the 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.

3 First experiment

3.1 Technical details

We implemented our first 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 description to familiarize with 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] 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.

3.2 Payments

On average participants earned ≈ $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 within three days after session 2, 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**
		(right after Session 0)	(2 days after session 1)	(within 2 days after Session 2)
Tasks	warm-up rounds of effort task	effort choices	effort task	
	consent	effort task		
	debrief survey			

3.3.1 The Tutorial

Session 0 is the tutorial[4] 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.[5]

There are two sizes of matrices in the experiment, a small one (7x12 cells), and a large one (10x15 cells), and subjects completed three 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][6]. 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.

0	1	1	1	1	0	1	0	0	1	1	0	0	1	0
1	0	1	0	0	0	1	1	1	0	0	0	1	1	1
1	0	1	1	1	1	0	0	1	1	0	1	1	0	1
0	0	0	0	1	0	1	1	1	1	1	0	1	0	0
0	0	1	0	0	0	0	0	1	1	0	1	0	0	1
1	0	1	0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	0	0	1	0
1	1	0	1	0	0	1	0	0	0	1	1	1	0	1
0	1	0	0	0	0	1	0	1	0	0	1	1	1	0
0	0	0	0	0	1	0	0	0	1	0	0	1	0	0

Figure 1: Example of a large matrix

[4] 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

[5] 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 one subject in our pilot who we think followed this strategy and we wanted to avoid it.

[6] 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"

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. [7] 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 subjects choose between 15 large matrices in two days with 90% probability for a $0.60 bonus or 22 large matrices in two days for sure for a bonus ranging from $0.00 to $3.40, increasing in $0.20 increments.

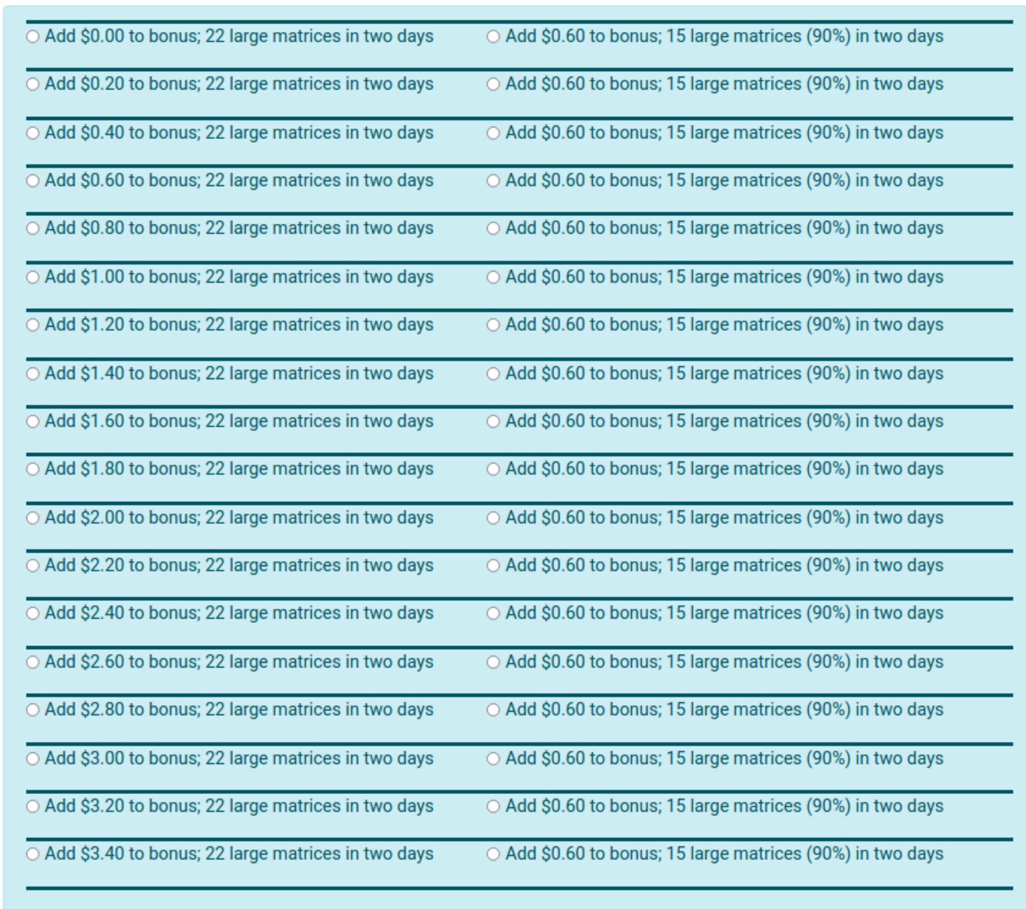

Figure 2: Example of a price list

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

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

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 two 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 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.[8] 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 2, with the excuse option being 15 large matrices in the future with $p = 0.9$, is the following:

$$\Delta\Delta = 0.9d(15) \cdot (\beta - \beta_E)$$

[8] 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 no-excuse and excuse alternatives.

(see the exact choices in the Appendix).

Based on Exley [2016], our hypothesis is that in the 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.

4 First 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.[9] 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 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.[10] 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 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).[11] Specifically, we run the following regression, separately by batch, and then with batches 1 and 2 combined:

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

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

[10] 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.

[11] 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.

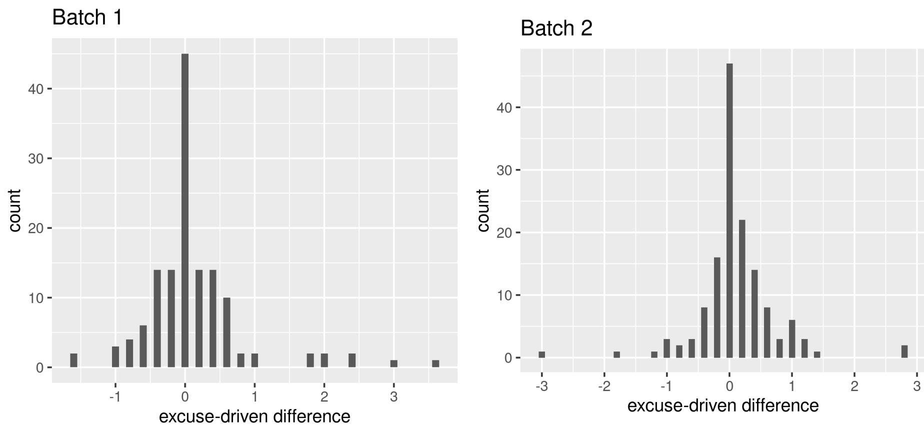

Figure 3: Excuse-driven difference in WTW in two days over WTW today

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

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.[12] 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.[13]

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
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 planned a follow-up study in a laboratory with students. We turn to the description of this lab experiment in the next section.

5 Second experiment

We wanted to replicate our results with students in a lab. We chose 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

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

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

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)[14], 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 wanted to replicate our findings for other potential excuses, rather than only for risk. Our idea was that we introduce two different types of tasks that are commonly used as effort tasks in the experimental literature – the matrix counting task, and a blurry Greek letter transcription task (see an example in Figure 4).[15] We then use a similar difference-in-differences strategy as with risk to identify excuse-making, using four choices. In the no-excuse condition people make choices about a *baseline* task, say, the matrix counting task. In the 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.

No-excuse condition:

$$\begin{aligned} x_1 \text{ matrices in 3 days} + \$X &\sim x_2 \text{ matrices in 3 days} \\ x_1 \text{ matrices today} + \$X' &\sim x_2 \text{ matrices in 3 days} \end{aligned}$$

Excuse condition:

$$\begin{aligned} x_1 \text{ matrices in 3 days} + \$Y &\sim x_2 \text{ Greek tasks in 3 days} \\ x_1 \text{ matrices today} + \$Y' &\sim x_2 \text{ Greek tasks in 3 days} \end{aligned}$$

The excuse-driven change in the willingness to do 10 matrices in 3 days over today is then, analogously to our identification with risk in Section 2, denoting the disutility of doing x Greek tasks with $d_G(x)$:

$$\Delta\Delta = \Delta_T - \Delta_C = d_G(x_2) \cdot (\beta - \beta_E)$$

[14] 0 was the least and 10 the most pleasant, while 5 was equally pleasant

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

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 preferring the other type of task from indeed using the other task as an excuse to choose to work in the future.[16]

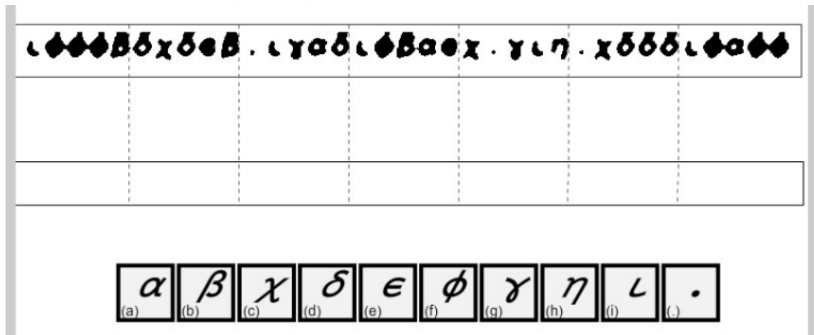

Figure 4: Example of a Greek letter transcription task

We pre-registered the second experiment the AEA RCT Registry as well [Drucker and Kaufmann, 2019]. The experiment ran between May and June 2020 with students subjects of the lab of the Luxembourg Institute of Socio-Economic Research (LISER-LAB). Due to the restrictions because of the COVID-19 pandemic, the experiment was conducted entirely online. It ran for four weeks with two sessions per week on Mondays and Thursdays. We moved from two days between sessions to three days to allow for a lower estimated β, as Augenblick [2018] finds a few hours' β of 0.94, a daily β of 0.91 and a weekly one of 0.87. We chose to run the experiment for four weeks to gain more data points, and to let participants become familiar with the setting. Participants received a €15 completion bonus each week if completing both sessions that week, but it was not required to participate in all weeks; a participant could complete one week then skip the next one and come back the third week, for example. There was an extra €15 bonus for completing all four weeks. The total payment if a student participated in all sessions throughout four weeks was around €95. Four waves of students started the experiment, each wave one week after the previous one. In total, 75 students started the study and 47 stayed until the end.

Similarly to the first experiment, participants made all decisions with price lists. We tested two excuses: risk and a different type of task (matrix task vs. Greek letter transcription task). Four price lists added up to a batch to identify $\Delta\Delta$ – the extent of excuse-making in present bias. In each week we asked four risk batches and four matrix-versus-Greek batches. All four questions of a risk batch involved the same task – Greek or matrix – but the type of task was randomized within participant

[16]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].

across batches. In the matrix-versus-Greek batches one task was the baseline and the other one the excuse task, also randomized across batches. Each Thursday, participants made work decisions for two weeks ahead, to elicit long-term preferences, and each Monday, they could revise their decisions for that week, to elicit short-term preferences. This allowed us to differentiate between present bias and exponential time discounting [as in Augenblick et al., 2015]. There were 10 required tasks in each session, and the choices this time were about *extra* tasks instead of hard versus easy tasks (see all choices in the Appendix.)

6 Second experiment results

In the second experiment we used the same simple regression as in the first one:

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

We report the results jointly for all batches within an excuse type. Table 4 shows the four main estimated $\Delta\Delta$-s. The first column shows the estimated $\Delta\Delta$ in the short-term work decisions made for that week when the excuse was risk. The second column shows the long-term $\Delta\Delta$ in the work decisions for two weeks ahead. The third and fourth columns report the estimated $\Delta\Delta$-s in the decisions where the excuse was a different type of task (matrix-versus-Greek). The third column is the short-term $\Delta\Delta$ and the fourth the long-term $\Delta\Delta$ of this excuse type. We expected the first and third $\Delta\Delta$-s to be significant and positive, and the second and fourth to be zero or smaller positive. Neither of the $\Delta\Delta$-s turned out to be significant, and the signs are not always as expected, either.

Table 4: Main results

	Dependent variable:			
	$\Delta\Delta$			
	Risk short-term	Risk long-term	M-v-G short-term	M-v-G long-term
	(1)	(2)	(3)	(4)
Constant	0.060	−0.086	−0.032	0.105
	(0.066)	(0.076)	(0.068)	(0.087)
Observations	411	293	259	166

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the individual level.
Clustered standard errors in parentheses.

In the first session, we asked the participants to fill in a survey about some aspects of the study. Two of the survey questions were to rate on a 0 to 10 scale how pleasant they find each type of task. From the two ratings we can see which task they prefer. Since the measured $\Delta\Delta$ depends on the disutility of the task, it might be possible that there is an excuse-driven present bias for the disliked task but not for the preferred task. In the matrix-vs-Greek decisions, it is possible that the disliked task is not a good excuse to postpone the preferred task, but the excuse could work the other way around. Table 5 looks at whether we get different results in the short-term risk decisions by whether the task is the preferred or the disliked one, and in the matrix-versus-Greek decisions by whether the baseline task is preferred or disliked. None of the $\Delta\Delta$-s are significant with this distinction, either. Based on these results we think that the different type of task does not work as an excuse. However, because of the promising MTurk results, we still think that risk might work.

Table 5: Short-term $\Delta\Delta$ by preferred/disliked task

	Dependent variable:			
	$\Delta\Delta$			
	Risk preferred	Risk disliked	M-v-G preferred	M-v-G disliked
	(1)	(2)	(3)	(4)
Constant	0.041	0.069	0.088	−0.159
	(0.086)	(0.087)	(0.080)	(0.103)
Observations	164	169	96	111

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the individual level.
Clustered standard errors in parentheses.

Although students are not doing these types of tasks often, it is still possible that their choices do not only depend on the disutility of our tasks and their time preferences, but they also substitute our tasks with other tasks unrelated to the experiment. To test whether they make different decisions in questions similar to the MTurk study, we recruited 30 new participants at the end of the second experiment. This follow-up was not pre-registered. In the follow-up we reintroduced the *intensive margin* questions that were about hard vs easy tasks instead of extra tasks. We also made the choices more simple,[17] to minimize the possibility that participants decide randomly. These participants made four batches of risk choices, only with matrix tasks (see the exact choices in the Appendix). Table 6 shows the $\Delta\Delta$ for these simpler, intensive margin choices. There is a large positive $\Delta\Delta$ for these choices that is significant at the 10 percent level:[18] The excuse-driven increase in the willingness to work in three days over work today is €0.24.

From these results we suspect that different mechanisms work on the intensive and the extensive

[17]E.g., 10 tasks today vs. 10 tasks in 3 days, instead of 12 tasks today vs. 8 tasks in 3 days.
[18]Since we have only 30 participants, the standard errors are not clustered.

Table 6: Follow-up risk choices

	Dependent variable:
	$\Delta\Delta$ Risk short-term
Constant	0.242* (0.146)
Observations	113
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses.

margins. In some of the batches participants had to choose between the same amount of work today or in three days, so from these batches we have a reduced-form measure for present bias. From the choices between the same amount of work for sure in three days and with 90% probability in three days we can estimate a reduced-form measure of the value of the 10% drop in probability. Table 7 shows these values at the extensive and intensive margin choices, only at the risk batches. Although we only had few such choices, it is clear that participants exhibit some present bias on both margins: they ask for 72 cents more for doing the same amount of extra tasks today instead of in three days, and 41 cents more for changing the same amount of tasks from easy to hard today instead of in three days. Unfortunately, we did not ask these questions two weeks into the future, so we do not have such reduced-form measures for δ only. In the third and fourth columns we can see the value of the 10% drop in the probability of future tasks. Participants clearly value the drop − therefore the excuse − positively: they ask 48 cents more for extra future tasks for sure instead of extra future tasks with 90% chance, and 47 cents more for hard instead of easy tasks for sure instead of the same option with 90% probability.

Table 7: Reduced-form direction of $\beta\delta$ and value of the 10% drop

	Dependent variable:			
	$(1-\beta\delta)d(x)$		$(1-0.9)\beta d(x)$	
	Extensive	Intensive	Extensive	Intensive
	(1)	(2)	(3)	(4)
Constant	0.717***	0.407***	0.476***	0.469***
	(0.181)	(0.120)	(0.124)	(0.150)
Observations	41	57	42	58
Note:				*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses.

Though there is present bias and a positive valuation of the excuse option on both margins, Table 8 suggests that the estimated excuse-driven present bias in these choices might be very different. In this experiment we do not have enough power to compare the extensive and intensive margins, so we plan to run a third experiment to test whether there is indeed an excuse-driven present bias at the intensive margin but not at the extensive margin. If the intensive margin results replicate but we find no effect at the extensive margin, that would suggest two different mechanisms: At the extensive margin, more work crowds out other tasks with similar disutility but lower pay, so the choices rather hinge on hourly wage instead of disutility of the task. At the intensive margin, on the other hand, harder tasks replace easier tasks with lower pay, so the choices hinge on disutility of the task, hence on time preferences.

Table 8: $\Delta\Delta$

	Dependent variable:	
	$\Delta\Delta$	
	Risk extensive	Risk intensive
	(1)	(2)
Constant	−0.000	0.439*
	(0.127)	(0.220)
Observations	41	57
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses.	

The other issue with our second experiment results is power. In the end, we only had 75 participants and 5-6 data points per participant on average from the short-term risk choices, even less from the long-term ones. To have 80 percent power to detect a $\Delta\Delta$ of 11 cents, with 5 data points per person we would have needed at least around 100 people, or the same amount of people but 10 data points per person. If we look at the short-term risk results separately by task (Table 9), we get a $\Delta\Delta$ of similar magnitude for the matrix task to that in the MTurk experiment, but a close to zero $\Delta\Delta$ for the Greek task, both insignificant. Therefore, we plan the final experiment to last only one week, to reduce attrition, and to have enough observations for 80 percent power to detect a $\Delta\Delta$ of similar magnitude. To be able to separate short-term excuse-driven behavior from long-term, we are going to have three sessions: a session 0 to make effort choices for Sessions 1 and 2, and 2 work sessions. In Session 1, participants will be able to revise their effort choices. We aim to test both extensive and intensive margin choices, to look at whether these choices indeed depend on different factors.

Table 9: Short-term $\Delta\Delta$ by type of task

	Dependent variable:	
	$\Delta\Delta$	
	Risk matrix	Risk Greek
	(1)	(2)
Constant	0.105	0.012
	(0.079)	(0.097)
Observations	212	199
Note:		*p<0.1; **p<0.05; ***p<0.01
		Standard errors clustered at the individual level.
		Clustered standard errors in parentheses.

7 Discussion

We tested whether people behave in a more present-biased way if they have excuses to do so. Our two experiments yielded mixed results. Still, the MTurk and the LISER-LAB results on risk suggest that a chance of no work/ easier work in the future makes people asymmetrically make different intertemporal work decisions than when all work is certain. We call this phenomenon excuse-driven present bias, but other mechanisms can work in the same direction, too. There is evidence that risk and time preferences are intertwined. E.g., people might exhibit present biased behavior because they think that future consumption is uncertain [Chakraborty et al., 2020], or they might trade off the probability and timing of consumption [Baucells and Heukamp, 2012]. Interrelated time and risk preferences would only explain our results if they were context-dependent: we identify excuse-driven present bias from a change in the valuation of an option depending on the alternative. We are unaware of such a theory that would predict participants treating risky future disutility differently by the alternative option.

It is possible that our participants have reference-dependent preferences [Kőszegi and Rabin, 2006] for the tasks, and they always set the fixed option in the price list as their reference point. In this case, if the excuse option is worse in itself than the no-excuse option, reference dependence also predicts a positive $\Delta\Delta$. However, if we set the excuse option to be better than the no-excuse option for most subjects, reference dependence works against us, predicting a negative $\Delta\Delta$. We made sure in our batches that the excuse option is at least as good as the no-excuse option, so if we find a positive $\Delta\Delta$, that is *despite* potentially reference-dependent preferences (see a formal derivation in Appendix Section A.3.4). In fact, the signs of the coefficients in Table 5 in the matrix vs. Greek choices can indicate reference-dependent preferences. It is also possible that the positive $\Delta\Delta$ we measure is a consequence of a framing effect [Tversky and Kahneman, 1981] or attention [Simon, 1971]. Still, reacting differently to a particular framing of a choice or differential attention to details of a choice depending on the alternative is really similar to a mechanism that we call excuse-making.

At the end of our second experiment, we asked participants to give their opinion on some topics related to procrastination and excuse-making. They had to indicate on a 0 to 10 scale how much they agree with different statements. The results are shown in Table 10. The first two questions aimed to elicit whether students think they are prone to some typical procrastination behaviors: ending up working out too little or always postponing household chores. On average, our students neither agree nor disagree with the corresponding statements. They are also indifferent in the question of whether they make excuses when postponing work to do. However, they rather agree with the statement that people in general make excuses for postponing work, and with people postponing work in the hope of not having to do it in the end. Based on these answers and the results of both experiments on the risk excuse, we hope to be able to corroborate in a third, final experiment that there can be excuse-making in present bias.

Table 10: Final survey results

Statistic	N	Mean	St. Dev.	Min	Max
I usually postpone household chores.	74	5.027	2.998	0	10
I usually work out too little.	74	5.392	3.355	0	10
People usually make excuses for postponing work they have to do.	67	7.463	1.995	1	10
I usually make excuses for postponing work I have to do.	67	4.881	2.826	0	10
People postpone their work in the hope of not having to do it later.	67	6.149	2.624	1	10

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.

Manel Baucells and Franz H. Heukamp. Probability and time trade-off. *Management Science*, 58(4): 831–842, 2012. doi: 10.1287/mnsc.1110.1450. URL https://doi.org/10.1287/mnsc.1110.1450.

Anujit Chakraborty, Yoram Halevy, and Kota Saito. The relation between behavior under risk and over time. *American Economic Review: Insights*, 2(1):1–16, March 2020. doi: 10.1257/aeri.20190051. URL http://www.aeaweb.org/articles?id=10.1257/aeri.20190051.

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, 2016.

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*, 2021.

Botond Kőszegi and Matthew Rabin. A Model of Reference-Dependent Preferences*. *The Quarterly journal of economics*, 121(4):1133–1165, 2006.

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.

Herbert A. Simon. Designing organizations for an information-rich world. In M. Greenberger, editor, *Computers, Communications, and the Public Interest*, pages 37–72. Johns Hopkins Press, 1971.

Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981. doi: 10.1126/science.7455683. URL https://www.science.org/doi/abs/10.1126/science.7455683.

A Appendix

A.1 Choices in batches 1 and 2 of the first experiment

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 11 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 ~ 20 hours weekly on MTurk, and earn ~ 132 dollars per week from this work.

Table 11: 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

A.3 Choices in the second experiment

A.3.1 Risk batches

The tasks are either all Greek tasks or all matrices within a batch. All tasks are additional to the 10 required tasks in the session. In the decisions about work two weeks ahead, the dates are "Monday in two weeks" and "Thursday in two weeks".

Batch 1

1. 20 tasks on Thursday or 13 tasks on Thursday + €2.20
2. 20 tasks on Thursday or 10 tasks on Thursday with $p = 0.9$ + €2.20
3. 19 tasks today or 13 tasks on Thursday + €1.20
4. 19 tasks today or 10 tasks on Thursday with $p = 0.9$ + €1.20

Batch 2

1. 18 tasks on Thursday or 14 tasks on Thursday + €1.80
2. 18 tasks on Thursday or 12 tasks on Thursday with $p = 0.9$ + €1.80
3. 21 tasks today or 14 tasks on Thursday + €0.80
4. 21 tasks today or 12 tasks on Thursday with $p = 0.9$ + €0.80

Batch 3

1. 22 tasks on Thursday or 16 tasks on Thursday + €1.00
2. 22 tasks on Thursday or 13 tasks on Thursday with $p = 0.9$ + €1.00
3. 25 tasks today or 16 tasks on Thursday + €0.80
4. 25 tasks today or 13 tasks on Thursday with $p = 0.9$ + €0.80

Batch 4

1. 25 tasks on Thursday or 12 tasks on Thursday + €1.60
2. 25 tasks on Thursday or 10 tasks on Thursday with $p = 0.9$ + €1.60
3. 23 tasks today or 12 tasks on Thursday + €1.00
4. 23 tasks today or 10 tasks on Thursday with $p = 0.9$ + €1.00

Batch 5

1. 11 tasks on Thursday or 11 tasks on Thursday + €1.20
2. 11 tasks on Thursday or 11 tasks on Thursday with $p = 0.9$ + €1.20
3. 11 tasks today or 11 tasks on Thursday + €1.20
4. 11 tasks today or 11 tasks on Thursday with $p = 0.9$ + €1.20

Batch 6

1. 17 tasks on Thursday or 10 tasks on Thursday + €0.40
2. 17 tasks on Thursday or 10 tasks on Thursday with $p = 0.9$ + €0.40
3. 10 tasks today or 10 tasks on Thursday + €0.20
4. 10 tasks today or 10 tasks on Thursday with $p = 0.9$ + €0.20

Batch 7

1. 15 tasks on Thursday or 15 tasks on Thursday + €1.00
2. 15 tasks on Thursday or 15 tasks on Thursday with $p = 0.9$ + €1.00
3. 15 tasks today or 15 tasks on Thursday + €1.00
4. 15 tasks today or 15 tasks on Thursday with $p = 0.9$ + €1.00

Batch 8

1. 12 tasks on Thursday or 8 tasks on Thursday + €0.20
2. 12 tasks on Thursday or 8 tasks on Thursday with $p = 0.9$ + €0.20
3. 12 tasks today or 8 tasks on Thursday + €0.00
4. 12 tasks today or 8 tasks on Thursday with $p = 0.9$ + €0.00

A.3.2 Matrix-versus-Greek batches

All tasks are additional to the 10 required matrix or Greek tasks in the session. In the decisions about work two weeks ahead, the dates are ”Monday in two weeks” and ”Thursday in two weeks”.

Batch 1

1. 20 matrix (Greek) tasks on Thursday or 12 matrix (Greek) tasks on Thursday + €1.80
2. 20 matrix (Greek) tasks on Thursday or 9 Greek (matrix) tasks on Thursday + €1.80
3. 18 matrix (Greek) tasks today or 12 matrix (Greek) tasks on Thursday + €1.20
4. 18 matrix (Greek) tasks today or 9 Greek (matrix) tasks on Thursday + €1.20

Batch 2

1. 19 matrix (Greek) tasks on Thursday or 13 matrix (Greek) tasks on Thursday + €1.60
2. 19 matrix (Greek) tasks on Thursday or 10 Greek (matrix) tasks on Thursday + €1.60
3. 23 matrix (Greek) tasks today or 13 matrix (Greek) tasks on Thursday + €0.60
4. 23 matrix (Greek) tasks today or 10 Greek (matrix) tasks on Thursday + €0.60

Batch 3

1. 25 matrix (Greek) tasks on Thursday or 14 matrix (Greek) tasks on Thursday + €1.00
2. 25 matrix (Greek) tasks on Thursday or 11 Greek (matrix) tasks on Thursday + €1.00
3. 22 matrix (Greek) tasks today or 14 matrix (Greek) tasks on Thursday + €0.80
4. 22 matrix (Greek) tasks today or 11 Greek (matrix) tasks on Thursday + €0.80

Batch 4

1. 21 matrix (Greek) tasks on Thursday or 15 matrix (Greek) tasks on Thursday + €1.40
2. 21 matrix (Greek) tasks on Thursday or 12 Greek (matrix) tasks on Thursday + €1.40
3. 24 matrix (Greek) tasks today or 15 matrix (Greek) tasks on Thursday + €1.20
4. 24 matrix (Greek) tasks today or 12 Greek (matrix) tasks on Thursday + €1.20

Batch 5

1. 11 matrix (Greek) tasks on Thursday or 11 matrix (Greek) tasks on Thursday + €0.80
2. 11 matrix (Greek) tasks on Thursday or 8 Greek (matrix) tasks on Thursday + €0.80
3. 11 matrix (Greek) tasks today or 11 matrix (Greek) tasks on Thursday + €0.80
4. 11 matrix (Greek) tasks today or 8 Greek (matrix) tasks on Thursday + €0.80

Batch 6

1. 16 matrix (Greek) tasks on Thursday or 16 matrix (Greek) tasks on Thursday + €0.60
2. 16 matrix (Greek) tasks on Thursday or 16 Greek (matrix) tasks on Thursday + €0.60
3. 16 matrix (Greek) tasks today or 16 matrix (Greek) tasks on Thursday + €0.60
4. 16 matrix (Greek) tasks today or 16 Greek (matrix) tasks on Thursday + €0.60

Batch 7

1. 20 matrix (Greek) tasks on Thursday or 14 matrix (Greek) tasks on Thursday + €2.00
2. 20 matrix (Greek) tasks on Thursday or 14 Greek (matrix) tasks on Thursday + €2.00
3. 20 matrix (Greek) tasks today or 14 matrix (Greek) tasks on Thursday + €1.20
4. 20 matrix (Greek) tasks today or 14 Greek (matrix) tasks on Thursday + €1.20

Batch 8

1. 15 matrix (Greek) tasks on Thursday or 15 matrix (Greek) tasks on Thursday + €1.20
2. 15 matrix (Greek) tasks on Thursday or 10 Greek (matrix) tasks on Thursday + €1.20
3. 15 matrix (Greek) tasks today or 15 matrix (Greek) tasks on Thursday + €1.20
4. 15 matrix (Greek) tasks today or 10 Greek (matrix) tasks on Thursday + €1.20

A.3.3 Follow-up risk

All tasks are matrix tasks. Batch 1

1. 11 hard tasks on Thursday or 11 hard tasks on Thursday + €1.20
2. 11 hard tasks on Thursday or 11 hard tasks on Thursday with $p = 0.9$ + €1.20
3. 11 hard tasks today or 11 hard tasks on Thursday + €1.20
4. 11 hard tasks today or 11 hard tasks on Thursday with $p = 0.9$ + €1.20

Batch 2

1. 17 hard tasks on Thursday or 10 hard tasks on Thursday + €0.40
2. 17 hard tasks on Thursday or 10 hard tasks on Thursday with $p = 0.9$ + €0.40
3. 17 hard tasks today or 10 hard tasks on Thursday + €0.20
4. 17 hard tasks today or 10 hard tasks on Thursday with $p = 0.9$ + €0.20

Batch 3

1. 15 hard tasks on Thursday or 15 hard tasks on Thursday + €1.00
2. 15 hard tasks on Thursday or 15 hard tasks on Thursday with $p = 0.9$ + €1.00
3. 15 hard tasks today or 15 hard tasks on Thursday + €1.00
4. 15 hard tasks today or 15 hard tasks on Thursday with $p = 0.9$ + €1.00

Batch 4

1. 12 hard tasks on Thursday or 8 hard tasks on Thursday + €0.20

2. 12 hard tasks on Thursday or 8 hard tasks on Thursday with $p = 0.9$ + €0.20

3. 12 hard tasks today or 8 hard tasks on Thursday + €0.20

4. 12 hard tasks today or 8 hard tasks on Thursday with $p = 0.9$ + €0.20

A.3.4 Ruling out reference dependence

Let us regard one general batch:

1. x_1 in three days + X or x_2 in three days + \$0

2. x_3 today + X' or x_2 in three days + \$0

3. x_1 in in three days + Y or x_4 in three days with $p = 0.9$ + \$$z$

4. x_3 today + Y' or x_4 in three days with $p = 0.9$ + \$$z$

Then, X, X', Y, and Y' are the differences between the utilities of the corresponding right-hand side (RHS) and the left-hand side (LHS) options.

$$X = u(RHS)_X - u(LHS)_X$$

Let us suppose that participants exhibit reference-dependent preferences such that $U(x|r) = u(x) + \mu(x)$. From having to do x tasks they get a (dis)utility of $u(x) = -d(x)$ plus a news-utility component that depends on whether they have to work more or less than their reference point:

$$\mu(x|r) = -\eta(d(r) - d(x)) \text{ if } x < r$$
$$\mu(x|r) = -\eta\lambda(d(x) - d(r) \text{ if } x \geq r$$

Since doing the tasks yield negative utility, having to do *more* tasks than the reference point is *negative* news. Also, having to do tasks today yields negative news utility compared to the reference point of no work today, but it also entails no tasks in three days, which has positive news utility compared to the expectation to work in three days. Let us assume that the participants set the fixed option in the price lists as their reference point. Assume that participants have present bias β but they are fully patient, so $\delta = 1$. In all of our batches, $x_1 \geq x_2, x_4$ and $x_3 \geq x_2, x_4$, to make sure that the LHS option is worse than the RHS option. $d(x)$ is the disutility of doing x tasks, and $d(x, p)$ is the disutility of a lottery where one has to do x task with probability p, and zero tasks otherwise. Let us further assume that $\eta = 1$ and the utility of money to be linear:

$$
\begin{aligned}
X &= (1+\lambda)\beta d(x_1) - (1+\lambda)\beta d(x_2) \\
X' &= (1+\lambda)d(x_3) - 2\beta d(x_2) \\
Y &= (1+\lambda)\beta d(x_1) - (1+\lambda)\beta d(x_4, 0.9) + \$z \\
Y' &= (1+\lambda)d(x_3) - 2\beta d(x_4, 0.9) + \$z
\end{aligned}
$$

Then,

$$\Delta\Delta = (\lambda - 1)\beta(d(x_4, 0.9) - d(x_2))$$

Therefore, the direction of $\Delta\Delta$ depends on the disutility difference between the excuse option and the no-excuse option. If the excuse option yields higher utility for most subjects than the no-excuse option, i.e. $d(x_4, 0.9) - d(x_2) < 0$, then reference dependence would predict a negative $\Delta\Delta$. In our first experiment, only Batch 2 met this requirement, but the results were similar in both batches. In the second experiment we made sure that the excuse option is always at least as good as the no-excuse option, so that if we find a positive $\Delta\Delta$, that means a strong excuse-driven present bias *despite* potentially reference-dependent preferences.