DISCUSSION PAPER SERIES

DP15716

Self-Set Goals Are Effective Self-Regulation Tools -- Despite Goal Revision

Jonas P. Kaiser, Alexander K Koch and Julia Nafziger

INDUSTRIAL ORGANIZATION

PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

Self-Set Goals Are Effective Self-Regulation Tools -- Despite Goal Revision

Jonas P. Kaiser, Alexander K Koch and Julia Nafziger

Discussion Paper DP15716
Published 25 January 2021
Submitted 13 January 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Self-Set Goals Are Effective Self-Regulation Tools -- Despite Goal Revision

Abstract

We test in an experiment whether and why non-binding, self-set goals are effective even though individuals can easily revise such goals. Depending on the treatment, subjects either set a goal a few days before working on an online task or right at the start of the task. In the former case, they may or may not be explicitly asked to revise their goal at the start of the task. Consistent with the hypothesis that goals are self-regulation tools, we observe that goals set before the task are larger than goals set at the start of the task. And they are effective: Subjects work more when a goal was set a few days before the task than when it was set at the start of the task. Importantly, these results arise even though subjects revise their initial goals. They do so no matter whether goal revision is made explicit or not -- suggesting that goal revision is an important factor for goal non-achievement.

JEL Classification: D03, D81, D91

Keywords: Self-Control, Present-Biased Preferences, Goals, goal revision, commitment devices, Real effort, Online Experiment

Jonas P. Kaiser - jkaiser@econ.au.dk
Aarhus University

Alexander K Koch - akoch@econ.au.dk
Aarhus University

Julia Nafziger - jnafziger@econ.au.dk
Aarhus University and CEPR

Acknowledgements
We would like to thank Elena Cettolin, Patricio Dalton, Daniele Nosenzo, Marco Schwarz, Ben Vollaard, seminar participants at the universities of Tilburg, Innsbruck and Aarhus, the Annual Meeting of the Verein fur Socialpolitik, and the CNEE Workshop for helpful comments. Financial support from the Interacting Minds Center is gratefully acknowledged.

Self-Set Goals Are Effective Self-Regulation Tools – Despite Goal Revision

Jonas Pilgaard Kaiser[a], Alexander Koch[b], and Julia Nafziger[c*]

[a]Aarhus University, [b]Aarhus University and IZA, [c]Aarhus University and CEPR

January 13, 2021

Abstract

We test whether and why non-binding, self-set goals are effective even though individuals can easily revise such goals. Depending on the treatment, subjects either set a goal a few days before working on an online task or right at the start of the task. In the former case, they may or may not be explicitly asked to revise their goal at the start of the task. Consistent with the hypothesis that goals are self-regulation tools, we observe that goals set before the task are larger than goals set at the start of the task. And they are effective: Subjects work more when a goal was set a few days before the task than when it was set at the start of the task. Importantly, these results arise even though subjects revise their initial goals. They do so no matter whether goal revision is made explicit or not – suggesting that goal revision is an important factor for goal non-achievement.

JEL Classification: D03, D81, D91

Keywords: Self-control, present-biased preferences, goals, goal revision, commitment devices, real effort, online experiment

*The study is pre-registered on OSF. In Kaiser et al. (2021), we report the plan, summarize changes compared to the plan and report missing analysis. Specifically, in Section 5 we mix reporting pre-registered mechanisms, pre-registered exploratory research and exploratory research – and we classify these analyses in the report. We would like to thank Elena Cettolin, Patricio Dalton, Daniele Nosenzo, Marco Schwarz, Ben Vollaard, seminar participants at the universities of Tilburg, Innsbruck and Aarhus, the Annual Meeting of the Verein für Socialpolitik, and the CNEE Workshop for helpful comments. Financial support from the Interacting Minds Center is gratefully acknowledged. Contact: Department of Economics and Business Economics, Aarhus University, Fuglesangsalle 4, 8210 Aarhus V, Denmark. [a]Email: jkaiser@econ.au.dk. [b]Email: akoch@econ.au.dk, [c]Email: jnafziger@econ.au.dk.

1 Introduction

When deciding whether and how much to study, work, diet, or exercise people often have a tendency to overemphasize present costs relative to future benefits. As a consequence, self-control problems arise in that people study, work, or exercise less and eat more than they initially thought is optimal. To engage in self-regulation, people can set goals for themselves. Yet, as these goals are non-binding and people can easily revise them when facing the temptation that these goals should help to counteract in the first place, two questions arise. Do people indeed use goals as self-regulation tools? And are they effective in regulating behavior?

The aim of this paper is to provide the first clean test on the use and effectiveness of self-set, non-binding goals as self-regulation tools. As we explain below, the existing experimental studies that consider whether self-set goals are effective motivators suffer from a problem of treatment migration that arises from unobserved goal setting. Further, as the question on the effectiveness of goals goes hand in hand with the role of goal revisions, we also analyze the extent of goal revision and the role that it plays for task performance. In doing so, we provide the first comprehensive study on goal revision in the light of self-control problems.

Our real-effort experiment mimics a typical work-leisure self-control problem by offering subjects a generous piece rate for doing the tedious, unpleasant task of counting zeros in tables of zeros and ones. To allow for exposure to the usual real-life temptations while subjects work, the experiment runs online and neither requires subjects to show up at a lab nor to obey a particular schedule.

To answer whether goals are self-regulation tools and whether they are effective, we compare a treatment where subjects set a goal five days before the task with a treatment where subjects set the goal immediately before the task. Applying the insights from a stylized model where the individual has present-biased preferences and sets goals,[1] we predict that the individual tries to counteract his present-bias when setting a goal five days in advance of the task, but not when setting it right before the task. Hence, if goals are self-regulation tools, subjects should set higher goals in the treatment where the goal is set five days in advance of the task compared to the treatment where goals are set immediately prior to the task.[2] Further, as higher goals should translate into higher effort, we should observe higher effort in the former treatment than in the latter treatment.

This design addresses the treatment migration problem that arises in the previous experimental literature on goal setting. Here, the typical experiment has some subjects set a goal before working on the task (treatment condition), while others simply work on the task (control

[1] The literature in economics on goal setting offers several theoretical models to capture how non-binding, personal goals help people to overcome self-control problems. The underlying idea in these models is that goals serve as reference points that make substandard performance painful (Suvorov and van de Ven, 2008; Jain, 2009; Koch and Nafziger, 2011; Hsiaw, 2013).

[2] In contrast to self-regulation tools, goals might be just expectations about effort or ordinary motivators that affect, for example, intrinsic motivation. In these cases, the goal schedule should be flat across dates, or possibly even increasing if subjects becomes more productive over time.

condition). Then, the researchers test the effectiveness of goals by comparing the performance in the two conditions. Yet, the self-regulation perspective of goal theory suggests that people set goals even if not explicitly asked to do so. Thus, subjects in the control condition may nevertheless be exposed to the treatment of setting goals. As treatment assignment is an imperfect proxy for treatment received, the intention-to-treat estimate understates the causal effect of treatment. As parts of the literature on goal setting in economics find insignificant effects or low effect sizes of goals on performance, it is important to find credible ways to identify the true effects.

Yet, when setting a goal in advance of a task, individuals may revise the goal when facing the task. Hence, in two further treatments, we explicitly allow subjects who set a goal five days before the task to revise their goal just before the task. These treatments enable us to examine whether and to what the extent people revise their goals when explicitly given the possibility to do so. Further, comparing the effort in these treatments with the treatment where subjects set a goal just before the start of the task enables us to test whether goals are effective despite (observed) goal revisions.[3]

To preview our results, we find, first, that goals are indeed used as self-regulation tools in our male-only-sample (in Section 2.2, we outline why we pre-registered to only recruit men). Goals set five days before the task are higher than goals set immediately before the task. This result also holds when considering goal revision as a within-subject comparison: Subjects on average revise their initial goal downwards. Second, goals are effective self-regulation tools. Subjects work more when they set a goal five days before the task compared to just setting the goal immediately before engaging in the task. Third, this holds even when subjects explicitly revise their goals. That is, having set a goal a few days in advance of the task and then revising it at the start of the task improves task performance relative to setting a goal only at the start of the task.

Examining the mechanisms behind these results makes it clear that setting an initial goal matters. People who set an initial goal are more likely to achieve the goal revised just before the task than people who only set a goal just before the task. That is, people seem to strive to some extend for their (high) initial goal and are thus more likely to achieve their revised goal than people who do not set an initial goal. However, we observe that the revised goal does not significantly differ from the goal people set just before the task. This speaks against the view that the initial goal serves as an anchor in goal revision.

The result that goals set some days before the task are larger than goals set at the start of the task is consistent with the interpretation that effort goals serve as self-regulation tools. We address a number of alternative explanations behind this results, such as changes in productivity, learning, or resolution of uncertainty about time shocks. Of these, only learning about the task may play a role. Yet, our results remain robust when we control for such learning.

[3] Asking explicitly subjects to revise goals may prompt more often goal revision than when these revisions are self-motivated. Hence, the comparison between these treatments may underestimate the effectiveness of goals.

Further, we test whether other goals than effort goals matter, but find no indication for that (e.g., income and time-based goals).

Finally, our design allows us to test whether framing can make goals more effective. Specifically, reminding subjects about a specific goal (either the revised or the initial goal) should make that goal more salient and thus the subject may strive more for it. Similarly, explicit goal revisions may make the revised goal more salient than private revisions and thus lead to a lower effort. Yet, we find that no matter which goal subjects are reminded about, they seem to care more about the recent, revised goal than about their initial goal. And no matter whether goal revision is explicit or not, subjects provide the same effort.

While the latter result goes against the framing hypothesis, it has important insights for the literature on goal setting. Combined with answers to our ex-post survey, it suggests that private, self-motivated goal revisions do take place. This suggests that the literature on the effort-goal relationship measures effort against the wrong goal (i.e., the stated initial goal rather than the revised goal). Thus, the literature may not reveal the true relationship.

The paper continues as follows. Next, we discuss the related literature. Section 2 lays out the experimental design and procedures. In Section 3, we present a stylized model to derive our main predictions. We test these in Section 4. In Section 5, we consider a number of possible mechanisms behind our findings and test alternative explanations behind the results. Section 6 discusses advantages and disadvantages behind certain design choices and outlines avenues for future research. Section 7 concludes the paper.

Related Literature. A number of papers in economics study how goal setting impacts performance by comparing treatments where subjects set a non-binding goal for performance with a control treatment where no goals are elicited. Most studies find that goal setting has a positive effect on performance (Fan et al., 2019; Goerg and Kube, 2012; Smithers, 2015, but not Akina and Karagozoglub, 2017). This is also the case for studies that consider the effects of goals for the performance in repeated tasks, such as weight loss (Toussaert, 2016), energy saving (Harding and Hsiaw, 2014), or studying (Clark et al., 2019; Himmler et al., 2019; van Lent, 2019; van Lent and Souverijn, 2020). Koch and Nafziger (2020) consider self-set, non-binding goals in repeated tasks and find that daily goals lead to higher effort than equivalent weekly goals. The psychology literature on goals often compares specific goals with "do your best" or no-goal setting; including externally assigned and participatory goals. See, for instance, Locke and Latham (1990, 2013, 2019) for references to additional studies.

Related is also the literature on goals in work environments, where managers can either tie goals to monetary rewards (Dalton et al., 2015; Goerg and Kube, 2012; Kaur et al., 2015; Gonzalez et al., 2020) or not (Brookins et al., 2017; Corgnet et al., 2015, 2018; Cettolin et al., 2020). With the exception of Dalton et al. (2015), these studies also suggest that goals have a positive impact on performance. However, Gonzalez et al. (2020) find that it can be counterproductive to tie self-set goals to monetary bonuses because loss aversion then can induce workers to set

lower goals.

While there is a large literature on goal setting and performance, less research has been done on goal revision. Studies in psychology focus on how people update their goals over multiple performance episodes *after* they start striving for a goal and receive feedback about performance (e.g., Campion and Lord, 1982; Donovan and Williams, 2003; Ilies and Judge, 2005). The typical finding is that goals are adjusted upwards following success or positive feedback and downwards following failure or negative feedback. In economics, van Lent (2019) provides, to our knowledge, the only experimental study on goal revision, which is similar in spirit to the studies in psychology. In a larger survey, he asks students whether they want to set a goal for their course grade, a non-grade goal, or no goal. After students get feedback about their performance through tutorials and a midterm exam, they can revise their goal(s) in a second survey. The novelty of our approach compared to this literature is that we study the revision of goals prior to engaging in goal pursuit. This means that goal revision is related to the temptation to work less when facing the task rather than by good or bad news about task performance.

The topic of goal revision also relates to the literature on reference-dependent preferences. Kőszegi and Rabin (2009) offer theoretical guidance on how to model revision of reference points and Koch and Nafziger (2016) apply these insights to modeling goal revision. Some experimental studies address how fast new information is incorporated in the reference point, with mixed findings. The tournament experiment of Gill and Prowse (2012) suggests that subjects rapidly update their reference points to both their own effort choice and that of their rival. Rapid adjustment of the reference point to an exogenous change in the current endowment is also consistent with the results in Smith (2019). However the field data of Card and Dahl (2011), DellaVigna et al. (2017), or Thakral and Tô (2020) suggest slow updating of the reference point in other domains. Our contribution to the empirical evidence on updating of reference points is to consider a context where individuals update reference points (goals) because of time-inconsistency.

Finally, our study relates to Augenblick et al. (2015), who estimate present bias in effort using a real-effort task similar to ours. Subjects have to specify several plans on how to allocate effort over two dates that are a few days into the future; and then they again specify plans right before providing effort. While in their setting subjects are committed to a selected effort plan, subjects make non-binding plans (expressed as goals) in our study. Augenblick et al. (2015) find evidence for present bias in the effort domain but not in the money domain. In a similar framework, Augenblick and Rabin (2019) furthermore elicit the beliefs that individuals hold about their future effort. They demonstrate that most individuals are (partially) naïve in that they overestimate how much effort they will provide. We incorporate such naïveté in our theoretical model.

Table 1: Treatments

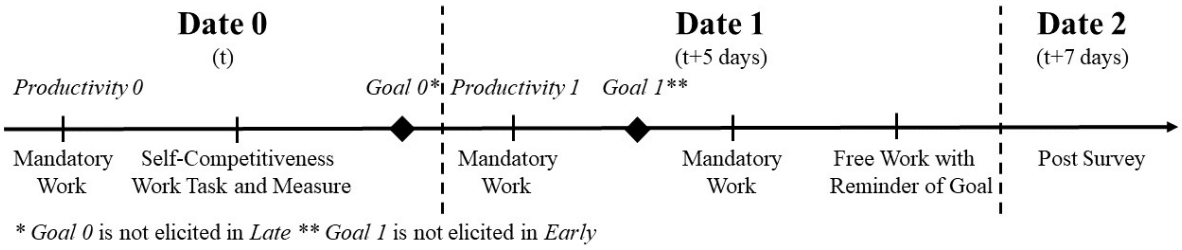
Treatment	Date 0 (t)	Date 1 ($t+5$ days)	Reminder during the free work phase
Early	Goal 0	-	Goal 0
Late	-	Goal 1	Goal 1
Revise0	Goal 0	Goal 1	Goal 0
Revise1	Goal 0	Goal 1	Goal 1

2 Experimental Design

The experiment has three parts that are conducted online on three different days, using the Qualtrics platform: A goal setting part at date 0 (t), a work part at date 1 ($t+5$ days), and a post survey at date 2 ($t+7$ days). We randomize subjects into four different treatments. In treatments *Early, Revise0*, and *Revise1*, subjects set a goal at date 0 (*goal 0*). In treatment *Late*, subjects only set a goal at date 1 (*goal 1*). Subjects in *Revise0* and *Revise1* can revise their goal at date 1. While working, we remind subjects in *Revise0* and *Early* about the goal they set at date 0. Conversely, in *Revise1* and *Late* we remind subjects about the goal that they have just set a few minutes earlier at date 1. Table 1 summarizes the four treatments. Figure 1 provides an overview of the different tasks and the timing of the experiment. Experimental instructions are in Online Supplement S. 7.

We next outline the experimental tasks in detail. In Section 6, we discuss possible disadvantages in contrast to advantages of certain design choices we make and address possible caveats.

Figure 1: Timeline of the experiment



2.1 Details of the Experimental Setup

2.1.1 Date 0: Goal Setting

The primary objective at date 0 is to elicit non-binding goals from the subjects in treatments *Early*, *Revise0*, and *Revise1* for the effort that they want to provide at date 1, i.e., in the free work phase of the experiment. For completing this part of the experiment, subjects receive DKK 35 (approx. USD 5.6) in addition to their earnings from three tasks.

Productivity Measure. Throughout the experiment, we measure effort in a real-effort task in which subjects count the number of zeros in a series of tables consisting of zeros and ones as in Abeler et al. (2011) and Koch and Nafziger (2020). To familiarize subjects with this real-effort task before they set goals, subjects count the zeros in as many tables as possible during three minutes (denoted mandatory work phase in Figure 1). For each table in which they count the number of zeros correctly (*completed table*, henceforth), subjects receive DKK .5. The total number of completed tables in these three minutes will serve as a measure of baseline productivity at date 0 (*productivity 0* for short). After the task, subjects answer a survey question on how much they like the task.

Self-Competitiveness Measure. To ensure that subjects in *Late* do not anticipate what will happen at date 1 of the experiment (and privately set goals), subjects perform an additional round of the real-effort task. Before doing so, we obtain a fine-grained measure of subjects' self-competitiveness (Apicella et al., 2017) based on the procedure of Saccardo et al. (2017). For the second round of counting zeros, subjects make a choice of what share of their pay shall be (i) determined by a fixed piece rate of DKK .5 for each completed table and (ii) determined based on their performance relative to the first round. In the latter scheme, subjects receive DKK 1 (DKK 0) for each completed table in case they complete more (fewer) tables than in the first round, and DKK .5 in case of a tie. This task solely serves as a 'smoke-screen'.

Goal Setting. This part is not relevant for subjects in *Late*. To avoid private goal setting, we provide subjects in *Late* with no details about the work to be performed at date 1 except the information that is necessary for informed consent.[4]

In all the other treatments, we inform subjects about the details of the free work phase at date 1 and the associated payment scheme (cf. Figure 2). We implement a declining piece rate to avoid corner solutions where a large fraction of the subjects counts all the available tables, which is likely with a constant piece rate (cf. Koch and Nafziger, 2020).

We then ask the subjects to set a goal for how many tables to complete in the free work phase (*goal 0*). Subjects know that the work takes place five days after the goal setting part. We fix the time interval so that present bias can create a discrepancy between desired effort in the goal setting and work parts. Specifically, Augenblick et al. (2015) and Augenblick and Rabin (2019) demonstrate how the discounting of future real-effort costs changes drastically within the first hours and days prior to the task, whereas it is almost constant $4 - 30$ days into the future.

Before setting goals, subjects have access to a slider tool that helps them reflect about how many tables they can realistically manage within their available time. Subjects indicate goal values with the slider, and the tool shows the estimated amount of time for reaching the goal

[4]Subjects fill out the consent form (Online Supplement S. 7) at least 24 hours before the experiment. It informs subjects of the overall structure of the study and that earnings depend on the number of tasks completed. The specific tasks are not described.

(based on the productivity of the subject) along with the associated earnings and the marginal piece rate.[5] We encourage subjects to experiment with the slider before entering a goal. We tell subjects that they will be reminded about the goal while working on the task with probability 2/3. We stress that how much they ultimately work is entirely up to themselves; under no circumstance will there be any punishment if they fail to reach their goal, and they may count more tables than their goal.

Survey Questions. Subjects answer a few survey questions at date 0. In the beginning, subjects fill in background information (age, type of degree, and field of study) and the number of upcoming exams and assignments in the next month. In addition, subjects answer the general risk aversion question from Dohmen et al. (2011) and the Cognitive Reflection Test (CRT, Frederick 2005). Subjects receive DKK 2 for each correct answer in the CRT.
After the mandatory work (but before setting goals), we ask subjects about their time schedule for date 1. Further, we ask them how likely they think it will be that they will end up having less than two hours of flexible time at date 1. These questions serve two purposes. First, they facilitate reflective goal setting by making subjects aware of how much time they realistically can devote to working on the task at date 1. Second, they allow us to control for possible time constraints and examine the effect of resolution of uncertainty about time shocks between dates 0 and 1.

2.1.2 Date 1: Work Part

Date 1 takes place five days after date 0. It consists of two phases. All subjects have to complete the first phase, while they can choose how much to work in the second phase.

Phase 1: Productivity Measure and Goal Setting. In the first phase, subjects have to count the number of zeros in a series of tables for two times three minutes with a break in between. They receive DKK .5 for each correctly counted table. The first three minutes serve as a baseline productivity measure at date 1 (*productivity 1*). In the break before the next three minutes, we inform/remind subjects in all treatments about the phase 2, the free work phase, that starts after the next three minutes of the counting task. In phase 2, they are free to work as much as they want under the payment scheme in Figure 2. Similar to date 0, we ask subjects to fill in their time schedule to see if (or how) the schedule for the day has changed since date 0.
Subjects in *Early* then go directly to the next three minutes of counting and thereafter to the free work phase.[6] In treatments *Late, Revise0*, and *Revise1*, we present the slider tool in the context of asking subjects to set a (new) goal. The tool is like the one at date 0 – with the

[5] There is a small difference between the mandatory work and the free work phase. In the latter, subjects have to reload the page for each table. Thus, subjects are slowed down slightly in the free work phase. The slider tool does not account for this or for potential improvements in productivity due to practice.

[6] As goal setting does not take very long, we do not include a filler task here.

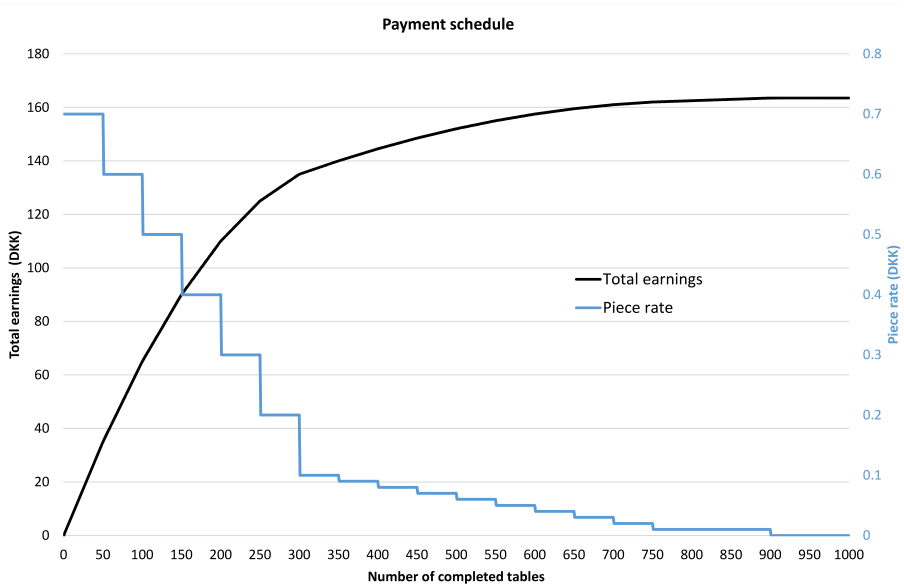


Figure 2: Payment scheme

only difference that it uses *productivity 1* as input. This way, we can see whether subjects in *Revise0* and *Revise1* adjust their goal in response to a change in their productivity between dates 0 and 1.

Subjects in *Late* set a non-binding goal for how much to work in the second phase (*goal 1*), and they know that they will be reminded about that goal when working. Subjects in *Revise0* and *Revise1* also set a goal, and we inform them that they will be reminded about their revised goal (*goal 1*) with probability 1/2 and about their previous goal (*goal 0*) with probability 1/2. To avoid noise due to differential memory, we tell subjects in both *Revise* treatments about the goal they have set at date 0 before they (potentially) adjust their goal.

In addition to the earnings from the mandatory work, subjects get a fixed payment of DKK 20 for completing the first phase.

Phase 2: Free Work. In the second phase, subjects are free to work as much as they like as long as they do not take more than 30 minutes between submitting answers and are paid according to the piece rate in Figure 2.

While working, they can always see the goal that we remind them about on the screen (*goal 0* in *Early* and *Revise0*, and *goal 1* in *Late* and *Revise1*), the number of correctly counted tables, the piece rate that applies, and their total earnings. Henceforth, we refer to the total number of completed tables in the free work phase as *effort*.

2.1.3 Date 2: Post Survey

Two days after the work part of the experiment, subjects receive an email with a link to the post survey. Subjects receive DKK 15 for completing it plus DKK 2 for each goal they remember. The survey (cf. Online Supplement S. 7) consists of several questions about goal

setting and goal commitment; both specific to the experiment and in general. The survey takes around 5 minutes to answer.

2.2 Sample

Several studies suggest that goals have a positive effect on the performance of men, while the effect sizes are very small or null for women (cf. Koch and Nafziger, 2020; Smithers, 2015; Clark et al., 2017). Thus, to achieve an appropriate power for the given budget, we only recruited men for the experiment (see Online Supplement S. 3 for the power analysis).

We recruited subjects from the subject pool of the Cognition and Behaviour Lab at Aarhus University and, during the COVID-19 lockdown, also from the general student population in Aarhus, Aalborg, Odense, and Copenhagen (the four largest Danish cities). In total, we recruited 499 subjects. Of these, 397 completed the date 0 part of the study, and 326 reached the free work part at date 1.[7] Crucially, there are no treatment differences in completing date 0 (80.8 percent in *Late* vs. 79.1 percent in *Early, Revise0*, and *Revise1*), and we find no evidence for selection on observables into the free work phase (see Online Supplement S. 4). A total of 277 subjects also completed the post survey (date 2), which we primarily use to test alternative mechanisms or for exploratory research.

Out of the 397 subjects who completed the date 0 part, 236 (60 percent) were bachelor students, 83 (21 percent) were master students, 13 (3 percent) were PhD or other types of students, and 65 (16 percent) were not students. Most students came from the largest study programs; Business and Economics (148 subjects, 37 percent of the sample). Subjects earned DKK 188 on average.

2.3 Procedures

We conducted all parts of the experiment online using the Qualtrics platform. When completing the consent form, subjects could select among a number of date 0,1,2 triplets for when to participate in the study. They then received an invitation email with a personalized link for the date 0 part of the study at midnight on the selected date. Similarly, subjects who completed date 0 (date 1) then received an email with access to the date 1 (date 2) part at midnight on the appropriate date. Subjects had to use a PC or tablet (access via smartphone was technically blocked). This was to enhance the feeling that the task is 'work'. To prevent participants from pasting table into a spreadsheet program to do the counting, we copy-protected tables.

We collected data in two time periods, the first running November-December 2019 and the second running March-May 2020. We implemented the break to ensure similar working conditions for all participants as exams took place in January and February. Subjects knew that

[7] Unfortunately, there was a technical error in the recording of the time schedule for 21 subjects in the *Late* treatment. In our primary specification, we thus exclude these subjects when controlling for whether subjects are time-constrained. Our results are qualitatively robust to including these subjects throughout or excluding them entirely.

they would receive payments 2-6 weeks after the experiment via a standard system that allows public bodies and companies to send money to people by means of their social security number. At date 0, we randomized subjects into either the *Late* treatment (with probability 1/4) or the other treatments (with probability 3/4). At date 1, we then randomized the latter subjects into either *Early, Revise0*, or *Revise1* with equal probabilities.

3 Theoretical Predictions

3.1 Model

Task and Preferences. We derive our predictions in a theoretical model where people have present-biased preferences (Laibson, 1997), and we incorporate (partial) naïvete (O'Donoghue and Rabin, 1999). In the absence of goal setting, the utility of self t (the incarnation of the individual at date $t \in \{0,1,2\}$) is $U_t = u_t + \beta \left[\sum_{\tau=t+1}^{T+1} \delta^\tau u_\tau\right]$, where u_t is the instantaneous utility. The individual faces a task that requires effort $e \in [0,\infty)$, causing immediate costs $c(e)$ (strictly increasing and strictly convex) and long-run benefits $b(e)$ (strictly increasing and concave). Thus, $u_0 = 0$, $u_1 = -c(e)$ and $u_2 = b(e)$. The present bias parameter $\beta \in (0,1]$ captures the extent to which the individual overemphasizes the immediate instantaneous utility relative to future instantaneous utilities. Without loss of generality, we set the exponential discount factor δ to one.

The present bias causes time inconsistency. For self 0, all costs and benefits are in the future. Hence, the optimal effort equates marginal costs and benefits:

$$b'(e_0^*) = c'(e_0^*). \tag{1}$$

Self 1 discounts future benefits by $\beta \leq 1$ but not the immediate costs. So, self 1 prefers effort such that

$$\beta\, b'(e_1^*) = c'(e_1^*). \tag{2}$$

Thus, a self-control problem arises because self 0 wants a higher effort than self 1: $e_0^* \geq e_1^*$.

Goals. To overcome this self-control problem, self 0 sets an effort goal g_0 at date 0. Following the literature, where we specifically build on the models of Koch and Nafziger (2016, 2020), we assume that a goal serves as a reference point in that the individual compares the actual effort e with the goal g. If the effort differs from the goal by z, the individual experiences a corresponding comparison utility $\mu(z) = z$ for $z < 0$, and $\mu(z) = 0$ for $z \geq 0$.

We allow self 1 to revise this goal at date 1 to g_1. We assume that an individual who (possibly) revises the goal at date 1 then compares the initial goal set at date 0, g_0, to the revised goal, g_1, and experiences comparison utility from this change. If the revised goal differs from the initial goal by z, the individual experiences a corresponding comparison utility $\nu(z) = \nu\, z$ for $z < 0$ and $\nu(z) = 0$ for $z \geq 0$. We assume $0 \leq \nu < 1$, implying that adjusting one's goal downward is psychologically less painful than failing to reach one's goal. The idea that changes in beliefs

about future outcomes are carriers of comparison utility, and the weighting of this comparison utility follows Kőszegi and Rabin (2009).

We assume that both the initial and the revised goals are 'sticky' in the sense that a combination of both goals enters the reference point to which the individual ultimately compares exerted effort. That is, the individual has g^* in mind with $g^* = \lambda^T g_0 + (1 - \lambda^T) g_1$, where λ^T and $1 - \lambda^T$ are the treatment-specific salience weights ($T \in \{Late, Early, Revise0, Revise1\}$) – see more on these weights below.

At the goal revision stage, the individual believes that he will evaluate performance against the reference point $\hat{g}^* = \hat{\lambda} g_0 + (1 - \hat{\lambda}) g_1$. To highlight the main driving forces, we assume for expositional purposes that at the goal revision stage the individual is myopic to the stickiness of the original goal g_0; therefore, he does not take it into account when setting the revised goal g_1. This amounts to assuming $\hat{\lambda} = 0$, i.e., the individual thinks he will compare effort to g_1. As shown in Online Supplement S. 1.1, the qualitative predictions of the model are robust to assuming correct anticipation of the stickiness of the original goal. Similarly, motivated by evidence on projection bias (cf. Loewenstein et al., 2003; Acland and Levy, 2015), we assume that self 0 is naïve about the possibility of revising goals. We discuss the implications of relaxing this assumption in Online Supplement S. 1.2.

Equilibrium. We assume that goals are 'quasi-rational'. Given his (erroneous) beliefs, the individual sets a goal that he believes he will achieve, and he chooses this goal level to maximize his expected utility. As mentioned above, we assume that (i) self 0 is naïve about the possibility of revising goal 0, (ii) self 1 (at the goal revision stage) is naïve about the stickiness of goal 0 (belief $\hat{\lambda} = 0$), and (iii) the individual might be fully or partially naïve about his present-bias ($1 \geq \hat{\beta} \geq \beta$). Consequently, the individual typically sets a goal that differs from the optimal goal under correct beliefs.

3.2 Main Hypotheses

Our theoretical framework yields a number of testable predictions, which we summarize in Table 2. The analysis behind the hypotheses is outlined in Appendix A. Here, we outline the main intuition.

Goal Setting. The present bias causes a wedge between the goals that the individual sets at date 0 compared to date 1. When setting a goal at date 0 (as in *Early*), self 0 weighs the future benefits and costs equally and thus wants a higher effort than self 1. Consequently, he sets a goal to counteract the present bias. Such a goal can motivate self 1 to provide more effort than he would in the absence of a goal because he fears suffering a psychological loss if he falls short of the goal. Note, however, that the goal cannot be too high – otherwise self 1 would deviate from it because then the gain in utility from a lower effort outweighs the loss from not reaching the goal. Thus, while the goal pushes self 1 to provide more effort than he

would in the absence of goal setting, self 0 may not always be able to implement his preferred effort – this depends on the extent of the present bias.

If the individual can only set a goal at date 1 (as in *Late*), the present bias makes him fully give in to his self-control problem. Thus, the goal in *Early* is larger than the goal in *Late*.

When the individual has the opportunity to revise his original goal at date 1, the individual discounts the future benefit with the present bias – in contrast to self 0. This is the case in *Revise0* and *Revise1* (and possibly privately in *Early*), where subjects can revise their goal before providing effort. Further, because of partial naïveté, self 0 might have set a goal that is too high in that it exceeds the highest effort that self 1 would be willing to provide. Both are reasons for revising the goal downward. Yet, lowering the goal triggers loss utility – similar to the loss one feels when failing to reach a goal, just that loss utility from goal revision possibly has less weight than the loss utility from actually falling short of the goal. As a consequence, the largest goal that is 'revision proof' is smaller than the largest goal that is implementable in the absence of goal revision. A direct implication is that the revised goal at date 1 is lower than the initial goal set at date 0. Yet, the revised goal is still higher than the effort that self 1 would choose in *Late*.

Summing up, we test for the wedge between goals set at date 0 and those set at date 1 both with a between-subject comparison (*Early* vs. *Late*) and a within-subject comparison (using *Revise0* and *Revise1*). Note, however, that subjects might become more productive from date 0 to date 1. This would imply that, mechanically, they set higher goals at date 1 than at date 0. We hence control for the productivities at dates 0 and 1, respectively, to test whether goals reflect self-regulation intentions.

Hypothesis 1 (Goals are self-regulation tools) *Controlling for the respective baseline productivities,*

1. *(Between-subjects) Goals set in Early are larger than goals set in Late.*
2. *(Within-subjects) Goals set in Revise0 and Revise1 are lower at date 1 than at date 0.*

Effort Provision. Higher goals translate into higher effort. If the individual only sets a goal at date 1 (as in *Late*), this goal is set at the preferred effort of self 1, and he then achieves this goal. Thus, when only setting a goal just before the task, the individual fully gives in to his self-control problem. In contrast, as both the initial and the revised goals in *Early, Revise0*, and *Revise1* are higher than the preferred effort of self 1, effort in these treatments should exceed the effort in *Late*. That is, goals are effective self-regulation tools – despite goal revision.

As effort may differ between *Early, Revise0*, and *Revise1* (see Hypothesis 3) and some treatments differ along several dimensions, we test the hypothesis that goals are effective self-regulation tools by making the following two comparisons. First, we compare effort between *Early* and *Late*. In both treatments, subjects are asked to set a goal only at a single date,

and they are later reminded about that goal. Second, to test if goals are effective despite explicit goal revision, we test whether the effort in *Revise1* exceeds the effort in *Late*. In both treatments, subjects are reminded about the goal they set at date 1, so treatment differences can only arise because subjects in *Revise1* set an initial goal at date 0 but those in *Late* did not.

Hypothesis 2 (Goals are effective self-regulation tools – despite goal revision)

1. *Goals are effective: Subjects provide more effort in Early than in Late.*
2. *Goals are effective despite explicit goal revision: Subjects provide more effort in Revise1 than in Late.*

If framing of goals and goal revision influences the salience of goals, then we should see differences in effort between treatments *Early, Revise0*, and *Revise1*. By Hypothesis 2.2 the initial goal should be larger than the revised goal; and by random assignment, and by the assumption that the individual is myopic about the possibility of goal revision, initial and revised goals should not differ between treatments *Revise0, Revise1* and *Early*.[8] Yet, the salience of the initial and revised goals may differ in these treatments and thereby effort. First, it seems plausible that the goal that is displayed while working on the task is the most salient in *Revise0* and *Revise1*. Second, making goal revision explicit in *Revise0* and *Revise1* results in greater salience for the revised goal than does the (possible) private goal revision in *Early*. From these two assumptions, we obtain that for the salience weights $0 = \lambda^{Late} \leq \lambda^{Revise1} < \frac{1}{2} < \lambda^{Revise0} < \lambda^{Early} \leq 1$.[9] The higher the salience weight on the (higher) initial goal, the higher the reference point g^*. As higher effective goals result in higher effort, we hence expect a higher effort in *Early* than in *Revise0*; and a higher effort in *Revise0* than in *Revise1*.

Hypothesis 3 (Framing can make goals more effective)

1. *Making goal revision explicit lowers effort: Subjects provide more effort in Early than in Revise0.*
2. *Making the initial goal salient increases effort relative to making the revised goal salient: Subjects provide more effort in Revise0 than in Revise1.*

Hypothesis 3 is not only informative about framing; it may also indicate to what extent private goal revision occurs in *Early*. If effort does not differ between *Early* and *Revise0*, and between

[8] This statement relies on the theoretical model. However, the explicit goal revisions in *Revise0* and *Revise1* may induce subjects to revise their goal differently (more often or to a greater extend) than subjects in *Early* do. The prediction below remains valid if this is the case.

[9] Our assumptions on these parameters are inspired by salience theory in social psychology (for an overview see Higgins, 1996) and the recent application of economic salience theory to memory (Bordalo et al., 2020). Salient information grabs the (limited) attention of the individual in that it receives disproportional weight relative to other pieces of information. A reminder makes an attribute salient (e.g., Karlan et al., 2016).

Table 2: Hypotheses and summary of findings

	Hypothesis	Finding	Effect size[a]
Goals are self-regulation tools			
H1.1	goal 0^{Early} > goal 1^{Late}	✓	.230
H1.2	goal $0^{Revise0,Revise1}$ > goal $1^{Revise0,Revise1}$ (within subject)	✓	.230
Goals are effective despite goal revision			
H2.1	effort^{Early} > effort^{Late}	✗	.169
H2.1+[b]	effort^{Early,Revise0,Revise1} > effort^{Late}	✓	.263
H2.2	effort^{Revise1} > effort^{Late}	✓	.399
Framing can make goals more effective			
H3.1	effort^{Early} > effort^{Revise0}	✗	-.075
H3.2	effort^{Revise0} > effort^{Revise1}	✗	-.146

[a] Standardized effect size Hedge's g_p.

[b] H2.1+ pools treatments with 'early' goal 0 as framing effects under Hypothesis 3 are rejected.

Revise0 and *Revise1*, while we observe that subjects revise their goal in *Revise0*, it suggests that the revised goal matters in *Revise0* and that subjects in *Early* revise their goal, and the revised goal is as salient as in *Revise0*.

4 Empirical Analysis

In the following, after describing the main variables and the analysis plan, we test our primary Hypotheses 1-3 by comparing effort and goals in the different treatments. We also comment on the robustness of the results in Section 4.4. In Section 5, we examine possible mechanisms and discuss the role of alternative explanations behind our results, such as alternative reference points and the role of learning, uncertainty and time shocks vs. present bias.

4.1 Main Variables and Analysis Plan

Our main outcome variables are *goal 0* (the goal set at date 0, except in *Late*; cf. Figure 1), *goal 1* (the goal set at date 1 in *Late* or the revised goal in *Revise0* and *Revise1*), and *effort*. Table 3 provides descriptive statistics of the average goals, effort, goal achievement, and baseline productivities in the different treatments.

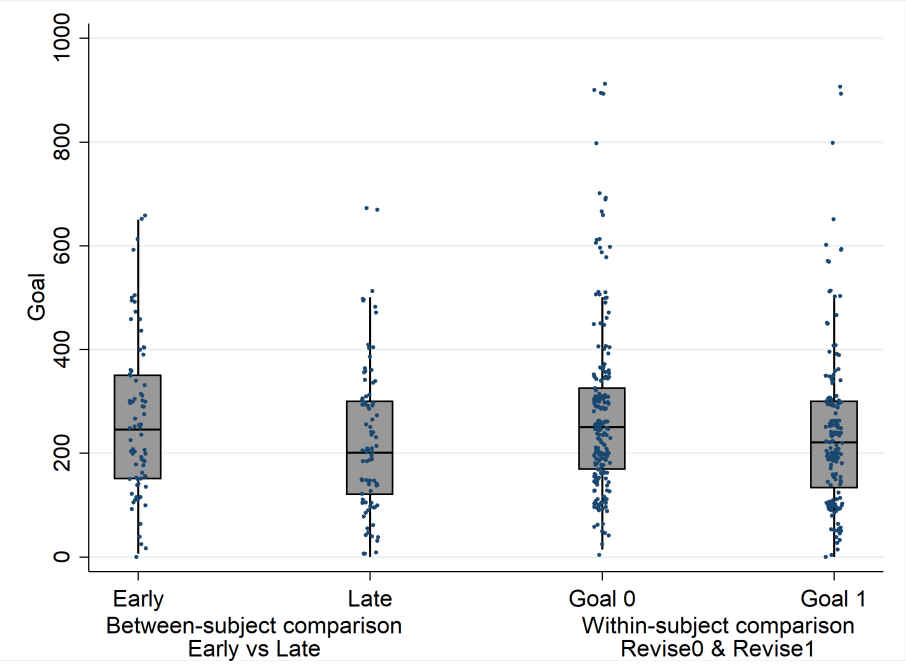
To test our hypotheses, we use OLS regressions (i) without control variables, (ii) with date-specific productivity measures as control variables, and (iii) with the full set of control variables specified in the pre-analysis plan (listed in Online Supplement S. 2). When effort is the outcome variable, we add specifications in which we control for the respective goals, both with and without other control variables. For robustness, we also use the nonparametric Mann-Whitney U-test (MWU) to examine differences in effort between treatments.[10] Throughout,

[10] A lack of control for productivity makes the MWU and Wilcoxon signed-rank tests ill-suited for comparing goals in Hypothesis 1.

Table 3: Descriptive Statistics

Treatment	N	Average			Share achieving		Average productivity	
		goal 0	goal 1	effort	goal 0	goal 1	date 0	date 1
Early	77	262.55	–	212.87	.58	–	14.44	17.25
Late	87	–	229.01	189.77	–	.67	14.60	17.39
Revise0	82	280.60	233.91	223.95	.59	.63	16.22	18.94
Revise1	80	274.57	249.95	246.39	.59	.74	15.68	18.24
All	326	272.77	237.35	217.72	.59	.68	15.23	17.95

Figure 3: Goals set by subjects



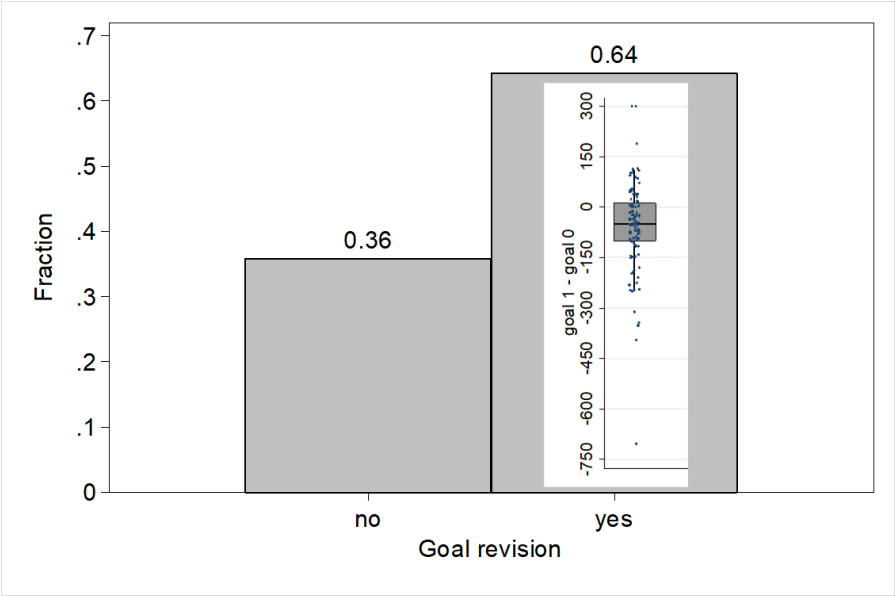
we report p-values for two-sided tests, and standardized effect sizes are summarized in Table 2. In Section 4.4, we perform a multiple hypothesis correction for the p-values.

4.2 Goal Setting (Test of Hypothesis 1)

In line with Hypothesis 1.1, the goal that subjects set in *Early* is on average 34 tables higher than the goal subjects set in *Late* (see also Figure 3). This difference is statistically significant once we control for the baseline productivity of subjects at the time of goal setting ($p = .01$, cf. Table 4). To understand why controlling for productivity is part of the hypothesis, despite random assignment to treatments, note that average productivity increases due to experience (cf. Table 3). This increase works against our prediction. Productivity explains 8 percent of the variance in goals between treatments.[11]

[11]The effect sizes for Hypothesis 1 in Table 3 are conservative estimates because Hedge's g_p does not take into account that subjects become more productive from date 0 to date 1.

Figure 4: Goal revision in *Revise0* and *Revise1*



Similarly, visual inspection of Table 3 and Figure 3 indicates that subjects in the *Revise0* and *Revise1* treatments revise their initial goal downwards at date 1. In these treatments, 64 percent of subjects revise their goal (cf. Figure 4) and for them the goal is on average revised downward by 56 tables. In line with Hypothesis 1.2, we observe in within-subject comparisons that *goal 1* is significantly smaller on average than *goal 0* ($p < 0.01$, cf. Specification (4) and (5) in Table 4).[12] Notably, there is some heterogeneity in goal revision. While 45 percent of the subjects revise their goal downwards (on average by 111 to a *goal 1* of 167), 36 percent of the subjects keep their early goal (average goal of 293), and 19 percent actually revise their goal *upwards* (on average by 73 to a *goal 1* of 320). These findings are robust to controlling for changes in productivity. It indicates that while most subjects have time-inconsistent goals, some people behave in a time consistent manner. Overall, the results support Hypothesis 1 that goals are used as self-regulation tools.

4.3 Effort Provision

Visual inspection of Table 3 and Figure 5 indicates that effort in *Late* is lower than effort in the other treatments, but there appear to be only few differences in effort between *Early, Revise0*, and *Revise1*. While the former pattern is in line with the view that goals are effective self-regulation tools (Hypothesis 2), the latter pattern goes against the predictions regarding

[12]Note that the intercept in Specification (6) in Table 4 has a different interpretation than in Specification (4) and (5). The intercept with full controls is not informative for the overall difference between the goals as it specifies the difference for a distinct baseline (including variables in a within-subject comparison that do not change between date 0 and date 1). Rather, it sheds light on possible mechanisms, which we return to in Section 5.

Table 4: Goal setting

	Hypothesis 1.1: Early vs. Late			Hypothesis 1.2: Revise0 & Revise1[a]		
	(1)	(2)	(3)	(4)	(5)	(6)
Late	-33.53	-58.47**	-63.97***			
	(22.90)	(22.44)	(23.24)			
Productivity[b]		8.45***	7.39***			
		(2.00)	(2.23)			
Change in productivity[c]					3.11**	2.79*
					(1.57)	(1.67)
Constant	262.55***	140.46***	-11.32	-35.79***	-44.01***	-164.29*
	(16.99)	(34.08)	(82.14)	(8.67)	(9.19)	(92.23)
Other controls[d]	No	No	Yes	No	No	Yes
N	164	164	143	162	162	162

Dependent variable: (1)-(3): goal (goal 0 for *Early* and goal 1 for *Late*); (4)-(6) goal 1 - goal 0.
Robust standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
[a] Within-subject comparison
[b] Baseline productivity at the date when the goal was set.
[c] Productivity 1 - productivity 0
[d] For the full specification, see Tables S.21 and S.22 in the online supplement. For 21 subjects, the time schedule was not available because of a technical problem (see Footnote 7)

Figure 5: Effort exerted by subjects in the four treatments

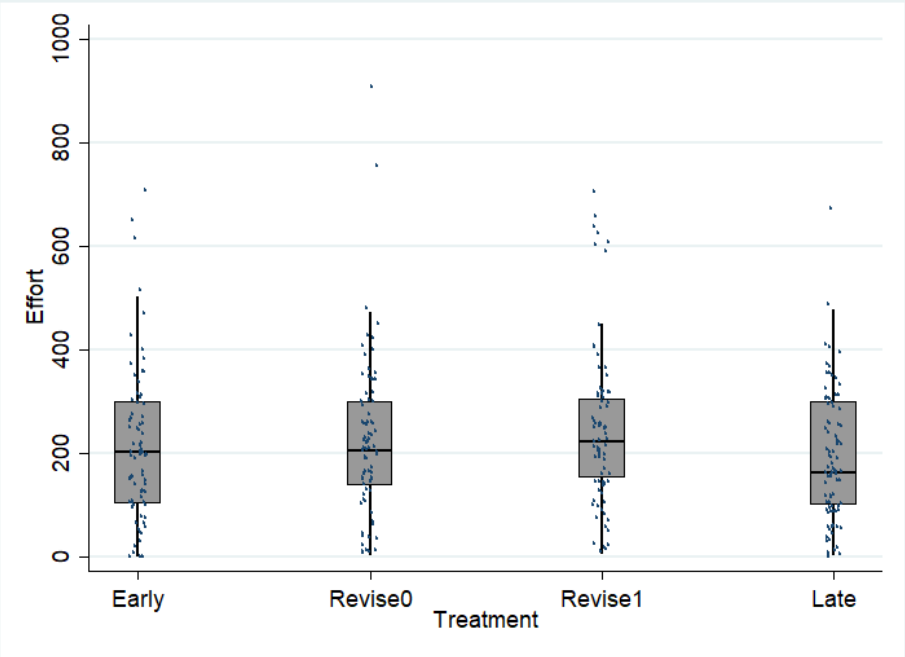


Table 5: Effort

No controls	Productivity	All controls
Hypothesis 2.1: Early vs. Late		
-23.10	-24.34	-31.24
(21.41)	(20.77)	(22.55)
Hypothesis 2.1+[a]: Late vs. Early & Revise0 & Revise1		
38.12**	31.53**	34.77**
(16.65)	(15.38)	(17.40)
Hypothesis 2.2: Late vs. Revise1		
56.62**	49.35**	51.51**
(22.10)	(20.39)	(21.14)
Hypothesis 3.1: Early vs. Revise0		
11.08	-3.44	-9.50
(23.43)	(22.51)	(22.28)
Hypothesis 3.2: Revise0 vs. Revise1		
22.44	28.46	29.78
(24.06)	(22.11)	(21.05)

Coefficients for the treatment mentioned last (with the treatment mentioned first as base category) in regressions with effort as dependent variable.

Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

[a] H2.1+ pools treatments with 'early' goal 0, as framing effects under Hypothesis 3 are rejected.

the framing of goal revision or goal reminders (Hypothesis 3). We test each of these in turn and report the results in Table 5.

4.3.1 Are Goals Effective Self-Regulation Tools? (Test of Hypothesis 2)

Regarding Hypothesis 2.1, we find that although effort is larger in *Early* than in *Late* (23 tables on average), this difference is not statistically significant ($p = .281$, cf. Table 5, MWU: $p = .405$). Subjects do, however, exert significantly more effort in *Revise1* than in *Late* ($p = .011$ cf. Table 5, MWU: $p = .023$), supporting Hypothesis 2.2.

As already indicated, we do not find any significant differences in effort between the treatments *Revise0, Revise1*, and *Early* (cf. Section 4.3.2), i.e., all treatments where subjects set a goal at date 0. Moreover, these treatments all have average efforts above treatment *Late*, i.e., above the treatment where subjects only set a goal at date 1. This observation prompts us to perform an additional (not pre-registered) analysis of Hypothesis 2.1, which also is in line with our theoretical model: We test the effectiveness of goals by pooling these three treatments.[13] If goals are effective despite goal revision, then effort in the pooled treatments *Early, Revise0*, and *Revise1* should exceed effort in *Late*. As seen in Table 5, the data support this prediction. Subjects in *Late* count on average 38 fewer tables, and this difference is statistically significant ($p = .023$ cf. Table 5, MWU: $p = .056$). With an effect size of Hedge's $g_p = .263$, this effect is economically significant as well, considering that goals are non-binding and that subjects perform a tedious task with no productive use and low intrinsic motivation. Moreover, this finding suggests that the comparison of *Early* and *Late* in Hypothesis 2.1 is imprecisely estimated.

Further, we observe that the coefficients for the *Late* treatment dummy become smaller and insignificant when goals are added to the specifications, suggesting that goal levels are the mediator (a Wald chi-square test for equality of the treatment dummy across models with and without goals yields $p = .058$, $p = .083$, and $p = .037$, respectively). This finding is predicted by the theory, according to which higher goals translate into higher effort and also consistent with earlier work on goal setting (cf, e.g., Koch and Nafziger, 2020).

4.3.2 Can Framing Make Goals More Effective? (Test of Hypothesis 3)

Regressions confirm the observation from Figure 5 that there are no differences in effort both between *Early* and *Revise0* and between *Revise0* and *Revise1* (cf. Table 5). Therefore, we reject Hypothesis 3. However, this rejection is not a threat to the overall hypothesis that goals are effective self-regulation tools. Rather, it informs on the occurrence of goal revision and the design of goal frames.

First, our findings suggest that subjects in *Early* privately revise their goal in the same way as subjects in *Revise0* and *Revise1* and that the revised goal matters for effort. Exploratory

[13] The additional test causes us to do multiple hypothesis correction for 7 rather than 6 tests in Section 4.4.

tests based on the responses from the post survey further confirm this view.[14] Among the 64 responses to the post survey in *Early*, 20 (31 percent) indicate that they privately revised their goal downward. On average, the subjects who adjust their goal do so by 62 tables, which explains almost all of their 66 table achievement gap relative to *goal 0*. In addition, the 44 subjects in *Early* who report no private revision exert effort statistically indistinguishable from their initial goal ($p = .150$). This evidence on private goal revision in *Early* is an important caveat to interpreting the effort-goal relationship in previous experiments.

Second, the results suggest that it is irrelevant what goal subjects are reminded about. While subjects pay attention to both goals, the more recent *goal 1* tends to matter more for subjects in both *Revise0* and *Revise1*. Specifically, across separate effort regressions for *Revise0*, we see that the coefficient on *goal 0* (.421) is borderline significantly smaller than the coefficient on *goal 1* (.686; Wald chi-square test for equality of coefficients across models, $p = .059$), and this also holds when adding controls ($p = .026$ and $p = .019$, respectively).[15] For *Revise1*, the coefficient on *goal 1* (.680) is larger than on *goal 0* (.618), but this difference is not statistically significant ($p = .681$; $p = .854$ and $p = .933$ when adding controls). Moreover, the recent *goal 1* appears to be equally important for subjects in *Revise0* and *Revise1*; reflected by an insignificant difference across treatments between the coefficients on *goal 1* ($p = .638$).

4.4 Robustness

We report several robustness tests in Online Supplement S. 5. Importantly, the result that setting an early goal leads to higher effort is qualitatively robust to using median regression, which is less affected by outliers than OLS. Considering other outcome variables, namely average mistakes or time spent per table, we find no difference between *Late* and the other treatments. However, these variables only correlate weakly to modestly with effort ($r = -.256$ and $r = -.449$, respectively), and this suggests that they might not be appropriate proxies for effort. For example, if a subject counts more tables, mistakes can easily go up due to fatigue. And subjects who spend more time per table may be more attentive to the task, with this not being affected by the goals set.

Multiple Hypothesis Testing. We perform 7 comparisons when testing our main hypotheses. Nevertheless, we present our findings in Section 4 without multiple hypothesis correction because the hypotheses are highly interdependent; also conceptually as Hypotheses 1 and 2 would both need to hold for the theory to be accurate (Hypothesis 3 concerns framing effects that are separate from the other hypotheses). The power for the joint comparison is below that for the individual tests. Thus, the experiment-wise type II error rate is inflated. Hence,

[14] Responses to the post survey should be interpreted with some caution. First, they could be influenced by self-justification. Second, not all subjects participated in the post survey. Third, most of the measures are not incentivized.

[15] Conceptually, one would regress effort on both *goal 0* and *goal 1*. However, such a test is hindered by collinearity of goal 0 and goal 1 ($r = .75$, $p < .001$). Instead, we compare coefficients across specifications.

some tests might turn up insignificant even though the theory is correct.

Here we investigate how the main hypotheses fare when correcting for multiple hypothesis testing. In doing so, we control the family-wise error rate (FWER) using the Holm-Šidák procedure (Šidák, 1967; Holm, 1979), and we control the false discovery rate (FDR) using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). FDR refers to the expected share of rejected null hypotheses that are wrongly rejected (type I errors), whereas the FWER is the probability of making *at least one* false rejection across *all* tests. Hence, controlling the FDR yields a higher power than the conservative FWER controlling method, but it does so by allowing more type I errors. As it is not obvious which of the two is most appropriate, we report both results in Table 6. Our main results remain at least borderline significant with one or both types of corrections.

Table 6: p-values corrected for multiple hypothesis testing

	Unadjusted	FWER-adjusted[a]	FDR-adjusted[b]
H1.1 goal 0^{Early} > goal 1^{Late}	0.0100	0.0587	0.0351
H1.2 goal $0^{Revise0\,\&\,Revise1}$ > goal $1^{Revise0\,\&\,Revise1}$	0.0000	0.0000	0.0000
H2.1 $\text{effort}^{Early} > \text{effort}^{Late}$	0.2423	0.4860	0.2827
H2.1+[c] $\text{effort}^{Early\,\&\,Revise0\,\&\,Revise1} > \text{effort}^{Late}$	0.0412	0.1548	0.0721
H2.2 $\text{effort}^{Revise1} > \text{effort}^{Late}$	0.0160	0.0777	0.0374

[a] Family-wise error rate (FWER) adjusted using the Holm- Šidák procedure, where $p_{adj} = 1 - (1 - p_{unadj})^{(8-\text{rank})}$, assigning the smallest p-value with rank 1.

[b] False discovery rate (FDR) adjusted using the Benjamini-Hochberg procedure, where $p_{adj} = p_{unadj} \cdot (7/\text{rank})$, assigning the smallest p-value with rank 1.

[c] H2.1+ pools treatments with 'early' goal 0, as framing effects under Hypothesis 3 are rejected.

5 Mechanisms

In the following, we consider possible mechanisms for why goals are effective self-regulation tools despite goal revision. Then, we test alternative mechanisms that could explain the results. Throughout, we often rely on the variable *goal achievement*, defined as the difference between a goal and effort. We also consider a binary goal achievement variable for *goal 0* and *goal 1* (equal to one if effort $\geq$ goal and zero otherwise), which provides the marginal effect that a treatment has on the probability of reaching a goal. Table 3 gives descriptive statistics.

5.1 Why Do Goals Work Despite Goal Revision? The Role of the Initial Goal

To examine more closely why goals are effective despite revision, we look at why setting an initial goal matters.

Does the the initial goal serve as a reference point in goal revision? In the theoretical model, the individual has the initial goal in mind and experiences loss utility if he revises the goal downward. Consequently, the model predicts that the initial goal pushes the revised goal up relative to the goal that the individual would set if he were to set a goal for the first time. That is, *goal 1*^{Revise} > *goal 1*^{Late}. However, we only find a small, positive difference across the treatments in the goals set at date 1 (*goal 1*^{Revise} = 241.83 and *goal 1*^{Late} = 229.01), and this difference is not statistically significant ($p = .562$, cf. Table S.29 in the Online Supplement). This result suggests that an individual experience no or a non-substantial loss utility when revising the goal, for example because the parameter ν in the model is close to zero.
Indeed, if setting an initial goal influenced effort entirely through a higher level of *goal 1*, then the treatment difference between *Revise1* and *Late* should disappear once we control for *goal 1*. Yet, subjects in *Revise 1* provide significantly more effort than subjects in *Late* even when controlling for *goal 1* (cf. Table S.26 in the Online Supplement).

How do the initial goal and the revised goal matter? The theory allows for another channel through which the initial goal impacts effort. Both the initial and the revised goals are assumed to be 'sticky' in the sense that the individual compares exerted effort with a linear combination of the initial goal and the revised goal. This predicts underachievement of the initial goal (*goal 0*) but overachievement of the revised goal (*goal 1*) – as long as the goals are not too high. We do find that subjects in *Revise0* and *Revise1* fall short of *goal 0*, namely by a statistically significant 43 tables on average ($p = .001$). But we find neither over- nor under-performance relative to *goal 1*. Subjects fall 7 tables short of their *goal 1* on average, but this is not significantly different from zero ($p = .470$).
Our findings thus indicate that the initial and revised goals do not enter as a linear combination to form a reference point, as assumed in the theoretical model. However, setting an initial goal does push-up effort and makes subjects more likely to achieve their revised goal even though it does not affect the revised goal. As noted above, subjects in *Revise0* and *Revise1* on average achieve their *goal 1*. In contrast, subjects in *Late* on average fall 39 tables short of their *goal 1* ($p < .001$).

5.2 Alternative Mechanisms

In the following, we test for some alternative mechanisms that could explain our results.

5.2.1 Alternatives to Present Bias as Explanations for Goal Revision

The result that *goal 0* is larger than *goal 1* is consistent with the explanation that individuals set a high goal ex ante to counteract the self-control problem that arises from their present bias. Our design allows us to examine alternative explanations for downward goal revision in the *Revise* treatments.

Uncertainty and Time Shocks. At both date 0 and date 1, we ask subjects to fill out their (expected) time schedule for date 1. This allows us to examine how unexpected time shocks and resolution of uncertainty about flexible time influence goal revision.[16] Notably, the downward goal revision remains highly statistically significant if we control for uncertainty and time shocks ($p < .001$, cf. Table S.30 in the Online Supplement) and if we restrict our analysis to those 87 percent of subjects who experience no time shocks ($p = .001$). In addition, uncertainty, time shocks, and difference in flexible time are not significant in any of the regressions, suggesting that these factors do not influence the wedge between the initial and revised goal.

Learning and overoptimism. Another potential mechanism that could drive a wedge between *goal 0* and *goal 1* is some sort of learning. *Learning about how to perform the task* goes against our result, because *goal 0* should be smaller than *goal 1* when the individual learns and gets better at the task as reflected by the productivity measures. There even is suggestive evidence that the difference between these two goals increases when we account for changes in productivity between dates 0 and 1 (Wald chi-square test for equality of constants across models, $p = .067$).

Learning about the cost of the task could play a role, however. When setting their initial goal, subjects may have insufficient experience with the task to anticipate how annoying or fatiguing it will become, or they may suffer from a projection bias. When setting their goal at date 1, they have more experience. To minimize the influence of such effects, subjects have only 3 more minutes of experience with the task when setting a goal at date 1 compared to date 0. Further, we have subjects perform the task immediately prior to setting both *goal 0* and *goal 1*. Also, previous research has found that experience does not affect goal setting or effort (cf. Koch and Nafziger, 2020).

To account for such learning effects, we proxy for learning by looking at subjects' (retrospective) enjoyment of the task reported in the surveys at dates 0 and 2, respectively.[17] Among

[16] Using the survey measures, we regard a subject with less than two hours of flexible time at date 1 as being severely time constrained. We have chosen two hours based on the idea that most subjects will work approximately one hour on the task – plus some time for the instructions and questions/goal setting. Based on this, we define that a subject faced a relevant negative (positive) time shock if he became (was no longer) severely time constrained when moving from planned to actual time schedule. Further, we use as our measure of uncertainty how likely subjects, at date 0, thought it was that they would have less than two hours of flexible time at date 1.

[17] A caveat is that pleasure in the task in the post survey is possibly influenced by other factors like satisfaction

the subjects who completed the post survey at date 2, we find that enjoyment declines from 3.34 to 3.12 from date 0 to date 2 (on a 5-point Likert scale), and this difference is statistically significant ($p < 0.001$ for both t-test and Wilcoxon signed-rank test). However, the change in enjoyment has no statistically significant effect on goal revision ($p = .179$, and $p = .247$ when controlling for change in productivity, cf. Table S.31). Thus, overall, our results go against these alternative explanations for the observed downward revision of the goal.

The individual also may set a higher goal at date 0 than at date 1 because goals reflect expectations and the individual is more optimistic at date 0 than at date 1 about what he can accomplish. Yet, the increase in productivity (and that subjects are informed about their productivity and can relate their goal to it with the slider tool) together with the insight that goal revisions do not relate to changes in enjoyment suggest against this view.

5.2.2 Reference Point Formation

Alternative Reference Points. The individual might be guided by other reference points than the elicited effort goal. In particular, a possible problem in inducing effort-based reference points through goals arises if people have time-based reference points. For example, if an experiment was set to last 1 hour, this time might be as salient/important as the effort goal. To minimize the effect of such time-based reference points, we provide subjects in the informed consent form only with information about broad time intervals ("the total time for participating in this study is between 60 and 240 minutes"). Yet, to help people set realistic goals, our slider tool at the goal setting stage translates effort into a time equivalent (based on baseline productivity at the date when the slider is used) and an earnings equivalent.

One way to see if subjects have a time or earnings reference point is to exploit people's tendency to focus on numbers that are 'round' in some way. Pope and Simonsohn (2011) and Allen et al. (2016) document a round number bias in goal setting across different contexts. In line with this insight, we observe bunching of effort goals on round numbers such as 100, 150, 200, etc. (cf. Figure S.8 in the Online Supplement). Yet, we see no clear bunching on focal numbers in the time or earnings dimensions (cf. Figures S.9 and S.10). This suggests that subjects indeed primarily focus on effort goals.

Private Goal Setting at Date 0 in Late. One possible concern is that subjects in *Late* already at date 0 form expectations (goals) about the task and effort. At date 0, subjects in *Late* do not know the tasks they are to perform at date 1. But for ethical reasons (informed consent and non-deception), we could not avoid all information, so subjects know that there will be *some* task. It is unlikely that subjects in *Late* guess what the task at date 1 is and set a private goal at date 0 – a claim that the data from the post survey supports. Here, only 8.6 percent of the subjects in *Late* indicate that they had a goal in mind at date 0, and the stated goals are virtually identical to their *goal 1*; ranging from 9 tables *fewer* than *goal 1* to

with the entire experiment, own performance/goal achievement, the payment obtained, and earnings per hour.

no difference (on average 1.5 tables fewer than *goal 1*).

5.2.3 Are Goals Rational?

While goals are effective on average, a share of people do not achieve them – neither in our experiment (cf. the descriptive statistics in Table 3) nor in other goal-setting studies. The theoretical model explains such goal non-achievement with (private) goal revision, and our results indeed show that such revision does take place. Another possible reason for goal non-achievement is that goals are non-rational. The (quasi-)rationality assumption in our model implies that subjects set goals that they believe they will achieve, and that they maximize their utility when setting goals.

As a first test of rationality, we examine the goals set by subjects right before starting to work on the task in *Late*. Unlike goals set at an earlier date, (i) they reflect the present bias of subjects (cf. Section 4.2), and (ii) uncertainty should not play a noticeable role because meaningful shocks to the free time available are unlikely at this stage. Thus, subjects in *Late* should achieve their goal if goals are rational. We observe that subjects in *Late* on average fall short of their goal by a statistically significant 39 tables ($p < .001$). Focusing on the 33 percent of subjects in *Late* who fail their goal (cf. Table 3), the average shortfall is 158 tables ($p < .001$).

As a second test of rationality, we consider goal achievement in the *Revise* treatments. For individuals who do not revise their goal (i.e., *goal 1=goal 0*), rationality predicts goal achievement because subjects would otherwise have adjusted their goal. In line with this, the 36 percent of subjects who do not revise their goal in *Revise 0* and *Revise 1* do not significantly deviate from their goal (12 tables shortfall on average; $p = .387$).[18]

Thus, the overall picture on the heterogeneity in goal achievement is that the majority of subjects achieve their goal, consistent with setting rational goals, but that a fraction of subjects fall short of their goal by a large margin. A possible explanation is that the latter are less deliberate when setting goals (see the discussion of 'Reflective Goal Setting' in Section 6). Another possible explanation is that some subjects deliberately set unrealistically high goals – often referred to as 'stretch goals' (see, e.g., Sitkin et al., 2011) – to push themselves to the edge.

5.2.4 Increased Goal Commitment Through More Frequent Goal Setting?

We observe significant difference in effort between treatment *Late* and *Revise1*, while the difference in effort between *Late* and *Early* is not significant. The latter result prompted us to induce an additional test – confirming our hypothesis that setting a goal early leads to a

[18] Note that if we consider all subjects in *Revise0* and *Revise1*, *goal 1* is on average achieved (7 tables shortfall, $p = .470$). Yet, the reasons for goal achievement in the two cases might be different. Subjects for whom *goal 0* > *goal 1* may aim to achieve to some extend also their higher initial goal – a driving force that is not present when *goal 0* = *goal 1*.

higher effort than setting it late. Yet, another interpretation for the observed pattern is that setting a goal twice (as in *Revise1*) in contrast to once (as in *Early* and *Late*) matters. Setting a goal twice could, for example, increase goal commitment.

We use the post survey to examine this channel. Here, we can identify 20 participants in *Early* who also set goals twice because they privately revised their initial goal; the remaining 44 participants in *Early* who completed the post survey set only one goal as they did not revise their goal. Regressing effort on a dummy for having set a goal twice yields an insignificant coefficient ($\beta = 9.33$, $se = 37.98$).

6 Discussion

In the following, we discuss some design choices and caveats of our study.

Motivation. Our study might underestimate the effect of goals because the real-effort task in our experiment does not have any productive use and subjects are likely to have low intrinsic motivation for it. Indeed, only 10 percent of the subjects report to like the task "a great deal" at date 0, and this drops to just 5 percent in the post survey. Further, subjects at date 1 receive already some earnings from the productivity measure and a show-up fee, which might further decrease their incentives to work at date 1.

It is plausible that goals have a stronger effect for tasks that are perceived as meaningful, either because individuals are intrinsically motivated for such tasks or because they are important for the individual in other ways (e.g., career goals). Furthermore, the effects of goals might be greater if one introduces extrinsic motivation, e.g., by making payment conditional on subjects reaching their goal as, for example, in Kaur et al. (2015).

Recalling vs. Revising. We remind subjects about the goals they set at date 0 before giving them the possibility to revise their goal at date 1. Clearly, such a reminder might serve as an anchor and work against our hypothesis that the revised *goal 1* is lower than *goal 0*. We opted to remind subjects because when not doing so one conflates measuring an intention to revise the goal with measuring whether subjects can remember their goal.

Our data provide some information on what subjects recall. While the most recent goal always appears to be salient, reminders seem to matter for how well subjects recall their initial goal. In the post survey, we have an incentivized question that asks subjects to recall their goals (denoted *recalled goal*). We measure the absolute error in recall as $|\textit{recalled goal } t - \textit{goal } t|$, $t \in \{0, 1\}$. In *Revise1*, subjects recall *goal 0* with less accuracy (mean error of 28.94) than *goal 1* (error of 8.53; t-test: $p = .044$). In contrast, for subjects in *Revise0*, there is no significant difference between the recall for *goal 0* (error of 27.39) and *goal 1* (26.93; t-test: $p = .953$).

Announcing Goal Revision. We do not announce at date 0 that subjects may revise goals at date 1. As we show in Online Supplement S. 1.2, one can extend our theory to allow for

anticipation of goal revision. The predictions of our main and such an extended setting are qualitatively similar. Yet, announcing goal revision in *Revise0* and *Revise1* makes comparisons between *Revise0* and *Early* potentially less straightforward, which complicates testing our main hypotheses.[19]

No Reminder of Goal Progress When Working. While working, subjects see both their completed number of tables and their goal. While paper reminders or numerous apps help individuals keep track of their goal achievement, in some tasks such self-monitoring may be more difficult or require additional effort. Yet, our finding in Section 4.3.2 that the displayed *goal 0* in *Revise0* matters less for effort than the non-displayed, revised *goal 1* suggests that the goal reminder may not play a great role.

Not Announcing the Task in *Late*. We did not announce the task at date 0 in *Late* to avoid private goal setting. But we announced it in the other treatments. This procedure may, for example, influence attrition or give subjects in *Early, Revise0*, and *Revise1*, but not in *Late*, the possibility to practice the task. Yet, we do not observe treatment differences in attrition (see Online Supplement S. 4). Also, there is no significant difference in how much subjects improve on the task between dates 0 and 1 between *Late* and the other treatments ($p = .861$).
Further, subjects in *Late* learn at a later date than the subjects in the other treatments that they will be reminded about their goal. This implies that anticipation of psychological losses from not reaching the goal may be more immediate for subjects in *Late*. Theoretically, this implies (if anything) that the highest goal that can potentially be implemented is larger in *Late* than in the other treatments – going against our findings.

Reflective Goal Setting. We failed to confirm Hypothesis 3 that framing can make goals more effective. Another approach for future research is to encourage individuals to take more time when setting goals. Doing so may induce individuals to reflect more upon their goals. We find that subjects who spend above median time setting their goal in *Early* and *Late* are 21 percentage points more likely to achieve it compared to those below the median (average partial effect from logit regression; $p = .005$, and $p = .006$ when adding controls). One possible reason for this result is that subjects may set more realistic goals if they are more deliberate when setting their goals – and this in turn makes them more likely to achieve their goals. In line with this idea, the psychology literature demonstrates that making "if-then" plans,

[19] Announcing goal revision has the advantage that it allows, in theory, to test the role of naïveté vs. sophistication for goal revision. If goal revision is announced, sophisticated individuals should never revise their goals as they already make the initial goal 'revision proof'. In contrast, naïve individuals may set an initial goal that is too high and may therefore revise it. Yet, these very subtle effects may have little practical relevance. In a field experiment where tutors encouraged students to set goals for academic performance, van Lent (2019) finds that making students aware of an explicit opportunity to revise goals did not significantly affect goal levels, assessed realism of goals, and motivation to achieve them; and students were just as likely to revise their goal.

also called implementation intentions, increases goal achievement (cf. Gollwitzer and Sheeran, 2006).

Women. As mentioned before, several studies find that goals are more effective at motivating men than women. To maximize power for a given budget, we therefore use a sample of men to study whether goals are effective in the light of goal revision. For women, a different focus is more interesting: Rather than asking whether goals stay effective, it is interesting to first understand why goals are less effective for women compared to men. To investigate this, many different mechanisms need to be tested. We consider this question to be an interesting avenue for future research.

7 Conclusion

In this study, we provide a clean test of whether individuals use self-set, non-binding goals as self-regulation tools, and whether goals are effective even though individuals may revise their goal. Our experiment addresses potential confounds of private goal setting and goal revision. We show that subjects do indeed use goals as self-regulation tools in the sense that they set a higher goal a few days before the task than immediately before the task. Many subjects revise their initial goal (also when not asked to do so). Nevertheless, goals are effective self-regulation tools because subjects who set a goal a few days in advance of the task exert more effort compared to subjects who only set a goal at the start of the task.

The effect sizes that we document are remarkable when taking into account that we provide no extrinsic motivation for subjects to achieve their (non-binding) goal, and subjects perform a tedious task with no productive use. This suggests that self-regulation through goal setting might have even more economic relevance in settings where people actually care about what they accomplish with their work.

References

Abeler, J., A. Falk, L. Götte, and D. Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

Aberson, C. L. (2019): *Applied Power Analysis for the Behavioral Sciences*, Routledge.

Acland, D. and M. R. Levy (2015): "Naiveté, Projection Bias, and Habit Formation in Gym Attendance," *Management Science*, 61, 146–160.

Akina, Z. and E. Karagozoglub (2017): "The Role of Goals and Feedback in Incentivizing Performance," *Managerial and Decision Economics*, 38, 193–211.

Allen, E. J., P. M. Dechow, D. G. Pope, and G. Wu (2016): "Reference-Dependent Preferences: Evidence from Marathon Runners," *Management Science*, 63, 1657–1672.

Apicella, C. L., E. E. Demiral, and J. Mollerstrom (2017): "No Gender Difference in Willingness to Compete When Competing Against Self," *American Economic Review*, 107, 136–40.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working over Time: Dynamic Inconsistency in Real Effort Tasks," *The Quarterly Journal of Economics*, 130, 1067–1115.

Augenblick, N. and M. Rabin (2019): "An Experiment on Time Preference and Misprediction in Unpleasant Tasks," *Review of Economic Studies*, 86, 941–975.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal statistical society: series B (Methodological)*, 57, 289–300.

Bordalo, P., N. Gennaioli, and A. Shleifer (2020): "Memory, Attention, and Choice," *Quarterly Journal of Economics*, 135, 1399–1442.

Brookins, P., S. J. Goerg, and S. Kube (2017): "Self-chosen goals, incentives, and effort," mimeo, Max Planck Institute for Research on Collective Goods.

Campion, M. A. and R. G. Lord (1982): "A Control Systems Conceptualization of the Goal-setting and Changing Process," *Organizational Behavior and Human Performance*, 30, 265–287.

Card, D. and G. B. Dahl (2011): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics*, 126, 103–143.

Cettolin, E., K. Cole, and P. Dalton (2020): "Developing Goals for Development: Experimental Evidence from Cassava Processors in Ghana," *Working Paper, Tilburg University.*

Clark, D., D. Gill, V. Prowse, and M. Rush (2017): "Using goals to motivate college students: Theory and evidence from field experiments," Tech. rep., National Bureau of Economic Research.

Clark, D., D. Gill, V. L. Prowse, and M. Rush (2019): "Using Goals to Motivate College Students: Theory and Evidence from Field Experiments," *Review of Economics and Statistics*, forthcoming.

Corgnet, B., J. Gómez-Miñambres, and R. Hernán-Gonzalez (2015): "Goal Setting and Monetary Incentives: When Large Stakes are not Enough," *Management Science*, 61, 2926–2944.

Corgnet, B., J. Gomez-Minambres, and R. Hernan-Gonzalez (2018): "Goal Setting in the Principal-Agent Model: Weak Incentives for Strong Performance," *Games and Economic Behavior*, 109, 311–326.

Dalton, P. S., V. Gonzalez, and C. N. Noussair (2015): "Paying with Self-Chosen Goals: Incentives and Gender Differences," CentER Discussion Paper Series No. 2016-036.

DellaVigna, S., A. Lindner, B. Reizer, and J. F. Schmieder (2017): "Reference-dependent job search: Evidence from Hungary," *The Quarterly Journal of Economics*, 132, 1969–2018.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, And Behavioral Consequences," *Journal of the European Economic Association*, 9, 522–550.

Donovan, J. J. and K. J. Williams (2003): "Missing the Mark: Effects of Time and Causal Attributions on Goal Revision in Response to Goal-performance Discrepancies." *Journal of Applied Psychology*, 88, 379.

Fan, J., J. Gómez-Miñambres, and S. Smithers (2019): "Make it too difficult, and I'll give up; let me succeed, and I'll excel: The interaction between assigned and personal goals," *Managerial and Decision Economics*.

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19, 25–42.

Gelman, A. and J. Carlin (2014): "Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors," *Perspectives on Psychological Science*, 9, 641–651.

Gill, D. and V. Prowse (2012): "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, 102, 469–503.

Goerg, S. J. and S. Kube (2012): "Goals (th)at Work [U+0096] Goals, Monetary Incentives, and Workers[U+0092] Performance," mimeo, Max Planck Institute for Research on Collective Goods.

Gollwitzer, P. M. and P. Sheeran (2006): "Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes," in *Advances in Experimental Social Psychology*, ed. by M. Zanna, New York: Academic Press, vol. 38, chap. 2, 69–119.

Gonzalez, V., P. S. Dalton, and C. Noussair (2020): "The Dark Side of Monetary Bonuses: Theory and Experimental Evidence," CentER Discussion Paper.

Goulet-Pelletier, J.-C. and D. Cousineau (2018): "A review of effect sizes and their confidence intervals, Part I: The Cohen's d family," *The Quantitative Methods for Psychology*, 14, 242–265.

Harding, M. and A. Hsiaw (2014): "Goal Setting and Energy Conservation," *Journal of Economic Behavior & Organization*, 107, 209–227.

Higgins, E. T. (1996): "Activation: Accessibility, and salience," *Social psychology: Handbook of Basic Principles*, 133–168.

Himmler, O., R. Jäckle, and P. Weinschenk (2019): "Soft Commitments, Reminders, and Academic Performance," *American Economic Journal: Applied Economics*, 11, 114–42.

Holm, S. (1979): "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 65–70.

Hsiaw, A. (2013): "Goal-Setting and Self-control," *Journal of Economic Theory*, 148, 601–26.

Ilies, R. and T. A. Judge (2005): "Goal Regulation across Time: The Effects of Feedback and Affect." *Journal of Applied Psychology*, 90, 453.

Jain, S. (2009): "Self-control and Optimal Goals: A Theoretical Analysis," *Marketing Science*, 28, 1027–1045.

Kaiser, J. P., A. K. Koch, and J. Nafziger (2021): "Results of the Pre-Analysis Plan: Are Self-Set Goals Effective Motivators? An Experiment," Working Paper, Aarhus University.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016): "Getting to the top of mind: How reminders increase saving," *Management Science*, 62, 3393–3411.

Kaur, S., M. Kremer, and S. Mullainathan (2015): "Self-control at Work," *Journal of Political Economy*, 123, 1227–1277.

Kőszegi, B. and M. Rabin (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99, 909–936.

Koch, A. K. and J. Nafziger (2011): "Self-regulation through Goal Setting," *Scandinavian Journal of Economics*, 113, 212–227.

——— (2016): "Goals and Bracketing under Mental Accounting," *Journal of Economic Theory*, 162, 305–351.

——— (2020): "Motivational Goal Bracketing: An Experiment," *Journal of Economic Theory*, 185, 104949.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–477.

Locke, E. A. and G. P. Latham (1990): *A Theory of Goal Setting & Task Performance*, Englewood Cliffs, NJ: Prentice Hall.

——— (2013): *New Developments in Goal Setting and Task Performance*, Routledge.

——— (2019): "The Development of Goal Setting Theory: A Half Century Retrospective." *Motivation Science*, 5, 93.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): "Projection Bias in Predicting Future Utility," *Quarterly Journal of economics*, 118, 1209–1248.

O'Donoghue, T. and M. Rabin (1999): "Doing It Now or Later," *American Economic Review*, 89, 103–124.

Pope, D. and U. Simonsohn (2011): "Round Numbers as Goals: Evidence from Baseball, SAT Takers, and the Lab," *Psychological science*, 22, 71–79.

Saccardo, S., A. Pietrasz, and U. Gneezy (2017): "On the size of the gender difference in competitiveness," *Management Science*, 64, 1541–1554.

Šidák, Z. (1967): "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions," *Journal of the American Statistical Association*, 62, 626–633.

Sitkin, S. B., K. E. See, C. C. Miller, M. W. Lawless, and A. M. Carton (2011): "The paradox of stretch goals: Organizations in pursuit of the seemingly impossible," *Academy of management review*, 36, 544–566.

Smith, A. (2019): "Lagged beliefs and reference-dependent utility," *Journal of Economic Behavior & Organization*, 167, 331–340.

Smithers, S. (2015): "Goals, Motivation and Gender," *Economics Letters*, 131, 75–77.

Suvorov, A. and J. van de Ven (2008): "Goal Setting as a Self-Regulation Mechanism," Working Papers w0122, Center for Economic and Financial Research (CEFIR).

Thakral, N. and L. T. Tô (2020): "Daily labor supply and adaptive reference points," *American Economic Review*.

Toussaert, S. (2016): "Connecting Commitment to Self-control Problems: Evidence from a Weight Loss Challenge," mimeo, NYU.

van Lent, M. (2019): "Goal Setting, Information, and Goal Revision: A Field Experiment," *German Economic Review*, 20, e949–e972.

van Lent, M. and M. Souverijn (2020): "Goal Setting and Raising the Bar: A Field Experiment," *Journal of Behavioral and Experimental Economics*, 101570.

A Analysis of the Theoretical Model

A.1 Goal Setting

The theory predicts that the present bias causes a wedge between the goal g_0 that the individual sets at date 0 and the goal g_1 that he sets at date 1. We start by characterising the goal g_0 that self 0 sets at date 0. As self 0 is myopic about the possibility of goal revision, this analysis is isomorphic to a situation where goals cannot be revised. To characterize the effort levels that self 0 believes he can achieve by setting an appropriate goal, we ask when he believes that his future self would have no incentive to deviate from goal g_0. For this, we consider the incentives of self 1 for a given goal g. If self 1 achieves the goal, i.e., provides some $e \geq g$, then his utility is $\beta\, b(e) - c(e)$. If self 1 fails the goal because $e < g$, then he suffers a loss and his utility is $\beta\, b(e) - c(e) - (g - e)$. Self 1 sticks to the goal if the utility from doing so exceeds the utility from falling short of it. This is the case if the latter utility is increasing in effort for any $e < g$. For this to hold, the goal must not be 'too high'; specifically, it must not exceed $e_{max}(\beta)$ defined by

$$\beta\, b'(e_{max}(\beta)) + 1 = c'(e_{max}(\beta)). \tag{3}$$

The *maximal achievable effort* $e_{max}(\beta)$, defined by Equation (3), is increasing in the present bias parameter β and exceeds the preferred effort of self 1, $e_{max}(\beta) > e_1^*$ (cf. Equations 2 and 3). This is because the fear of a loss makes self 1 strive harder than he would in the absence of comparison utility. Further, as there is no psychological gain from overachieving the goal ($\mu(z) = 0$ for $z > 0$), the lowest possible effort level that self 1 provides is his preferred effort level, e_1^*.

Self 0 understands the incentives of self 1 to stick to the goal. Yet, a partially naïve self 0 thinks that the present bias parameter is $\hat{\beta} > \beta$ and calculates the maximal achievable goal with this wrong estimate in mind. Consequently, he calculates the maximal achievable goal to be $e_{max}(\hat{\beta})$, which is strictly larger than the true maximal achievable goal $e_{max}(\beta)$. Thus, he picks his goal to maximize his utility $\beta\,[b(g_0) - c(g_0)]$ subject to $g_0 \in [e_1^*, e_{max}(\hat{\beta})]$. This gives $g_0^* = \min\{e_0^*, e_{max}(\hat{\beta})\}$ for treatments *Revise0, Revise1* and *Early*.

In contrast, if only self 1 was to set a goal, as it is the case in the *Late* treatment, he would set the goal that maximizes his utility, and he would achieve this goal, i.e., set $g_1^* = e_1^*$ which is smaller than the goal set at date 0, g_0^*.

A.2 Goal Revision

At date 1, in treatments *Revise0* and *Revise1* (and possibly also in *Early*), the individual first revises the goal and then provides effort. Both decisions reflect his true present bias because self 1 faces immediate effort costs. Yet, goal revision is constrained by the fact that the initial goal acts as a reference point. As noted above, lowering the goal triggers loss utility – similar to the loss one feels when failing to reach a goal, just that loss utility from goal revision has weight $\nu \in [0, 1)$. Hence, self 1 has no incentive to revise g_0 as long as it does not exceed the

revision-proof goal $g_{rev}(\beta)$ given by:

$$\beta\, b'(g_{rev}(\beta)) + \nu = c'(g_{rev}(\beta)). \tag{4}$$

Thus, the individual will only revise the goal in case the initial goal is 'too high'. From $\nu < 1$ it follows that the revision-proof goal is smaller than the maximal achievable goal; but it is still larger than the preferred effort of self 1:

$$e_1^* \leq g_{rev}(\beta) < e_{max}(\beta) \leq e_{max}(\hat{\beta}). \tag{5}$$

In sum, goal revision weakens the power of goals, but it does not completely hamper the ability of the individual to self-regulate. Specifically, self 1 will not revise the goal and provide the desired effort level of self 0 if $g_0^* = e_0^* \leq g_{rev}(\beta)$, but he will revise the goal if $g_0^* = \min\{e_0^*, e_{max}(\hat{\beta})\} > g_{rev}(\beta)$, in which case he sets $g_1^* = g_{rev}(\beta)$. Overall, we thus have $g_1^* = \min\{g_0^*, g_{rev}(\beta)\}$ – the revised goal at date 1 is equal or lower than the initial goal set at date 0.

A.3 Goal Achievement and Effort Provision

If the individual only sets a goal at date 1 (as in *Late*), then he achieves this goal, i.e., $e^* = g_1^* = e_1^*$. Thus, when only setting a goal just before the task, the individual fully gives into his self-control problem.

When setting a goal at date 0 (as in *Revise0, Revise1* and *Early*) then, when making the effort decision, the individual is reminded either about the initial goal g_0^* or the revised goal g_1^*, depending on the treatment. Recall that both the initial and the revised goals are 'sticky' in the sense that a combination of both goals enters the reference point g^* to which effort is compared, defined by $g^*(\lambda^T) = \lambda^T\, g_0^* + (1 - \lambda^T)\, g_1^*$, where the size of λ^T depends on the treatment as explained in the main text.

Self 1 provides effort to achieve the reference point g^* as long as the goal does not exceed the maximal achievable goal $e_{max}(\beta)$; otherwise effort is capped at $e_{max}(\beta)$. The latter case can arise when the individual is sufficiently naïve so that $g_0^* > e_{max}(\beta)$, and it is more likely to occur for higher values of λ. That is, $e^* = \min\{g^*, e_{max}(\beta)\} \geq e_1^*$. The equality only arises if the individual does not care about his initial goal ($\lambda^T = 0$) and the individual perceives no comparison utility from revising the initial goal ($\nu = 0$). Thus, having set an initial goal alleviates the self-control problem – even if the individual revises the goal.

Overall, the effort provided by self 1 lies between the initial goal set by self 0 and the revised goal: $g_1^* \leq e^* \leq g_0^*$. The individual (weakly) overperforms relative to g_1^* and (weakly) underperforms relative to g_0^*. More precisely, if the goal set at date 0 is sufficiently low so that $g_0^* \leq g_{rev}(\beta)$, the goal is not revised and $e^* = g_0^* = g_1^* = g^*$.[20] In contrast, if $g_0^* > g_{rev}(\beta)$, we have $g_1^* < g_0^*$ and e^* is an increasing function of λ, bounded between the two goals.

[20] Also, note that $g_1^* = e^*$ can occur in *Revise1* if $\lambda^{Revise1} = 0$, and $g_0^* = e^*$ can occur in *Early* if $\lambda^{Early} = 1$.

Online Supplement for Self-Set Goals Are Effective Self-Regulation Tools – Despite Goal Revision

January 13, 2021

S. 1 Extensions of the theoretical model

S. 1.1 Anticipation of the salience parameter

In the analysis in Section 3, we assumed that the individual does not anticipate the stickiness of his original goal when revising his goal at date 1. In the following, we demonstrate that we obtain the same predictions if we allow individuals to anticipate a salience parameter $\hat{\lambda}$ at the goal revision stage. To keep *Revise0* and *Revise1* comparable, subjects were told that they would be reminded about each goal with probability $\frac{1}{2}$, which suggests that $\hat{\lambda} = \frac{1}{2}$. In *Early*, goal revision is private so it is plausible to assume that $\hat{\lambda} = \lambda$.

S. 1.1.1 Goals

Maximal implementable goals at date 1. The individual believes that when providing effort he will face $\hat{g}^* = \hat{\lambda}\, g_0^* + (1-\hat{\lambda})\, g_1^*$, where g_0^* is the optimal goal set at date 0 and g_1^* the revised goal from date 1. Goals are quasi-rational, i.e., $e_1 = \hat{g}^*$. Thus, $\hat{g}^* \leq e_{max}(\beta)$ has to hold. Define

$$g_{max}(\beta, \hat{\beta}, \hat{\lambda}) = \frac{e_{max}(\beta) - \hat{\lambda}\, g_0^*}{1-\hat{\lambda}}.$$

This is the highest goal that can be set at date 1 such that self 1 believes he will not deviate from it when facing $\hat{g}^*$. By construction, as long as $g_0^* \leq e_{max}(\beta)$ we have $g_{max}(\beta, \hat{\beta}, \hat{\lambda}) \geq e_{max}(\beta)$. Further, $\frac{\partial g_{max}(\beta,\hat{\beta},\hat{\lambda})}{\partial \hat{\lambda}} \sim (e_{max}(\beta) - g_0^*)$. That is, whenever $g_0^* > e_{max}(\beta)$ (which can only arise if the individual is partially naïve), then $g_{max}(\beta, \hat{\beta}, \hat{\lambda})$ decreases in $\hat{\lambda}$. The individual sets $g_1^* = \min\{g_0^*, g_{rev}(\beta), g_{max}(\beta, \hat{\beta}, \hat{\lambda})\}$, where $g_{rev}(\beta)$ is defined as in Section 3.

Goal chosen at date 1. The individual faces the initial goal, $g_0^* = \min\{e_0^*, e_{max}(\hat{\beta})\}$. He revises g_0^* if $g_0^* > \min\{g_{rev}(\beta), g_{max}(\beta, \hat{\beta}, \hat{\lambda})\}$. If he revises, then he sets as new goal $\min\{g_{rev}(\beta), g_{max}(\beta, \hat{\beta}, \hat{\lambda})\}$. Thus, $g_1^* = \min\{g_0^*, g_{rev}(\beta), g_{max}(\beta, \hat{\beta}, \hat{\lambda})\}$. As long as $g_0^* \leq e_{max}(\beta)$, $g_1^* = \min\{g_0^*, g_{rev}(\beta), g_{max}(\beta, \hat{\beta}, \hat{\lambda})\} = \min\{g_0^*, g_{rev}(\beta)\}$ because $g_{rev}(\beta) < e_{max}(\beta)$ and, in this case, $g_{max}(\beta, \hat{\beta}, \hat{\lambda}) \geq e_{max}(\beta)$. Note that $g_0^* = e_0^*$ in this case. If $g_0^* > e_{max}(\beta)$ (which only arises if the individual is partially naïve), then $g_{max}(\beta, \hat{\beta}, \hat{\lambda}) < e_{max}(\beta)$ and $g_1^* = \min\{g_0^*, g_{rev}(\beta), g_{max}(\beta, \hat{\beta}, \hat{\lambda})\} = g_{max}(\beta, \hat{\beta}, \hat{\lambda})$ may arise if $\hat{\beta} - \beta$ is large enough, otherwise $g_1^* = g_{rev}(\beta)$.

Reference point at date 1. As in the main analysis, the individual is then reminded about *goal 0* or *goal 1*, depending on the treatment, and takes as references point g^* defined by: $g^* = \lambda\, g_0^* + (1 - \lambda)\, g_1^*$.

S. 1.1.2 Goal achievement and effort provision at date 1

Treatment *Early*. In *Early*, $\hat{\lambda} = \lambda$, so $\hat{g}^* = g^* \leq e_{max}(\beta)$ is always achieved: $e_1^{Early} = g^* = \lambda^{Early}\, g_0^* + (1 - \lambda^{Early})\, g_1^*$. As $e_1^{Early} = g^* \leq g_0^*$, the individual may underperform relative to his goal. How much he underperforms depends on the unobserved salience parameter.

Treatment *Revise0*. In *Revise0*, $\lambda > \frac{1}{2} = \hat{\lambda}$. Fixing g_1^*, it follows that $g^* > \hat{g}^*$. Suppose first $g^* \leq e_{max}(\beta)$. Then the individual provides $e_1 = g^*$. This case arises if $g_0^* \leq e_{max}(\beta)$ or if $g_0^* > e_{max}(\beta)$ and $g_1^* = g_{rev}(\beta) \leq g_{max}(\beta, \hat{\beta}, \lambda) < g_{max}(\beta, \hat{\beta}, \hat{\lambda})$ (note that for $g_0^* > e_{max}(\beta)$, $g_{max}(\beta, \hat{\beta}, \hat{\lambda})$ is decreasing in λ). In both cases, the individual underperforms relative to g_0^* (as $e_1 = g^* < g_0^*$) and overperforms relative to g_1^* (as $e_1 = g^* > g_1^*$).

Suppose next $g^* > e_{max}(\beta)$. Then the individual provides $e_2 = e_{max}(\beta)$, i.e., underperforms relative to g^*. This case arises if $g_0^* > e_{max}(\beta)$ and $g_1^* = \min\{g_{max}(\beta, \hat{\beta}, \hat{\lambda}), g_{rev}(\beta)\} > g_{max}(\beta, \hat{\beta}, \lambda)$. The individual underperforms relative to g_1^* and g_0^*.

Overall, in *Revise0*, the individual provides $e_1^{Revise0} = \min\{e_{max}(\beta), \lambda^{Revise0}\, g_0^* + (1 - \lambda^{Revise0})\, g_1^*\}$.

Treatment *Revise1*. In *Revise1*, $\lambda < \frac{1}{2} = \hat{\lambda}$. As $\lambda < \frac{1}{2} = \hat{\lambda}$, for fixed g_1^*, we have that $g^* < \hat{g}^* \leq e_{max}(\beta)$. Thus, $e_1^{Revise1} = g^* = \lambda^{Revise1}\, g_0^* + (1 - \lambda^{Revise1})\, g_1^*$. The individual underperforms relative to g_0^* (as $e_1 = g^* < g_0^*$) and overperforms relative to g_1^* (as $e_1 = g^* > g_1^*$).

Comparison of efforts. As $\lambda^{Revise1} < \lambda^{Revise0} < \lambda^{Early}$, we have that $e_1^{Revise1} < e_1^{Revise0} \leq e_1^{Early}$.

Comparison of goals. $g_1^{Revise0} = g_1^{Revise1} \leq g_0^*$, with equality for $e_0^* \leq g_{rev}(\beta)$.[1]

Comparison of goal achievement. For goal 0, we have that $g_0^* - e_1^{Revise1} > g_0^* - e_1^{Revise0} > g_0^* - e_1^{Early}$. Goals are achieved (and not revised) whenever $g_0^* = g_1^* = e_0^* < g_{rev}(\beta)$.
For goal 1, we have that $e_1^{Revise1} - g_1^* < e_1^{Revise0} - g_1^*$. Further, more subjects fail to achieve (in the sense of working more or equal) goal 1 in *Revise0* than in *Revise1*.

S. 1.2 Anticipated goal revision.

Self 0 selects $g_0^* = \min\{e_0^*, g_{rev}(\hat{\beta})\}$. Whenever $g_0^* \leq g_{rev}(\beta)$, self 1 will not revise the goal, otherwise he will revise it downward. Thus, when the individual is sophisticated, there will never be goal revision. In contrast, a partially naïve individual still may revise the goal because it was set too high as $\hat{\beta}$ was too optimistic.

S. 2 List of control variables

We use the following control variables.[2] As described above, some measures are controls for specific hypotheses, or are not included together in some analyses because they are likely to be collinear.

- Productivity – depending on the analysis, we use one of the following:
  - Baseline productivity 0 (from the first 3-minute round of mandatory work at date 0).
  - Baseline productivity 1 (from the first 3-minute round of mandatory work at date 1).
  - Change in productivity.

  As described under the hypotheses, baseline productivity at dates 0 and 1 allows to control for possible adjustment of goals to changes in productivity. To avoid collinearity issues, only the baseline productivity at date 0 (date 1 if appropriate) is included in the general analysis. The change in baseline productivity between date 0 and date 1 however allows us to assess some mechanisms (adjusting goals in response to learning about the task).

[1] Whenever $e_0^* \leq g_{rev}(\beta)$, then also $e_0^* \leq e_{max}(\beta)$ and so $g_{max}(\beta, \hat{\beta}, \hat{\lambda}) \geq e_{max}(\beta)$.

[2] We collected a number of other variables that were not included in our planned analyses. For example, we had no ex ante plan to control for the age and study area of participants (because we did not have an expectation that they would be strongly related to goal setting or effort behavior and because we, anyway, expected little variation in age and small cells for the subject area). Nevertheless, we collected this information because such background information about the subject pool in the experiment is of general interest.

- CRT: The number of questions the subjects answer correctly in the 3-item cognitive reflection test.
- Slider moved: A binary variable capturing interaction with the goal setting tool. We record whether the slider position in the tool was different from zero at page submission.
- Response time: Time until submission of the goal setting page. Unless otherwise noted, the measure uses the first time a goal is set.
- Self-competition: The percent allocated to the self-tournament pay option B measures self-competitiveness (based on Saccardo et al., 2017).
- Risk tolerance: Willingness to take risk question from Dohmen et al. (2011).
- Pleasure in task: The response to the question how much subjects like the task (Like a great deal (1) - Dislike a great deal (5)); from date 0 unless otherwise noted.
- Time constrained – depending on the analysis, we use one of the following:
  - Time constrained(P): Dummy = 1 if ≤ 2 hours of flexible time in the planned time schedule for date 1, reported at date 0.
  - Time constrained(A): Dummy = 1 if ≤ 2 hours of flexible time in the actual time schedule for date 1.
- Uncertainty: Perceived likelihood of being time constrained at date 1 (Extremely likely (1) - Extremely unlikely (5), reported at date 0.

Table S.20 provides summary statistics for key control variables.

S. 3 Power Analysis

The following analysis of the ex-ante power of our experiment drew on a pilot study of our reward schedule and previous (laboratory) experiments on goal setting.

Pilot study. Before conducting the experiment, we tested whether corner responses in effort and goals could be avoided by applying a declining piece rate for counting tables. We thus ran a pilot study with 28 subjects, testing the payment schemes finally implemented ($N_1 = 19$) and a variant of it with only slight differences ($N_2 = 9$). There was no goal setting in the pilot, and the 28 subjects counted on average 242 tables (standard deviation 150). 242 tables thus was our best guess of the average tables in *Late*. We had no prediction for how the standard deviation differs between treatments, so we simply assumed it to be 150 for all treatments.

Previous evidence. To get a view of what differences between the treatments could be expected, we drew on the related literature on goal setting that applies the same or similar real effort tasks. Firstly, using the same real-effort task as in our study, Koch and Nafziger (2020) look at the difference in goals and effort for subjects who set either a daily or a weekly goal. They find that subjects who set daily goals set higher goals (Effect size = .35, OLS) and provide more effort (Effect size = .42, OLS) than subjects who set weekly goals.[3] Secondly, in the original real-effort experiment involving counting zeros in tables, Abeler et al. (2011) pay subjects a fixed amount with probability .5 or based on a piece rate with probability .5. By varying the fixed payment (LO = 3 euros or HI = 7 euros, respectively), they induce different reference points. Thus, they find that subjects in the HI treatment count 46.33 tables on average ($SD = 25.25$) whereas subjects in LO count 37.05 tables on average ($SD = 25.07$), yielding an effect size of Hedges' $g_p = .37$.[4] Thirdly, in a within-subject comparison, Augenblick and Rabin (2019) examine preferences of subjects for the unpleasant task of transcribing blurry foreign letters either immediately or at future dates. Using their main sample of 68 participants (i.e., subjects without ML estimation issues), they obtain a correlation between preferences for immediate effort and effort 4-7 days into the future of .883,[5] and an effect size of Hedges' $g_D = .43$.[6]

Hence, it did not seem unrealistic to anticipate effect sizes between .3 and .4 (Hedges' g_p and Hedges' g_D for between- and within-subject comparisons, respectively). When considering effect sizes in the literature, however, we also recognize that it is often more likely to see overestimation than underestimation of population effects (see, e.g., Gelman and Carlin, 2014; Aberson, 2019).

The current study. For practical and financial reasons, it was only possible for us to recruit around 400 participants in total, i.e., 100 per treatment. For the between-subject comparison of subjects in *Early* and *Late*, we thus need an average difference of 60 tables (Hedges' $g_p = .40$) to obtain power of .8 in our main hypotheses (two-sided test, $\alpha = .05$, and $SD = 150$). Figure S.6 shows the relation between the power of this test, the sample size, and the difference in tables counted between the treatments.

When comparing goals within subjects in *Revise*, we hypothesize that subjects adjust their

[3] Effect sizes are calculated and reported as Effect size $= \frac{\text{Margin.effect(daily goals)}}{\text{Standard deviation in } \textit{Weekly} \text{ treatment}}$.

[4] Following Goulet-Pelletier and Cousineau (2018), we use Hedges' $g_p = \frac{M_2 - M_1}{S_p} \cdot J(\nu)$, where M_1 and M_2 are the means of effort decisions immediately and 4-7 days into the future, respectively, $J(\nu)$ is Hedges' correction factor, and S_p is the pooled standard deviation.

[5] The experiment involves multiple measurement for each individual for immediate and future effort (with varying number of observations for each individual), so the correlation is calculated using the average effort decision for each individual at $t = 0$ and $t \in \{4, 5, 6, 7\}$, respectively.

[6] Again following Goulet-Pelletier and Cousineau (2018), we use Hedges' $g_D = \frac{M_2 - M_1}{S_D} \cdot J(\nu)$, where S_D is the standard deviation of the differences. Note that this approach to standardizing the effect size of within-subject comparisons (Hedges' g_D) is not directly comparable to the above effect size of the between-subject comparison (Hedges' g_p) as the standard deviation of differences tends to be smaller than the pooled standard deviation. The comparable effect size is Hedges' $g_p = \frac{M_2 - M_1}{S_p} \cdot J(\nu) = .21$.

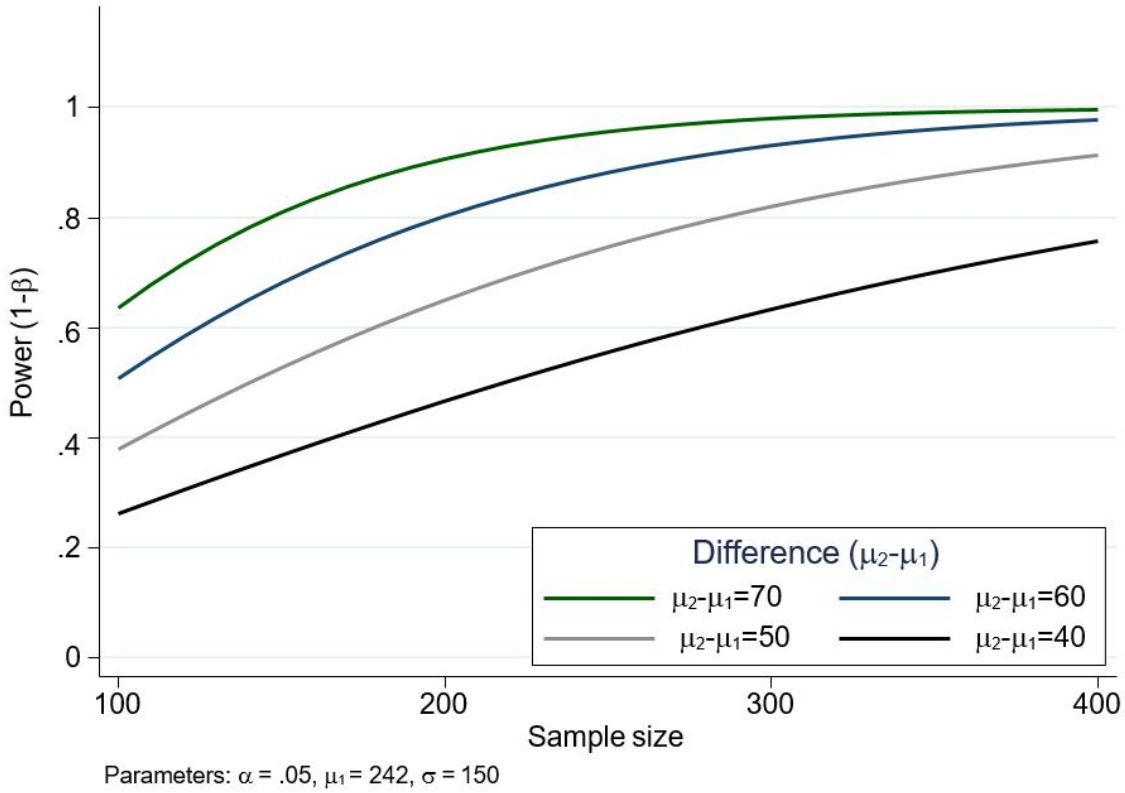


Figure S.6: Power for between-subject comparisons, two-sample t-test

goal downward at date 1. However, our theory builds on the notion that goals are sticky, and we therefore did not expect the difference to be as large as the difference between subjects in *Early* and *Late*. Furthermore, the assumption that goals are sticky implies that there is some positive correlation between the goals that subjects set at dates 0 and 1, but the size of this correlation was ex ante uncertain. In Figure S.7, we therefore examine the power for different samples sizes using both $\rho = .5$ and $\rho = .8$. In the case with $\rho = .5$ ($\rho = .8$), a sample size of 200 yields power of .8 to detect a difference of 30 (19) tables (two-sided test, $\alpha = .05$, and $SD = 150$), i.e., Hedges' $g_D = .20$.

As seen in Figure S.6 and Figure S.7, the expected effect size matters greatly for the ex ante power of our experiment. Note, however, that the figures do not account for the additional explanatory power provided by our control variables, some of which had previously been found to be statistically significant in other studies (e.g., Koch and Nafziger, 2020). So, the calculations above are conservative with regards to the power of our full model specification.[7] Finally, as explained in Section 2.2 of the paper, studies have found that goals are not as effective for women as for men (Koch and Nafziger, 2020; Smithers, 2015; Clark et al., 2017).

[7] We also test robustness of the results for the main hypotheses using non-parametric tests; a Mann-Whitney U-test for the between-subjects comparisons and a Wilcoxon signed-rank test for the within-subject comparisons. While the power of these tests depends on the specific data distributions, the tests do not perform much worse (assuming normality, for instance, both non-parametric approaches have asymptotic relative efficiencies of .955 compared to two-sample and paired t-tests, respectively).

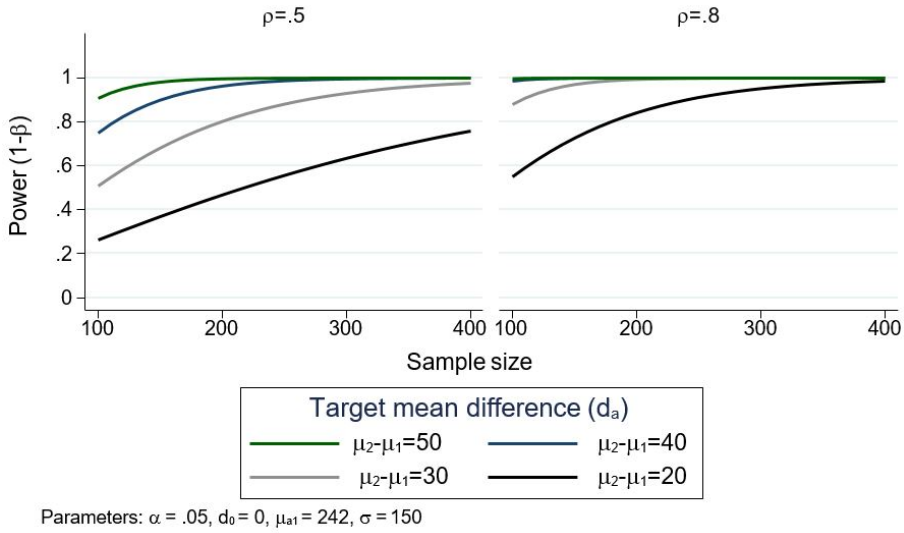


Figure S.7: Power for within-subject comparisons, two-sample paired t-test

For example, Koch and Nafziger (2020, Working Paper Version) find effect sizes of daily goals vs. no goals for women to be -0.1/-0.08. Thus, to achieve an appropriate power for the given budget, we only recruited men for the study.

S. 4 Analysis of Attrition

To examine whether there is systematic selection, we compare subjects who completed the date-1 and date-2 parts of the study with those who only completed the date-0 part. In Table S.9, we report the results of logit and linear probability models using as the dependent variable whether a subject completed the date-1 and date-2 parts. For the date-1 part, we fail to reject that the variables are jointly insignificant ($p = .331$ and $p = .297$, respectively), indicating no selection on observables. Similarly, for the date-2 part – conditioned on subjects completing the date-1 part, since this allows for a comparison of all treatments – we fail to reject that the variables are jointly insignificant ($p = .162$ and $p = .140$, respectively). However, economics and business students are around 13 percentage points more likely to complete the date-2 part, and people who enjoy the task of counting zeros more are more likely to complete the date-2 part (one point on the five-point Likert scale corresponds to 4 percentage points). Note that the result that the treatment does not influence selection into the work part is interesting in its own: It shows that setting goals does not make it more likely that people will show up for the task.

Table S.7: Attrition for Date 0

	Emails Sent for Date 0	Completed Date 0	Percentage
Early, Revise0, & Revise 1[a]	374	296	79.14%
Late	125	101	80.80%
Total	499	397	79.56%

[a] Randomization into Early, Revise0, & Revise 1 occurred at date 1.

Table S.8: Attrition for dates 1 and 2

	Emails Sent for Date 1	Completed Date 1	Percentage	Completed Date 2	Percentage
Early	98	77	78.57%	64	65.31%
Late	101	87	86.14%	70	69.31%
Revise0	99	82	82.83%	72	72.73%
Revise1	99	80	80.81%	70	70.71%
Total	397	326	82.12%	276	69.52%

Percentages reflect the share of subjects who completed the date-1 and date-2 parts out of the participants who received a link for the respective part.

Table S.9: Selection on observables

	Probability of completing				
	Date 1			Date 2	
	Logit	OLS		Logit	OLS
Late	0.069	0.064		-0.011	-0.015
	(0.05)	(0.04)		(0.06)	(0.06)
Revise0				0.052	0.053
				(0.05)	(0.05)
Revise1				0.038	0.039
				(0.06)	(0.06)
Productivity 0	-0.001	-0.001		-0.006	-0.006
	(0.00)	(0.00)		(0.00)	(0.00)
Uncertainty	-0.019	-0.020	0.004	0.003	
	(0.02)	(0.02)		(0.02)	(0.02)
Time-constrained(P)	0.027	0.026		-0.010	-0.009
	(0.07)	(0.07)		(0.07)	(0.07)
CRT	0.003	0.003		-0.017	-0.019
	(0.02)	(0.02)		(0.02)	(0.02)
Pleasure in task	0.014	0.014		0.037**	0.040**
	(0.02)	(0.02)		(0.02)	(0.02)
Risk tolerance	-0.021*	-0.020*		-0.002	-0.003
	(0.01)	(0.01)		(0.01)	(0.01)
Self-competition	0.001	0.001		0.001	0.001
	(0.00)	(0.00)		(0.00)	(0.00)
Share of Economics/Business	0.062	0.061		0.135***	0.126***
	(0.04)	(0.04)		(0.05)	(0.04)
Constant		0.865***		0.734***	
		(0.12)			(0.14)
N	397	397		326	326
Adj. R^2		0.00			0.02

Robust standard errors in parentheses.

Logit estimates are Average Partial Effects.

* $p < .10$, ** $p < .05$, *** $p < .01$.

S. 5 Robustness tests

In the following, we summarize the results of the alternative specifications with which we have tested the robustness of our main findings.

Hypothesis 1: Goals are self-regulation tools. In testing whether goals are different between *Early* and *Late* (H1.1), our primary specification uses OLS. Because OLS tests for differences in means, it is more sensitive to outliers, i.e., subjects who count particularly many tables. In Table S.10, we show that subjects in *Late* complete fewer tables than subjects in *Early* also when looking at a median regression, and that this becomes borderline significant when all controls are included.

In the date-0 part, subjects specified their expected time schedule for date 1. For 21 subjects, however, a technical error meant that subjects filled in their time schedule without the page showing them the specific day they had to complete date 1. As it is likely that not everyone recalled the day they chose while filling out the consent form, we exclude these 21 subjects from the main specification with all controls. In Table S.11, we show that the results to the regression without control and with productivity as the only control are qualitatively the same when these 21 subjects are excluded. In Table S.12, we furthermore show that the results are similar when we include the 21 subjects to the regression with the full set of controls.

In our main specification, we use the above mentioned time schedule to control for whether subjects were time-constrained. Nevertheless, we also elicit the number of exams and assignments that the subject needs to complete during the four weeks after date 0. In Table S.13, we show that subjects in *Late* also complete significantly fewer tables than in *Early* when we control for exams and assignments rather than whether subjects are time-constrained and that this holds for both the full sample and when excluding the subjects for whom the time schedule does not apply. Note that in these specifications, we only include students.

Hypothesis 2: Goals are effective self-regulation tools – despite goal revision. In Table S.14, we compare effort between treatments *Early/Revise* and *Late* using median regression instead of OLS. Again, we find that subjects in *Late* provide significantly lower effort than subjects in the other treatments, but that this effect becomes insignificant once we control for subjects' chosen goal (suggesting that goals are indeed the mediator).

Furthermore, we show that the results are qualitatively robust to excluding the 21 subjects for whom the time schedule is not available (Table S.15) and for including the entire sample (Table S.16). In Table S.17, we show that the results are also robust to using the number of exams and assignments rather than whether the subject is time-constrained, and this holds both when including and excluding the 21 subjects, respectively.

In Tables S.18 and S.19, we show that there is no effect when we use mistakes pr. table or time spent pr. table as proxies for effort.

Table S.10: Comparison of goals set in *Early* and *Late* using median regression

	(1)	(2)	(3)
Late	-45.00	-32.00	-52.94*
	(29.95)	(28.15)	(27.79)
Productivity		9.50***	5.58**
		(2.73)	(2.76)
CRT			-2.76
			(12.82)
Slider moved			74.95
			(61.67)
Response time			0.05
			(0.16)
Self-competition			-0.48
			(0.41)
Risk tolerance			12.99*
			(7.62)
Pleasure in task			27.94**
			(11.61)
Time-constrained(P)			32.07
			(41.79)
Constant	245.00***	85.00*	-57.67
	(21.82)	(44.04)	(95.38)
N	164	164	143
Pseudo R^2	0.01	0.06	0.13

* $p < .10$, ** $p < .05$, *** $p < .01$.

For 21 subjects, the time schedule was not available because of a technical problem (see Footnote 7).

Table S.11: Comparison of goals set in *Early* and *Late*, excluding in specifications (1) and (2) also the 21 subjects for whom the time schedule is not available

	(1)	(2)	(3)
Late	-41.45*	-67.03***	-63.97***
	(24.10)	(23.41)	(23.24)
Productivity		8.16***	7.39***
		(2.18)	(2.23)
CRT			-5.07
			(12.92)
Slider moved			60.04
			(53.93)
Response time			0.33**
			(0.14)
Self-competition			-0.44
			(0.33)
Risk tolerance			6.37
			(7.40)
Pleasure in task			22.08**
			(9.64)
Time-constrained(P)			28.39
			(32.38)
Constant	262.55***	144.68***	-11.32
	(17.00)	(36.57)	(82.14)
N	143	143	143
Adj. R^2	0.01	0.08	0.14

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

For 21 subjects, the time schedule was not available because of a technical problem (see Footnote 7).

Table S.12: Comparison of goals set in *Early* and *Late* using all who completed the date-1 part

	(1)	(2)	(3)
Late	-33.53	-58.47**	-49.43**
	(22.90)	(22.44)	(22.88)
Productivity		8.45***	7.65***
		(2.00)	(2.04)
CRT			-8.19
			(12.11)
Slider moved			84.76**
			(40.63)
Response time			0.29**
			(0.14)
Self-competition			-0.18
			(0.32)
Risk tolerance			-0.23
			(6.83)
Pleasure in task			20.33**
			(9.55)
Time-constrained(P)			22.05
			(29.12)
Constant	262.55***	140.46***	0.88
	(16.99)	(34.08)	(66.60)
N	164	164	164
Adj. R^2	0.01	0.08	0.12

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.13: Compare goals set in *Early* and *Late* using exams and assignments instead of time-constrained

	(1)	(2)
Late	-57.27**	-80.51***
	(23.65)	(23.35)
Productivity	5.72***	5.66**
	(2.01)	(2.23)
CRT	-6.61	-1.26
	(13.11)	(14.16)
Slider moved	72.85*	36.14
	(42.76)	(63.59)
Response time	0.16	0.16
	(0.14)	(0.14)
Self-competition	-0.12	-0.43
	(0.35)	(0.34)
Risk tolerance	0.90	11.18*
	(6.48)	(6.32)
Pleasure in task	20.09**	23.04**
	(9.94)	(9.79)
Exams & Assignments	9.20	8.45
	(8.12)	(7.66)
Constant	27.15	2.41
	(63.79)	(83.12)
N	136	118
Adj. R^2	0.10	0.15

Robust standard errors in parentheses.

These specifications use markedly smaller sample sizes since exams and assignments only apply to students.

Specification (1) reports data among all who completed the date-1 part whereas specification (2) reports data only among those for whom the time schedule applies.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.14: Comparison of effort between *Early/Revise* and *Late* using median regression

	(1)	(2)	(3)	(4)	(5)	(6)
Late	-40.00*	-41.22**	-33.34*	-7.64	-8.61	-13.67
	(20.98)	(17.64)	(19.11)	(15.11)	(14.63)	(16.53)
Productivity 1		7.61***	6.96***		3.29***	3.27***
		(1.35)	(1.39)		(1.15)	(1.22)
CRT			0.59			0.64
			(7.72)			(6.65)
Self-competition			0.13			0.04
			(0.24)			(0.21)
Risk			2.47			-2.04
			(4.29)			(3.70)
Pleasure in task			24.74***			7.71
			(7.43)			(6.54)
Time-constrained(A)			-11.00			-6.09
			(23.38)			(20.16)
Displayed goal				0.83***	0.69***	0.63***
				(0.04)	(0.04)	(0.05)
Constant	203.00***	77.09***	-19.46	35.55**	0.04	-0.76
	(10.84)	(26.20)	(46.94)	(13.86)	(22.57)	(40.37)
N	326	326	305	326	326	305
Pseudo R^2	0.01	0.06	0.09	0.25	0.27	0.27

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.15: Comparison of effort between *Early/Revise* and *Late*, excluding the 21 subjects for whom the time schedule is not available

	(1)	(2)	(3)	(4)	(5)	(6)
Late	-39.51**	-34.51**	-34.77**	-15.05	-14.01	-15.14
	(18.16)	(16.74)	(17.40)	(12.49)	(12.18)	(12.67)
Productivity 1		8.58***	7.95***		5.37***	5.23***
		(1.51)	(1.56)		(1.24)	(1.31)
CRT			10.01			6.14
			(7.17)			(5.90)
Self-competition			-0.07			0.01
			(0.25)			(0.20)
Risk			6.01			2.99
			(4.48)			(3.88)
Pleasure in task			24.44***			9.88
			(6.92)			(6.15)
Time-constrained(A)			-32.21*			-15.91
			(17.65)			(15.09)
Displayed goal				0.56***	0.52***	0.49***
				(0.08)	(0.08)	(0.08)
Constant	227.89***	72.10***	-47.98	78.93***	-5.97	-60.63
	(9.76)	(26.12)	(45.32)	(17.56)	(22.44)	(38.20)
N	305	305	305	305	305	305
Adj. R^2	0.01	0.12	0.16	0.36	0.40	0.40

Robust standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.16: Comparison of effort between *Early/Revise* and *Late* using all who completed the date-1 part

	(1)	(2)	(3)	(4)	(5)	(6)
Late	-38.12**	-31.53**	-31.10**	-18.22	-15.87	-16.36
	(16.65)	(15.38)	(15.68)	(11.98)	(11.72)	(11.92)
Productivity 1		8.57***	7.89***		5.28***	5.09***
		(1.45)	(1.48)		(1.20)	(1.24)
CRT			9.31			5.84
			(6.88)			(5.68)
Self-competition			-0.03			-0.00
			(0.24)			(0.19)
Risk			3.95			2.27
			(4.29)			(3.64)
Pleasure in task			24.17***			10.35*
			(6.70)			(5.74)
Time-constrained(A)			-32.50*			-13.88
			(16.67)			(14.44)
Displayed goal				0.56***	0.51***	0.49***
				(0.07)	(0.07)	(0.07)
Constant	227.89***	72.19***	-34.21	79.68***	-3.54	-53.61
	(9.76)	(25.12)	(44.47)	(16.51)	(21.52)	(36.99)
N	326	326	326	326	326	326
Adj. R^2	0.01	0.12	0.16	0.36	0.40	0.40

Robust standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.17: Comparison of effort between *Early/Revise* and *Late* using exams and assignments rather than time-constrained

	(1)	(2)
Late	-64.43***	-31.96**
	(16.09)	(12.69)
Productivity 1	7.51***	5.07***
	(1.75)	(1.42)
CRT	11.35	6.04
	(7.73)	(6.67)
Self-competition	0.03	0.09
	(0.27)	(0.22)
Risk	10.74**	5.12
	(4.45)	(3.83)
Pleasure in task	27.12***	12.86**
	(7.20)	(6.46)
Exams & Assignments	1.12	-2.82
	(5.84)	(5.46)
Displayed goal		0.47***
		(0.09)
Constant	-91.81*	-77.40*
	(48.38)	(42.01)
N	255	255
Adj. R^2	0.19	0.40

Robust standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.18: Comparison of effort between *Early/Revise* and *Late* using mistakes pr. table as proxy for effort

	(1)	(2)	(3)
Late	0.01	0.01	-0.02
	(0.03)	(0.02)	(0.01)
Productivity 1		-0.01***	-0.01***
		(0.00)	(0.00)
CRT			-0.00
			(0.01)
Self-competition			0.00
			(0.00)
Risk			0.00
			(0.00)
Pleasure in task			-0.01
			(0.01)
Time-constrained(A)			0.05
			(0.03)
Constant	0.16***	0.33***	0.32***
	(0.01)	(0.04)	(0.04)
N	324	324	303
Adj. R^2	-0.00	0.10	0.11

Robust standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.19: Comparison of effort between *Early/Revise* and *Late* using average seconds pr. table as proxy for effort

	(1)	(2)	(3)
Late	0.66	0.32	0.21
	(0.63)	(0.54)	(0.60)
Productivity2		-0.44***	-0.42***
		(0.05)	(0.05)
CRT			-0.64***
			(0.24)
Self-competition			0.00
			(0.01)
Risk tolerance			0.01
			(0.14)
Pleasure in task			-0.38
			(0.23)
Time-constrained2			0.01
			(0.75)
Constant	14.75***	22.82***	24.86***
	(0.34)	(1.06)	(1.51)
N	324	324	303
Adj. R^2	0.00	0.25	0.26

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

S. 6 Additional Tables and Figures

Table S.20: Summary statistics of control and key background variables

	Early	Late	Revise0	Revise1	Total
CRT	2.03	2.29	2.34	2.40	2.27
Self-Competition	63.75	60.13	55.28	65.20	61.01
Risk Tolerance	6.09	6.01	6.05	6.09	6.06
Pleasure in Task (Date 0)	3.25	3.24	3.39	3.30	3.29
Time Constrained (A)	0.14	0.11	0.15	0.13	0.13
Exams & Assignments	3.45	3.11	3.14	3.03	3.18
Age	24.70	24.60	24.02	25.26	24.64
Share of Econ/Business	0.34	0.31	0.44	0.46	0.39

Table S.21: Between-subject comparison of goals set in *Early* and *Late*

	(1)	(2)	(3)
Late	-33.53	-58.47**	-63.97***
	(22.90)	(22.44)	(23.24)
Productivity[a]		8.45***	7.39***
		(2.00)	(2.23)
CRT			-5.07
			(12.92)
Slider moved[b]			60.04
			(53.93)
Response time[c]			0.33**
			(0.14)
Self-competition[d]			-0.44
			(0.33)
Risk tolerance			6.37
			(7.40)
Pleasure in task			22.08**
			(9.64)
Time-constrained(P)[e]			28.39
			(32.38)
Constant	262.55***	140.46***	-11.32
	(16.99)	(34.08)	(82.14)
N	164	164	143
Adj. R^2	0.01	0.08	0.14

Dependent variable: goal (goal 0 for *Early* and goal 1 for *Late*)
Robust standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
[a] Baseline productivity at the date when the goal was set.
[b] Dummy = 1 if the slider tool on the goal setting page was used.
[c] Time spent on the page with the slider tool, truncated at 5 minutes.
[d] Percentage allocated to self-competitive option at date 0.
[e] Dummy = 1 if ≤ 2 hours of flexible time in the planned time schedule for date 1.

Table S.22: Within-subject comparison of goal 0 and goal 1 in *Revise0 & Revise1*

	(1)	(2)	(3)
Change in productivity		3.11**	2.79*
		(1.57)	(1.67)
CRT			4.98
			(9.05)
Slider moved, Date 0			61.98
			(73.87)
Slider moved, Date 1			2.47
			(33.22)
Response time, Date 0			0.13
			(0.15)
Response time, Date 1			0.18
			(0.14)
Self-competition			-0.28
			(0.27)
Risk tolerance			5.73
			(4.73)
Pleasure in task			-4.02
			(7.50)
Uncertainty			12.07
			(7.76)
More time			-17.32
			(36.76)
Less time			-18.63
			(28.32)
Constant	-35.79***	-44.01***	-164.29*
	(8.67)	(9.19)	(92.23)
N	162	162	162
Adj. R^2	0.00	0.01	-0.00

Dependent variable: goal 1 - goal 0.

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

[a] Productivity 1 - productivity 0

[b] Percentage allocated to self-competitive option at date 0.

[c] Belief how likely to have ≤ 2 hours of flexible time at date 1.

[d] Dummy = 1 if planned (actual) time schedule for date 1 had $\leq$ ($>$) 2 hours of flexible time

[e] Dummy = 1 if planned (actual) time schedule for date 1 had $>$ ($\leq$) 2 hours of flexible time

Table S.23: Comparison of effort in *Late* vs. *Early, Revise0,* and *Revise1*

	(1)	(2)	(3)	(4)	(5)	(6)
Late	-38.12**	-31.53**	-34.77**	-18.22	-15.87	-15.14
	(16.65)	(15.38)	(17.40)	(11.98)	(11.72)	(12.67)
Productivity 1[a]		8.57***	7.95***		5.28***	5.23***
		(1.45)	(1.56)		(1.20)	(1.31)
CRT			10.01			6.14
			(7.17)			(5.90)
Self-competition[b]			-0.07			0.01
			(0.25)			(0.20)
Risk tolerance			6.01			2.99
			(4.48)			(3.88)
Pleasure in task			24.44***			9.88
			(6.92)			(6.15)
Time-constrained(A)[c]			-32.21*			-15.91
			(17.65)			(15.09)
Displayed goal[d]				0.56***	0.51***	0.49***
				(0.07)	(0.07)	(0.08)
Constant	227.89***	72.19***	-47.98	79.68***	-3.54	-60.63
	(9.76)	(25.12)	(45.32)	(16.51)	(21.52)	(38.20)
N	326	326	305	326	326	305
Adj. R^2	0.01	0.12	0.16	0.36	0.40	0.40

Dependent variable: effort. Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

[a] Baseline productivity at date 1.

[b] Percentage allocated to self-competitive option at date 0.

[c] Dummy = 1 if $\leq$ 2 hours of flexible time in the actual time schedule for date 1.

[d] Goal that was displayed in the work part at date 1.

Table S.24: Goal achievement in *Revise0* and *Revise1*

	(1)	(2)	(3)	(4)	(5)	(6)
Productivity 1[a]		5.27**	5.52**		1.70	1.65
		(2.12)	(2.22)		(1.84)	(1.94)
CRT			8.54			6.42
			(14.45)			(11.74)
Self-competition[b]			-0.08			0.16
			(0.35)			(0.22)
Risk tolerance			11.54*			5.27
			(6.11)			(6.19)
Pleasure in task			-4.93			2.73
			(10.11)			(9.13)
Time-constrained(A)[c]			17.56			4.61
			(29.85)			(19.62)
Constant	-42.59***	-140.53***	-216.78***	-6.80	-38.46	-103.76
	(12.08)	(44.95)	(77.06)	(9.38)	(37.28)	(70.92)
N	162	162	162	162	162	162
Adj. R^2	0.00	0.03	0.03	0.00	0.00	-0.02

Dependent variable: (1)-(3) effort - goal 0, (3)-(6) effort - goal 1

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

[a] Baseline productivity at date 1.

[b] Percentage allocated to self-competitive option at date 0.

[c] Dummy = 1 if ≤ 2 hours of flexible time in the actual time schedule for date 1.

Table S.25: Goal achievement of goal 1 in *Late* vs. *Revise0* and *Revise1*

	(1)	(2)	(3)
Late	-32.44**	-30.91**	-23.99*
	(14.17)	(14.44)	(14.42)
Productivity 1[a]		1.27	1.20
		(1.35)	(1.47)
CRT			8.81
			(8.76)
Self-competition[b]			0.29
			(0.19)
Risk tolerance			1.81
			(4.81)
Pleasure in task			-1.89
			(6.93)
Time-constrained(A)[c]			12.02
			(14.91)
Constant	-6.80	-30.46	-73.68
	(9.39)	(28.49)	(53.17)
N	249	249	228
Adj. R^2	0.01	0.01	0.00

Dependent variable: effort - goal 1

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

[a] Baseline productivity at date 1.

[b] Percentage allocated to self-competitive option at date 0.

[c] Dummy = 1 if $\leq$ 2 hours of flexible time in the actual time schedule for date 1.

Table S.26: Comparison of effort across treatments

	No controls	Productivity	All controls	No controls and goal	Productivity and goal	All controls and goal
Early vs. Late	-23.10	-24.34	-31.24	-4.11	-6.45	-7.57
	(21.41)	(20.77)	(22.55)	(16.40)	(16.21)	(17.37)
Late vs. Early & Revise0 & Revise1	38.12**	31.53**	34.77**	18.22	15.87	15.14
	(16.65)	(15.38)	(17.40)	(11.98)	(11.72)	(12.67)
Late vs. Revise1	56.62**	49.35**	51.51**	44.76***	41.29*	39.43**
	(22.10)	(20.39)	(21.14)	(16.25)	(15.70)	(16.20)
Early vs. Revise0	11.08	-3.44	-9.50	0.86	-7.19	-9.92
	(23.43)	(22.51)	(22.28)	(19.85)	(19.64)	(20.25)
Revise0 vs. Revise1	22.44	28.46	29.78	39.79**	42.04**	41.79**
	(24.06)	(22.11)	(21.05)	(19.40)	(18.56)	(18.47)

Coefficients for the treatment mentioned last (with the treatment mentioned first as base category) in regressions with effort as dependent variable.
Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.27: Effect of *goal 0* in *Revise* treatments

	Revise0	Revise1	Revise
Goal 0	0.31***	0.53***	0.39***
	(0.09)	(0.11)	(0.07)
Productivity, Date 1	8.16***	9.02***	8.45***
	(2.91)	(2.11)	(1.73)
CRT	-4.28	32.07**	10.88
	(14.85)	(13.83)	(9.91)
Slider moved, Date 0	119.16	-123.32*	-29.39
	(103.68)	(69.27)	(57.20)
Response time, Date 0	0.36	0.39**	0.44***
	(0.25)	(0.19)	(0.15)
Self-competition	-0.32	-0.03	-0.20
	(0.45)	(0.42)	(0.29)
Risk tolerance	9.05	14.34**	13.37***
	(9.18)	(6.23)	(5.07)
Pleasure in task	13.28	5.56	11.64
	(16.07)	(13.02)	(10.09)
Time-constrained (A)	-20.28	16.76	-19.11
	(43.83)	(45.62)	(29.88)
Uncertainty	-11.49	-3.28	-7.36
	(14.40)	(12.18)	(9.00)
Constant	-217.14	-167.06	-165.08**
	(130.86)	(109.37)	(83.05)
N	82	80	162
Adj. R^2	0.30	0.47	0.37

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.28: Effect of *goal 1* in *Revise* treatments

	Revise0	Revise1	Revise
Goal 1	0.63***	0.56***	0.61***
	(0.08)	(0.10)	(0.06)
Productivity, Date 1	3.46	5.91***	4.89***
	(2.42)	(2.07)	(1.53)
CRT	-0.71	16.93	6.15
	(11.92)	(13.24)	(8.57)
Slider moved, Date 1	-16.24	-177.66***	-53.98
	(39.61)	(66.01)	(33.29)
Response time, Date 1	0.26	0.28	0.20
	(0.21)	(0.43)	(0.19)
Self-competition	0.19	0.03	0.05
	(0.36)	(0.40)	(0.25)
Risk tolerance	5.02	11.31*	8.72**
	(6.84)	(5.98)	(4.32)
Pleasure in task	19.67	7.28	12.08
	(12.63)	(12.36)	(8.55)
Time-constrained (A)	-20.36	9.28	-7.56
	(36.58)	(42.40)	(26.03)
Uncertainty	-9.42	-17.98	-14.03*
	(10.98)	(12.17)	(7.74)
Constant	-72.58	54.18	-45.29
	(84.75)	(100.48)	(62.21)
N	82	80	162
Adj. R^2	0.55	0.52	0.54

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.29: Comparison of *goal 1* in *Late* vs. *Revise0 & Revise1*

	(1)	(2)	(3)
Late	-12.82	-2.41	-11.31
	(19.69)	(18.57)	(19.45)
Productivity1 [a]		8.66***	8.43***
		(1.84)	(1.96)
CRT			-3.24
			(10.21)
Self-competition[b]			-0.32
			(0.29)
Risk tolerance			4.84
			(5.63)
Pleasure in task			23.33***
			(8.85)
Time-constrained(A)[c]			-48.48**
			(21.85)
Constant	241.83***	80.81**	11.41
	(12.35)	(33.87)	(62.56)
N	249	249	228
Adj. R^2	-0.00	0.10	0.13

Dependent variable: goal 1 Robust standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.

[a] Baseline productivity at date 1.

[b] Percentage allocated to self-competitive option at date 0.

[c] Dummy = 1 if $\leq$ 2 hours of flexible time in the actual time schedule for date 1.

Table S.30: Goal revision when controlling for uncertainty and time shocks

	(1)[a]	(2)	(3)
Change in productivity	2.67	2.94*	3.11*
	(1.82)	(1.73)	(1.72)
Uncertainty	11.03	10.84	
	(9.16)	(8.25)	
More time		-3.38	
		(43.86)	
Less time		-12.25	
		(33.31)	
Difference in flexible time			0.11
			(2.29)
Constant	-61.94***	-62.30***	-43.95***
	(18.72)	(17.34)	(9.84)
N	143	162	162
Adj. R2	0.01	0.01	0.01

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

[a] Sample restricted to those with no time shock.

Table S.31: Effect of change in enjoyment on goal revision

	(1)	(2)
Change in enjoyment	9.88	8.50
	(7.31)	(7.32)
Change in productivity		2.97**
		(1.40)
Constant	-28.23***	-36.50***
	(8.19)	(9.35)
N	142	142
Adj. R2	0.00	0.02

Robust standard errors in parentheses.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Figure S.8: Goals and effort

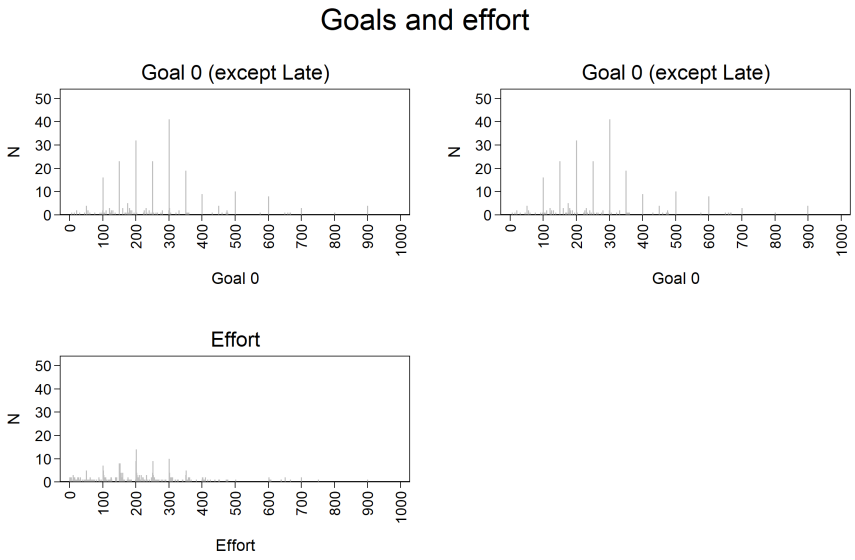
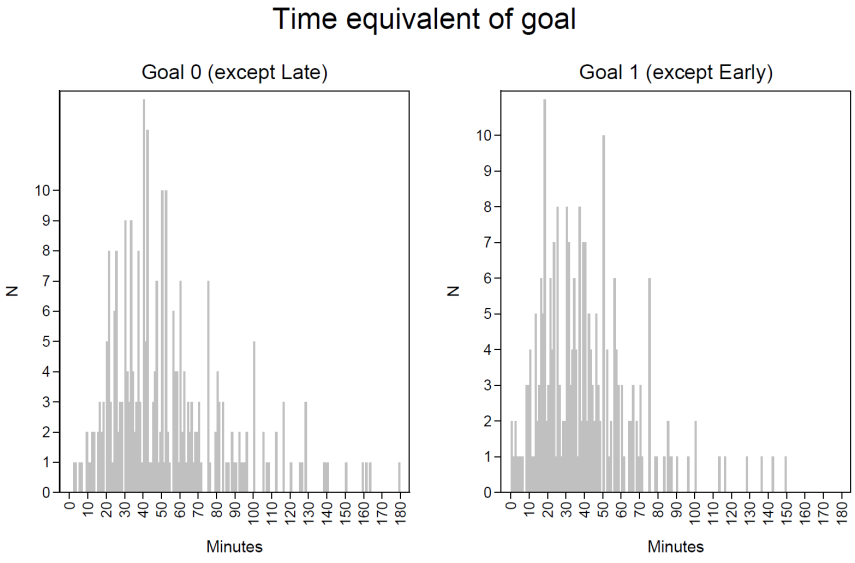
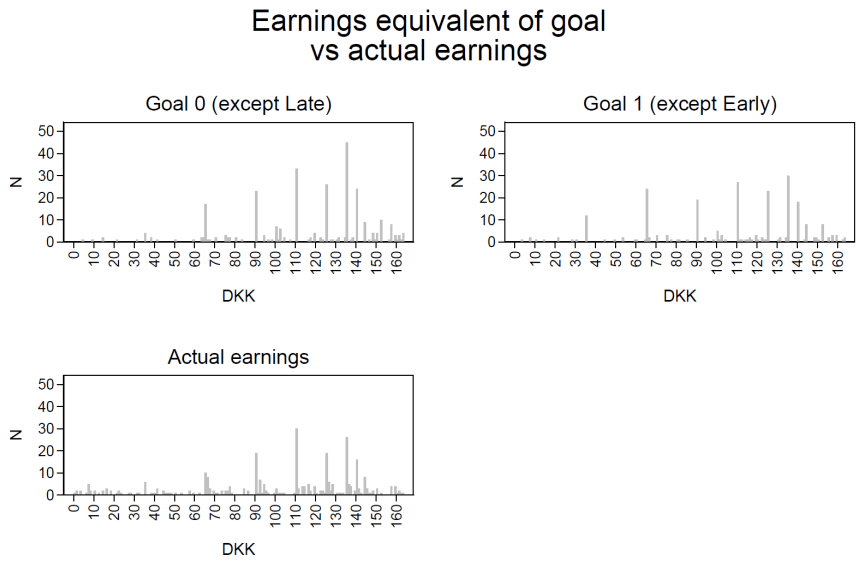


Figure S.9: Time equivalent of goals and actual time spent



Note: Time is truncated at 180 minutes.

Figure S.10: Earnings equivalent of goals set by subjects vs. actual earnings



S. 7 Experimental instructions

Consent form

1. Title of research

You are being invited to take part in the research study "Working on online tasks", and we would like to ask you for your consent to participate in the study and for us to treat your data in agreement with data protection legislation. Before you decide to participate in this study, it is important that you understand why the research is being conducted, and what it will involve. Please take the time to read the following information carefully. Please contact the researchers if there is anything that is not clear or if you need more information (see contact information below).

2. Project description and aim of the study

The aim of this study is to investigate how people work on online tasks.

3. Data controller, research group, and principal investigator

Data controller and principal investigator: Alexander Koch, Aarhus University, akoch@econ.au.dk.

Other researchers: Jonas Pilgaard Kaiser, Aarhus University, jkaiser@econ.au.dk, and Julia Nafziger, Aarhus University, jnafziger@econ.au.dk

4. Study procedure

The study consists of three parts, which are all completed online using a computer or notepad. The study does not work on mobile phones. The first part takes place on a Wednesday, Thursday, or Friday in <month, year>. You will choose the date from a list. The second part takes place 5 days after part 1 and the third part 7 days after part 1. You will be required to participate in each part of the study in "one go". That is, once you get started with a part, if you are inactive for more than 30 minutes, the computer interface will sign you out and stop collecting data for that part of the study. It is not possible to restart a part of the study once you are signed out.

Part 1 (date you chose from a list): if you give consent to participate in the study by pressing the button below, you will choose a date from a list and receive an invitation email to part 1 shortly before 0:00 on that date. You can use that link until 23:59 on the same day. Following the link in the email will lead you to a web page where you will get detailed instructions. After receiving information about how you will get paid for working on a task, you will be given time to work on the task. The task is to count the number of zeros in a series of tables. The task does not require any prior training or ability. In addition, we will ask you several survey questions, for example, related to your background (e.g., gender, age, and study area), your attitudes, the task, and your time planning. Part 1 will take around 30 minutes and will have to be completed by 23:59.

Part 2 (5 days after part 1): if you complete the first part, then 5 days later, shortly before 0:00, you will receive an invitation email to the second part of the study. During the second part, you will again be given time to work on some online tasks after having received information about how you will be paid for working on the task. Again, we will ask you several survey questions. Depending on how long you want to work on the task, part 2 will take between around 25 minutes and 200 minutes. Part 2 has to be completed by 23:59.

Part 3 (7 days after part 1): if you complete the second part, then 7 days after part 1 you will receive an invitation email to the third part of the study shortly before 0:00. Following the link in the email leads you to a survey. Filling out this survey will take around 5 minutes. Part 3 has to be completed by 23:59.

Participation links will be sent from jkaiser@econ.au.dk or akoch@econ.au.dk. Please add these addresses to your address book so that the emails do not end in your spam folder.

5. Benefits and risks

There are no risks beyond those encountered in normal everyday life.

The total time for participating in this study is between around 60 and 240 minutes. If you complete all parts, you earn at least DKK 65 and you can earn up to approximately DKK 300.

Earning Part 1: if you complete the entire first part, you will receive DKK 35. Further, you will get paid for working on the online task. Your payment here depends on how much you work. In addition, you can earn up to DKK 6 depending on the accuracy of your answers on some of the survey questions. *Your total expected total earnings from part 1 are around DKK 55.* Overall, *part 1 takes around 30 minutes.*

Earnings Part 2:

1. In the first block of part 2, you will answer some questions and work on some tasks. You will get paid DKK 20 for completing this part. In addition, you will get paid for the number of tasks you solve. *Your expected total earnings from this block are approximately DKK 35.* The *time commitment for this block is approximately 20 minutes*. Please note that you can stop working at any time, but if you do so or if you do not answer the questions, then your earnings for this part are zero and you cannot go on to the second block of part 2.
2. In the second block of part 2, your earnings depend on the number of tasks you solve. Again, you can stop working at any time you like. Once you stop working, your earnings will be determined by the number of tasks you have solved up to this point. **Your maximal earnings in the second block are DKK 163.**

Earnings Part 3: you will receive *DKK 15 plus up to DKK 4* depending on the accuracy of your answers. The survey *takes approximately 5 minutes*.

Payments will be into the NEM account linked to your cpr number. Alexander Koch and his team will start registering the payments with the administration of Aarhus University once the study is concluded. Then the administrative process might take between 2-6 weeks. You can contact Alexander Koch by email (akoch@econ.au.dk) if you want information on the payment process. Please write this email address down, so that you have his contact details in case you later have any questions!

Taxes: according to Danish law, Aarhus University reports payments to the tax authorities. Please note that taxes might be deducted from the amount of money you earn.

6. Type of personal data and when it is deleted/anonymized

We process normal personal information in form of your CPR number, email address, and your name. The email address is used to contact you and provide the links needed to access the different parts of the study. To determine the payments that you will receive for participation in the study, we need to link your name and CPR number with your data from the experiment through a participant ID number. Once the payment process is finalized, your name, email address, and CPR number are deleted (approximately 3 months from now).

This study collects and processes other normal personal information in form of, for example, your gender, age, and study area. These data are collected for the scientific analysis. The survey software that we use for this study collects, like most webpages, your IP address and estimates your location based on the IP address. This information will be used to produce some aggregate statistics on the background of the participants; thereafter, it will be deleted (approximately 6 months from now).

In sum, we will only temporarily store and process your name, CPR number, email address, IP address, and estimated location. After a period of approximately 6 months, this information will be deleted and the data will be anonymized.

7. External data processors

Your data (including your CPR number) will be collected using the survey software Qualtrics. Aarhus University has a data processing agreement with the company Qualtrics. The data processing agreement documents that the cooperation between Aarhus University and Qualtrics complies with the rules concerning the protection of personal data.

Any publication of the research in this study will be based on anonymized data (i.e., the data without personal identifiers). As part of such a publication, the anonymized data set will be made publically available to allow other researchers to reproduce the statistical analysis.

8. Withdrawal of consent

Participation is voluntary, and you may withdraw your consent at any time. This is done by contacting Alexander Koch by email. Please note that your data can only be deleted before the data from the study are anonymized. Thereafter, your entries can no longer be identified in the data.

Please note that you can only participate in this study once. We reserve the right to cancel participation in case the study gets oversubscribed before your date of participation. In that case, we will of course inform you by email to the address that you provide us with.

Acceptance Button

By answering "Yes" below, I confirm to have received, read, and understood the above information and that:

A. My participation is voluntary, and I may withdraw my consent and discontinue participation in the project at any time as specified in point 8. My refusal to participate will not result in any penalty.

B. By accepting this agreement, I do not waive any legal rights or release Aarhus University, its agents, or you from liability for negligence.

C. I give my consent to treat my name and CPR number for payment purposes and to participate as a subject in the study as described above.

Instructions for part 1

Page 1: Welcome to part 1 of the research study "Working on online tasks".

This part will take around 30 minutes. You need to complete this part by 23:59 today (<date string>) to be eligible to participate in the next parts of the study. Go to the next page to get started.

Page 2: Please enter your **CPR number** (or your "midlertidigt"/temporary CPR-number), which will be transmitted by a secure internet connection. Write it in without spaces or hyphen (e.g., 0112401234):

We cannot pay you for your participation in the study without a correct and complete CPR number! Your CPR number will only be used for the payment process and will be deleted after. <entry field>

Please confirm your CPR number: <entry field>

Page 3: What is your **age** (in years)? <entry field>

Page 4: What type of **faculty** are you studying at?

- o Arts/Humanities/Theology
- o BSS (Business and Social Sciences)/Social Sciences/Law
- o Health
- o Science and Technology
- o Other
- o I am not a student

(If not a student) Page 5: What best describes **your situation**?

- o University employee
- o Employed in other public sector
- o Employed in the private sector
- o Self-employed
- o Unemployed
- o Other

(If a student) Page 5: What **type of degree** are you studying for?

- o Bachelor
- o Master
- o PhD
- o Other

(If a student in Arts/Humanities/Theology) Page 6: What best describes your field of study?

- o Archaeology
- o Anthropology
- o Languages
- o Information studies

- o Theology
- o Other

(If a student in BSS (Business and Social Sciences)/Social Sciences/Law) Page 6: What best describes your field of study?

- o Business Administration/Economics
- o Law
- o Political Sciences (Statskundskab)
- o Psychology
- o Anthropology
- o Sociology
- o Other

(If a student in Health) Page 6: What best describes your field of study?

- o Dentistry
- o Medicine
- o Public health
- o Sports sciences
- o Other

(If a student in Science and Technology) Page 6: What best describes your field of study?

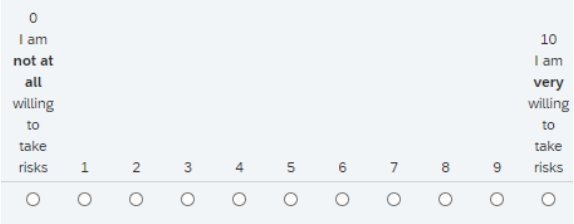
- o Agrobiology
- o Biology
- o Chemistry
- o Computer Science
- o Data Science
- o Engineering
- o Physics
- o Geoscience
- o IT Product Development
- o Chemistry
- o Mathematics
- o Mathematics - Economics
- o Nanoscience
- o Other

(If a student) Page 7: Do you have a **thesis, project report, or other assignments** to hand in during the **next 2 weeks**?

- o no
- o yes - one
- o yes - two
- o yes - three
- o yes - four or more

Page 8: How do you see yourself? **Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?**

Please select a value between 0 and 10, where the value 0 means: 'not at all willing to take risks' and the value 10 means: 'very willing to take risks'



Page 9: A bat and a ball cost DKK 110 in total. The bat costs DKK 100 more than the ball. How much does the ball cost (in DKK)? <entry field>

Page 10: If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets (in minutes)? **<entry field>** <entry field>

Page 11: In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake (in days)? <entry field>

Page 12: Task

Your task will now be to **count zeros in a series of tables**. Such a table looks like follows and once you have counted the number of zeros in a table, you should enter the number of zeros in that table into a field below the table.

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

How many zeros are in the table?
(17 is the correct answer for this table)

On the next page you will have **3 minutes** to count zeros in up to 40 tables. **You earn DKK 0.5 for each table where you counted the number of zeros correctly.**

Once you finished a table, please scroll down to access the next table. Use the tab key to jump to the next data entry field, or select the field with a mouse click. The remaining time will be displayed on the right-hand side of the screen. **After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen**.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

When you are ready to start, press the -> button.

Page 13: You have 3 minutes to count the number of zeros in up to 40 tables.

After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

<Tables>

Page 14: Thanks. Your answers have been recorded.

Page 15: How much do you like the task of counting zeros?

- o Like a great deal
- o Like somewhat
- o Neither like nor dislike
- o Dislike somewhat
- o Dislike a great deal

Page 16: You will again have **3 minutes** to count zeros in up to 40 tables. Now **your earnings may, depending on your choices, depend on whether you do better than in the first round.**

You are asked to choose what portion of your earnings for this task (between 0 and 100 percent, inclusive) you wish to be determined by either of the following two options.

Option A:
You earn **DKK 0.5** for each table.

Option B:
- You earn **DKK 1** for each table **if you count more tables** than you did in the first round.
- You earn **zero** for each table **if you count fewer tables** than you did in the first round.
- You earn **DKK 0.5** for each table **if you count exactly the same number of tables** as in the first round.

Enter a number into the text box to adjust the percent of earnings determined according to each option. The two numbers must add up to 100.

<entry field> percent according to option A

<entry field> percent according to option B

Page 17: Once you finished a table, please scroll down to access the next table. Use the tab key to jump to the next data entry field, or select the field with a mouse click. The remaining time will be displayed on the right-hand side of the screen. **After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen.**

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

When you are ready to start, press the -> button.

Page 18: You have 3 minutes to count the number of zeros in up to 40 tables.

After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

<Tables>

Page 19: Thanks. Your answers have been recorded

Page 20: We would like to know about **your time schedule for <date string for part 2>** (the date when you will participate in part 2 of the study).

Please indicate **what best describes your plans for each 1-hour block** by ticking the appropriate box. (Any time planned for participating in part 2 of the study should count as "flexible time".)

	Please select one option for each time slot					
	Sleep	**Work** (f.ex. student job)	**Classes or tutorials**	**Scheduled studying** (f.ex. self-studying or study group)	**Scheduled leisure activities**	**Flexible time**
0:00- 1:00	○	○	○	○	○	○
1:00- 2:00	○	○	○	○	○	○
2:00- 3:00	○	○	○	○	○	○
3:00- 4:00	○	○	○	○	○	○
...						
18:00-19:00	○	○	○	○	○	○
19:00-20:00	○	○	○	○	○	○
20:00-21:00	○	○	○	○	○	○
21:00-22:00	○	○	○	○	○	○
22:00-23:00	○	○	○	○	○	○
23:00-24:00	○	○	○	○	○	○

Page 21: How likely do you think it is that you will end up having **less** than 2 hours of flexible time on **<date string for part 2>** ? (Any time planned for participating in part 2 of the study should count as "flexible time".)

- o Extremely likely
- o Somewhat likely
- o Neither likely nor unlikely
- o Somewhat unlikely
- o Extremely unlikely

(all treatments, except Late) Page 22:

In **part 2 of the study, on <date string for part 2>** between 0:00 and 23:59, you will have the **opportunity to count the number of zeros in as many tables as you like**.

You will earn a piece rate, that is, a payment for each table in which you count the numbers of zeros correctly (for simplicity we call this a "correctly counted table"). **The piece rate varies with the number of tables that you count** as follows:

- For tables **1 to 50**, you earn **DKK 0.7** per correctly counted table
- For tables **51 to 100**, you earn **DKK 0.6** per correctly counted table
- For tables **101 to 150**, you earn **DKK 0.5** per correctly counted table
- For tables **151 to 200**, you earn **DKK 0.4** per correctly counted table
- For tables **201 to 250**, you earn **DKK 0.3** per correctly counted table
- For tables **251 to 300**, you earn **DKK 0.2** per correctly counted table
- For tables **301 to 350**, you earn **DKK 0.1** per correctly counted table
- For tables **351 to 400**, you earn **DKK 0.09** per correctly counted table
- For tables **401 to 450**, you earn **DKK 0.08** per correctly counted table
- For tables **451 to 500**, you earn **DKK 0.07** per correctly counted table
- For tables **501 to 550**, you earn **DKK 0.06** per correctly counted table
- For tables **551 to 600**, you earn **DKK 0.05** per correctly counted table
- For tables **601 to 650**, you earn **DKK 0.04** per correctly counted table
- For tables **651 to 700**, you earn **DKK 0.03** per correctly counted table
- For tables **701 to 750**, you earn **DKK 0.02** per correctly counted table
- For tables **751 to 900**, you earn **DKK 0.01** per correctly counted table
- For tables **901 and beyond**, you earn **zero** per correctly counted table

Click **here** to see a graph of how your earnings depend on the number of tables you complete (opens a new window)

You will need to work on the task in "one go". That is, once you get started **on <date string for part 2>**, if you are inactive for more than 30 minutes, the computer interface will record the number of correctly counted tables, sign you out, and stop collecting data for part 2 of the study.

(all treatments, except LATE) Page 23: Set a goal!

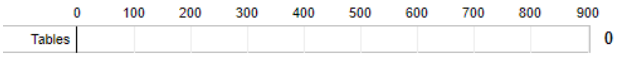
We ask you to set yourself a goal for how many tables to count on <date string for part 2>. We will remind you of the goal you set with a probability of 2/3. But, of course, you are free to work as much as you want.

Below, we give you feedback on your performance on the task today. Before you set your goal, play around a bit with the slider below.

Use the slider to indicate different goals or click on the number to the right of the slider to type in a goal. The text above will then explain how much time you would need to reach your goal and what your earnings would be (if you worked at the same speed as when you tried out the task before).
Note: The slider stops at 900 because if you count more tables your earnings do not change.

What if I set a goal of <value> tables?

- When trying out the task, **you managed to complete <value> tables in 3 minutes**.
- At this speed, **reaching a goal of <value> tables would take approximately <value> minutes and <value> seconds.**
- Your **total earnings would be DKK <value> .** The piece rate for the last table would be DKK **<value>**.



My goal for how many tables to complete on <date string for part 2>: <entry field>

Reminder:
1. **You will need to work on the task in "one go"**.
That is, once you get started **on <date string for part 2>**, if you are inactive for more than 30 minutes, the computer interface will record the number of correctly counted tables, sign you out and stop collecting data for part 2 of the study.

2. Click **here** to see the **table with the piece rates** from the previous screen (opens a new window)
3. Click **here** to see a **graph** of how your earnings depend on the number of tables you complete (opens a new window)

Page 24: Thank you for completing part 1 of the study.

On <date string for part 2>, you will receive an invitation email with a link for accessing the second part of the study. The link will work between 0:00 and 23:59 on <date string for part 2>.

Your earnings in this study so far are DKK <value>.

Details:
1. You receive DKK <value>for completing part 1.
2. You receive DKK <value> because you correctly answered <value> out of the 3 questions paid DKK 2 for each correct answer
3. You receive: DKK <value>from the first round of the counting task.
You managed to complete <value> tables in 3 minutes. The piece rate was DKK 0.5.
4. You receive: DKK <value>from the second round of the counting task.
You managed to complete <value>tables in 3 minutes. You selected to be paid <value> percent according to option A (piece rate of DKK 0.5) and <value> percent according to option B (piece rate of DKK 1/0.5/0 if more/the same number/fewer tables correctly counted than in the first round of the counting task).

Move to the next page to finish.

Instructions for part 2

Page 1: Welcome to part 2 of the research study "Working on Online Tasks".

First, you will spend two times three minutes working on the counting task. In between, you will answer a few questions. Thereafter, you will have the opportunity to increase your earnings by working as much as you like on some tasks.

Go to the next page to get started.

Page 2: Task

Your task will now be to **count zeros in a series of tables**. Such a table looks like follows and once you have counted the number of zeros in a table, you should enter the number of zeros in that table into a field below the table.

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

How many zeros are in the table?
(17 is the correct answer for this table)

On the next page you will have **3 minutes** to count zeros in up to 40 tables. **You earn DKK 0.5 for each table where you counted the number of zeros correctly.**

Once you finished a table, please scroll down to access the next table. Use the tab key to jump to the next data entry field, or select the field with a mouse click. The remaining time will be displayed on the right-hand side of the screen. **After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen**.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

When you are ready to start, press the -> button.

Page 3: You have 3 minutes to count the number of zeros in up to 40 tables.

After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

<Tables>

Page 4: Thanks. Your answers have been recorded.

Page 5: We would like to know about **your time schedule for today**.

Please indicate **what best describes your plans for each 1-hour block** by ticking the appropriate box. Count as "flexible time" any time planned for participating in today's part of the study.

	Please select one option for each time slot					
	Sleep	**Work** (f.ex. student job)	**Classes or tutorials**	**Scheduled studying** (f.ex. self-studying or study group)	**Scheduled leisure activities**	**Flexible time**
0:00- 1:00	○	○	○	○	○	○
1:00- 2:00	○	○	○	○	○	○
2:00- 3:00	○	○	○	○	○	○
3:00- 4:00	○	○	○	○	○	○
...						
18:00-19:00	○	○	○	○	○	○
19:00-20:00	○	○	○	○	○	○
20:00-21:00	○	○	○	○	○	○
21:00-22:00	○	○	○	○	○	○
22:00-23:00	○	○	○	○	○	○
23:00-24:00	○	○	○	○	○	○

Page 6: Next, you will answer some questions and spend another 3 minutes working on the task. Once you are done with this, you will have the **opportunity to count the number of zeros in as many tables as you like** until 23:59 today. However, **you must work on the task in "one go"**. That is, once you get started with counting, if you are inactive for more than 30 minutes, the computer interface will record the number of correctly counted tables, sign you out, and stop collecting data for part 2 of the study.

You will earn a piece rate, that is, a payment for each table in which you count the numbers of zeros correctly (for simplicity we call this a "correctly counted table"). **The piece rate varies with the number of tables that you count** as follows:

- For tables **1 to 50**, you earn **DKK 0.7** per correctly counted table
- For tables **51 to 100**, you earn **DKK 0.6** per correctly counted table
- For tables **101 to 150**, you earn **DKK 0.5** per correctly counted table
- For tables **151 to 200**, you earn **DKK 0.4** per correctly counted table
- For tables **201 to 250**, you earn **DKK 0.3** per correctly counted table
- For tables **251 to 300**, you earn **DKK 0.2** per correctly counted table
- For tables **301 to 350**, you earn **DKK 0.1** per correctly counted table
- For tables **351 to 400**, you earn **DKK 0.09** per correctly counted table
- For tables **401 to 450**, you earn **DKK 0.08** per correctly counted table
- For tables **451 to 500**, you earn **DKK 0.07** per correctly counted table
- For tables **501 to 550**, you earn **DKK 0.06** per correctly counted table
- For tables **551 to 600**, you earn **DKK 0.05** per correctly counted table

- For tables **601 to 650**, you earn **DKK 0.04** per correctly counted table
- For tables **651 to 700**, you earn **DKK 0.03** per correctly counted table
- For tables **701 to 750**, you earn **DKK 0.02** per correctly counted table
- For tables **751 to 900**, you earn **DKK 0.01** per correctly counted table
- For tables **901 and beyond**, you earn **zero** per correctly counted table

Click **here** to see a graph of how your earnings depend on the number of tables you complete (opens a new window).

Remember that **if you are inactive for more than 30 minutes, the computer interface will sign you out.**

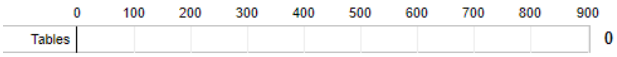
(If treatment Late) Page 7: Set a goal!

We ask you to set yourself a goal for how many tables to count today. We will remind you of the goal you set with a probability of 2/3. But, of course, you are free to work as much as you want.

Below, we give you feedback on your performance on the task today. Before you set your goal, play around a bit with the slider below.
Use the slider to indicate different goals or click on the number to the right of the slider to type in a goal. The text above will then explain how much time you would need to reach your goal and what your earnings would be (if you worked at the same speed as when you tried out the task before).
Note: The slider stops at 900 because if you count more tables your earnings do not change.

What if I set a goal of <value> tables?

- When trying out the task, **you managed to complete <value> tables in 3 minutes**.
- At this speed, **reaching a goal of <value> tables would take approximately <value> minutes and <value> seconds.**
- Your **total earnings would be DKK <value> .** The piece rate for the last table would be DKK **<value>**.



My goal for how many tables to complete today: <entry field>

Reminder:
1. **You will need to work on the task in "one go".**
That is, once you get started, if you are inactive for more than 30 minutes, the computer interface will record the number of correctly counted tables, sign you out and stop collecting data for part 2 of the study.
2. Click **here** to see the **table with the piece rates** from the previous screen (opens a new window)
3. Click **here** to see a **graph** of how your earnings depend on the number of tables you complete (opens a new window)

(If treatment Revise0 or Revise1) Page 7: In part 1, you set yourself the goal of counting <value> tables today.

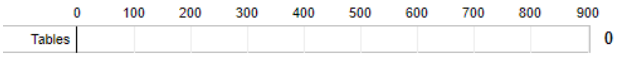
You now again have the opportunity to set a goal for how many tables to count today. We will remind you about either the goal you set now or the goal you set in part 1, each with probability 1/2. But, of course, you are free to work as much as you want.

Below, we give you feedback on your performance on the task today. Before you set your goal, play around a bit with the slider below.

Use the slider to indicate different goals or click on the number to the right of the slider to type in a goal. The text above will then explain how much time you would need to reach your goal and what your earnings would be (if you worked at the same speed as when you just worked on the task).
Note: the slider stops at 900 because if you count more tables your earnings do not change.

What if I set a goal of <value> tables?

- When trying out the task, **you managed to complete <value> tables in 3 minutes**.
- At this speed, **reaching a goal of <value> tables would take approximately <value> minutes and <value> seconds.**
- Your **total earnings would be DKK <value> .** The piece rate for the last table would be DKK **<value>**.



My goal for how many tables to complete today: <entry field>

Reminder:
1. **You will need to work on the task in "one go"**.
That is, once you get started, if you are inactive for more than 30 minutes, the computer interface will record the number of correctly counted tables, sign you out and stop collecting data for part 2 of the study.
2. Click **here** to see the **table with the piece rates** from the previous screen (opens a new window)
3. Click **here** to see a **graph** of how your earnings depend on the number of tables you complete (opens a new window)

(If treatment Early) Page 7: In part 1, you set yourself the goal of counting <value> tables today.

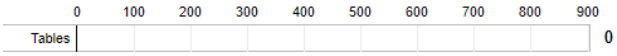
We will remind you of this goal. But, of course, you are free to work as much as you want.

Below, we give you feedback on your performance on the task today. Before you set your goal, play around a bit with the slider below.

Use the slider to indicate different goals or click on the number to the right of the slider to type in a goal. The text above will then explain how much time you would need to reach your goal and what your earnings would be (if you worked at the same speed as when you tried out the task before).
Note: The slider stops at 900 because if you count more tables your earnings do not change.

What if I complete <value> tables?

- When trying out the task, **you managed to complete <value> tables in 3 minutes**.
- At this speed, **reaching a goal of <value> tables would take approximately <value> minutes and <value> seconds.**
- Your **total earnings would be DKK <value> .** The piece rate for the last table would be DKK **<value>**.



Reminder:
1. **You will need to work on the task in "one go"**.
That is, once you get started, if you are inactive for more than 30 minutes, the computer interface will record the number of correctly counted tables, sign you out and stop collecting data for part 2 of the study.
2. Click **here** to see the **table with the piece rates** from the previous screen (opens a new window)
3. Click **here** to see a **graph** of how your earnings depend on the number of tables you complete (opens a new window)

Page 8: On the next page, you will again have **3 minutes** to count zeros in up to 40 tables. **You earn DKK 0.5 for each table where you counted the number of zeros correctly.**

Once you finished a table, please scroll down to access the next table. Use the tab key to jump to the next data entry field, or select the field with a mouse click. The remaining time will be displayed on the right-hand side of the screen. After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

When you are ready to start, press the -> button.

Page 9: You have 3 minutes to count the number of zeros in up to 40 tables.

After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

<Tables>

Page 10: Thanks. Your answers have been recorded.

In the two 3-minute rounds of the counting task you managed to complete
- ${e://Field/p2productivity1} tables (first round)
- ${e://Field/p2productivity2} tables (second round)
The piece rate was DKK 0.5 in both rounds.

In addition, you receive DKK ${e://Field/fixedpay2} because you completed the first block of today's part of the study.

Please move to second block now.

Page 11: You now have the opportunity to **count the number of zeros in as many tables as you like until 23:59 today**.

You set yourself the goal of counting <value> tables.

From the next page on, if you are inactive for more than 30 minutes, you cannot resume working.
Important: Once you continue to the next page, you will have to do all the tasks that you wish to complete without any breaks that last longer than 30 minutes. If you accidentally close your browser, you can use your survey link to open the study again and continue where you stopped, as long as you were not inactive for more than 30 minutes. You need to use the same computer and browser (this feature works by having the survey software place a cookie on your browser that keeps track of how far you got). If you do not wish to start with the study at this time point, close your browser and use your survey link to open the study again at a later time point, but before the deadline of 23:59 today.

Page 12 - : Your goal is to complete <value> tables.

So far, you have completed <value> tables.

For the next table you complete, you earn DKK <value>.

Your total earnings for part 2 of the study so far are DKK <value>.

Please count the number of zeros in the following table.

Once you counted the table, please click "->" to save your response. If you miscount the table, you will be asked to count it again.

<Table>

How many zeros are in the table?

<entry field>

Reminder:

1. You need to submit an entry before <current time + 30 min>. If you remain inactive beyond that time, you will not be able to continue with part 2 of the study, and your earnings will be DKK <value>. (DKK <value> for the first block and DKK <value> for the second block of part 2). If you accidentally close your browser, you can use your survey link to open the study again and continue where you stopped - as long as you were not inactive for more than 30 minutes. You need to use the same computer and browser (this feature works by having the survey software place a cookie on your browser that keeps track of how far you got).

2. Click **here** to see the **table with the piece rates** for tables completed (opens a new window)
3. Click **here** to see a **graph** of how your earnings depend on the number of tables you complete (opens a new window)

Instructions for part 3

Page 1: Welcome to the final part of the research study "Working on Online Tasks".

This part consists of several survey questions and will take around 5 minutes. Go to the next page to get started.

Page 2: How much do you like the task of counting the number of zeros in tables?

- o Like a great deal
- o Like somewhat
- o Neither like nor dislike
- o Dislike somewhat
- o Dislike a great deal

Page 3:

(All treatments, except Late) We now ask you to **recall the goal that you set yourself in part 1 (on <date>).**

You receive **DKK 2 if you correctly recall** the goal that you set. <entry field>

(All treatments, except Early) We now ask you to **recall the goal that you set yourself in part 2 (on <date>).**

You receive **DKK 2 if you correctly recall** the goal that you set. <entry field>

(If treatment Revise0 or Revise1) Which of the two goals did you care more about?

- o The goal that I set myself in part 1 (on <date>)
- o The goal that I set myself in part 2 (on <date>)
- o I cared equally about both goals

(If treatment Late) Page 4: Early in part 2 of the study, you were asked to set yourself a goal for how many tables to count in part 2.

Did you already have a goal in mind before starting with part 2?

- o Yes, before starting part 2 I had already set a goal for how many tables to count in part 2.
- o No, I first thought about what goal to set in part 2 when asked to set a goal.

If you answered yes, please recall the goal you had already set. Otherwise leave this field empty. <entry field>

(If treatment Late) Page 5: Consider **how you felt at the start of part 2 (on >date>) when setting yourself a goal for how many tables to count a few minutes later.**

	Not at all	Slightly	Moderately	Quite a bit	Extremely
How committed were you to this goal?	○	○	○	○	○
To what extent did you care about this goal?	○	○	○	○	○
To what extent did you think that you would replace this goal with a new one?	○	○	○	○	○

(If treatment Late) Page 6: Consider **how you felt when counting tables in part 2 (on <date>).**

To what extent did any of the items below **influence how many tables you counted**?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
The **goal I was reminded about on the screen** while counting (which I had set for myself early in part 2)	○	○	○	○	○
A **different goal for how many tables to count**	○	○	○	○	○
A **goal for the time** that I wanted to use on the task	○	○	○	○	○
A **goal for how much I wanted to earn**	○	○	○	○	○
The **piece rate** for the tables correctly counted	○	○	○	○	○
Other factors	○	○	○	○	○

(If treatment Early) Page 4: Consider **how you felt in part 1 (on <date>) when setting yourself a goal for how many tables to count in part 2.**

	Not at all	Slightly	Moderately	Quite a bit	Extremely
How committed were you to this goal?	○	○	○	○	○
To what extent did you care about this goal?	○	○	○	○	○
To what extent did you think that you would replace this goal with a new one?	○	○	○	○	○

(If treatment Early) Page 5: In part 1 (on <date>), you set yourself a goal for how many tables to count in part 2. **Before starting to count tables** in part 2, did you set yourself a **new goal** for how many tables to count?

- o Yes, I set myself a new goal after having set a goal in part 1
- o No, I did not set myself a new goal after having set a goal in part 1

If you answered yes, please recall the new goal you set. Otherwise leave this field empty. <entry field>

(If treatment Early) Page 6: Consider **how you felt when counting tables in part 2 (on <date>).**

To what extent did any of the items below **influence how many tables you counted**?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
The **goal I was reminded about on the screen** while counting (which I had set for myself in part 1)	○	○	○	○	○
A **different goal for how many tables to count**	○	○	○	○	○
A **goal for the time** that I wanted to use on the task	○	○	○	○	○
A **goal for how much I wanted to earn**	○	○	○	○	○
The **piece rate** for the tables correctly counted	○	○	○	○	○
Other factors	○	○	○	○	○

(If treatment Revise0 or Revise1) Page 4:

Consider **how you felt in part 1 (on <date>) when setting yourself a goal for how many tables to count in part 2.**

	Not at all	Slightly	Moderately	Quite a bit	Extremely
How committed were you to this goal?	○	○	○	○	○
To what extent did you care about this goal?	○	○	○	○	○
To what extent did you think that you would replace this goal with a new one?	○	○	○	○	○

Consider **how you felt at the start of part 2 (on <date>) when setting yourself a goal for how many tables to count a few minutes later.**

	Not at all	Slightly	Moderately	Quite a bit	Extremely
How committed were you to this goal?	○	○	○	○	○
To what extent did you care about this goal?	○	○	○	○	○
To what extent did you think that you would replace this goal with a new one?	○	○	○	○	○

(If treatment Revise0) Page 5: Consider **how you felt when counting tables in part 2 (on <date>).**

To what extent did any of the items below **influence how many tables you counted**?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
The **goal I was reminded about on the screen** while counting (which I had set for myself in part 1)	○	○	○	○	○
The **goal I was not reminded about** while counting (which I had set for myself a few minutes before starting to count)	○	○	○	○	○
A **different goal for how many tables to count**	○	○	○	○	○
A **goal for the time** that I wanted to use on the task.	○	○	○	○	○
A **goal for how much I wanted to earn**	○	○	○	○	○
The **piece rate** for the tables correctly counted	○	○	○	○	○
Other factors	○	○	○	○	○

(If treatment Revise1) Page 5: Consider **how you felt when counting tables in part 2 (on <date>).**

To what extent did any of the items below **influence how many tables you counted**?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
The **goal I was reminded about on the screen** while counting (which I had set for myself a few minutes before starting to count)	○	○	○	○	○
The **goal I was not reminded about** while counting (which I had set for myself in part 1)	○	○	○	○	○
A **different goal for how many tables to count**	○	○	○	○	○
A **goal for the time** that I wanted to use on the task.	○	○	○	○	○
A **goal for how much I wanted to earn**	○	○	○	○	○
The **piece rate** for the tables correctly counted	○	○	○	○	○
Other factors	○	○	○	○	○

Page 7: Please read the following sentences and state how well they describe you.

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
When setting a goal, I carefully think about what I want to achieve and when to achieve it	○	○	○	○	○
I feel angry with myself when I give up a goal	○	○	○	○	○
I sometimes do not set goals because I am afraid that I will not be able to achieve them	○	○	○	○	○
I set goals in my daily life (e.g., for the number of hours you want to study, for saving money, . . .)	○	○	○	○	○

Page 8: **You have now completed the study.**

Your total earnings in this study are DKK <value>.
(DKK <value> from part 1, DKK <value> from part 2, and DKK <value> from part 3)

Thank you for helping us with our research.

Move to the next page to finish.