Does Goal Revision Undermine Self-Regulation Through Goals? An Experiment

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

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

June 5, 2022

Abstract

We offer a novel test of whether non-binding goals set ahead of a task are effective motivators, taking into account that individuals in principle could easily revise these goals. In our setting, subjects either set a goal some days prior to an online task (early goal) or right at the start of the task (late goal). Two further treatments allow for (unanticipated) explicit revision of the early goal. Consistent with early goals being self-regulation tools, we observe that (i) early goals are larger than late goals; (ii) subjects in early goals treatments work more than those in the late goal treatment, despite evidence of substantial downward goal revision in the former. A secondary contribution of our paper is that our design addresses a treatment migration problem present in earlier studies on goals that stems from the fact that subjects in a 'no goals' control condition may privately set goals.

JEL Classification: D90, C91, D01

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

Statements and Declarations: The authors have no competing interests. Financial support from the Interacting Minds Seeds at Aarhus University, grant 2019-120, is gratefully acknowledged.

*The study is pre-registered on OSF (https://osf.io/3q82y) and was approved by the Human Subjects Committee of the Cognition and Behavior lab, Aarhus University (21.10.2019, ID 272). In Kaiser et al. (2021), we report additional pre-registered analyses and document any changes from the pre-analysis plan. In particular, in Section 5, we mix reporting pre-registered mechanisms, pre-registered exploratory research and post hoc exploratory research; and we classify these analyses in the report. We thank Elena Cettolin, Patricio Dalton, Daniele Nosenzo, Victor Gonzales Jiminez, Marco Schwarz, Ben Vollaard, seminar participants at the universities of Tilburg, Innsbruck, Wien, and Aarhus, the Annual Meeting of the Verein für Socialpolitik, and the CNEE Workshop for helpful comments. 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 some time before facing the task (early goals) when they are not yet tempted to shirk. But such personal goals are non-binding. Thus, when people actually face the task and the temptation to shirk, they may simply change their mind and revise their goal. This raises two empirical questions that the literature has not directly addressed (see the literature review below) and which we tackle in this paper: Are early goals designed as self-regulation tools? And are early goals effective in regulating behavior despite goal revision?

Our real-effort experiment mimics a typical work-leisure self-control problem by offering male subjects a generous piece rate for doing the tedious, unpleasant task of counting zeros in tables of zeros and ones.[1] 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 study whether subjects design early goals as self-regulation tools, and whether early goals are effective, we compare the goal and the effort in a treatment where subjects set a goal five days before the task (treatment *Early*) with a treatment where subjects set the goal immediately before the task (*Late*). Based on a stylized model in which the individual has present-biased preferences and sets goals, we predict that the individual tries to counteract his present-bias when setting a goal in advance of the task, but not when setting it right before the task.[2] Hence, subjects should set higher goals in *Early* compared to *Late*.[3] Further, as higher goals should translate into higher effort, we expect

[1] To maximize power for a data collection with a high cost per subject, we pre-registered to only recruit men (see Section 2.2 for details).

[2] 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 basic 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)

[3] Rather than being self-regulation tools, goals might just be expectations about effort or ordinary

a higher effort in *Early* than in *Late*. The latter hypothesis presumes that, when facing the task, the individual does not revise the early goal downward too much and/or cares at least to some extent about the early goal set days in advance. A higher effort in *Early* than in *Late* therefore suggests that goals are effective despite goal revision.

To examine precisely whether and to what extent subjects revise their goals, we implement two further treatments. Like in *Early*, subjects set a goal five days before the task. But now we explicitly allow subjects to revise their goal just before engaging in the task. Comparing the effort in these treatments, where there is explicit goal revision, with *Early* enables us to get a sense of the extent of private goal revision in *Early*. And comparing the effort in these treatments with the treatment *Late* enables us to test whether goals are effective despite (observed) goal revisions.[4]

A second contribution of our design is that it addresses the treatment migration problem that arises in most experimental studies on goal setting (see the literature review). The typical experiment has some subjects set a goal before working on the task (treatment condition), while others simply work on the task (control 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. Indeed, the results of Sackett et al. (2014) show that they do so. Consequently, a problem of treatment migration arises, because subjects in the control condition may nevertheless be exposed to the 'treatment' of setting goals. While these prior studies are valuable for learning whether explicitly eliciting personal goals has a beneficial impact on performance, the treatment migration problem means that the intention-to-treat estimate may understate the causal effect of goal setting. As parts of the literature on goal setting find insignificant effects or low effect sizes of goals on performance, addressing the treatment migration problem is important for understanding the extent to which goals are effective self-regulation tools. The comparison of

motivators that affect, for example, intrinsic motivation. In these cases, the goal schedule should be flat across dates, or possibly even increasing if subjects become more productive over time.

[4]One caveat is that explicitly asking subjects to revise goals may prompt more often goal revision than when such revisions are self-initiated. Hence, the comparison between these treatments may underestimate the effectiveness of early goals.

treatments *Early* and *Late* avoids the treatment migration problem because we observe the goals from the subjects in both treatments.

To preview the results, we find, first, that early goals are higher than late goals. This result is consistent with the hypothesis that subjects design early goals as self-regulation tools. When considering the treatments where we observe explicit goal revision, we also observe this pattern within subject: Subjects on average revise their early goal downwards. Second, early goals are effective self-regulation tools. Subjects who set an early goal work more compared to when they just set a late goal. Third, this result also holds for the treatments where subjects explicitly revise their goals. That is, having set an early goal and then revising it at the start of the task improves task performance relative to only setting a late goal.

Examining the mechanisms behind these results makes it clear that setting an early goal matters. A theoretically plausible mechanism is that the early goal serves as an anchor in goal revision in the way one would expect if changes in the goal triggered gain-loss utility – in analogy to the model of Kőszegi and Rabin (2009). Yet, somewhat surprisingly, the evidence goes against this mechanism for why early goals matter: We observe that the revised goal does not significantly differ from the late goal. That is, subjects almost seem to be behaving as if they set a new goal, rather than revising an old one. Nevertheless, early goals matter because people still seem to strive to some extent for their high early goal. Thus, they are more likely to achieve their revised goal than people who do not set an early goal – consistent with the view that both early and revised goals enter the reference point to which the individual compares performance.

The result that early goals are larger than late goals is consistent with the hypothesis that early goals are self-regulation tools. We address a number of possible alternative explanations behind this result, 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. Further, we test whether other goals than effort goals matter (e.g., income and time-based goals), but we find no evidence for this.

Finally, our design allows us to contribute to an additional research question: Can certain frames make goals more effective? Specifically, reminding subjects about a specific goal (either the revised or the early goal) should make that goal more salient and thus the subject may be more likely to strive for it. Similarly, explicit goal revision may make the revised goal more salient than private revision and thus lead to a lower effort. We test for these effects in two additional treatments, *Revise0* and *Revise1*, in which we explicitly provide subjects the opportunity to revise their goals and subsequently remind them about either the goal that they set at date 0 or date 1. We find that no matter which goal subjects are reminded about, the effort-goal relationship tends to be larger for the recent, revised goal than for their early 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, together with answers to our ex-post survey, provide evidence that private, self-initiated goal revisions do take place. This underscores that a proper understanding of the effort-goal relationship requires eliciting not only early goals but also revised goals, as we do in our study.

The paper proceeds 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 for alternative explanations behind the results. Section 6 discusses advantages and disadvantages of our design choices and outlines avenues for future research. Section 7 concludes the paper.

Related Literature. Our study relates to the literature that asks how goals influence performance. Industrial and organizational psychology studies on task performance in the work place laid the foundations for a vast literature on goals (cf. Locke and Latham, 1990, 2013, 2019). With employees traditionally operating with vague 'do-your-best' goals, research focused on examining whether employers can improve task performance with specific performance goals and by letting employees participate in setting these goals. Meta-analyses indicate that task performance increases with goal difficulty, is

higher for specific compared to 'do-your-best' goals, and is higher for participatory set or self-set goals compared to assigned goals (Epton et al., 2017; Tubbs, 1986; Mento et al., 1987; Chidester and Grigsby, 1984; Wood et al., 1987).

Next to comparing specific goals to do-your-best goals, a number of studies compare treatments where subjects choose non-binding goals themselves with a control treatment where no goals are elicited.[5] Most studies find that self-set goals have a positive effect on performance (Anshel et al., 1992; Erbaugh and Barnett, 1986; Fan et al., 2019; Goerg and Kube, 2012; McCalley and Midden, 2002; Schunk, 1985; Smith and Lee, 1992; Smithers, 2015; West et al., 2001),[6] but some do not (Akina and Karagozoglub, 2017; Goudas et al., 1999; Hayes et al., 1985; Hinsz, 1995; Tanes and Cho, 2013). This mixed picture arises also for studies that consider the effects of goals for the performance in repeated tasks, such as weight loss (Chapman and Jeffrey, 1978; Toussaert, 2016), energy saving (Harding and Hsiaw, 2014), or studying (Clark et al., 2020; 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.

While there is a large literature on goal setting and performance, less research has been done on goal revision. Sackett et al. (2014) elicit time goals two weeks prior to a marathon. They observe that eliciting goals increases performance relative to a condition where goals were not elicited. They suggest that asking runners two weeks before the task to explicitly state the goal locks them into their early, high goal, i.e., hinders goal revisions. Yet, they do not test for such goal revisions. Extant 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

[5]Goals can be made binding by tying them 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.

[6]In Smith and Lee (1992), 6 out of the 17 subjects in the no-goal treatment reported in an ex-post survey that they had set a goal. Consistent with the treatment migration problem described in the introduction, excluding the 6 subjects who had privately set a goal, the performance gap actually was larger for the goal treatments vs the no-goal treatment.

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 the economics literature, van Lent (2019) provides, to our knowledge, the only experimental study on goal revision. It is similar in spirit to the studies in psychology. As part of 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 is that we study the revision of goals prior to engaging in goal pursuit. This allows us to capture goal revision related to being tempted to work less when facing a task rather than goal revision due to 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 in a theoretical framework on which we build here. Some experimental studies address how fast new information is incorporated into the reference point, and their findings are mixed. 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. Similarly, Smith (2019) finds rapid adjustment to an exogenous change in current endowments. Nevertheless, the field data of Card and Dahl (2011), DellaVigna et al. (2017), and Thakral and Tô (2021) suggest slow updating of the reference point in other domains. Our contribution to the empirical evidence on updating of reference points is to provide evidence on the 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 binding plans on how to allocate effort over two dates that are a few days into the future; then, they again specify plans right before providing effort. The key difference to our study is that in their setting subjects are committed to a selected effort plan, so they have to provide

the stated effort to receive a completion bonus. In contrast, subjects make non-binding plans (expressed as goals) in our study, and we test whether such non-binding plans can motivate effort. 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) 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.

2 Experimental Design

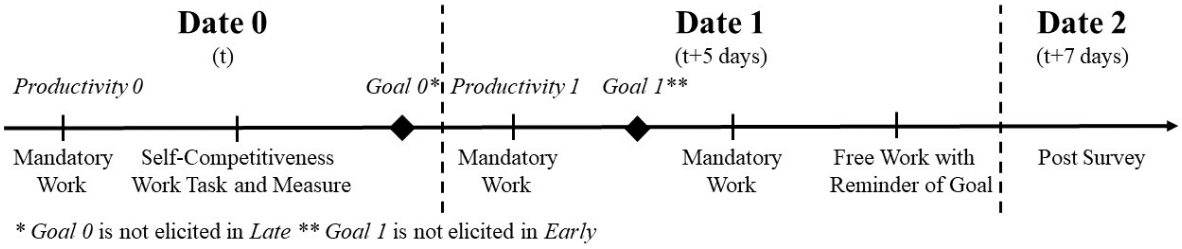
The experiment has three parts that are conducted online on three different days, using the Qualtrics platform: A goal setting part at date 0 (t), a work part at date 1 ($t+5$ days), and a post survey at date 2 ($t+7$ days). We randomize subjects into four different treatments. In treatments *Early, 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 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.11.

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

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

Figure 1: Timeline of the experiment



2.1 Details of the Experimental Setup

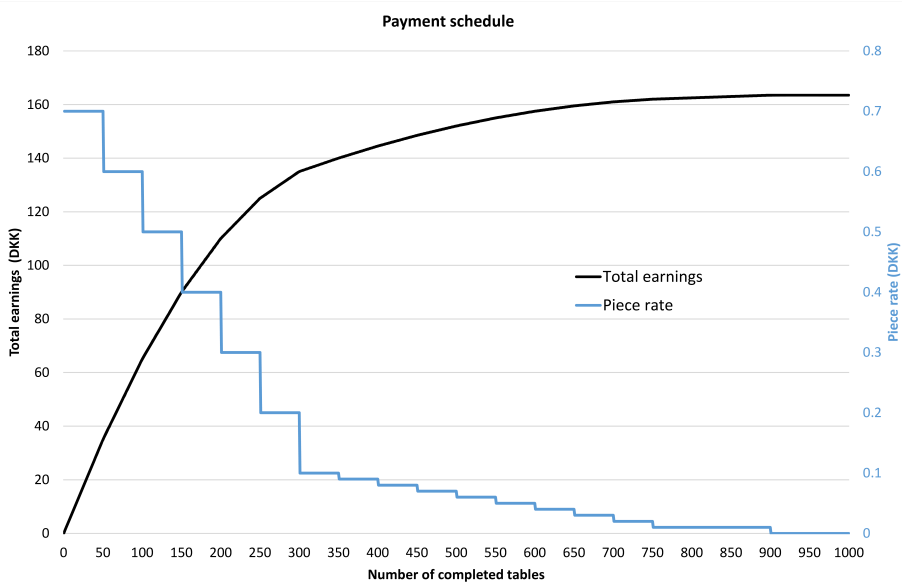
2.1.1 Date 0: Goal Setting

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 the date-0 part of the experiment, subjects in all treatments 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 in 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 provides us with 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

Figure 2: Payment scheme



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', so subjects in *Late* do not anticipate the connection between the tasks at date 0 and date 1.

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.[7]

In all the other treatments, we at this stage 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 so that the marginal benefit is decreasing 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

[7]Subjects fill out the consent form (see Online Supplement S.11) 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.

work phase (*goal 0*). That is, goals are self-set, but engagement in goal setting is exogenously induced, and all subjects state a goal. 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.

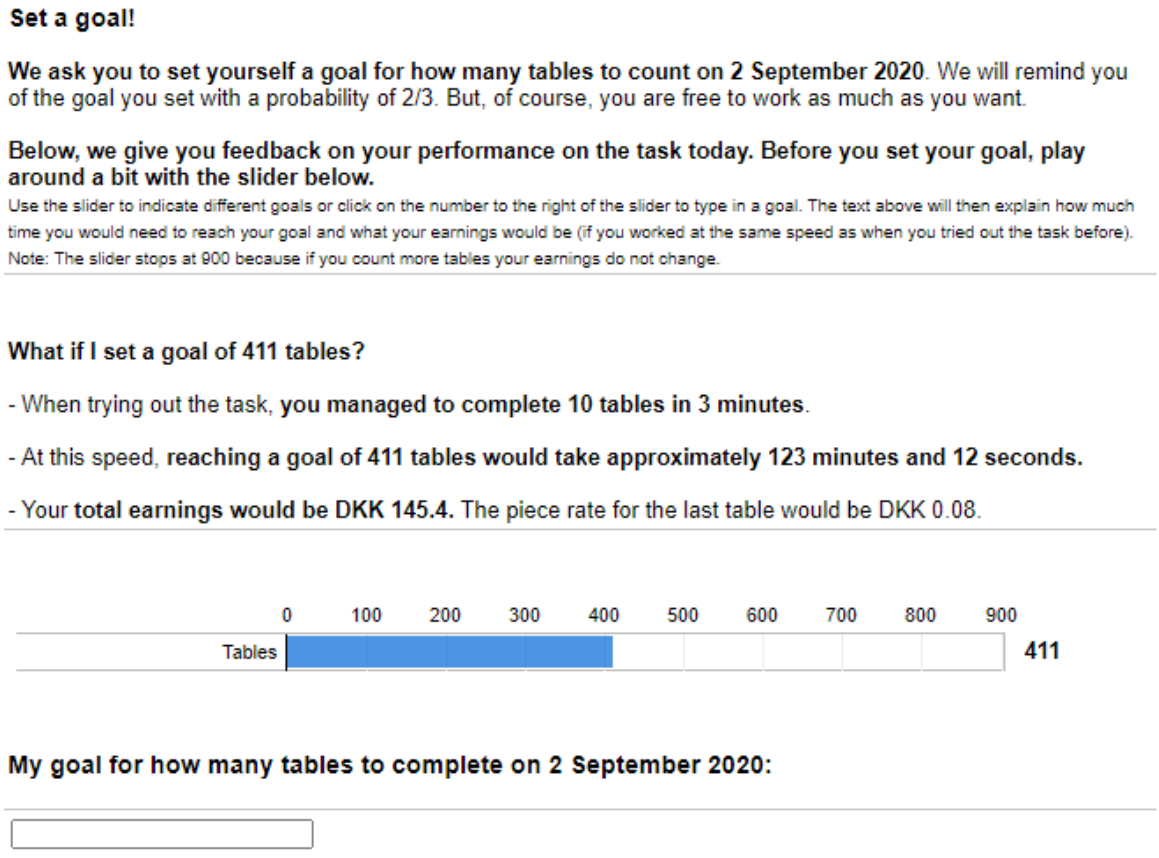
Before setting goals, subjects have access to a slider tool that should help them to reflect about how much time it would take them to achieve a certain goal (see Figure 3). Subjects indicate goal values with the slider, and the tool shows the estimated amount of time for reaching the goal (based on the productivity of the subject) along with the associated earnings and the marginal piece rate.[8] 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 – the probability reflecting the random assignment to treatments that takes place after setting goals. 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

[8]There is a small difference between the mandatory work phase 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. But since it only takes a few milliseconds to reload the page, it is unlikely that this difference drives goal non-achievement. As the free work phase takes place after goals have been set/revised, this issue cannot drive goal revisions.

Figure 3: Slider tool and goal setting



schedule for date 1. Further, we ask them how likely they think it is that they will end up having less than two hours of flexible time at date 1. These questions serve two purposes: First, they should make 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, but they can freely choose whether and 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 provide us with 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 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. Further and similar to date 0, we ask subjects in this break 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.[9] 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 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

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

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 early 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. They 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.11) consists of several questions about goal setting and goal commitment; both specific to the experiment and in general. In particular, it gives us an indication to what extent subjects in *Early* privately revised their goals and allows us to check that subjects in *Late* did not anticipate the free work task for date 1 and set a goal before date 1. 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 smaller or null for women (cf. Koch and Nafziger, 2020; Smithers,

2015; Clark et al., 2020). At the same time, our study has a high cost per participant. Thus, to achieve an appropriate power for the given budget, we only recruited men for the experiment (see Section 6 for a further discussion and Online Supplement S.3 for the power analysis).

We recruited subjects from the subject pool of the COBElab at Aarhus University and, during the COVID-19 lockdown, also from the general student population in the four largest Danish cities. In total, we recruited 499 subjects. Of these, 394 completed the date 0 part of the study, and 326 reached the free work part at date 1. A total of 277 subjects also completed the post survey (date 2), which we primarily use for exploratory research.[10]

Our main sample consists of the 326 subjects who reached the free work part at date 1. Out of these, 192 (59 percent) were bachelor students, 71 (22 percent) were master students, 10 (3 percent) were PhD or other types of students, and 53 (16 percent) were not students. Most students came from the largest study programs; Business and Economics (126 subjects, 39 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 tables into a

[10]The ethics rules of COBElab at Aarhus University did not allow us to enforce participation in all parts of the study by making all payments conditional on the completion of the post survey. We incentivized participation in the post survey by paying far more than the average student wage. Given the fixed budget, a further increase in incentives would have implied lower payments for earlier parts and possibly attrition problems there. In Section 5 we address attrition in more detail.

spreadsheet program to do the counting, we copy-protected tables.

We collected data November-December 2019 and March-May 2020. The break during the exam period in January and February ensured similar working conditions for all participants. Subjects knew that 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

In the following, we develop a stylized theoretical model to form hypotheses for our experiment. We provide the formal analysis of the model in Appendix A.

3.1 Model

Task and Preferences. We derive our predictions in a theoretical model where people have present-biased preferences (Laibson, 1997). We allow for partial naïveté (O'Donoghue and Rabin, 1999). That is, the individual is aware of facing a self-control problem but not necessarily of its full extent. 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. The individual might be fully or partially naïve about his present bias; that is, he holds a belief $1 \geq \hat{\beta} \geq \beta$ about his present bias. 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. Consistent with the evidence from psychology on goals (e.g. Heath et al., 1999; Locke and Latham, 2002; Wu et al., 2008) and building 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 $z = e - g$, the individual experiences a corresponding comparison utility $\mu(z) = z$ for $z < 0$, and $\mu(z) = 0$ for $z \geq 0$.[11]

Self 1 may revise the goal g_0 at date 1 to g_1. We assume that an individual who (possibly) revises the goal at date 1 then compares the early 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 early 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 early and the revised goals are 'sticky' in the sense that a

[11] As in Koch and Nafziger (2020), comparison utility is defined over effort and we abstract from gains from overachieving the goal for reasons of parsimony. See Koch and Nafziger (2016) for a model where the individual experiences comparison utility over both gains and losses in the benefit and cost domains.

combination of both goals enters the reference point to which the individual ultimately compares exerted effort. Specifically, 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 and because it is a plausible scenario, we assume for expositional purposes that at the goal revision stage the individual is myopic about the stickiness of the original goal g_0. Consequently, he does not take stickiness into account when setting the revised goal g_1 and thinks believes that he will compare effort only to g_1. This amounts to assuming $\hat{\lambda} = 0$. 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 (relaxed in Online Supplement S.1.2), (ii) self 1 (at the goal revision stage) is naïve about the stickiness of goal 0 (belief $\hat{\lambda} = 0$; relaxed in Online Supplement S.1.1), and (iii) the individual might be fully or partially naïve about his present bias ($1 \geq \hat{\beta} \geq \beta$). Consequently, the individual may set 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 presented 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 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. What effort level self 0 thinks he can implement depends on his (possibly naïve) belief about his 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 early goal at date 1, the individual discounts the future benefit with the true present bias – in contrast to self 0. This is the case in *Revise0* and *Revise1*, where subjects can revise their goal before providing effort. In *Early*, they possibly revise their goal privately. 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. This is similar to the loss one feels when failing to reach a goal, but the loss from goal revision possibly has less weight than the loss from actually falling short of the goal. Hence, loss aversion is a weaker motivator in the goal revision stage than at the effort stage. 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 early goal set at date 0. Yet, the revised goal is still higher than the goal that self 1

would choose in *Late*.

Summing up, we test for the wedge between goals set at date 0 and goals set at date 1, both with a between-subject comparison (*Early* vs. *Late*) and a within-subject comparison (*Revise0* and *Revise1*). If goals set at date 0 are larger than those set at date 1, we speak of early goals as self-regulation tools. 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 early goals reflect self-regulation intentions.

Hypothesis 1 *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 early 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, individuals do not only design early goals as self-regulation tools, but they are also effective – despite goal revision.

As effort may differ between *Early, Revise0*, and *Revise1* (see Hypothesis 3), we test the hypothesis that early 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 early goal setting is 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 early goal at date 0 but those in *Late* do not.

Hypothesis 2

1. *Subjects provide more effort in Early than in Late.*

2. *Subjects provide more effort in Revise1 than in Late.*

If framing of goals and goal revision influence the salience of goals, then we should see differences in effort between treatments *Early*, *Revise0*, and *Revise1*. By random assignment to treatments *Revise0*, *Revise1*, and *Early*, the early and revised goals should not differ between these treatments.[12] Yet, the salience of the early and revised goals may differ in these treatments. First, it seems plausible that the goal that is displayed while working on the task is the most salient. Second, making goal revision explicit in *Revise0* and *Revise1* may result in greater salience for the revised goal than does the (possible) private goal revision in *Early*. Thus, we assume salience weights $0 = \lambda^{Late} \leq \lambda^{Revise1} < \frac{1}{2} < \lambda^{Revise0} < \lambda^{Early} \leq 1$.[13] Because (by Hypothesis 1.2) the early goal is larger than the revised goal, the reference point g^* should be higher in *Early* than in *Revise0* and higher in *Revise0* than in *Revise1*. As higher effective goals result in higher effort, we hence expect a higher effort in *Early* than in *Revise0*, and we expect a higher effort in *Revise0* than in *Revise1*.

Hypothesis 3

1. *Subjects provide more effort in Early than in Revise0.*

2. *Subjects provide more effort in Revise0 than in Revise1.*

[12] 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 extent) than subjects in *Early* do. The prediction below remains valid if this is the case.

[13] 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. Moreover, a reminder makes an attribute salient (e.g., Karlan et al., 2016).

Table 2: Hypotheses and summary of findings

	Hypothesis	Finding	Effect size
Goals are self-regulation tools			
H1.1	goal 0^{Early} > goal 1^{Late}	✓	.229
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+	$\text{effort}^{Early,Revise0,Revise1}$ > effort^{Late}	✓	.263
H2.2	$\text{effort}^{Revise1}$ > effort^{Late}	✓	.399
Framing can make goals more effective			
H3.1	effort^{Early} > $\text{effort}^{Revise0}$	✗	-.075
H3.2	$\text{effort}^{Revise0}$ > $\text{effort}^{Revise1}$	✗	-.146

Standardized effect size: Hedge's g_p based on mean comparisons. H2.1+ pools all treatments with 'early' goal 0, as framing effects under Hypothesis 3 are rejected.

4 Empirical Analysis

In this section, we first describe the main variables and the analysis plan. Then, we test our primary hypotheses (H.1-H3) by comparing effort and goals in the different treatments. Finally, we comment on the robustness of the results. In Section 5, we examine possible mechanisms and discuss alternative explanations that could influence the results, such as alternative reference points, learning, uncertainty, and time shocks. Tables and Figures with prefix S. are in the online supplement.

4.1 Main Variables and Analysis Plan

Our main outcome variables are *goal 0* (the goal set at date 0, except in *Late*), *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 follow the pre-analysis plan and 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 (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 non-parametric Mann-Whitney U-test (MWU) to examine differences in effort between treatments. Note, however, that a lack of control for productivity makes the

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

Averages are taken over all subjects in a treatment. The share of subjects achieving *goal 0* and *goal 1* is calculated by averaging over the indicator that is one (zero) if the subject achieves (does not achieve) the respective goal. A technical error prevented recording the time schedule for 21 subjects in *Late*. In our later analyses, these subjects are excluded when controlling for whether subjects are time constrained. Our results are qualitatively robust to including these subjects throughout or excluding them entirely.

MWU and Wilcoxon signed-rank tests ill-suited for comparing goals in Hypothesis 1 as these do not take into account the increase in productivity between dates 0 and 1 that mechanically increases goals. Throughout, we report p-values for two-sided tests. Standardized effect sizes are summarized in Table 2. In Section 4.4, we discuss multiple hypothesis correction for the p-values.

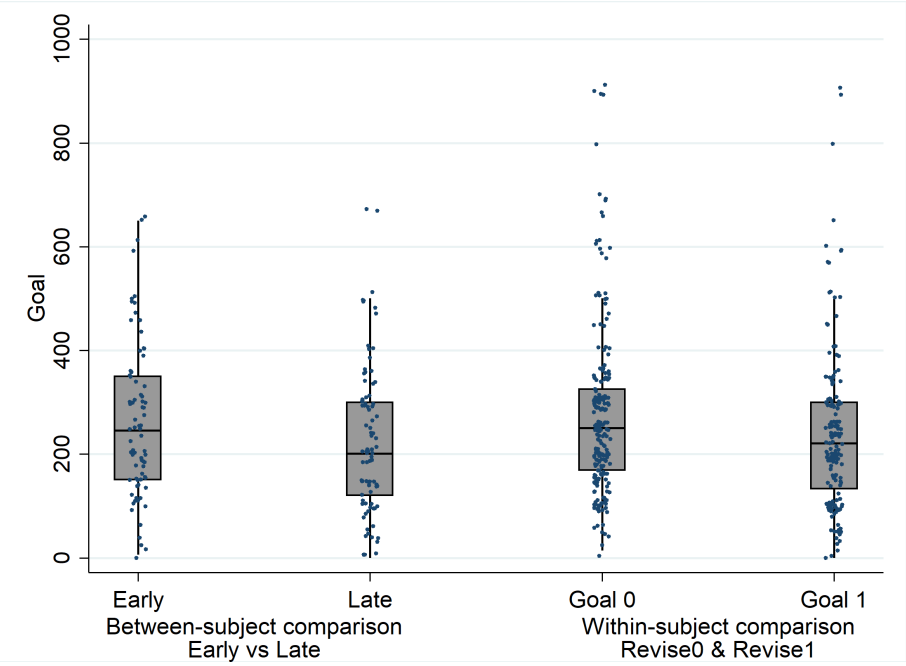
4.2 Goal Setting (Test of Hypothesis 1)

In line with Hypothesis 1.1, the goal that subjects set in *Early* is on average 34 tables higher than the goal that subjects set in *Late*. This difference is statistically significant when we control for the baseline productivity of subjects at the time of goal setting ($p = .01$, cf. Specifications (1)-(3) in Table 4). To understand why controlling for productivity is part of the pre-registered hypothesis despite random assignment to treatments, note that average productivity increases due to experience (cf. Table 3). This increase works against our prediction because it tends to increase the *Late* goal. Productivity explains 8 percent of the variance in goals between treatments.[14]

Similarly, visual inspection of Table 3 and Figure 4 indicates that subjects in the *Revise0* and *Revise1* treatments revise their early goal downwards at date 1. Figure 5 shows the extensive margin of goal revision (64 percent of subjects revise their goal) and the intensive margin (box plot of *goal 1-goal 0*). Conditional on goal revision occurring, the

[14]The effect sizes for Hypothesis 1 in Table 2 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: Goals set by subjects

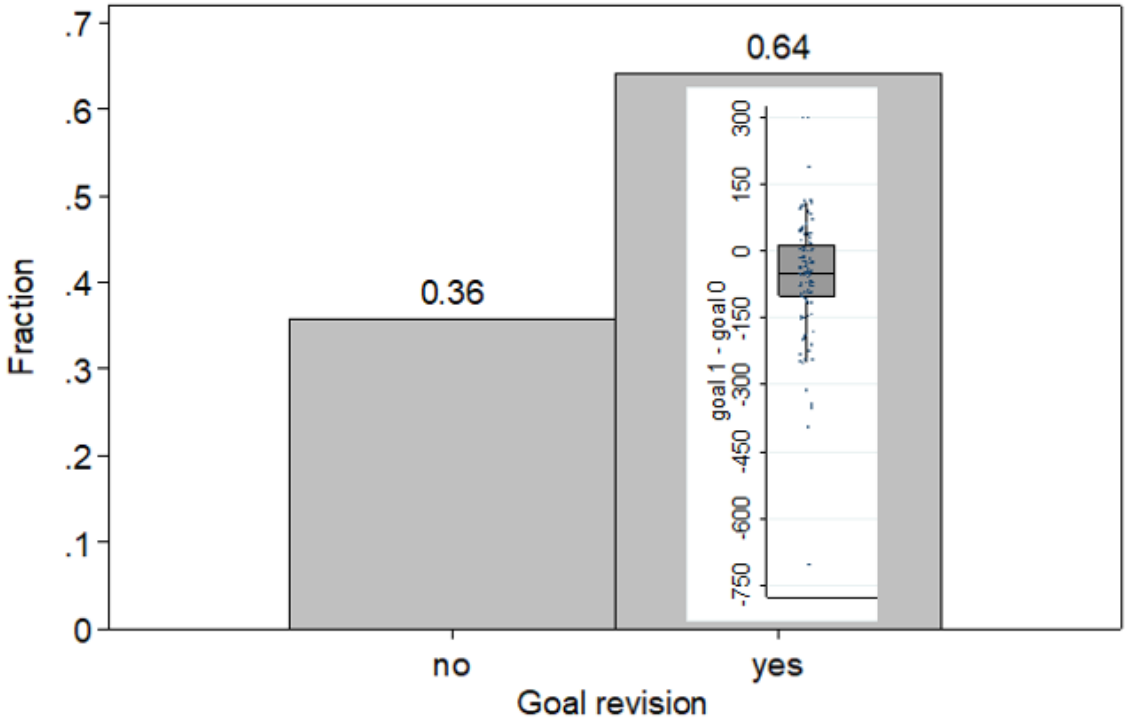


Box plots show the median as well as upper and lower quartiles of goals in the data. Spikes extend to the largest or smallest values within 1.5 times the upper or lower quartiles, respectively.

average subject revises his goal downward by 56 tables (49 tables when excluding the outlier with goal revision -700). In line with Hypothesis 1.2, we observe in a within-subject comparison that *goal 1* is significantly smaller on average than *goal 0* ($p < 0.01$, cf. the intercept in Specifications (4) and (5) in Table 4; results are robust to adding controls – notably productivity – and to excluding outliers, cf. Table S.10).[15] 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). Our results indicate that while most subjects have time-inconsistent goals, some people do behave in a time consistent manner.

[15]In Specifications (4) and (5) in Table 4, the intercept shows the average (fitted) goal revision in the case of no change in productivity. In Specification (6), however, it has a different interpretation since the regression is estimated with the full set of controls. This intercept is not informative about the overall difference between the goals but instead captures the difference for a distinct baseline (including variables in a within-subject comparison that do not change between date 0 and date 1). Hence, it sheds light on possible mechanisms, which we return to in Section 5.

Figure 5: Goal revision in *Revise0* & *Revise1*



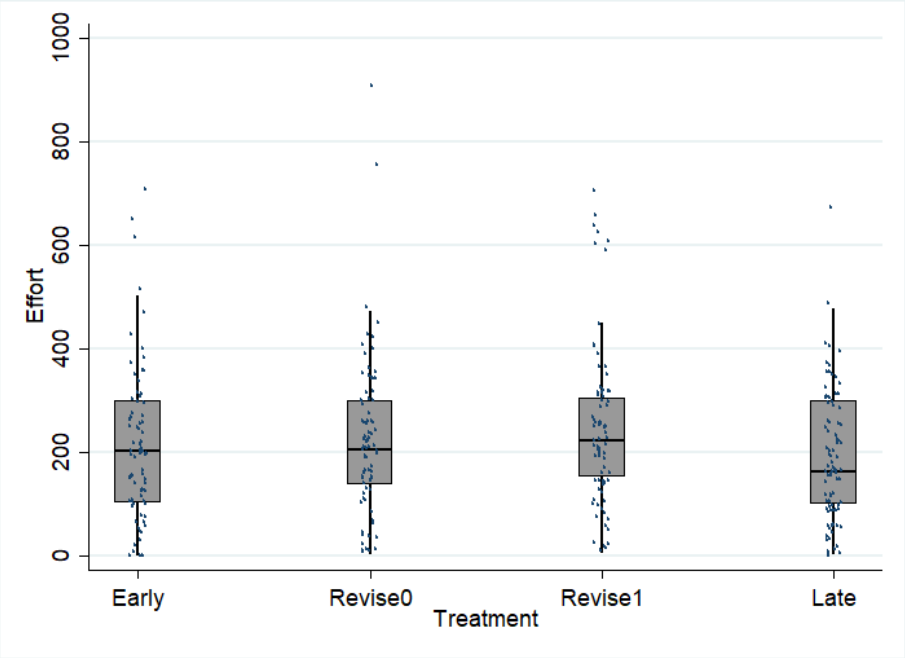
Extensive margin: Bar chart. Intensive margin: Box plot of the within-subject differences between goal 1 and goal 0, conditional on goal revision. It shows the median as well as upper and lower quartiles. Spikes extend to the largest or smallest values within 1.5 times the upper or lower quartiles, respectively.

Table 4: Goal setting

	Hypothesis 1.1 (goal): Early vs. Late			Hypothesis 1.2 (goal 1 -goal 0): Revise0 & Revise1		
	(1)	(2)	(3)	(4)	(5)	(6)
Late	-33.53	-58.47**	-63.97***			
	(22.90)	(22.44)	(23.24)			
Productivity		8.45***	7.39***			
		(2.00)	(2.23)			
Change in productivity[a]					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	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*. Specifications: (1)-(3): OLS regressions of the dependent variable on a treatment dummy (that is equal to one if the subject was randomly assigned to treatment *Late* and zero otherwise) and (1) a constant; (2) a constant and productivity (which refers to baseline productivity at the date when the goal was set); (3) a constant, productivity, and the set of controls listed in Online Supplement S.2. See Table S.7 for coefficients on the controls. The sample size in (3) is smaller because a technical problem prevented recording the time schedule control for 21 subjects in *Late* (see note in Table 3). (4)-(6): Within-subject comparison using OLS regressions of *goal 1 - goal 0* on (4) a constant; (5) a constant and change in productivity (*productivity 1 - productivity 0*); (6) a constant, change in productivity, and the set of controls. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.
[a] Productivity at date 1 - productivity at date 0.

Figure 6: Effort exerted by subjects



Box plots show the median as well as upper and lower quartiles of effort in the data. Spikes extend to the largest or smallest values within 1.5 times the upper or lower quartiles, respectively.

Table 5: Comparison of effort across treatments

	(1)	(2)	(3)	(4)	(5)	(6)
Hypothesis 2.1: Early vs. Late						
	-23.10	-24.05	-28.10	-2.78	-4.29	-1.34
	(21.41)	(20.72)	(23.24)	(16.18)	(15.94)	(17.12)
Mean level of effort in Early: 212.87						
Hypothesis 2.1+: Early & Revise0 & Revise1 (pooled) vs. 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)
Mean level of effort in Early & Revise0 & Revise1 (pooled): 227.89						
Hypothesis 2.2: Late vs. Revise1						
	56.62**	48.68**	50.43**	42.71***	39.56**	34.73**
	(22.09)	(20.35)	(21.21)	(16.11)	(15.74)	(16.03)
Mean level of effort in Late: 189.77						
Hypothesis 3.1: Early vs. Revise0						
	11.08	-1.95	-9.30	2.42	-5.66	-9.09
	(23.44)	(22.55)	(22.69)	(19.67)	(19.71)	(20.90)
Mean level of effort in Early: 212.87						
Hypothesis 3.2: Revise0 vs. Revise1						
	22.44	29.87	31.51	38.78**	41.56**	41.85**
	(24.06)	(22.02)	(20.89)	(19.14)	(18.29)	(18.29)
Mean level of effort in Revise0: 223.95						
Productivity	No	Yes	Yes	No	Yes	Yes
Other controls	No	No	Yes	No	No	Yes
Goal	No	No	No	Yes	Yes	Yes

Coefficients for the treatment mentioned last (with the treatment mentioned first as base category) in OLS regressions with effort as dependent variable. The first column reports the coefficient in a regression where no further controls are added, the second where productivity is added as a control, and the third where all controls are added (cf. Online Supplement S.2). See Table S.8 for coefficients on the controls. The last three columns control also for the goal level shown in the free work phase: *goal0* in *Early* and *Revise0*, *goal1* in *Late* and *Revise1*. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$. H2.1+ pools all treatments with 'early' goal 0, as framing effects under Hypothesis 3 are rejected.

4.3 Effort Provision

Visual inspection of Table 3 and Figure 6 indicates that effort in *Late* is lower than effort in the treatments where subjects set an early goal, but there appears to be little difference in effort between *Early, Revise0*, and *Revise1*. While the former pattern is in line with the view that early goals are effective self-regulation tools (Hypothesis 2), the latter pattern goes against the predictions regarding the framing of goal revision or goal reminders (Hypothesis 3). We test each of these in turn and report the results in Table 5 (results are robust to excluding outliers, cf. Table S.11).

4.3.1 Test of Hypothesis 2

Regarding Hypothesis 2.1, we find that effort indeed is larger in *Early* than in *Late* (23 tables on average), but this difference is not statistically significant ($p = .281$, cf. Table 5, MWU: $p = .405$). The effect size of .169 is meaningful, but our ex ante power analysis suggests that we are not sufficiently powered to detect effects of this magnitude (see Online Supplement S.3). Yet, our additional treatments with an early goal allow us to perform a more highly powered test of the hypothesis, as explained below. Subjects exert significantly more effort in *Revise1* than in *Late* ($p = .011$, cf. Table 5, MWU: $p = .023$). This is in line with Hypothesis 2.2.

As already indicated, we do not find any significant differences in effort between the treatments where subjects set a goal at date 0: *Revise0, Revise1*, and *Early* (cf. also Section 4.3.2). 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 analysis of Hypothesis 2.1, which is also in line with our theoretical model (but not pre-registered): By pooling the three treatments where an early goal is set, we increase the power to pick up also more modest effect sizes when testing the effectiveness of early goals. If setting early goals is 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.

Further, we observe that the coefficients for the *Late* treatment dummy become smaller and insignificant when goals are added (cf. Specifications (4)-(6) in Table 5), suggesting that goal levels are the mediator (a Wald chi-square test for equality of the treatment dummy across model specifications 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 it is consistent with earlier work on self-set, non-binding goals (e.g., Koch and Nafziger, 2020).

4.3.2 Test of Hypothesis 3

Regressions confirm the observation from Figure 6 that there are no differences in effort both between *Early* and *Revise0* and between *Revise0* and *Revise1* (cf. Table 5), leading us to reject Hypothesis 3. However, this rejection is not a threat to the overall hypothesis that setting early goals is an effective self-regulation tool: Hypothesis 3 relies on assumptions on exogenous parameters that are not central for the theory. Thus, the rejection only shows that certain frames cannot make goals more effective. Namely, the results suggest that it is irrelevant what goals 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*. Across separate effort regressions for *Revise0*, the coefficient on *goal 0* (.421; Specification (1) in Table S.9) is borderline significantly smaller than the coefficient on *goal 1* (.686; Specification (7); 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).[16] For *Revise1*, the coefficient on *goal 1* (.680; Specification (10) in Table S.9) is larger than on *goal 0* (.618; Specification (4)), 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

[16] 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.

Revise0 and *Revise1*; reflected by an insignificant difference across treatments between the coefficients on *goal 1* ($p = .638$; Specifications (7) and (10)).

The non-significant difference in effort between *Early* and *Revise0* suggests that explicitly asking subjects to revise their goal does not matter for effort. One plausible explanation for this is that subjects in *Early* privately revise their goals and that such privately updated goals are as important as explicitly updated goals. Exploratory analysis of the responses from the post survey support this explanation.[17] Among the 64 responses 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 goal ($p = .150$).

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 (cf. Tables S.27 and S.30). The piece rate being zero for any effort larger than 900 suggests that any goal or effort beyond 900 is irrational. Results are robust to excluding subjects who set a goal equal to or larger than 900 (3 subjects) or provide an effort equal to or larger than 900 (1 subject) in a pre-registered robustness check (cf. Tables S.10 and S.11).

Considering other outcome variables, namely average mistakes or time spent per table, we find no difference between *Late* and the other treatments (cf. Table S.12). However, these variables do not correlate strongly with effort ($r = -.256$ and $r = -.449$, respectively), which suggests that they might not be appropriate proxies for effort. For example, if a subject counts more tables, such effort may increase mistakes due to fatigue. And the impact on time spent is unclear as subjects who exert much effort in

[17] 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.

counting tables may be fast (proficient) or slow (attentive) in doing so.

Table 6: Multiple hypothesis corrected p-values

Unadjusted	FWER-adjusted[a]	FDR-adjusted[b]
H1.1: goal 0^{Early} > goal 1^{Late}		
.0100	.0587	.0351
H1.2: goal $0^{Revise0\,\&\,Revise1}$ > goal $1^{Revise0\,\&\,Revise1}$		
<.0001	<.0001	<.0001
H2.1: effort^{Early} > effort^{Late}		
.2423	.4860	.2827
H2.1+: effort$^{Early\,\&\,Revise0\,\&\,Revise1}$ >effort^{Late}		
.0412	.1548	.0721
H2.2: effort^{Revise1} >effort^{Late}		
.0160	.0777	.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.
H2.1+ pools all treatments with 'early' goal 0, as framing effects under Hypothesis 3 are rejected.

Multiple Hypothesis Testing. We present our findings in Section 4 without multiple hypothesis correction because the hypotheses are highly interdependent. In particular, 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).

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. Note that the power for the joint comparison of our 7 main hypotheses is below that for the

individual tests. Thus, the experiment-wise type II error rate is inflated and some tests might turn up insignificant even though the theory is correct. Our main results remain at least borderline significant with one or both types of corrections.

5 Mechanisms

In the following, we consider possible mechanisms for why early goals are effective self-regulation tools despite goal revision. We start by mechanisms that are based on our model. 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 subject's effort and goal. To estimate the marginal effect that a treatment has on the probability of reaching a goal, we use a binary goal achievement variable for *goal 0* and *goal 1* (equal to one if effort $\geq$ goal and zero otherwise). Table 3 provides descriptive statistics.

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

We now examine more closely why setting early goals is effective despite the fact that subjects revise such goals.

Does the Early Goal Serve as a Reference Point in Goal Revision? We developed our hypotheses based on a theoretical model where the individual has the early goal in mind and experiences loss utility if he revises the goal downward. Because early goals are higher than late goals, goal revision should not go all the way down to the level of what the late goal would have been. That is, the theory offers the between-subjects prediction that $goal\ 1^{Revise} > goal\ 1^{Late}$. We do find a positive difference across the treatments in the goals set at date 1 ($goal\ 1^{Revise\,0\,\&\,Revise\,1} = 241.83$ and $goal\ 1^{Late} = 229.01$), but this difference is not statistically significant ($p = .562$, cf. Table S.13). This result suggests that an individual experiences no substantial loss utility when revising the goal.

In terms of our model, this suggests that the parameter ν is close to zero. Indeed, if setting an early 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 5).

How do the Early Goal and the Revised Goal Matter? The theory allows for another channel through which the early goal impacts effort. Both the early and the revised goals are assumed to be 'sticky' in the sense that the individual compares exerted effort to a reference point that is a function of the early goal and the revised goal. The model used the parsimonious assumption that these two goals enter as a linear combination of the early and the revised goals – predicting underachievement of the early 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 ($p = .001$, cf. Specification (1) in Table S.14). 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 ($p = .470$, cf. Specification (4) in Table S.15).

Our findings thus indicate that the early and revised goals do not enter as a linear combination to form a reference point, as assumed in the theoretical model. However, the early goal still seems to affect the reference point that the individual has in mind when working: Setting an early goal does increase effort and it makes subjects more likely to achieve their revised goal (even though it does not affect the level of the revised goal). Specifically, while subjects in *Revise0* and *Revise1* on average achieve their *goal 1*, subjects in *Late* on average fall 39 tables short of their *goal 1* ($p < .001$).

An effective goal given by $g^*(g_0, g_1) = a\, g_0 + b\, g_1$, with $0 < a < b$, could rationalize these findings. In *Late*, only the $b\, g_1$ part is relevant, and if $b\, g_1 < g_1$, the individual does not achieve *goal 1* in *Late* (see also Section 5.2.4 on the rationality of goals). But if $g^*(g_0, g_1) = a\, g_0 + b\, g_1$ is close to $g_1 < g_0$ (requiring $a\, g_0 > b\, g_1$, which can hold if g_0 is

large enough) the individual achieves *goal 1*, but not *goal 0* in *Revise0* and *Revise1*. In sum, while early goals do not appear to serve as a reference point in goal revision, they matter because individuals appear to still strive for them to some extent even after goal revision.

5.2 Alternative Mechanisms

We derived our predictions in a model where individuals are present biased. Further, in this model, we assumed that goals serve as reference points measured in the effort dimension and that goals are (quasi-)rational. Consequently, the model predicts that treatment differences in effort should relate to treatment differences in goals, and our main results are consistent with this. Yet, other mechanisms could potentially drive the results, and we test for and discuss these in the following.

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. In the following, we examine alternative explanations to present bias for downward goal revision in the *Revise* treatments.

Uncertainty and Time Shocks. At both date 0 and date 1, after taking the productivity measures and before setting goals, we ask subjects to fill out their (expected) time schedule for date 1. This allows us to examine the influence of unexpected time shocks (arising, for example, from unforeseen contingencies or from mispredicting future time pressure as described by the planning fallacy of Kahneman and Tversky, 1979) and resolution of uncertainty about flexible time.[18] Notably, the downward goal revision re-

[18]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 the planned to the 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.

mains statistically significant if we control for uncertainty and time shocks ($p < .001$, cf. Specifications (2) & (3) in Table S.16) and if we restrict our analysis to those 87 percent of subjects who experience no time shocks ($p = .001$, cf. Specification (1) in Table S.16). 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 *goal 0* and *goal 1*.

Learning and Overoptimism. Another potential mechanism that could drive a wedge between *goal 0* and *goal 1* is some sort of learning. Consider first *learning about how to perform the task*. If the individual learns and gets better at the task (as reflected by the productivity measures) *goal 0* should be smaller than *goal 1*, which is the opposite of our result. When we account for changes in productivity between dates 0 and 1, there even is suggestive evidence that the difference between *goal 1* and *goal 0* increases (Wald chi-square test for equality of constants across models, $p = .012$).

Consider next *learning about the cost of the task*. When setting their early 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 (e.g., Acland and Levy, 2015). Previous research found that experience did not affect goal setting or effort in a closely related setting (cf. Koch and Nafziger, 2020). Nevertheless, to limit the possible influence of experience at date 1 compared to date 0, our design gives subjects only 3 more minutes of experience with the task when setting a goal at date 1. And our design ensures that subjects have fresh experience when setting a goal, because they perform the task immediately prior to setting *goal 0* or *goal 1*.

In the analysis, we proxy for learning about the costs by looking at subjects' (retrospective) enjoyment of the task reported in the surveys at dates 0 and 2, respectively.[19] Among 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 ($p = .001$ for both t-test and Wilcoxon signed-rank

[19] A caveat is that pleasure in the task in the post survey is possibly affected by other factors like satisfaction with the entire experiment, own performance/goal achievement, the payment obtained, and earnings per hour.

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.17). Thus, overall, there is little evidence for this alternative explanation for the observed downward revision of the goal.

Finally, the individual also may set a higher goal at date 0 than at date 1 because *goals reflect expectations* and the individual is overly optimistic at date 0 about the productivity gains from date 0 to date 1. Accordingly, an individual may revise the goal downward when these expectations are disappointed at date 1. This interpretation would imply that the individual reacts differently to the feedback about the productivity at date 0 and date 1. If the individual was overconfident in this way, the individual would boost the productivity feedback at date 0 when setting goal 0 to be, say, $z \cdot productivity\, 0$. But when the productivity feedback at date 1 reveals that the expected productivity gains failed to materialize, goal 1 is set with no (or lower) anticipated further increases in productivity at $y \cdot productivity\ 1$, where $y < z$. That is, when regressing *goal 0* on *productivity 0*, the coefficient on productivity should be higher than when when regressing *goal 1* on *productivity 1* in *Revise0* and *Revise1*. We observe the opposite with the coefficient on *productivity 1* (8.73, $p = .001$) being larger than on *productivity 0* (5.28, $p = .126$; cf. Table S.18), but they are not significantly different from each other (Wald chi-square test for equality of coefficients across models, $p = .173$).

5.2.2 Reference Point Formation

Alternative Reference Points. A possible problem in inducing effort-based reference points through goals arises if people have time- or earnings-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").

One way to see if subjects have a time or earnings reference point is to use the fact that

the slider tool at the goal setting stage translates effort into a time- and an earnings-equivalent, and 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.9), to which the thresholds for changes in the piece rates also contribute. Yet, we see no clear bunching on focal numbers in the time or earnings dimensions (cf. Figures S.10 and S.11). This suggests that subjects indeed primarily focus on effort goals.

Nevertheless, we can replicate our analysis of the treatment effects on goals also in terms of time- and earnings-equivalents (cf. Table S.19). According to the goals that they set, subjects in *Late* aim to work about 17 minutes less than subjects in *Early*, a 30 percent drop in work time. For the earnings-equivalents, we observe that goals imply DKK 9-18 lower earnings in *Late* compared to *Early*, corresponding to an 8-16 percent gap. Comparing *goal 1* and *goal 0* in the time- and earnings dimensions in the *Revise0* and *Revise1* treatments reveals goal revision of a similar extent.

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 yet know the tasks they are to perform at date 1. 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* (on average 1.5 tables fewer than *goal 1*, which is not significantly different from zero, t-test, $p = .36$).[20]

[20] A possible explanation is that subjects make up a post hoc *goal 0* and use *goal 1* as an anchor. Noise in memory is also apparent in *Early*, where the goal that subjects recall to have set at date 0 is on average 4 tables smaller than their actual *goal 0*.

5.2.3 Alternative Explanations for Treatment Differences in Effort

In the regressions, effort levels are significantly related to goals (cf. Table S.20). Further, the difference in effort between *Early/Revise0/Revise1* and *Late* becomes small and insignificant when controlling for the goal displayed whiled working (*goal 0* in *Early, Revise0* and *goal 1* in *Revise1, Late*), in line with the explanation that this goal drives the treatment differences (cf. Table 5). When controlling for *goal 1* and *productivity 1*, the treatment difference between *Revise1* and *Late* is significant ($p = .013$; cf. Table S.21). This result may arise because subjects in *Revise1* also strive for *goal 0*. Indeed, the treatment difference becomes insignificant when controlling for the first goal that subjects set in the two treatments (*goal 0* in *Revise1* and *goal 1* in *Late*) and productivity ($p = .120$; cf. Table S.21). Altogether, these results are thus in line with the theoretical prediction that differences in goal levels explain treatment differences in effort.

In the following we discuss robustness checks that we performed to examine possible other factors than goals that may also influence treatment differences in effort (for details see Online Supplement S.6).

First, a concern might be that learning about the task and setting goals early vs. late could influence attrition and in doing so affect treatment differences. Yet, we find no evidence for this.

Second, setting goals in advance could increase how meaningful the task appears (Hackman and Oldham, 1976; Grant, 2008) or prompt additional practicing. Yet, the data on enjoyment and changes in productivity provide no evidence for this channel.

Third, being asked to reflect twice about the goal could increase goal commitment compared to only setting it once. Yet, we do not find evidence that setting a goal twice (as in *Revise1* or *Revise0*) matters in contrast to setting a goal only once (as in *Early* and *Late*).

5.2.4 Are Goals Rational?

The (quasi-)rationality assumption in our model means that subjects set goals that they believe they will achieve and that they maximize their utility when setting goals.

Consistent with setting rational goals, the majority of subjects achieve their goal, (cf. the descriptive statistics in Table 3). Nevertheless, a share of subjects do not achieve their goals – neither in our experiment 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 (see Online Supplement S.7). Such individuals may put less reflection into the goal setting process or may deliberately set unrealistically high goals – often referred to as 'stretch goals' (see, e.g., Sitkin et al., 2011) – to push themselves to the edge.

6 Discussion of Design Choices

In the following, we discuss some of the major design choices of our study and how they may influence the results.

Motivation. As mentioned in the introduction, the real-effort task in our experiment was chosen to mimic features of typical self-control problems. Subjects are likely to have low intrinsic motivation for it, also because it does not have any productive use. 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, at date 1, subjects already know that they will receive earnings from the mandatory work phase and the show-up fee, and the sense that they "already have been paid" may lower subjects' inclination to work further.

Goals might have a different 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 further extrinsic motivation, e.g., by making payment conditional on subjects reaching their goal as 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 we would otherwise conflate measuring an intention to revise the goal with measuring whether subjects can remember their goal. Our post survey provides some information on what subjects recall by incentivizing subjects to remember their goals. While the most recent goal always appears to be salient, reminders seem to matter for how well subjects recall their early goal. Specifically, we measure the absolute error in recall as $|$*recalled goal* $t -$ *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 their goal at date 1. As we show in Online Supplement S.1.2, one can extend our theory to allow for anticipation of goal revision. This yields predictions that are qualitatively similar to the predictions of our main model. Yet, announcing goal revision in *Revise0* and *Revise1* complicates comparisons between *Revise0* and *Early* and would interfere with testing our main hypotheses, which is why we do not announce goal revision in our design.[21]

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 practice, in some tasks such self-monitoring may be more difficult or require additional effort. Yet, our finding

[21] A potential advantage from announcing goal revision is 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 early goal 'revision proof'. In contrast, naïve individuals may set an early 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 as in a treatment where the opportunity to revise the goal was a surprise.

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.

Time Span between Date 0 and Date 1. We have observed that individuals who set a goal a few days in advance of the task set higher goals and exert more effort compared to subjects who only set a goal at the start of the task. In the experiment, we have chosen the time span between date 0 and date 1 to be five days – following the insights of Augenblick et al. (2015) that 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. Yet, for future research it is interesting to investigate different time spans between goal setting and the task to examine whether the time span matters and, if so, what time span maximizes performance. The optimal time span could depend, for example, on task characteristics (how difficult or boring it is), personal characteristics, or the interaction of these two. When investigating the optimal time span one could also let subjects choose when they want to set a goal. Doing so would reveal whether people understand that setting a goal early on is beneficial.

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 in the light of goal revision, 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 test for a sample of male subjects, whether self-set, non-binding early goals are effective self-regulation tools even though subjects can easily revise these goals.

A secondary contribution of our paper is that it addresses potential confounds of private goal setting and goal revision. Specifically, our design avoids the treatment migration problem that might be responsible for the mixed evidence found in studies comparing performance with self-set goals compared to a no-goals condition. This treatment migration problem will cause researchers to underestimate treatment effects if they do not take into account that even people in the no-goals condition may privately set goals.

Our results highlight that it is important to set goals some time before the task: Subjects who set a goal a few days in advance of the task set higher goals and exert more effort compared to subjects who only set a goal at the start of the task. While our results highlight that goal revision does occur (also when individuals are not asked to revise their goal), they also show that these revisions do not make goals ineffective. Further, our results show that one cannot prevent or alleviate goal revisions by highlighting the early goal or by "hiding" the opportunity to revise goals.

These findings are important for organizations and individuals (with the caveat that the results are obtained for a male only sample). Organizations may be sceptical about the use of non-binding goals as motivators to increase performance. Our results suggest that it is important to set goals in advance of the task and that such early goals 'work' even though people may revise their goals. For individuals, the results demonstrate the potential for early goals in self-regulation. Lastly, our results highlight the need for researchers to recognize private goal revision. For example, when examining goal achievement, instead of just relying on initially stated goals researchers can elicit revised goals to avoid comparing performance to a different goal than the one people have in mind.

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.

Anshel, M. H., R. Weinberg, and A. Jackson (1992): "The Effect of Goal Difficulty and Task Complexity on Intrinsic Motivation and Motor Performance," *Journal of Sport Behavior*, 15, 159.

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," *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*.

Chapman, S. L. and D. B. Jeffrey (1978): "Situational Management, Standard Setting, and Self-reward in a Behavior Modification Weight loss Program." *Journal of consulting and clinical psychology*, 46, 1588.

Chidester, T. R. and W. C. Grigsby (1984): "A Meta-Analysis of the Goal Setting-Performance Literature," in *Academy of Management Proceedings*, Academy of Management Briarcliff Manor, NY 10510, 1, 202–206.

Clark, D., D. Gill, V. L. Prowse, and M. Rush (2020): "Using Goals to Motivate College Students: Theory and Evidence from Field Experiments," *Review of Economics and Statistics*, 102.

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," *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.

Epton, T., S. Currie, and C. J. Armitage (2017): "Unique Effects of Setting Goals on Behavior Change: Systematic Review and Meta-analysis." *Journal of Consulting and Clinical Psychology*, 85, 1182.

Erbaugh, S. J. and M. L. Barnett (1986): "Effects of Modeling and Goal-setting on the Jumping Performance of Primary-grade Children," *Perceptual and Motor Skills*, 63, 1287–1293.

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: Goals, Monetary Incentives, and Workers's 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.

Goudas, M., N. Ardamerinos, S. Vasilliou, and S. Zanou (1999): "Effect of Goal-setting on Reaction Time," *Perceptual and Motor Skills*, 89, 849–852.

Goulet-Pelletier, J.-C. and D. Cousineau (2018): "A Review of Effect Sizes and their Confidence Intervals, Part I: The Cohen's d Family," *Quantitative Methods for Psychology*, 14, 242–265.

Grant, A. M. (2008): "The Significance of Task Significance: Job Performance Effects, Relational Mechanisms, and Boundary Conditions." *Journal of Applied Psychology*, 93, 108.

Hackman, J. R. and G. R. Oldham (1976): "Motivation through the Design of Work: Test of a Theory," *Organizational Behavior and Human Performance*, 16, 250–279.

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

Hayes, S. C., I. Rosenfarb, E. Wulfert, E. D. Munt, Z. Korn, and R. D. Zettle (1985): "Self-reinforcement Effects: An Artifact of Social Standard Setting?" *Journal of Applied Behavior Analysis*, 18, 201–214.

Heath, C., R. P. Larrick, and G. Wu (1999): "Goals as Reference Points," *Cognitive Psychology*, 38, 79–109.

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.

Hinsz, V. B. (1995): "Goal Setting by Groups Performing an Additive Task: A Comparison With Individual Goal Setting 1," *Journal of Applied Social Psychology*, 25, 965–990.

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.

Kahneman, D. and A. Tversky (1979): "Intuitive Prediction: Biases and Corrective Procedures," *TIMS studies in Management Science*.

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.

Klein, H. J., J. C. Molloy, and C. T. Brinsfield (2012): "Reconceptualizing Workplace Commitment to Redress a Stretched Construct: Revisiting Assumptions and Removing Confounds," *Academy of Management Review*, 37, 130–151.

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.

——— (2002): "Building a Practically Useful Theory of Goal Setting and Task Motivation," *American Psychologist*, 57, 705–717.

——— (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.

McCalley, L. and C. J. Midden (2002): "Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation," *Journal of Economic Psychology*, 23, 589–603.

Mento, A. J., R. P. Steel, and R. J. Karren (1987): "A meta-analytic study of the effects of goal setting on task performance: 1966–1984," *Organizational behavior and human decision processes*, 39, 52–83.

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.

Sackett, A. M., G. Wu, R. J. White, and A. Markle (2014): "Harnessing Optimism: How Eliciting Goals Improves Performance," *Available at SSRN 2544020*.

Schunk, D. H. (1985): "Participation in Goal Setting: Effects on Self-efficacy and Skills of Learning-disabled Children," *Journal of Special Education*, 19, 307–317.

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.

Smith, M. and C. Lee (1992): "Goal setting and performance in a novel coordination task: Mediating mechanisms," *Journal of Sport and Exercise psychology*, 14, 169–176.

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).

Tanes, Z. and H. Cho (2013): "Goal Setting Outcomes: Examining the Role of Goal Interaction in Influencing the Experience and Learning Outcomes of Video Game Play for Earthquake Preparedness," *Computers in Human Behavior*, 29, 858–869.

Thakral, N. and L. T. Tô (2021): "Daily Labor Supply and Adaptive Reference Points," *American Economic Review*, forthcoming.

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

Tubbs, M. E. (1986): "Goal Setting: A Meta-analytic Examination of the Empirical Evidence." *Journal of Applied Psychology*, 71, 474.

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.

West, R. L., D. C. Welch, and R. M. Thorn (2001): "Effects of Goal-setting and Feedback on Memory Performance and Beliefs Among Older and Younger Adults." *Psychology and Aging*, 16, 240.

Wood, R. E., A. J. Mento, and E. A. Locke (1987): "Task Complexity as a Moderator of Goal Effects: A Meta Analysis." *Journal of Applied Psychology*, 72, 416.

Wu, G., C. Heath, and R. Larrick (2008): "A Prospect Theory Model of Goal Behavior," mimeo, University of Chicago.

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 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 early 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 early 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 early 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 early goal g_0^* or the revised goal g_1^*, depending on the treatment. Recall that both the early 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 early goal ($\lambda^T = 0$) and the individual perceives no comparison utility from revising the early goal ($\nu = 0$). Thus, having set an early goal alleviates the self-control problem – even if the individual revises the goal.

Overall, the effort provided by self 1 lies between the early 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 $e^* = g_0^* = g_1^* = g^*$.[22] 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.

[22] 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

June 5, 2022

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 early 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*.

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

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: Productivity 1 - productivity 0.

  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).

[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.26 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

[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.

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.7 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 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.8, 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.7 and Figure S.8, 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

[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.7: Power for between-subject comparisons, two-sample t-test

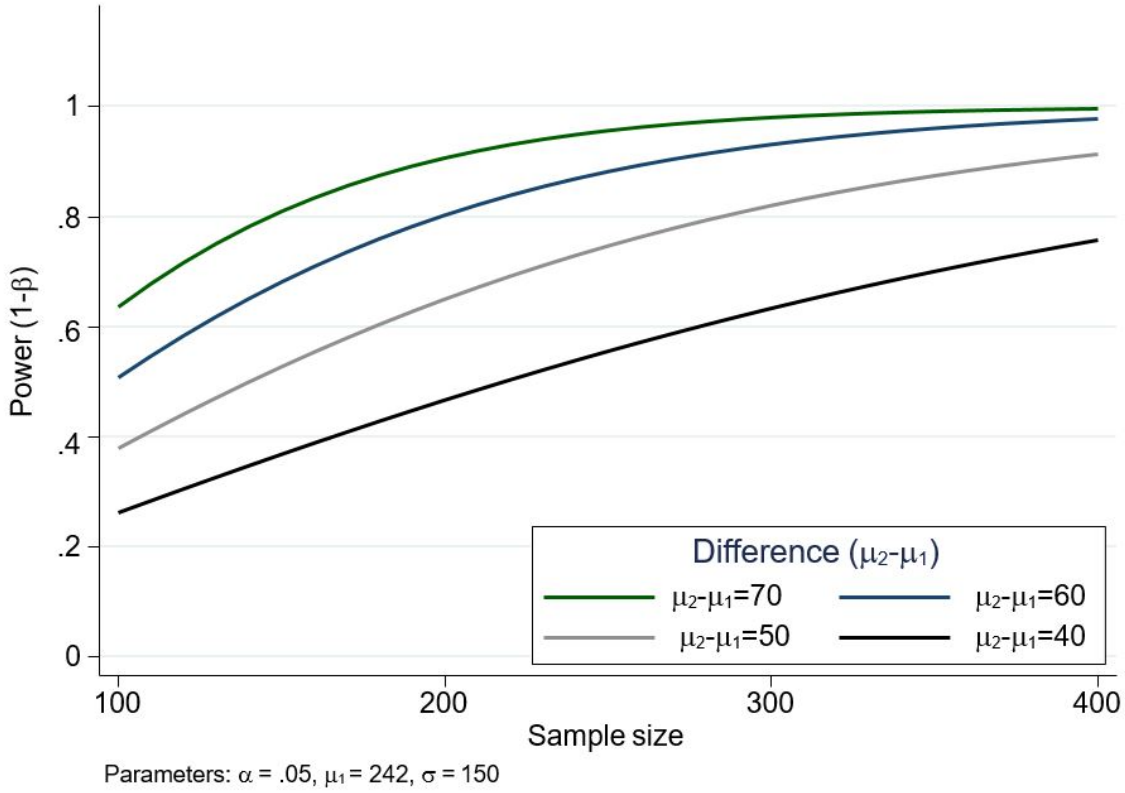
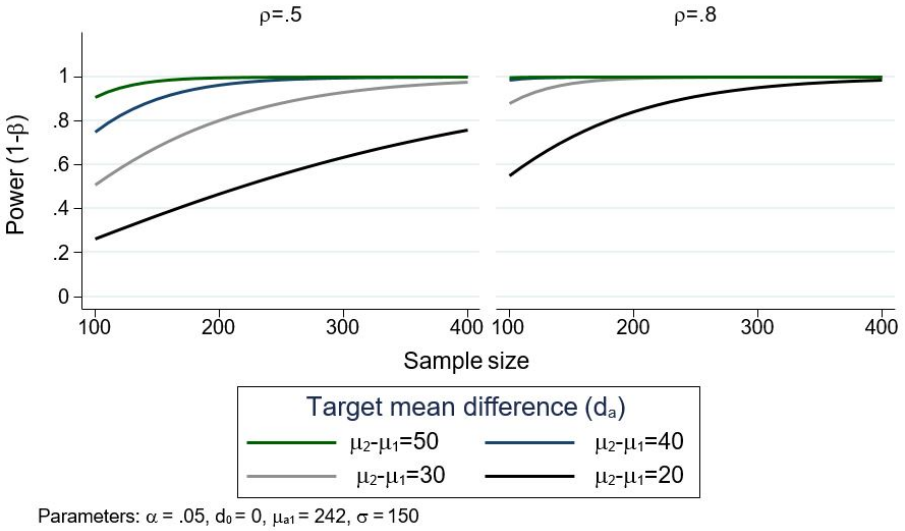


Figure S.8: Power for within-subject comparisons, two-sample paired t-test



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., 2020). For example, Koch and Nafziger (2020, Working Paper Version) find effect sizes of daily goals vs. no goals for women to be -.1/-.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.24, 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 (Wald χ^2, $p = .331$), 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 ($p = .162$). 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.

[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).

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.27, 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 in *Late*, 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 the top panel of Table S.28, 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 the bottom panel of Table S.28, 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.29, 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.30, 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 (top panel of Table S.31) and for including the entire sample (bottom panel of Table S.31). In Table S.32, 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 Table S.12, we show that there is no effect when we use mistakes per table or time spent per table as proxies for effort.

S.6 Alternative Explanations for Treatment Differences in Effort

Attrition. At date 0 there are two treatment groups: subjects who are later (at date 1) randomized into *Early, Revise0*, and *Revise1* (i.e., subjects who know that the task at date 1 will be to count zeros) vs. subjects in *Late* (i.e., subjects who do not know what happens at date 1). A concern might be that learning about the task and setting goals early vs. late could influence attrition and in doing so affect treatment differences. Yet, we observe no treatment differences in completing date 0 (80.8 percent in *Late* vs. 78.3 percent in *Early, Revise0*, and *Revise1*, $p = .560$; cf. Table S.22). Similarly, we observe no significant treatment differences in the probability of completing parts 1 and 2 (cf. Tables S.23 and S.24) and little indication of selection on observables (cf. Online Supplement S.4).

Even though not significant, there might be a concern that the 4.8 percentage point

difference in raw numbers for completion of date 1 suggests that subjects in *Late* are more likely to participate at date 1 than subjects in the other treatments (cf. Specification (1) in Table S.24), and that this difference may influence treatment differences in effort and goals. Specifically, unproductive subjects or subjects who set low goals may select-out of the study in *Early, Revise0*, and *Revise1* after getting to know the task at date 0, but not in treatment *Late* where they do not get to know the task at date 0. Yet, comparing subjects who return at date 1 versus not for *Early, Revise0*, and *Revise1*, there are only small and insignificant differences in the goal 0 set (273 vs 270; $p = .927$, cf. Table S.25) and the productivity at date 0 (15.46 vs 15.11; $p = .595$, cf. Table S.25) – alleviating this concern. Hence, overall, there is no indication that attrition drives treatment differences.

Increasing Motivation by Setting Goals Early or Announcing the Task Early. One additional explanation for the difference in effort between *Early/Revise0/Revise1* and *Late* could be that the meaning of the task (Hackman and Oldham, 1976; Grant, 2008) is enhanced by asking people to set an early goal: The individual might think that the task at date 1 is especially important when being asked to provide a goal for it 5 days in advance. And he might also think more about the task when he learns about it in advance, which might make the task seem more relevant. If so, higher effort in treatments with an early goal could be due to these subjects being more motivated. One might also suspect that this channel could give subjects in *Early, Revise0*, and *Revise1* (but not in *Late*) the possibility to practice the task.[8]

If setting goals in advance indeed increased task meaning, or if subjects practiced the task more in the treatments with an early goal, we should observe that the change in productivity between dates 0 and 1 and the change in enjoyment from date 0 to date 2 is larger in *Early, Revise0,* and *Revise1* than in *Late*. This is not what we observe (t-test, $p = .861$ and $p = .757$, respectively).

[8]Subjects in *Late* also 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 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 – opposite to what we find.

Finally, setting an early goal could give subjects the possibility to make "if-then" plans in order to achieve their goals. Such implementation intentions have been shown to increase goal commitment and performance (Gollwitzer and Sheeran, 2006). If people use time to form implementation intentions when setting an early goal, goal setting should take longer than when people for the first time set a goal at date 1. However, we find no difference in the average time that subjects in *Late* spend on setting a goal at date 1 (103.09 seconds) compared to subjects in *Early*, *Revise0*, and *Revise1* at date 0 (104.73 seconds, t-test: $p = .871$).[9]

Increased Goal Commitment Through More Frequent Goal Setting? We observe a significant difference in effort between *Late* and *Revise1*, while the difference in effort between *Late* and *Early* has the predicted sign but is not significant (cf. Table 5). The fact that there were no treatment differences across the treatments with early goal setting (*Early*, *Revise0*, and *Revise1*), prompted us to pool these treatments for a better powered additional test of Hypothesis 2.1 – which indicates that setting a goal early leads to a higher effort than setting it late. Yet, another possible explanation for the observed pattern could be that setting a goal twice (as in *Revise1* or *Revise0*) matters in contrast to setting a goal only once (as in *Early* and *Late*). Setting a goal twice could, for example, encourage people to reflect on what really is the optimal goal for them and, in doing so, increase goal commitment.

Higher goal commitment could be reflected in more people achieving their goal or in higher effort in treatments where goals are set twice than when they are only set once. We have seen that effort as well as goal achievement is higher in *Revise1* than in *Late* (cf. Tables 5 and S.15). These result are consistent with both the interpretation that more frequent goal setting increases goal commitment and the interpretation that *goal 0* also enters the reference point in *Revise1*. Yet, further evidence suggests against the first interpretation.

First, looking at the proportions of subjects who achieve their (displayed) goal, we do not

[9]Time is truncated at 5 minutes to account for possible outliers where people do not consistently pay attention, e.g. due to people taking a break.

find significant differences between treatments *Early vs. Revise0* (logit marginal effect .001, $p = .990$) and *Late vs. Revise1* (logit marginal effect .071, $p = .317$). Second, we find no significant difference in effort between *Early*, where subjects are asked to set a goal once, and *Revise0*, where subjects are asked to set a goal twice (cf. Table 5).

Third, evidence from the post survey also suggests against increased goal commitment due to setting a goal twice. Here, we can identify 20 participants in *Early* who also set goals twice because they privately revised their early goal; the remaining 44 participants in *Early* who completed the post survey set only one goal as they did not privately revise their goal. Regressing effort on a dummy for having set a goal twice yields an insignificant coefficient ($\beta = 9.33$, $p = .807$). This result thus speaks against the alternative explanation. Further, in the post survey, we also measured goal commitment using Klein et al. (2012)'s unidimensional target-free scale ranging from 1 to 5 (cf. Online Supplement S.8). We observe that commitment to the goal chosen at date 0 in *Early* (mean 2.97) is not significantly different from that for the goal chosen at date 1 in *Late* (3.15; t-test, $p = .127$). And comparing commitment for *goal 0* (*goal 1*) in *Revise0* (mean 3.13) (in *Revise1*, mean 3.10) with that in *Early* (*Late*), we find no significant differences ($p = .132$ and $p = .631$, respectively).[10]

S.7 Rationality of 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. Still, we observe that subjects in *Late* on average fall short of their goal by a statistically significant 39 tables (t-test, $p < .001$). Focusing on the 33 percent of subjects in *Late* who fail their goal (cf.

[10] Another way to test whether setting a goal twice versus once matters would be to let subjects in *Late* and *Early* also set a goal twice. As Augenblick et al. (2015) demonstrate that discounting of future real effort costs changes drastically within the first hours prior to a task, such goal setting thus would have to be very close to the task – making the comparison to *Revise0* and *Revise1* difficult.

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; $p = .387$).[11] Thus, overall, a picture of heterogeneity in goal achievement emerges. The majority of subjects achieve their goal, consistent with setting rational goals, but a fraction of subjects fall short of their goal by a large margin. A possible explanation is that the latter group is less deliberate when setting goals.[12] Another possible explanation is that some subjects value setting high goals, even if they do not literally believe in accomplishing the goal, because such 'stretch goals' (see, e.g., Sitkin et al., 2011) provide a strong motivation for effort.

S.8 Goal commitment scale

We measured goal commitment in the post survey using Klein et al. (2012)'s unidimensional target-free (KUT) scale.[13] To avoid possible confounds for the main purpose of our study, we included the goal commitment questions in the post survey rather than at the goal setting stage.

[11] Note that if we consider all subjects in *Revise0* and *Revise1*, *goal 1* also is on average achieved (7 tables shortfall; $p = .470$). If we exclude outliers (subjects with deviations larger/smaller than 300/-300), then the shortfalls are .33 ($p = .976$) and -.49 ($p = .942$), respectively. 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 extent also their higher early goal – a driving force that is not present when *goal 0* = *goal 1*.

[12] Exploratory analysis of who revises their early goal downwards (available upon request) reveals that a one standard deviation increase in the cognitive reflection test score (mean 2.30, std.dev. 1.05) is associated with an 8.7 percentage point lower probability of downward goal revision (logit marginal effect, $p = .045$). The time used in setting goal 0 also has a significant coefficient suggesting that more time spent reduces the probability of downward revision, but the standardized marginal effect is close to zero. A similar exercise for who revises their early goal upwards, produces no significant effects.

[13] The following items were assessed on a 5-point Likert scale: 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? Because the fourth item of the original KUT scale makes little sense for self-chosen goals, we did not include it in the questionnaire (The fourth item is: To what extent have you chosen to be committed to this goal?).

S.9 Tables

Table S.7: Coefficients on controls for the goal setting regressions

	(1)	(2)
Late	-63.97***	
	(23.24)	
Productivity 0	7.39***	
	(2.23)	
Change in productivity		2.79*
		(1.67)
CRT	-5.07	4.98
	(12.92)	(9.05)
Slider moved	60.04	
	(53.93)	
Slider moved, Date 0		61.98
		(73.87)
Slider moved, Date 1		2.47
		(33.22)
Response time	0.33**	
	(0.14)	
Response time, Date 0		0.13
		(0.15)
Response time, Date 1		0.18
		(0.14)
Self-competition	-0.44	-0.28
	(0.33)	(0.27)
Risk tolerance	6.37	5.73
	(7.40)	(4.73)

Continued on next page

Table S.7 – continued from previous page

	(1)	(2)
Pleasure in task	22.08**	-4.02
	(9.64)	(7.50)
Time-constrained(P)	28.39	
	(32.38)	
Uncertainty		12.07
		(7.76)
More time		-17.32
		(36.76)
Less time		-18.63
		(28.32)
Constant	-11.32	-164.29*
	(82.14)	(92.23)
N	143	162

Coefficients for the specifications in Table 4 with the full set of controls listed in Table S.7.

(1) Hypothesis 1.1 (goal): Early vs. Late

(2) Hypothesis 1.2 (goal): Revise0 & Revise1

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

Table S.8: Coefficients on controls for the effort regressions

	(1)	(2)	(3)	(4)	(5)
Late	-28.10	-34.77**			
	(23.24)	(17.40)			
Revise0				-9.30	
				(22.69)	
Revise1			50.43**		31.51
			(21.21)		(20.89)
Productivity 0					
Productivity 1	5.60***	7.95***	8.85***	7.11***	10.21***
	(2.01)	(1.56)	(1.92)	(2.45)	(2.31)
CRT	3.96	10.01	9.88	8.88	12.35
	(10.94)	(7.17)	(9.81)	(10.71)	(9.00)
Self-competition	-0.06	-0.07	0.02	-0.24	-0.16
	(0.37)	(0.25)	(0.29)	(0.41)	(0.36)
Risk tolerance	-2.86	6.01	6.86	6.04	11.60**
	(7.14)	(4.48)	(6.01)	(7.00)	(5.60)
Pleasure in task	21.27**	24.44***	20.94**	26.85**	26.69**
	(8.94)	(6.92)	(9.37)	(10.43)	(10.90)
Time-constrained(A)	-2.63	-32.21*	-41.93	-22.18	-43.79*
	(25.81)	(17.65)	(26.55)	(25.36)	(25.89)
Constant	60.97	-47.98	-95.94	-33.38	-143.70**
	(63.43)	(45.32)	(66.51)	(68.93)	(68.29)
N	143	305	146	159	162

Coefficients for the specifications in column (3) of Table 5 with the full set of controls listed in Table S.8.
(1) Hypothesis 2.1: *Early* vs. *Late*.
(2) Hypothesis 2.1+: *Early & Revise0 & Revise1 (pooled)* vs. *Late*
(3) Hypothesis 2.2: *Late* vs. *Revise1*
(4) Hypothesis 3.1: *Early* vs. *Revise0*
(5) Hypothesis 3.2: *Revise0* vs. *Revise1*
Robust standard errors in parentheses. * $p<.10$, ** $p<.05$, *** $p<.01$

Table S.9: Coefficients on *goal 0* or *goal 1* in *Revise 0* & *Revise1*

	Revise0			Revise1		
	(1)	(2)	(3)	(4)	(5)	(6)
Goal 0	0.42**,†	0.34**,††	0.31*,††	0.62***	0.60***	0.53***
	(0.17)	(0.16)	(0.17)	(0.20)	(0.17)	(0.16)
Constant	105.82**	-24.77	-217.14*	76.76	-82.96*	-167.06*
	(40.42)	(58.15)	(113.80)	(47.23)	(47.82)	(98.91)
	(7)	(8)	(9)	(10)	(11)	(12)
Goal 1	0.69***,†	0.64***,††	0.61***,††	0.68***	0.62***	0.56***
	(0.11)	(0.11)	(0.13)	(0.12)	(0.12)	(0.13)
Constant	63.40***	11.34	-121.72	76.53***	-22.00	-24.90
	(22.52)	(34.32)	(76.91)	(28.41)	(39.57)	(93.48)
Productivity	No	Yes	Yes	No	Yes	Yes
Other controls[a]	No	No	Yes	No	No	Yes
N	82	82	82	80	80	80

OLS Regression of *effort* on *goal 0* or *goal 1* in the respective treatments. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$. Wald chi-square test for equality of *goal 0* and *goal 1* coefficients across models within a treatment: † $p < .10$, †† $p < .05$, ††† $p < .01$. A Wald chi-square test fails to reject equality of the coefficients on *goal 1* across treatments.
[a] See Online Supplement S.2.

Table S.10: Goal setting, excluding outliers

	Hypothesis 1.1 (goal): Early vs. Late			Hypothesis 1.2 (goal 1 -goal 0): Revise0 & Revise1[a]		
	(1)	(2)	(3)	(4)	(5)	(6)
Change in productivity[a]	No outliers			3.41**	3.22**	
					(1.35)	(1.43)
Constant				-33.95***	-42.67***	-138.89*
				(7.47)	(8.91)	(82.09)
Other controls				No	No	Yes
N				159	159	159

Regressions as described in Table 4 – excluding outliers: subjects who set a goal $\geq$ 900, provide effort $\geq$ 900, or have goal revision *goal0-goal1*$\leq$-700. Dependent variable: (4)-(6) *goal 1 - goal 0*. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

[a] Productivity at date 1 - productivity at date 0.

Table S.11: Comparison of effort in the different treatments, excluding outliers

No controls	Productivity	All controls	
	(1)	(2)	(3)
Hypothesis 2.1: Early vs. Late			
		No outliers	
Hypothesis 2.1+: Late vs. Early & Revise0 & Revise1 (pooled)			
	-34.06**	-28.59*	-31.57*
	(16.37)	(15.26)	(17.16)
Hypothesis 2.2: Late vs. Revise1			
	51.65**	44.18**	47.10**
	(21.69)	(19.99)	(21.10)
Hypothesis 3.1: Early vs. Revise0			
	4.13	-5.56	-10.16
	(22.09)	(21.94)	(22.47)
Hypothesis 3.2: Revise0 vs. Revise1			
	24.42	29.84	30.69
	(22.36)	(20.93)	(20.33)
Productivity	No	Yes	Yes
Other controls	No	No	Yes

Regressions as described in Table 5 – excluding outliers: subjects who set a goal ≥ 900 or provide an effort ≥ 900 (in *Revise 0* and 1 in *Revise 1*). 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. * $p < .10$, ** $p < .05$, *** $p < .01$. H2.1+ pools all treatments with 'early' goal 0, as framing effects under Hypothesis 3 are rejected.

Table S.12: Alternative outcome measures in *Early* & *Revise0* & *Revise1* (pooled) vs *Late*

Dep.variable	Mistakes per table			Average seconds per table		
	(1)	(2)	(3)	(4)	(5)	(6)
Late	0.01	0.01	-0.02	0.66	0.32	0.21
	(0.03)	(0.02)	(0.01)	(0.63)	(0.54)	(0.60)
Productivity 1		-0.01***	-0.01***		-0.44***	-0.42***
		(0.00)	(0.00)		(0.05)	(0.05)
Constant	0.16***	0.33***	0.32***	14.75***	22.82***	24.86***
	(0.01)	(0.04)	(0.04)	(0.34)	(1.06)	(1.51)
Other controls	No	No	Yes	No	No	Yes
	324	324	303	324	324	303

Regression as described in Table 5 for Hypothesis 2.1+ with alternative outcome measures. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.13: 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)
Productivity 1		8.66***	8.43***
		(1.84)	(1.96)
Constant	241.83***	80.81**	11.41
	(12.35)	(33.87)	(62.56)
Other controls	No	No	Yes
N	249	249	228

OLS regressions of *goal 1* on a treatment dummy (that is equal to one if the subject was randomly assigned to treatment *Late* and zero otherwise) and (1) a constant; (2) a constant and productivity (which refers to baseline productivity at the date when the goal was set); (3) a constant, productivity, and the set of controls listed in Online Supplement S.2. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.14: Goal achievement in *Revise0* & *Revise1*

	Goal achievement					
	Effort-goal 0			Effort-goal 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Productivity 1		5.27**	5.52**		1.70	1.65
		(2.12)	(2.22)		(1.84)	(1.94)
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)
Other controls	No	No	Yes	No	No	Yes
N	162	162	162	162	162	162

OLS regression of the dependent variable on (1),(4) a constant (2),(5) a constant and productivity at date 1, and (3),(6) a constant, productivity, and the set of controls listed in Table S.8. Dependent variable: (1)-(3) *effort - goal 0*, (3)-(6) *effort - goal 1*. Robust standard errors in parentheses.

Table S.15: Goal achievement for *goal 1* in *Late* vs. *Revise0* & *Revise1*

	(1)	(2)	(3)
Late	-32.44**	-30.91**	-23.99*
	(14.17)	(14.44)	(14.42)
Productivity 1		1.27	1.20
		(1.35)	(1.47)
Constant	-6.80	-30.46	-73.68
	(9.39)	(28.49)	(53.17)
Other controls	No	No	Yes
N	249	249	228

OLS regression of the dependent variable on (1) a constant (2) a constant and productivity at date 1, and (3) a constant, productivity, and the set of controls listed in Table S.8. Dependent variable: *effort - goal 1*. Robust standard errors in parentheses.

Table S.16: 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

Regressions for dependent variable *goal 1 - goal 0* as described in in Table 5, now controlling for uncertainty and time shocks. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.
[a] Sample restricted to subjects with no time shock.

Table S.17: 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

Regressions for dependent variable *goal 1 - goal 0* as described in in Table 5, now controlling for controlling for change in enjoyment. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.18: Goal-productivity relation for *Revise0* and *Revise1* (pooled)

Dep. variable	Goal 0		Goal 1	
	(1)	(2)	(3)	(4)
Productivity 0		5.28		
		(3.43)		
Productivity 1				8.74***
				(2.48)
Constant	277.62***	193.35***	241.83***	79.39*
	(12.08)	(56.40)	(12.34)	(45.11)
N	162	162	162	162

OLS regression of the dependent variable on productivity at the date when the respective goal was set. Dependent variable: (1)-(2) *goal 0*, (3)-(4) *goal 1*. Robust standard errors in parentheses.

Table S.19: Goal setting (time- and earnings-equivalents)

	Hypothesis 1.1 (goal): Early vs. Late			Hypothesis 1.2 (goal 1 -goal 0): Revise0 & Revise1		
Dep. variable	Time-equivalent of goals					
	(1)		(2)	(3)		(4)
Late	-16.49***		-17.17***			
	(5.08)		(5.56)			
Constant	57.62***		18.22	-12.03**		-36.60*
	(4.18)		(18.84)	(5.23)		(20.80)
Productivity	No		No	No		No
Other controls	No		Yes	No		Yes
N	164		143	161		161
Dep. variable	Earnings-equivalent of goals					
	(5)	(6)	(7)	(8)	(9)	(10)
Late	-9.27	-19.37***	-17.86***			
	(5.95)	(6.14)	(6.10)			
Constant	112.44***	73.78***	12.78	-11.29***	-14.44***	-37.05*
	(4.09)	(10.17)	(23.01)	(2.09)	(2.80)	(21.32)
Productivity	No	Yes	Yes	No	Yes	Yes
Other controls	No	No	Yes	No	No	Yes
N	164	143	143	162	162	162

Regression as described in Table 4, now with time- or earnings-equivalents of goals as dependent variable. Time-equivalents are computed by dividing the goal by the productivity/3 (tables per minute at time when goal was set), and hence productivity is not added as a control. In specifications (3) and (4), the time-equivalent goal for one subject is undefined because *productivity 1*=0. Earnings-equivalents are obtained by plugging the goal into the payment schedule (cf. Figure 2). Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.20: Coefficients on the displayed goal in the effort regressions

	(1)	(2)	(3)	(4)	(5)
Late	-1.34	-15.14			
	(17.12)	(12.67)			
Revise1			34.73**		41.85**
			(16.03)		(18.29)
Revise0				-9.09	
				(20.90)	
Displayed goal	0.59***	0.49***	0.63***	0.42***	0.42***
	(0.08)	(0.08)	(0.09)	(0.11)	(0.11)
Constant	16.36	-60.63	-98.44*	-40.42	-139.86**
	(46.67)	(38.20)	(56.03)	(57.18)	(59.58)
Productivity	Yes	Yes	Yes	Yes	Yes
Other controls[a]	Yes	Yes	Yes	Yes	Yes
N	143	305	146	159	162

Coefficients on the goal level shown in the free work phase (displayed goal): *goal0* in *Early* and *Revise0*, *goal1* in *Late* and *Revise1* for the specifications in column (6) of Table 5 with the full set of controls listed in Table S.8.
(1) Hypothesis 2.1: *Early* vs. *Late*.
(2) Hypothesis 2.1+: *Early & Revise0 & Revise1 (pooled)* vs. *Late*
(3) Hypothesis 2.2: *Late* vs. *Revise1*
(4) Hypothesis 3.1: *Early* vs. *Revise0*
(5) Hypothesis 3.2: *Revise0* vs. *Revise1*
Robust standard errors in parentheses. * p <.10, ** p <.05, *** p <.01

Table S.21: Comparison of effort in *Late* vs. *Revise1*

	(1)	(2)	(3)	(4)	(5)	(6)
Revise1	42.71***	39.56**	37.88**	27.71*	24.41	21.81
	(16.11)	(15.74)	(15.65)	(16.08)	(15.60)	(15.43)
Goal 1	0.66***	0.61***	0.60***			
	(0.07)	(0.08)	(0.08)			
First goal				0.63***	0.58***	0.59***
				(0.10)	(0.10)	(0.09)
Productivity 1		5.12***	4.80***		6.75***	6.26***
		(1.74)	(1.74)		(1.78)	(1.77)
Constant	37.70**	-38.43	-89.09*	44.48*	-60.77**	-137.23***
	(16.22)	(23.43)	(51.55)	(23.27)	(26.53)	(52.30)
Other controls[a]	No	No	Yes	No	No	Yes
N	167	167	167	167	167	167

Regression as described in Table 5 for *Late* vs *Revise1*, but now in (1)-(3) controlling for *goal 1*, in (3)-(6) controlling for the first goal that subjects set (*goal 0* in *Revise1* and *goal 1* in *Late*). Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.
[a] See Online Supplement S.2.

Table S.22: Participation at Date 0

Treatment	Emails Sent for Date 0	Completed Date 0	Percentage
Early, Revise0, & Revise 1[a]	374	293	78.34%
Late	125	101	80.80%
Total	499	394	78.95%

[a] Randomization into Early, Revise0, & Revise 1 occurred at date 1. There is no treatment difference in completion of date 0 (logit- regression, $p = .560$)

Table S.23: Attrition for Dates 1 and 2

Treatment	Emails Sent for Date 1	Completed Date 1	Percentage	Completed Date 2	Percentage
Early, Revise0, Revise1[a]	293	239	81.57%	206	86.19%
Late	101	87	86.14%	70	80.46%
Early		77		64	83.12%
Revise0		82		72	87.80%
Revise1		80		70	87.50%
Total	394	326	82.74%	276	84.66%

Percentages reflect the share of subjects who completed the date-1 and date-2 parts, respectively, out of the participants who received a link for the respective part. There are no treatment difference in completion (cf. Table S.24).

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

Table S.24: Selection on observables

	Probability of completing					
	Date 1			Date 2		
	(1)	(2)	(3)	(4)	(5)	(6)
Late	0.048	0.049	0.061	-0.027	-0.026	-0.011
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Revise0				0.047	0.054	0.052
				(0.06)	(0.06)	(0.05)
Revise1				0.044	0.050	0.038
				(0.06)	(0.06)	(0.06)
Productivity 0		0.001	-0.001		-0.004	-0.006
		(0.00)	(0.00)		(0.00)	(0.00)
Uncertainty			-0.023			0.004
			(0.02)			(0.02)
Time-constrained(P)			0.021			-0.010
			(0.07)			(0.07)
CRT			0.003			-0.017
			(0.02)			(0.02)
Pleasure in task			0.014			0.037**
			(0.02)			(0.02)
Risk tolerance			-0.021*			-0.002
			(0.01)			(0.01)
Self-competition			0.000			0.001
			(0.00)			(0.00)
Economics/Business			0.072*			0.135***
			(0.04)			(0.05)
N	394	394	394	326	326	326
Wald $\chi^2(11)$			10.55			15.47
p-value			.308			.162

Logit regressions (average partial effects). Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$. For variable definitions, see Online Supplement S.2.

Table S.25: Selective attrition: Goal 0 and productivity 0

Dep. variable	Goal 0			Productivity 0	
	(1)	(2)	(3)	(4)	(5)
Completed date 1	2.67	1.11	-2.30	0.35	0.18
	(28.95)	(29.23)	(27.85)	(0.66)	(0.67)
Constant	270.09***	203.36***	70.79	15.11***	12.48***
	(27.22)	(48.05)	(60.50)	(0.60)	(1.56)
Productivity 0	No	Yes	Yes	–	–
Other controls	No	No	Yes	No	Yes
N	293	293	293	293	293

OLS regressions with (1)-(3) goal 0 (4)-(5) productivity 0 as dependent variable, using the treatments with 'early' goal 0 (*Early, Revise 0* & *Revise 1*). (1) and (4) report the coefficient on a dummy for completion of part 1 in a regression where no further controls are added, (2) where productivity 0 is added as a control and (3) and (5) where other controls are added (cf. Online Supplement S.2). Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.26: 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

For variable definitions, see Online Supplement S.2.

Table S.27: Comparison of goals set in *Early* vs. *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)
Constant	245.00***	85.00*	-57.67
	(21.82)	(44.04)	(95.38)
Other controls[a]	No	No	Yes
N	164	164	143[a]

Regressions as described in Table 4, (1)-(3) – using median regressions instead of OLS. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.
[a] See Online Supplement S.2.

Table S.28: Robustness checks for comparison of goals set in *Early* vs. *Late*

	(1)	(2)	(3)
Excluding subjects for whom the time schedule is not available[a]			
Late	-41.45*	-67.03***	-63.97***
	(24.10)	(23.41)	(23.24)
Productivity		8.16***	7.39***
		(2.18)	(2.23)
Constant	262.55***	144.68***	-11.32
	(17.00)	(36.57)	(82.14)
Other controls	No	No	Yes
N	143	143	143
Using only subjects who completed the date-1 part			
Late	-33.53	-58.47**	-49.43**
	(22.90)	(22.44)	(22.88)
Productivity		8.45***	7.65***
		(2.00)	(2.04)
Constant	262.55***	140.46***	0.88
	(16.99)	(34.08)	(66.60)
Other controls	No	No	Yes
N	164	164	164

Regression as described in Table 4, (1)-(3). Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.
[a] For 21 subjects, the time schedule is not available (see note in Table 3).

Table S.29: Comparison of goals set in *Early* vs. *Late*, using exams and assignments instead of 'time constrained' variable

	(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

Regression as described in Table 4, (1)-(3) – using exams and assignments instead of 'time constrained' variable. Sample sizes are smaller because exams and assignments are only relevant for students. Specification (1) uses all subjects who completed the date-1 part. Specification (2) uses only those subjects for whom the time schedule is available. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table S.30: Comparison of effort in *Early & Revise0 & Revise1 (pooled)* vs. *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)
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)
Other controls[a]	No	No	Yes	No	No	Yes
N	326	326	305	326	326	305

Regressions as described in Table 5, (1)-(3) – using median regressions instead of OLS. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.
[a] See Online Supplement S.2.

Table S.31: Robustness check for comparison of effort in *Early/Revise* (pooled) vs. *Late*

	(1)	(2)	(3)	(4)	(5)	(6)
Excluding subjects for whom the time schedule is not available[a]						
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)
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)
Other controls[b]	No	No	Yes	No	No	Yes
N	305	305	305	305	305	305
Using only subjects who completed the date-1 part						
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)
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)
Other controls[b]	No	No	Yes	No	No	Yes
N	326	326	326	326	326	326

Regression as described in Table 5. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

[a] For 21 subjects, the time schedule is not available (see notes in Table 3).

[b] See Online Supplement S.2.

Table S.32: Comparison of effort in *Early/Revise* and in *Early/Revise* (pooled) vs. *Late*, using exams and assignments instead of 'time constrained' variable

	(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

Regression as described in Table 5, (3),(6) – using exams and assignments instead of 'time constrained' variable. Sample sizes are smaller because exams and assignments are only relevant for students. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

S.10 Figures

Figure S.9: Goals and effort

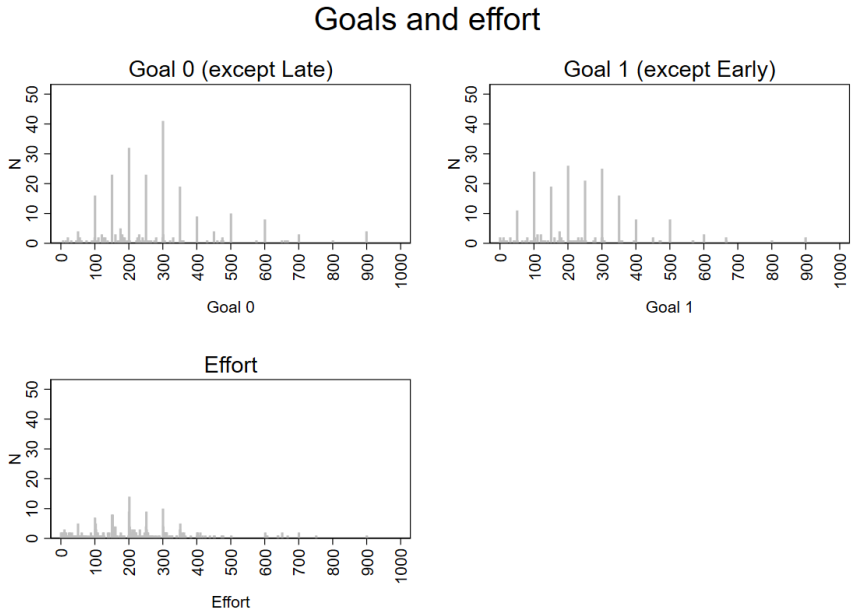
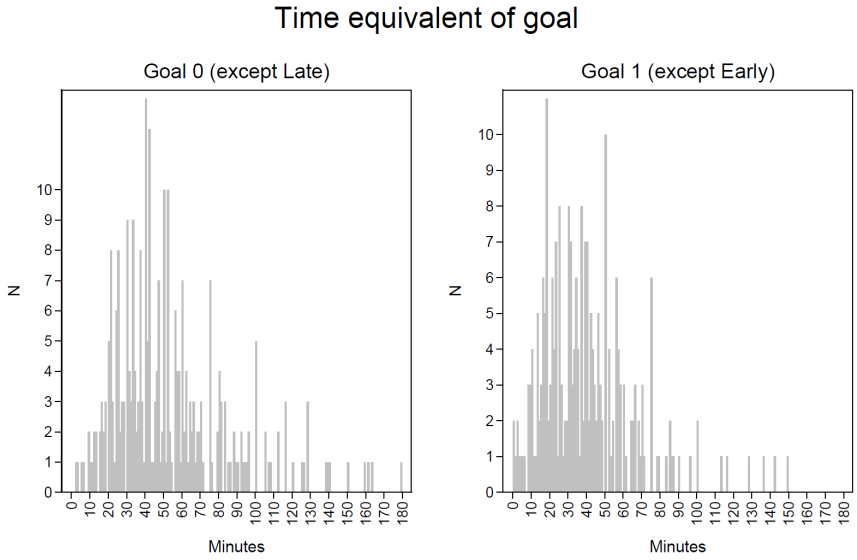
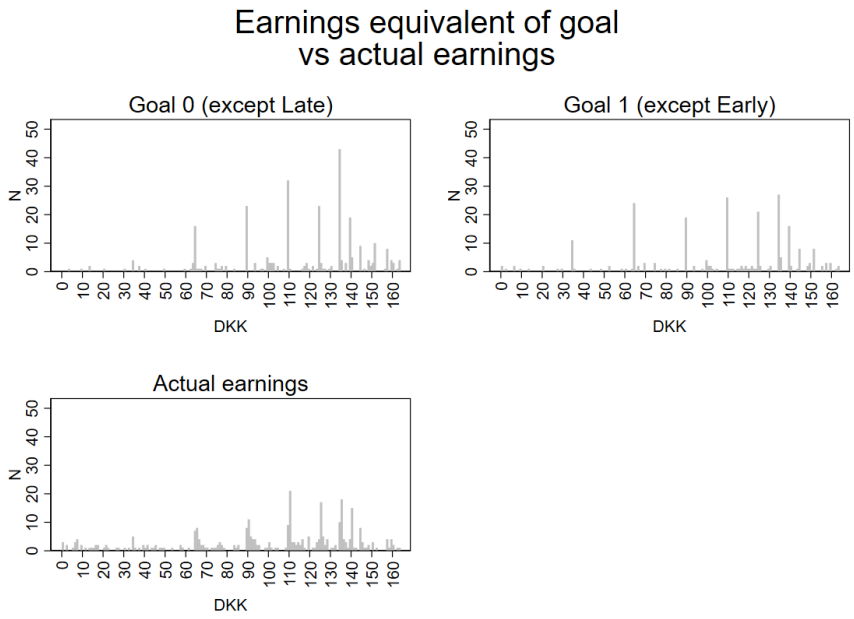


Figure S.10: Time equivalent of goals



Note: Time is truncated at 180 minutes.

Figure S.11: Earnings equivalent of goals set by subjects vs. actual earnings



S.11 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?

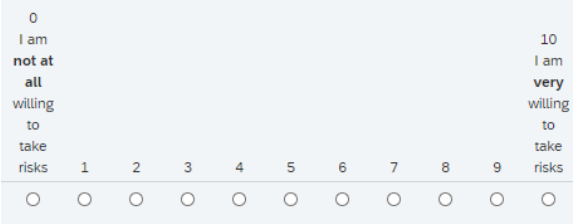
- o Agrobiology
- o Biology
- o Chemistry
- o Computer Science
- o Data Science
- o Engineering
- o Physics
- o Geoscience
- o IT Product Development
- o Chemistry
- o Mathematics
- o Mathematics - Economics
- o Nanoscience
- o Other

(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!

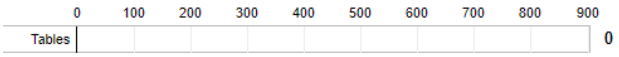
We ask you to set yourself a goal for how many tables to count on <date string for part 2>. We will remind you of the goal you set with a probability of 2/3. But, of course, you are free to work as much as you want.

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.**

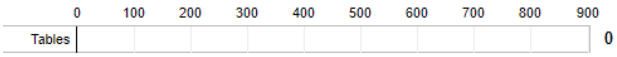
(If treatment Late) Page 7: Set a goal!

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.

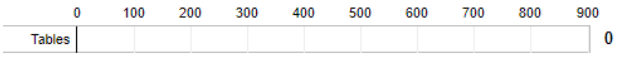
You now again have the opportunity to set a goal for how many tables to count today. We will remind you about either the goal you set now or the goal you set in part 1, each with probability 1/2. But, of course, you are free to work as much as you want.

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.

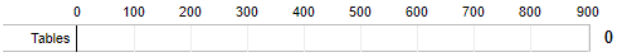
We will remind you of this goal. But, of course, you are free to work as much as you want.

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.