Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 185 (2020) 104949

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Motivational goal bracketing: An experiment

Alexander K. Koch [a,*,1], Julia Nafziger [a,b,1]

[a] Aarhus University, Denmark
[b] CEPR, United Kingdom

Received 13 March 2018; final version received 20 September 2019; accepted 5 October 2019
Available online 22 October 2019

Abstract

We study theoretically and experimentally how the bracketing of non-binding goals in a repeated task affects the level of goals that people set for themselves, the actual effort provided, and the pattern of effort over time. In our model, a sophisticated or partially naïve individual sets either daily or weekly goals to overcome a motivational problem caused by present-biased preferences. In an online, real-effort experiment, we randomly allocated subjects to treatments where they set either daily goals for how much to work over a one-week period or a single weekly goal. Consistent with the theoretical predictions, in the treatment with daily goals, the aggregate goal level for the week was higher and subjects provided more effort compared to the treatment with a weekly goal. The higher effort was driven by the higher aggregate goal level. Additional treatments complemented internal commitment through goals with an externally enforced minimum work requirement to get started working each day.

JEL classification: D03; D81; D91

Keywords: Self-control; Goals; Narrow bracketing; Commitment; Real effort; Online experiment

* Corresponding author.
E-mail addresses: akoch@econ.au.dk (A.K. Koch), jnafziger@econ.au.dk (J. Nafziger).

1 Contact: Department of Economics and Business Economics, Aarhus University, Fuglesangsalle 4, 8210 Aarhus V, Denmark.

https://doi.org/10.1016/j.jet.2019.104949
0022-0531/

1. Introduction

Every day people face motivational problems in repeated tasks such as working, studying, dieting, exercising, or saving. While decades of research in psychology document that goals play an important role in helping people to overcome their motivational problems (e.g. Locke and Latham 2015), it is still poorly understood how goals work in repeated tasks. In such settings, a person may focus on single instances of the task and evaluate tasks relative to narrowly bracketed goals or, instead, evaluate the aggregate performance over a longer time period relative to a broadly bracketed goal. How goals are bracketed can often be linked to the way feedback about performance is given (e.g. Asch, 1990; Cadena et al., 2011), the availability of salient reference points (e.g. Pope and Schweitzer, 2011; Allen et al., 2016), or explicit advice about how to set goals.[2] But how does the bracketing of goals affect the level of goals that people set for themselves, the actual effort provided, and the pattern of effort over time? In this paper, we study these open questions theoretically and experimentally.

We develop a model where a sophisticated or partially naïve individual works repeatedly on a task. He sets non-binding goals to overcome the motivational problems caused by present-biased preferences. Goals can either be narrow (daily goals) or broad (a weekly goal). Our model predicts that the aggregated daily goals are higher than the weekly goal. As a consequence, individuals with daily goals work more than those with the weekly goal. The reason is the following. A weekly goal tempts individuals to put in low effort at the beginning of the week and to compensate with higher effort later (*effort substitution*). This asymmetric effort profile is suboptimal (from an ex ante perspective). Because effort costs are convex, the individual would prefer a constant effort pattern. Under plausible assumptions, a reduction in the weekly goal level leads to less variation in effort over time. Thus, taking into account evidence on the distribution of present bias, our model predicts that on average individuals adopt a lower weekly goal compared to the aggregated goal level that they would chose with daily goals.

We tested the predictions of the model using an online, real-effort experiment. Subjects were randomly assigned to set either non-binding daily goals for how much to work online in the following week or a non-binding weekly goal. The experiment mimicked a typical work-leisure self-control problem. Work was desirable (the piece-rate pay was generous) but involved unpleasant effort. Subjects faced the usual real-life temptations because our study neither required them to show up at a lab nor to obey a particular schedule.

By exogenously varying the goal bracket, we provide a clean test of the *motivational bracketing* hypothesis that narrowly bracketed goals help individuals to address their self-control problems (e.g. Read et al., 1999). It was first suggested as an explanation for why individuals who can choose their working hours, such as taxi drivers, often appear to have daily income targets (e.g. Camerer et al. 1997, Dupas and Robinson 2016). While there is much suggestive evidence for this hypothesis, clean evidence is missing.

We find support for the motivational bracketing hypothesis. Our data reveal that subjects with daily goals set a higher aggregated goal for the week than subjects with a weekly goal, and that they worked more than those with a weekly goal. The latter effect largely disappears when we control for the goal level. That is, in line with the theory, the higher effort with daily goals seems

[2] For example, the UK National Health Service advises daily calorie targets and weekly weighing (https://www.nhs.uk/Tools/Pages/Losing-weight.aspx, accessed June 2019), and 150 minutes of exercise per week (https://www.nhs.uk/live-well/exercise/, accessed June 2019).

to be related to the higher goals that subjects set under the daily bracket and not to the daily bracket per se.

We extended our two baseline treatments in two directions. First, we conducted treatments that manipulated whether goals were framed as daily goals or a weekly goal, but that left the overall goal level unaffected by the framing.[3] These treatments allow us to directly test the prediction that the higher effort with daily goals is driven by the higher aggregated goal level compared to a weekly goal and not the framing of the goal. We confirm the prediction. Further, we test and confirm the prediction that subjects with a weekly goal work less at the beginning of the week than subjects with daily goals who face the same aggregated goal level, and that they work more at the end of the week to make up for the shortfall.

Second, to examine whether the positive effect of daily goals stems from their ability to get people started working each day, we ran additional treatments which complemented goals with a minimum work requirement. To receive any payment, subjects had to complete at least one real-effort task per day, which took less than a minute. An innovative feature of the requirement is that it forced subjects to 'get started' working each day, but otherwise gave full flexibility of how to allocate work and how much to work. In the treatments with the work requirement, effort and goal levels no longer differed across treatments where subjects set daily goals or a weekly goal, and effort patterns were similar. This result is suggestive for the interpretation that forcing subjects to 'get started' each day mitigates problems of effort substitution with a weekly goal.

With these treatments we make another novel contribution by addressing the question of how externally enforced work requirements interact with internal commitment through goals. Surprisingly, subjects worked less if they were forced to 'get started' working each day (in addition to setting daily goals) than if they just set daily goals. This result can be explained by a large fraction of subjects dropping out when they were forced to work each day. Focusing only on those subjects who did not drop out, performance did not differ across treatments with and without the work requirement. This is consistent with the interpretation that daily goals on their own already get people started. The pattern is reversed for the treatments with a weekly goal. For subjects who did not drop out, effort was significantly higher with the requirement than without it - as one would expect if the requirement gets people started working. But due to an increase in dropout, the overall effort with the requirement was no different from that without the requirement.

The paper is structured as follows. After reviewing the related literature, we present the experimental design in Section 2. Section 3 presents the theoretical model and predictions. In Section 4, we empirically compare daily and weekly goals and effort under them. Section 5 presents findings from the two extensions to our baseline treatments. Section 6 discusses alternative mechanisms and possible extensions. Section 7 concludes. The online appendix contains proofs, robustness checks regarding the theoretical predictions and the empirical analysis, a description of control variables, further results, and the experimental instructions.

Related literature The narrow bracketing literature goes back to Tversky and Kahneman (1981). Much of it considers simultaneous risky choices, where narrow bracketing is a choice error (e.g. Rabin and Weizsäcker, 2009). We contribute to the literature strand that considers narrow bracketing as a tool to overcome self-control problems (e.g. Shefrin and Thaler, 1988; Fudenberg and Levine, 2006). Our theoretical contribution is to provide the first model of how

[3] This treatment manipulation also is of independent interest because it allows us to test predictions of our model that also apply to settings where the goal level is unaffected by the way that goals are bracketed – for example, because the level is externally determined by an employer.

the bracketing of goals in repeated tasks affects the level of goals that people set for themselves, the actual effort provided, and the pattern of effort over time. Previous work gives conditions under which narrow bracketing is optimal with simultaneous tasks (Koch and Nafziger, 2016) or in a twice repeated optimal stopping problem (Hsiaw, 2018).[4] The phenomenon of effort substitution was previously noted in a simple two-period model by Jain (2009) and for two tasks by Koch and Nafziger (2016). These studies provide many important insights, but they do not capture situations where a task has to be performed repeatedly over some time and where the decision is not about stopping, but about how much effort to provide.

Our theoretical predictions build on the premise that people have a present bias in the real-effort task. Augenblick et al. (2015) are the first to estimate present bias in effort. They find evidence of present biased preferences in the effort domain, but not in the money domain. Further, they find that the present bias relates to demand for commitment devices. Augenblick and Rabin (2018) enrich the former study by eliciting the beliefs of the individuals about their future effort. They find that most subjects are (partially) naïve about their present bias.

Our empirical contribution is to compare behavior for narrowly and broadly bracketed goals in repeated tasks. Two early studies from psychology (Bandura and Simon, 1977; Bandura and Schunk, 1981) suggest that narrow ('proximal') goals are better at motivating effort than broad ('distal') goals. Yet, they have several conceptual problems and small sample sizes.[5] Other studies of goal setting in repeated tasks are distinct from our study because they either focus on (repeated) daily goals only or have a single goal for the entire time span. In Kaur et al. (2015), workers could set individual, daily work targets that were then externally enforced by the firm with a penalty for low output. Kaur et al. observed the effort and daily goals of the workers repeatedly over a period of 13 months, but did not vary the goal bracket. Setting goals was a dominated option for rational workers because goals penalized low output without any additional reward for high output. Nevertheless, in 36 percent of the cases workers set positive goals. They produced more and earned more than when not offered the opportunity to set goals. Uetake and Yang (2018) provide descriptive evidence from a weight loss app that daily goals can help achieve a long-run goal. The app suggests daily calorie targets for achieving a self-chosen, long-run weight loss target. Other studies consider the impact of a single, self-set goal on the outcome of a repeated task, such as weight loss (Toussaert, 2016), energy saving (Harding and Hsiaw, 2014), or studying (van Lent and Souverijn, 2017; van Lent, 2018; Clark et al., 2019; Himmler et al., 2019). Most of these studies find a positive effect of goals.

[4] Fischer and Ghatak (2016) study the effects of frequent vs. infrequent repayment installments of loans in microfinance. Frequent repayment requirements have the disadvantage of delaying the reward (e.g. a new loan), but the advantage of providing better incentives because parts of the repayment are shifted to a future period. Neither effect is present with a narrow goal.

[5] In a 4-week weight loss program, Bandura and Simon (1977) assigned 27 subjects either to a condition with 'distal' goals for food consumption over one week or 'proximal' goals for each of four time periods during each day. Yet, more than half of the subjects apparently set proximal goals in the 'distal' treatment. While those subjects with proximal goals lost the most weight, the results must be interpreted with caution because they do not rely on exogenous variation. Bandura and Schunk (1981) ran a remedial program with children who had severe deficits in math. They assigned 10 subjects the narrow goal of completing a certain number of problems in each of seven daily, 30-minute sessions. 10 other subjects were assigned the broad goal of completing the equivalent total number of problems over all sessions. Those assigned the fixed, narrow goals performed better in a number of dimensions than those assigned the fixed, broad goal. This study, however, has been criticized because only children in the narrow goal condition had been able to evaluate their progress toward their goal (cf. Kirschenbaum, 1985, p. 494).

Table 1
Overview of treatments.

	Treatment[a]	Goals set	Goal feedback	Min. work requirement[c]	N
Sections 2-4	Daily	daily	daily	no	78
	Weekly	weekly	weekly	no	77
Section 5.1	Daily(R)[b]	daily	daily	no	75
	Aggregated	daily	weekly	no	75
Section 5.2	DailyRequirement	daily	daily	yes	47
	WeeklyRequirement	weekly	weekly	yes	45
Appendix I	NoManipulation	no goals elicited		no	71

Notes. Total number of subjects 468.
[a] Subjects got daily reminders about goals by email (except NoManipulation).
[b] Daily(R) replicates Daily.
[c] Complete at least one table per day.

By comparing goals and effort under different goal frames, we also contribute to the literature in economics that studies how to optimally design and set goals. Suvorov and van de Ven (2008), Jain (2009), Koch and Nafziger (2011), and Hsiaw (2013) model non-binding personal goals in one-time tasks. Goals help to overcome self-control problems by serving as reference points that make substandard performance psychologically painful. Empirical work yields mixed results on the impact of asking subjects to set goals in one-time tasks. While Akina and Karagozoglub (2017) observe that goals have no effect on performance, Smithers (2015) and Goerg and Kube (2012) find that goal setting increases performance. Other studies consider non-binding goals in work environments. These goals can either be tied to monetary rewards (Dalton et al., 2015; Goerg and Kube, 2012; Kaur et al., 2015) or not (Brookins et al., 2017; Corgnet et al., 2015, 2018). The latter find that goals increase performance. Evidence on the former is mixed. While Dalton et al. (2015) find an overall null result, Goerg and Kube (2012) and Kaur et al. (2015) find positive effects. Overall, most studies point to goals having a positive impact on performance – specifically in the presence of monetary incentives. Several studies point to goals being more effective for men than for women (Smithers, 2015; Dalton et al., 2015; Clark et al., 2019) – a finding that we replicate (cf. Appendix J). Clark et al. (2019) find that effort goals outperform performance goals.

Finally, we contribute to the literature on externally enforced commitment devices (for an overview see Bryan et al., 2010). Ariely and Wertenbroch (2002) observe that imposing binding deadlines on students improves their academic performance. Bisin and Hyndman (2014) and Burger et al. (2011), however, find no such effects. The novel feature of our experiment is to study whether a flexible, externally enforced minimum work requirement can complement self-enforced goals.

2. Experimental design and procedures

Our study included seven treatments, summarized in Table 1. In total 468 subjects participated (see Appendix E for details on the recruitment of subjects). We focus on the main treatments, *Daily* and *Weekly*, in Sections 2-4. In Section 5, we discuss four additional treatments that address certain mechanisms. One treatment (*NoManipulation*) does not contribute directly to the research question and is discussed in Appendix I. Experimental instructions are reproduced in Appendix M.

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

Fig. 1. Example of a table for the counting task.

2.1. Design and treatments

Treatments consisted of a goal setting part and a work part. All parts were conducted online.

Goal setting part On a Wednesday at midnight, subjects received an email informing them that they could earn up to 500 Danish kroners ($83) by completing a short online questionnaire before Friday midnight and performing some online tasks in the following week from Monday to Friday. A reminder was sent out on Friday 9 am to those who had not responded.

Subjects completed a task based on Abeler et al. (2011). The task required them to count correctly the number of zeros in a series of tables with zeros and ones as in Fig. 1.

Subjects had three minutes to complete as many tables as possible and earned DKK 0.5 ($0.08) per completed table. If they miscounted the number of zeros in a table, an error message appeared. A table was not recorded as completed until the correct number was entered.[6] There was no punishment for miscounting. This stage ensured that subjects had a good understanding of how difficult the task was and provided us with a baseline measure of how easy the task was for a subject initially, referred to as *baseline productivity* in the following.

Subjects were then informed that they could complete up to 1,000 such tables at any time during the following week from Monday to Friday, and that they would receive DKK 0.5 per completed table. Each day they would receive an email with a web link to complete the tables. They were asked a set of questions designed to make them think about the benefits of working on the task and their work week ahead.

Subjects completed the part by setting non-binding goals. The goal bracket differed across treatments. In *Daily*, subjects set for each day of the following week a separate goal for how many tables to complete (adding up to at most 1000 tables for the entire week). In *Weekly*, they set a weekly goal of up to 1000 for the number of tables to complete from Monday to Friday. Subjects were informed that we would remind them of their goals in the following week. On the final screen, subjects were told that they would receive an email at 0:00h on Monday with a link to the work screen.

Work part In the following week, each day at 0:00h, subjects received an email with a link to the work screen. Subjects were reminded that they could work anytime until Friday 23:59h. In *Daily* (*Weekly*), the email additionally informed subjects about the goal they had set for that day (the week). The only other treatment difference was how the first two lines above the table to be counted were presented (each table is on a separate screen). In *Daily* (*Weekly*), they showed the goal for the current day (week) and how many tables the subject had already counted on that day (during the week so far). Subjects always saw how many of the 1,000 tables remained, a reminder about the earnings, and a reminder that they could use the link to come back as

6 Subjects were not told that we allowed an error margin of ± 1.

often as they liked. Each time a subject completed a table, a new table appeared and the screen information was updated. If someone miscounted, an error message appeared and the same table was presented again. Upon reaching the maximum of 1,000 tables, a ‘Thank you’ screen appeared and no further counting was possible.

Our design aimed to create a work-leisure self-control problem. It featured generous pay to make it desirable to complete the task. Specifically, our pay was above the usual hourly wage for students of around DKK 130 per hour (completing all 1,000 tables required about 3 hours of work for DKK 500). But once a subject faced the task, its tedious nature made the leisure alternative tempting.

2.2. Procedures

Subjects were informed that payments would be made 2-6 weeks after the experiment by bank transfer via the Danish payment system through which public bodies and companies can send money to a person using their social security number. The procedure was required by Aarhus University and is perceived as normal by Danish citizens. The experiment ran online using the Qualtrics software.[7] It could be accessed via desktop, notebook, or touch-pad. A software filter blocked access via smartphone. This was to avoid that subjects would solve a bit of the task here and there on their smartphone (say, while waiting for the bus), which might not be perceived as costly. Tables were copy protected to prevent pasting them into a spreadsheet program to do the counting. Sample sizes were determined by a rule of thumb, subject availability, and budget (Section 6 discusses power).

3. Theoretical predictions

Our theoretical predictions are based on a setting where a quasi-hyperbolic discounter (Laibson, 1997) works repeatedly on a task and can set daily or weekly goals to motivate himself. We first describe the model before we proceed to the theoretical predictions.

3.1. Model

Task The individual works repeatedly on a task at $t \in \{1, \dots, T\}$. The activity requires effort $e_t \in [0, \infty)$, causing immediate costs $c(e_t)$ (strictly increasing and strictly convex) and long-run benefits $b(e_t)$ (strictly increasing and concave).

Preferences We assume that the individual is a quasi-hyperbolic discounter. Self t (the incarnation of the individual at date $t \in \{0, 1, \dots, T+1\}$) has utility $U_t = u_t + \beta \left[\sum_{\tau=t+1}^{T+1} u_\tau\right]$, where u_t is the instantaneous utility. In the absence of goal setting, the instantaneous utility is given by $u_0 = 0$, $u_t = -c(e_t)$ for $t \in \{1, \dots, T\}$, and $u_{T+1} = \sum_{t=1}^{T} b(e_t)$. The present bias parameter $\beta \in (0, 1)$ captures the extent to which the individual overemphasizes immediate utility flows relative to future utility flows. The exponential discount factor δ is set to one.[8] The present bias causes a work-leisure self-control problem.

[7] Patterns in IP addresses suggest that task outsourcing (e.g. to MTurkers) did not occur (Appendix G).

[8] Using a real-effort task similar to ours, Augenblick and Rabin (2018) find that subjects are present-biased and have an estimated daily discounting parameter $\delta \approx 1$.

Self 0 weighs equally future costs and benefits and thus prefers effort to equate marginal costs and benefits for all dates:

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

Each self $t \geq 1$ discounts future benefits by $\beta < 1$. So for all $t = 1, \ldots, T$, self t prefers effort such that

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

As $\beta < 1$, self 0 wants a higher effort than self t: $e_0^* > e_t^*$. To overcome this self-control problem, self 0 sets effort goals: Either in the form of a narrow goal g_t for each day $t \in \{1, \ldots, T\}$, or a broad goal G for the sum of effort over all T days. In the context of our experiment, $T = 5$ and treatment *Daily* elicits daily goals, whereas treatment *Weekly* elicits a weekly goal.

We allow the individual to hold an overly optimistic belief about his present bias $\hat{\beta} \geq \beta$ (O'Donoghue and Rabin, 1999), encompassing the cases of sophistication ($\hat{\beta} = \beta$), partial naïveté ($\hat{\beta} \in (\beta, 1)$), and full naïveté ($\hat{\beta} = 1$).

Goals as reference points Consistent with the evidence from psychology on goals (e.g. Heath et al., 1999; Locke and Latham, 2002; Wu et al., 2008), we assume that future selves take their goals as reference points (for a model that allows for goal revision see Koch and Nafziger, 2016). With narrow goals, the individual compares the actual effort e_t with the goal g_t. With a broad goal, the individual compares the overall effort $\sum_{t=1}^{T} e_t$ with the goal G. If effort differs from the goal by z, the individual experiences a corresponding comparison utility $\mu(z) = z$ for $z < 0$, and $\mu(z) = 0$ for $z \geq 0$.[9] The individual experiences the comparison utility in the last period when the benefits accrue (this assumption can be relaxed; see Appendix D). That is, with a broad goal, we have $u_{T+1} = \sum_{t=1}^{T} b(e_t) + \min\{0, \sum_{t=1}^{T} e_t - G\}$, and with narrow goals we have $u_{T+1} = \sum_{t=1}^{T} [b(e_t) + \min\{0, e_t - g_t\}]$.

Equilibrium The equilibrium concept is that of preferred personal equilibrium (Kőszegi and Rabin, 2006). Goals are assumed to be rational in the sense of perception perfection (O'Donoghue and Rabin 1999; 2001). The goal(s) that self 0 sets have to be consistent with the (possibly erroneous) beliefs $\hat{e}_{t,0}$ that self 0 holds about his future effort at dates $t = 1, \ldots, T$. That is, $\hat{e}_{t,0} = g_t$ for narrow goals and $\sum_{t=1}^{T} \hat{e}_{t,0} = G$ for a broad goal.

3.2. Analysis: sophisticated individual

We characterize the implementable effort profiles and optimal goals for daily and weekly goal setting formats for a sophisticated individual. In the next section, we discuss what changes when we introduce naïvité.

3.2.1. Daily goals (narrow goals)

Implementable effort profiles To characterize the effort levels that self 0 can implement with daily goals, we need to ask when his future self, who works on task t, does not have an incentive to deviate from goal g_t. If self t puts in at least the effort required by his goal, his utility is

[9] Defining comparison utility over effort matches the frame of the experiment. For an alternative approach that assumes that the goal induces reference standards for costs and benefits, see Koch and Nafziger (2016).

$\beta\, b(e_t) - c(e_t)$. If effort falls short of the goal, he suffers a loss and the utility after a deviation $e_t < g_t$ is:

$$\beta\, b(e_t) - c(e_t) - \beta\,(g_t - e_t). \tag{3}$$

For a goal to be implementable, the utility from sticking to the goal has to exceed the utility from falling short of it. That is, (3) has to be increasing in e_t for any $e_t < g_t$. This is the case for any goal that is not 'too high', i.e., that does not exceed $e_{max}(\beta)$ defined by

$$\beta\,\left[b'(e_{max}(\beta)) + 1\right] = c'(e_{max}(\beta)). \tag{4}$$

The maximal implementable effort $e_{max}(\beta)$, defined by (4), is increasing in β. Further, note that (2) and (4) imply $e_{max}(\beta) > e_t^*$. The maximal implementable effort exceeds the preferred goal of self t because the fear of a loss makes self t strive harder than he would in the absence of comparison utility. Similarly, the effort cannot fall short of e_t^*, because self t will always choose at least this effort level.

Goal setting Self 0 picks his daily goals to maximize his utility subject to the goal being implementable. The following result summarizes the findings. Proofs are in Appendix A.

Proposition 1. *Suppose a sophisticated individual ($\hat{\beta} = \beta$) sets daily goals. Self 0 picks the implementable goal that maximizes his utility:* $\max_{g_t \in [e_t^*, e_{max}(\beta)]}\ \beta\,[b(g_t) - c(g_t)]$.

1. *For β large enough, $e_{max}(\beta) \geq e_0^*$. Self 0 sets daily goals equal to his preferred effort, $g_t^* = e_0^*$ for $t = 1, \ldots, T$, and self t provides $e_t = g_t^*$.*
2. *For lower values of β, $e_{max}(\beta) < e_0^*$. Self 0 sets daily goals $g_t^* = e_{max}(\beta)$ and selves $t = 1, \ldots, T$ provide effort $e_t = g_t^* = e_{max}(\beta)$, where $e_0^* > e_t > e_t^*$. That is, effort exceeds the effort that self t would pick in the absence of comparison utility but still lies below the level e_0^* that self 0 would prefer.*

3.2.2. Weekly goal (broad goal)

Implementable effort profiles A first insight is that a weekly goal cannot improve self-regulation relative to daily goals. This is because the incentives to deviate from the goal in a single period are the same under a weekly goal as under daily goals – provided all other selves stick to the plan. Thus, $e_{max}(\beta)$ defined in (4) is also the maximal implementable effort in a given period with a weekly goal. If self t believes that the weekly goal can still be reached with some $e_t \leq e_{max}(\beta)$, he provides at least such an effort. If self t believes that G no longer will be achieved for any $e_t \leq e_{max}(\beta)$, he provides $e_{max}(\beta)$ (see Lemma 1 and its proof in Appendix A.2).

But a weekly goal can harm self-regulation. Self 0 may not be able to implement certain effort profiles with a weekly goal that are implementable with a daily goal, because future selves would deviate, for example, by lowering effort today and compensating with increased effort tomorrow. We refer to such behavior as *effort substitution*. To provide some intuition, suppose self 0 sets a weekly goal that equals the sum of his desired daily effort, $G = T\,e_0^*$, and $e_0^* < e_{max}(\beta)$. Further, suppose that all selves provided e_0^*, except that self $T-1$ worked less hard than e_0^*. Now if self T just provided e_0^*, the individual would suffer a loss from falling short of G. To avoid this loss, self T will increase his effort up to $e_{max}(\beta)$. For self $T-1$ it indeed pays off to work less than e_0^*, knowing that self T will work harder to make up for the shortfall. Because of the present bias, self $T-1$ prefers (on the margin) to shift effort costs into the future. Consequently, the individual does not provide e_0^* in every period when facing the broad goal $G = T\,e_0^*$ (see also

Proposition 4 in Appendix A.3.3). That is, even though $e_0^* < e_{max}(\beta)$ self 0 cannot implement his preferred daily effort e_0^* in each period. Instead, he has to implement an increasing effort profile that prevents him from pushing (even more) effort to the future. In contrast, if $e_0^* = e_{max}(\beta)$, each self $t \in \{1, \ldots, T-1\}$ is committed not to lower his effort because future selves will not compensate. The next result summarizes and extends these insights for any weekly goal $G \leq T\, e_{max}(\beta)$. The effort profile can either be increasing, $e_1 \leq e_0^* < e_T$ (with $e_T < e_{max}(\beta)$ or a corner solution $e_T = e_{max}(\beta)$), or flat with $e_t = e_{max}(\beta)$ in all periods $t = 1, \ldots, T$.

Proposition 2. *Suppose a sophisticated individual ($\hat{\beta} = \beta$) sets a weekly goal $G \leq T\, e_{max}(\beta)$. The goal is achieved: $G = \sum_{t=1}^{T} e_t$.*

1. *For $e_{max}(\beta) > e_0^*$, effort is (weakly) increasing over time: $e_1 \leq e_0^* < e_T$. There exists $\underline{t} \in \{2, \ldots, T+1\}$ such that $e_1 < e_2 < \cdots < e_{\underline{t}-1} < e_{\underline{t}} = e_{\underline{t}+1} = \cdots = e_T = e_{max}(\beta)$. If $\underline{t} > 2$, $e_1 < e_0^* < e_{\underline{t}-1}$. If $\underline{t} = 2$, $e_1 = e_0^*$.*
2. *For $e_{max}(\beta) \leq e_0^*$, $G^* = T e_{max}(\beta)$, $e_t = e_{max}(\beta)$ in all periods $t = 1, \ldots, T$.*

Goal setting How does effort substitution affect the goal that self 0 sets? Our next result imposes a technical assumption on the third derivatives of the benefit and cost functions, which is satisfied, for example, if the benefits are linear and costs are quadratic.[10] Under this assumption, we can compare the optimal weekly goal to the optimal daily goals. For $e_{max}(\beta) \leq e_0^*$, there is no difference between the two goal setting formats in that $G^* = T\, g^* = T\, e_{max}(\beta)$. So we focus on the case where $e_{max}(\beta) > e_0^*$.

Proposition 3. *Suppose a sophisticated individual ($\hat{\beta} = \beta$) sets a weekly goal, $\beta\, b'''(e) - c'''(e) \geq 0$ and $b'''(e) - c'''(e) \geq 0$. Define $\check{\beta} : e_{max}(\check{\beta}) = e_0^*$.*

1. *There exists $\bar{\beta} \in (\check{\beta}, 1)$, such that for $\beta \geq \bar{\beta}$ self 0 sets a total goal $G^* < T\, e_0^*$ and anticipates interior effort $\hat{e}_{t,0} < e_{max}(\beta)$ for every period $t = 1, \ldots, T$.*
2. *There exists $\underline{\beta} \in (\check{\beta}, \bar{\beta})$, such that for $\beta \in [\check{\beta}, \underline{\beta}]$ self 0 sets $G^* = (T-1)\, e_{max}(\beta) + e_0^*$ and anticipates a corner solution for periods $t > 1$: $(e_0^*, e_{max}(\beta), \ldots, e_{max}(\beta))$.*

Given that $e_{max}(\beta) > e_0^*$, self 0 could in principle implement the sum of desired daily effort levels with a weekly goal $G = T\, e_0^*$ and thereby achieve the same overall effort as with daily goals $g_t = e_0^*$. Yet, because of effort substitution, effort would be asymmetrically allocated over the days, starting below and ending above the desired daily effort of self 0. Because of the strictly convex effort cost, the utility of self 0 under the weekly goal $G = T\, e_0^*$ is lower than the utility under the equivalent daily goals $g_t = e_0^*$. Part 1 of Proposition 3 shows that self 0 chooses a lower weekly goal than $T\, e_0^*$ if β is sufficiently large. The reason is that lowering the goal relative to $T\, e_0^*$ reduces the spread in effort costs across periods and leads to a lower average cost per unit of effort. Part 2 of Proposition 3 shows that for a relatively severe present bias (low β) it may pay to commit at least self 1 to provide e_0^* by setting a weekly goal that forces all selves $t > 1$ to provide $e_t = e_{max}(\beta)$. Essentially, this follows from the continuity of $e_{max}(\beta)$: If β is such that $e_{max}(\beta)$ only slightly exceeds e_0^*, the cost of excessive effort $e_{max}(\beta)$ is negligible

[10] Augenblick and Rabin (2018) find an approximately quadratic cost function when estimating effort costs for a real-effort task similar to ours.

and the utility is close to the self 0 optimum. Note that part 1 corresponds to the case where $\underline{t} = T + 1$ in Proposition 2 (self 0 anticipates interior effort $\hat{e}_{t,0} < e_{max}(\beta)$ for every period $t = 1, \ldots, T$) and part 2 corresponds to the case where $\underline{t} = 2$. These two cases are the essential ones for understanding the main driving forces. The other possibilities of partial corner solutions $\underline{t} \in \{3, \ldots T\}$ are included in the proof and show up as a 'staircase' pattern in the left panel of Fig. 2 for our parametric example below.

3.3. Analysis: partially naïve individual

We next discuss the implications of partial naïvité. We make two assumptions. First, the individual cannot revise the goal. Second, given past effort and his wrong belief about the present bias, the individual is able to update beliefs about future effort and goal achievement. We discuss these assumptions in Section 6.

3.3.1. Daily goals

A partially naïve self 0 picks his daily goals for tasks to maximize his utility $\hat{\beta}\,[b(g_t) - c(g_t)]$ subject to $g_t \in [e_t^*, e_{max}(\hat{\beta})]$. The individual overestimates what goals are realistic, because the perceived maximal implementable effort $e_{max}(\hat{\beta})$ exceeds the actual $e_{max}(\beta)$ for $\hat{\beta} > \beta$ (for details see Appendix A.1). Yet, this mistake has no consequences as long as the goal satisfies $g_t \leq e_{max}(\beta)$. In this case, the goal is achieved. If the individual sets an overly ambitious goal $g_t > e_{max}(\beta)$, self t will underperform relative to that goal.

3.3.2. Weekly goal

A partially naïve individual sets weekly goals in the same way as a sophisticated individual, except that the individual applies $\hat{\beta}$ (the belief about his present bias) when predicting future behavior rather than β (his true present bias). As a consequence, in Proposition 3, whether case 1 or case 2 applies does not depend on the actual β, but on the belief $\hat{\beta}$. For example, self 0 chooses a weekly goal $G < T\,e_0^*$ if $\hat{\beta}$ is sufficiently large.

However, the actual effort pattern differs from the one that self 0 anticipates. First, self 0 anticipates that $\hat{\beta}$ will be applied whereas, when deciding on the effort e_τ, self $\tau > 0$ applies the actual $\beta < \hat{\beta}$. Second, while all selves hold the same wrong belief that future selves will provide effort up to $e_{max}(\hat{\beta})$, at some point the individual will observe a different history of past effort than self 0 anticipated. At that point, self τ will plan to compensate for the short-fall and beliefs will have to shift upward to adjust for lagging behind the original expectations. As a result, actual effort falls short of the one-period ahead expectation and expectations about future effort increase relative to those held in the previous period.[11] For a formal statement and proof of this result, see Proposition 5 in Appendix A.6.

Naïvité exacerbates the problem of effort substitution. Because the individual incorrectly predicts the extent to which a future self will increase his effort in response to him providing less effort today, the individual might lower his effort 'too much' and fail to meet the goal. Note that the individual believes that his future selves will provide effort up to $e_{max}(\hat{\beta})$. So even if the goal is no longer achievable, the individual will not realize this until the point comes when the goal cannot be achieved even with $e_{max}(\hat{\beta})$ in every following period. In our parametric example in Fig. 2, for example, the individual only realizes in the final period that he will fail the goal.

[11] Except in the case of a corner solution where self 0 expects effort $\hat{e}_{1,0} = e_0^*$ for period 1 and $\hat{e}_{t,0} = e_{max}(\hat{\beta})$ for periods $t = 2, \ldots, T$. Here actual effort in period 1 may match expectations: $e_1 = \min\{e_0^*, e_{max}(\beta)\}$.

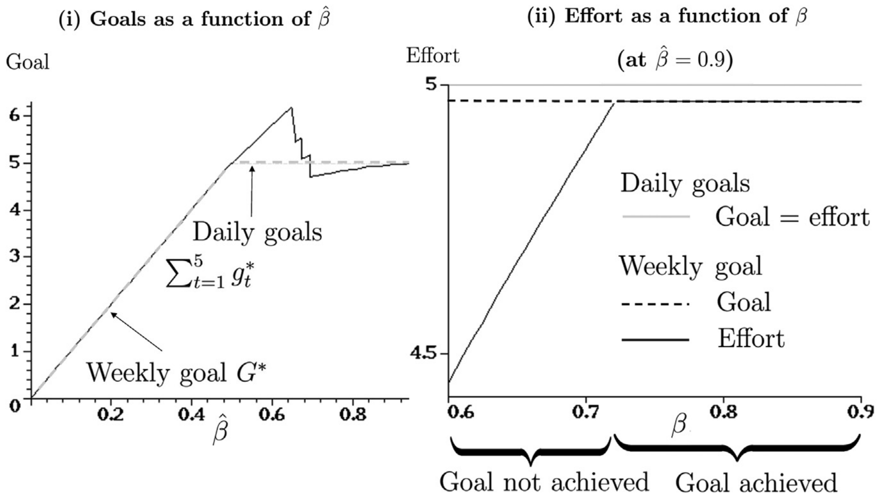

Fig. 2. Goals and effort in the parametric example with $T = 5$.

Naïvité also has consequences for goal achievement. A partially naïve individual is weakly more likely to fail the weekly goal compared to a situation where he sets daily goals. For $e_{max}(\beta) < e_0^*$, the individual fails to reach his goals under both goal setting formats. For $e_{max}(\beta) \geq e_0^*$, daily goals would be achieved ($g_t^* = e_0^* = e_t$). Here the weekly goal may not be achieved though: The individual only achieves the goal if self T faces a remaining part of the goal $G_T = G - \sum_{t=1}^{T-1} e_t \leq e_{max}(\beta)$. Clearly, this is violated in case of a corner solution where self 0 anticipates effort in period T of $\hat{e}_{T,0} = e_{max}(\hat{\beta})$, because $e_{max}(\hat{\beta}) > e_{max}(\beta)$. In case of an interior solution with $\hat{e}_{T,0} < e_{max}(\hat{\beta})$, we know from Proposition 3 that self 0 anticipates $\hat{e}_{T,0} > e_0^*$. In addition, because of naïveté, the actual effort falls short of the one-period ahead expected effort. Consequently, expectations about future effort increase. This further pushes-up the remaining goal G_T. So even though self 0 sets what looks like a 'feasible' goal ($G^* < T\, e_0^* < T\, e_{max}(\beta)$), self T faces $G_T > e_0^*$ which may exceed $e_{max}(\beta)$.

3.4. Parametric example

To illustrate which cases are most relevant we provide a parametric example with $b(e) = e$ and $c(e) = \frac{1}{2} e^2$. Appendix B provides the detailed calculations. We have $e_0^* = 1$, $e_1^* = \beta$, and $e_{max}(\beta) = 2\,\beta$. With daily goals, self 0 sets as goal his preferred effort as long as $\hat{\beta} \geq \frac{1}{2}$, because then $e_0^* \leq e_{max}(\hat{\beta})$. The left panel of Fig. 2 plots the optimal weekly goal G^* as a function of $\hat{\beta}$ if $T = 5$, like in our experiment. For $\hat{\beta} \leq \frac{1}{2}$ the individual implements $e_{max}(\hat{\beta})$, as he would for daily goals. For $\hat{\beta} \in (\frac{1}{2}, \bar{\beta})$, where $\bar{\beta} \approx 0.694$, the individual has a higher weekly goal than the sum of daily goals would have been, because of a corner solution like in Part 2 of Proposition 3. For $\hat{\beta} \geq \bar{\beta}$, Part 1 applies and the individual has a lower weekly goal than the sum of daily goals would have been.

Augenblick et al. (2015) estimate a population present bias parameter of $\beta = 0.9$ using a real-effort task similar to ours and find that almost all mass of present-biased individuals lies on $\beta \in [0.6, 1)$ (see their Figure VI).[12] Augenblick and Rabin (2018) estimate $\beta = 0.83$ and $\hat{\beta} \approx 1$. Individual estimates for present-biased individuals are concentrated on $\beta \in [0.5, 1)$ and $\hat{\beta}$ is concentrated around 1 (see their Figure 6). They find only a weak correlation between β and $\hat{\beta}$.

In the right panel of Fig. 2, we therefore consider a partially naïve individual with $\hat{\beta} = 0.9$ and the actual effort response for $\beta \in [0.6, 1]$. Note that the goals for a partially naïve individual depend on $\hat{\beta}$ and do not vary with β. With daily goals, the individual would always achieve his goal and the implemented effort would equal the preferred effort of self 0 ($\sum_t^T g_t^* = \sum_t^T e_t = Te_0^* = 5$). For a weekly goal, the goal G^*, and hence the effort, are always below what they would have been with daily goals. The weekly goal is achieved except in the case where the individual severely overestimates β. Non-achievement occurs for $\beta < \tilde{\beta}$, where $\tilde{\beta} \approx 0.722$.

To sum up, the evidence on β and $\hat{\beta}$ from Augenblick et al. (2015) and Augenblick and Rabin (2018) indicates that most people have $\hat{\beta}$ close to 1 and that few people have a severe present bias. This suggests that Part 1 of Proposition 3 is the relevant case for most subjects, i.e., subjects set a lower goal and provide less effort under a weekly goal than under daily goals. Further, few subjects are likely to so severely overestimate β that they drive a wedge between *Weekly* and *Daily* in terms of goal achievement. With these observations, we reach the following hypotheses:

Hypothesis 1 *(Goals and effort).*

H1a *The total goal in Daily is larger than in Weekly.*
H1b *The total effort in Daily is larger than in Weekly. Controlling for the total goal, the total effort in both treatments is equal.*

Hypothesis 2 *(Goal achievement).*

H2a *The goal achievement rates in Daily and Weekly are equal.*
H2b *Goal non-achievers are closer to their total goal in Daily than in Weekly.*

Hypothesis 3 *(Effort substitution).*

H3a *The effort on Monday in Daily is higher than in Weekly.*
H3b *The effort on Friday in Daily is lower than in Weekly.*

4. Empirical findings: daily goals vs. weekly goal

This section analyzes the observed treatment differences between *Daily* and *Weekly*, considering in turn goals, effort, goal achievement, and effort substitution. For each hypothesis, we

[12] Some of our subjects participated in a prior study (Epper et al., 2018), where we implemented the elicitation task of Augenblick et al. (2015). We estimate a population β of 0.94, restricting our sample to those subjects whose choices were monotonic. Compared to Augenblick et al. (2015), a larger proportion of subjects in our data make choices that are non-monotonic in the 'efficiency' of effort (conversion rate between effort today vs. effort in one week). Around 45 percent of subjects is present biased, 15 percent dynamically consistent, and 40 percent future biased, compared to 33, 47, and 21 percent, respectively in Augenblick et al. (2015). The small number of observations where we are able to estimate an individual $\beta < 1$ prevents us from exploiting this measure in our empirical analysis.

Table 2
Descriptive statistics.

Treatment	N	Average total		Fraction of subjects with			Average		
		goal	effort	goal =1000	effort =1000	effort <goal	effort −goal	number of logins[a]	effort per login
Daily	78	789	690	0.51	0.41	0.45	−98.54	6.72	131.45
		(304)	(370)				(399.47)	(4.78)	(109.63)
Weekly	77	682	521	0.47	0.30	0.47	−161.01	5.52	104.06
		(336)	(392)				(398.14)	(4.32)	(97.90)

Notes. Standard deviations in parentheses.
[a] New login: new day or >30 min. since last entry.

present descriptive statistics, non-parametric tests, and use regression analysis to test for treatment differences. Tables and figures prefaced with 'A' are in the online appendix.

Estimation procedure Our design aimed to make high effort desirable from the perspective of self 0. This led to substantial censoring of goals and effort at the maximum of 1,000 tables.[13] We deal with censoring using tobit regressions. Robustness checks indicate that results are robust to relaxing assumptions of the tobit model (Appendix L).

We regress the respective outcome variables on *daily goals*, which is a dummy for the treatment being *Daily*. The following primary outcome variables are available for all 155 subjects in *Daily* and *Weekly*. To test Hypothesis H1a (Section 4.1), we use the *total goal* (the aggregated daily goals in *Daily*, the weekly goal in *Weekly*). To test H1b (Section 4.2), we use the *total effort* (the total number of correctly counted tables). To test H2a,b (Section 4.3), we use *goal achievement* (a dummy variable whether *total effort*≥*total goal*). Finally, to test for the effort substitution hypotheses H3a,b (Section 4.4), we use *effort* on a given weekday, with specific focus on Monday and Friday. Secondary outcome variables, such as logins and task completion time allow us to study some mechanisms (Sections 5 and 6).

Control variables available for all treatments are a *gender* dummy and *baseline productivity* (balance tests are in Table A4).[14] Data from a prior study (Epper et al., 2018) provide us with a wider range of control variables (*full set of controls*), explained in Appendix F. For some of the extension treatments, these are not available though.

4.1. Goals

We test Hypothesis H1a by comparing the total goal in *Daily* and *Weekly*. In line with the hypothesis, subjects in *Daily* set a total goal of 789 tables on average, whereas those in *Weekly* set a goal of 682 (Table 2; Permutation test for two independent samples, $p = 0.041$).[15]

---

13 Roughly half of the subjects in *Daily* and *Weekly* chose a goal of completing all 1,000 tables. 30 and 41 percent of subjects, respectively, completed all 1,000 tables (Table 2).

14 Before subjects set goals, we asked them how many hours they had available for the task in the following week, how many tables they thought they could realistically solve within this time, and how much money they wanted to earn. These questions were meant to make subjects think about the task and their goals. They are collinear with the goals that subjects set and therefore not used as control variables.

15 A non-parametric test of difference in distributions. Permutation tests perform well even with censoring (Neuhaus, 1993). We ran Monte Carlo simulations with 10,000 repetitions, using Kaiser's (2007) implementation `permtest2`. Permutation tests and Fisher's exact tests reported below are all against the two-sided null hypothesis of no difference.

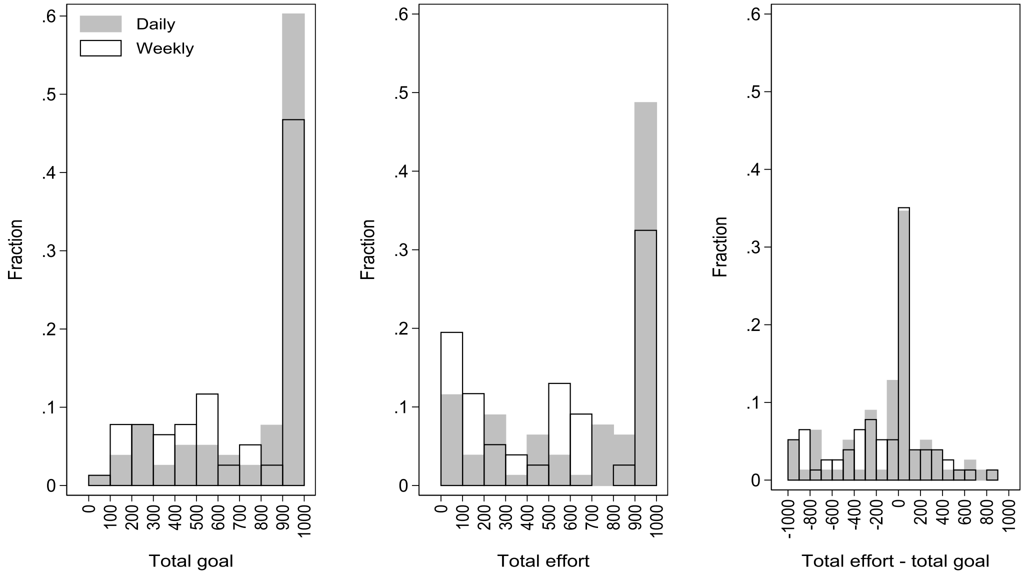

Fig. 3. Distribution of goals and effort (*Daily* & *Weekly*).

In tobit regressions, this difference is significant only after controlling for gender and productivity – indicating that the coefficients are imprecisely estimated without the controls (Table 3). Looking at the distribution of goals shown in the left panel of Fig. 3, *Daily* had more mass on high goals than *Weekly*. For example, 65 percent aimed for at least 800 tables in *Daily* compared to 48 percent in *Weekly*. This can also be seen in the left panel of Fig. 4, which plots the quantiles of goals. All quantiles below the median appear to be shifted upward in *Daily* relative to *Weekly*. Quantile regressions (controlling for baseline productivity and gender; with bootstrapped standard errors, 1000 replications) reveal significant treatment effects for the 20th and 30th percentile of 172.18 (se =88.51, p =0.054) and 245.10 ($se = 116.44$, $p = 0.037$), respectively. For the 40th and 50th percentile, we obtain 161.11 ($se = 121.24$, $p = 0.186$) and 40.4 ($se = 112.76$, $p = 0.721$), respectively.

4.2. *Effort*

First, we provide some descriptive statistics on the effort measure, and then we test Hypothesis H1b by comparing the total effort in *Daily* and *Weekly*.

Subjects took on average 15 seconds to complete a table and 90 percent of subjects required 20 seconds or less on average per table.[16] More productive subjects were slightly faster (OLS coefficient -0.390, $se = 0.124$, $p = 0.002$). A one standard deviation higher baseline productivity cut completion time by 1.6 seconds. Subjects made 38 mistakes in the range of $\pm$ 1 around the correct solution on average. For technical reasons, we could not record mistakes that required recounting of a table.

In line with Hypothesis H1b, subjects provided more effort in *Daily* than in *Weekly* on average: 690 vs. 521 tables (Table 2; permutation test, $p = 0.005$). This difference is significant in tobit

[16] We truncate completion time at 120 seconds to take out effects of breaks/stopping to log on again later.

Table 3
Impact of goal setting format on total goal and total effort (*Daily* vs. *Weekly*).

Dependent variable	Total goal			Total effort				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Daily goals[a]	150.95	173.14*	186.89**	235.46**	228.39**	180.91*	151.44*	100.89
	(96.82)	(89.37)	(92.25)	(96.97)	(94.86)	(95.09)	(91.31)	(92.44)
Total goal							0.60***	0.60***
							(0.14)	(0.15)
Baseline productivity		24.17**	10.88		33.52***	67.04***	24.49**	57.82**
		(11.08)	(19.00)		(10.67)	(23.59)	(10.38)	(22.41)
Female		-417.57***	-448.27***		-93.66	-176.73	53.56	-26.97
		(90.65)	(102.71)		(99.86)	(117.39)	(99.11)	(114.60)
Constant	866.26***	677.16***	805.54	611.57***	109.27	-99.35	-230.05	-349.58
	(78.33)	(184.81)	(562.15)	(71.05)	(176.97)	(645.24)	(168.23)	(616.96)
Full controls[b]	no	no	yes	no	no	yes	no	yes
Margin.effect(daily goals)[c]	75.14	89.44*	97.36**	135.79**	134.38**	110.07*	93.93*	64.15
Effect size[d]	0.22	0.27	0.29	0.35	0.34	0.33	0.24	0.19
N	155	155	153	155	155	153	155	153

Notes. Tobit coefficient (marginal effect on the latent dependent variable) with robust standard error in parenthesis.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

[a] Dummy for *Daily*.

[b] See Appendix F.

[c] Tobit marginal effect on the censored latent variable (at the means of control variables).

[d] $\frac{\text{Margin.effect(daily goals)}}{\text{Standard deviation of total goal in } \textit{Weekly}}$.

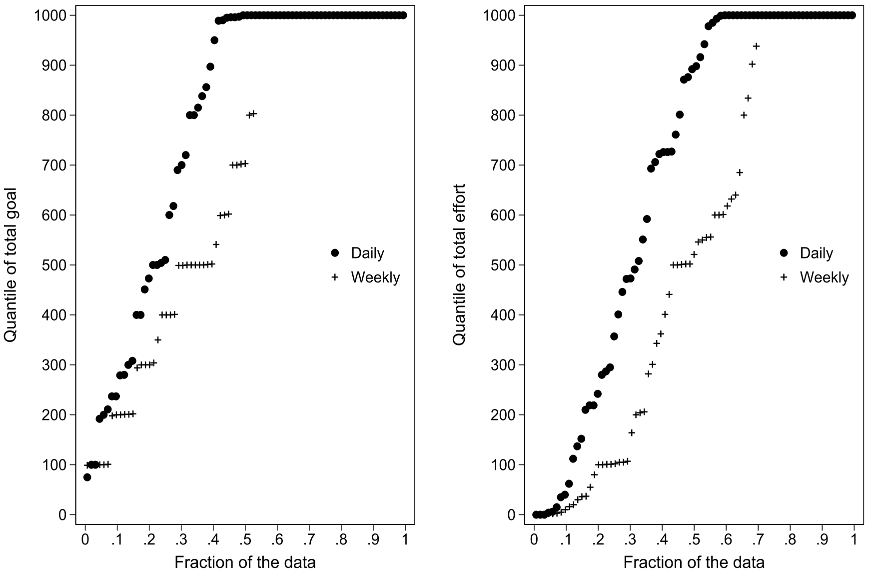

Fig. 4. Quantiles of goals and effort (*Daily* & *Weekly*).

regressions (Table 3) and is also reflected in the distribution of effort. The middle panel of Fig. 3 shows that *Daily* has more mass on high effort than *Weekly*. In the same vein, the right panel of Fig. 4 shows that quantiles shift upward in *Daily* relative to *Weekly*. Quantile regressions reveal (borderline) significant treatment effects for the $20th$, $30th$, $40th$, and $50th$ percentile of 209.33 ($se = 101.08$, $p = 0.040$), 223.00 ($se = 35.18$, $p = 0.101$), 230.71 ($se = 134.06$, $p = 0.087$), and 273.31 ($se = 123.60$, $p = 0.029$), respectively.

As the goal is endogenous, the effort gap between treatments either could be driven by the different goal frame or by differences in goal setting or goal achievement. We examine the latter two explanations next.

Controlling for the total goal Once we control for the total goal, the treatment difference in effort becomes smaller and is no longer consistently significant (Table 3). The effort gap between *Daily* and *Weekly* appears mainly to be related to the higher total goal in *Daily* compared to *Weekly*, in line with Hypothesis H1b. Another indication that *Daily* increased effort by exogenously shifting goals upwards relative to *Weekly* is that the instrumental variables tobit coefficient in a 2SLS regression is larger than the regular tobit coefficient on total goal (Tobit 0.66, $se = 0.14$, $p < 0.001$; IV 1.89, $se = 0.92$, $p = 0.039$).

4.3. Goal achievement

In this section, we test for treatment differences in goal achievement (Hypotheses H2a and H2b). According to our model, naïve individuals overestimate the maximum amount of effort they will be able to put in. However, as summarized in Hypothesis H2a, this should typically not lead to a treatment difference in goal achievement. Indeed, we find no treatment difference in goal achievement (Fisher's exact test, $p = 0.872$; Table A5.)

In addition, according to Hypothesis H2b, goal non-achievers are predicted to be closer to their total goal in *Daily* than in *Weekly*. Indeed, the right panel of Fig. 3 shows that more mass is

Table 4
Transition from goal profiles to effort profiles in *Daily*.

Goal profiles	Effort profiles							
	Flat[d]		High-low[e]		Low-high[f]		Other	
	N	Percent	*N*	Percent	*N*	Percent	*N*	Percent
Flat[a]	1	5.88	12	70.59	3	17.65	1	5.88
High-low[b]	0	0.00	24	77.42	5	16.13	2	6.45
Low-high[c]	0	0.00	13	48.15	13	48.15	1	3.70
Other	0	0.00	2	66.67	1	33.33	0	0.00

Notes. [a] $g_{mon} = g_{tue} = \cdots = g_{fri}$. [b,c] $g_{mon} + g_{tue} >^{b} (<^{c}) g_{thu} + g_{fri}$. [d,e,f] Analogous to goal profiles.

concentrated just to the left of zero in *Daily* than in *Weekly*. Tobit regressions of *total effort-total goal* for those subjects who fell short of the total goal (but attempted working) reveal a significant treatment effect once controlling for gender and productivity (Table A6).[17]

4.4. *Effort substitution*

We examine the emergence of effort substitution by looking at effort patterns over time. Specifically, we test Hypothesis H3a (H3b) by comparing effort on Monday (Friday) for *Daily* vs. *Weekly*. Fig. 5 reveals that the average daily goals in *Daily* exhibited a downward sloping pattern (also found in the other treatments with daily goals discussed in Section 5). Subjects might have wanted to work less at the end of the week, because student parties or other leisure options increased opportunity costs. Extending our model with marginal costs that increase in t leads to a downward sloping pattern of goals for *Daily*, because both the self-0 preferred effort for date t, $e^*_{t,0}$, and the maximal implementable effort, $e_{max,t}(\beta)$, then are decreasing in t. For *Weekly*, this extension affects effort substitution (Hypotheses H3a,b). While effort on Monday still is likely to be lower in *Weekly* than in *Daily*, the prediction of higher effort on Friday in *Weekly* than in *Daily* is weakened because a corner solution with $e_T = e_{max,T}(\beta)$ in both treatments is more likely (see Appendix C).

The data from *Daily* reveal some heterogeneity in goal profiles (Table A7). 86 percent of subjects set a non-zero goal for each day, 22 percent set the same goal for each day, 40 percent aimed to 'start high and end low' (the average goal for the first two days of the week exceeded the average goal for the last two days), and 35 percent aimed to 'start low and end high'. Considering actual effort in *Daily*, most subjects ended up 'starting high and ending low', and those who aimed to 'start high and end low' where the most likely to follow their planned profile of effort (Table 4). The tendency is similar for goal achievers and non-achievers (Table A9). This suggests that some subjects failed to predict either that other things might get in the way, or that they might simply be less attentive to the study later in the week. For *Weekly* we do not have information on the planned profile of effort. Looking at the actual effort profile, most subjects ended up 'starting high and ending low' – like in *Daily* (Table A10).

Fig. 5 suggests that subjects in *Weekly* worked less on Monday than subjects in *Daily*, as predicted by Hypothesis H3a. Yet, a permutation test yields $p = 0.218$ and the effect is not robustly significant in the regressions. In addition, we do not see in *Weekly* the pattern of catching

[17] Because, $e_{max}(\beta) > 0$ for $\beta > 0$, the argument behind Hypothesis H2b only applies if any attempt at effort is made (total effort>0). This excludes three subjects from each treatment, with an average goal of 999 for *Daily* and 734 for *Weekly*.

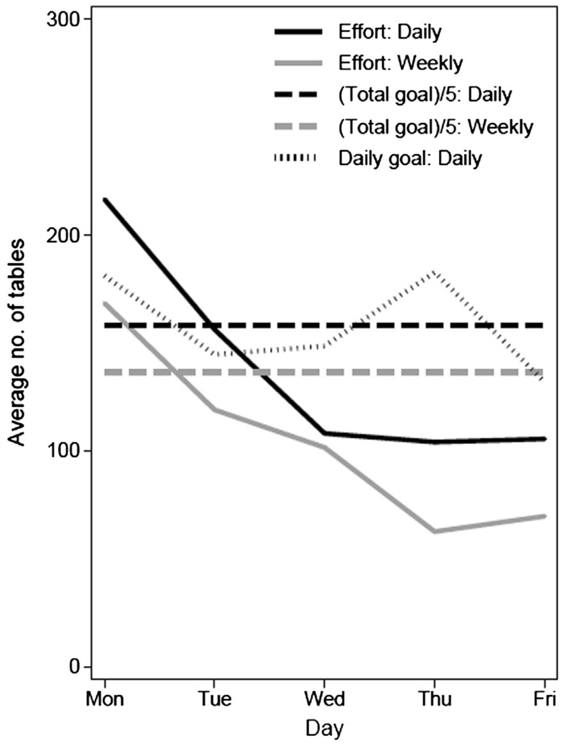

Fig. 5. Average daily effort and goals (*Daily* and *Weekly*).

up on Friday predicted by Hypothesis H3b (Tables A13 and A14). One explanation could be that effort costs are increasing toward the end of the week. As explained above, this would weaken the prediction of a treatment difference in effort on Friday. In addition, as noted before, the patterns of daily goals suggest heterogeneity in effort costs across days. In our comparison of *Daily* with *Weekly* we can only control for the total goal but have no control for effort costs on a given weekday, which might not be balanced across treatments. Our further treatments *Aggregated* and *Daily(R)*, discussed in the next section, address this issue.

5. Extensions

In this section, we consider two extensions of our main treatments. The first allows us to cleanly test for effort substitution. The second explores the effects of adding a minimum work requirement in addition to goal setting.

*5.1. Goal feedback format (*Daily(R) *vs.* Aggregated*)*

By comparing *Daily* and *Weekly* we examined the overall effect of the goal format on goals and effort. Yet, as the goal level was endogenous, the control for the *total goal* in Hypothesis H1b and the corresponding regressions provided only an imperfect test of whether the difference in effort was caused by the goal bracket (daily vs. weekly) or by the difference in goals that the different treatments induced. Further, the goal setting patterns in *Daily* suggested non-constant effort costs over time, for which we could not control in our examination of effort substitution (Hypotheses H3a,b). To address these issues, we conducted treatments *Daily(R)* and *Aggregated*. We first describe the design and hypotheses.

5.1.1. Design and hypotheses

Treatments All subjects set daily goals like in *Daily*. Then they were randomized either to *Daily(R)* (where subjects got feedback about their daily goals, thus replicating *Daily*), or to *Aggregated* (where they got feedback about their weekly goal derived by aggregating the daily goals). There was daily feedback about goals in both treatments, as in *Daily* and *Weekly*.

Hypotheses First, the treatments allow us to compare effort with daily and weekly goals 'holding fixed' the total goal (Hypothesis H1b). By design, the total goal should not vary across *Daily(R)* and *Aggregated*, so we can directly compare effort across treatments. Second, the treatments provide us with daily goals for both treatments that we can use as controls to obtain a cleaner test of effort substitution (H3a,b).[18] In addition, the treatments are of independent interest because showing the effect of exogenously shifting the framing (but not the level) of goals is highly relevant to organizations (see the discussion of the study by Cadena et al., 2011, in Section 7).

5.1.2. Analysis

We follow the structure of Section 4, comparing across treatments the goals, effort, goal achievement, and effort patterns over time. For our hypotheses to be valid, we need to check that the data do not contradict the assumptions behind the hypotheses. In Appendix H we confirm (i) that there was no treatment difference in the goals, and (ii) that the framing of the goal feedback was successful in the sense that subjects in *Aggregated* responded to the weekly goal about which feedback was given and not to the daily goals they had initially set.

Total effort We first test Hypothesis H1b. In line with the hypothesis, we find no treatment effect on total effort (permutation test, $p = 0.247$; Table A12). Yet, some caution should be taken when interpreting this result as the non-significant effect might be due to the sample size.[19]

Goal achievement In line with Hypothesis H2a, we find no treatment difference in the likelihood of achieving the total goal (Fisher's exact test, $p = 0.101$; the logit coefficient becomes insignificant once adding controls in Table A5.) In line with H2b, the right panel of Fig. 6 shows that more mass is concentrated just to the left of zero in *Daily(R)* than in *Aggregated*. Yet, tobit regressions of *total effort-total goal* for subjects who fell short of the total goal but attempted working (total effort> 0) reveal no significant treatment effect (Table A6).

Effort substitution Our second objective was to cleanly test for effort substitution (H3a,b) by exploiting daily goals as a control for possible heterogeneity in effort costs over the course of the week. In line with Hypothesis H3a, Fig. 7 suggests that subjects in *Aggregated* worked less on Monday than subjects in *Daily*. While a permutation test has $p = 0.154$, the tobit coefficient

[18] It is straightforward from the proofs of Propositions 2 and 5 that H3a,b continue to apply, as do H2a,b.

[19] Table A12 indicates that the treatment dummy is not statistically significant. Without controls, the coefficient appears large (-108.26). Given the large coefficient, there might be some concern that the result is imprecisely estimated. However, the effect size is negligible (-0.13). With controls, the coefficient is small and the effect size is close to zero.

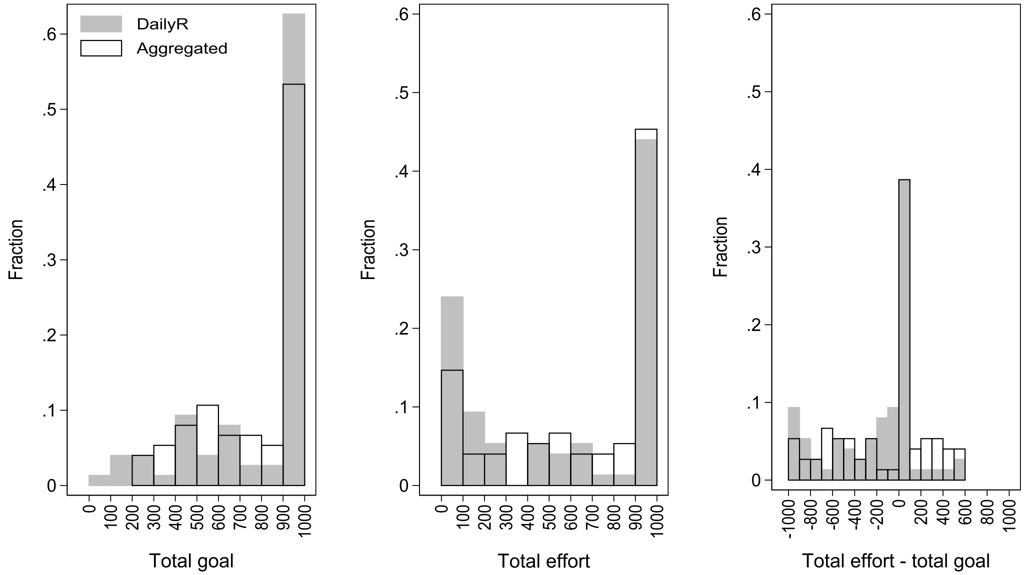

Fig. 6. Distribution of goals and effort (*Daily(R)* & *Aggregated*).

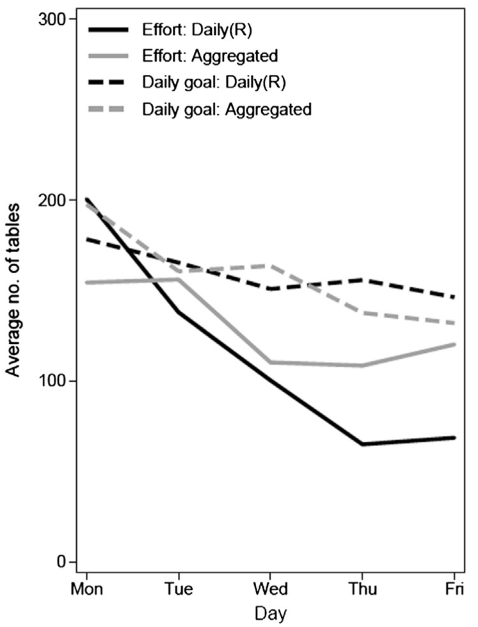

Fig. 7. Average daily effort and goals.

becomes significant once adding controls (Table A13). Further, the figure suggests that subjects made up for the shortfall by working harder on Friday in *Aggregated* than subjects in *Daily(R)*, in line with Hypothesis H3b. This is confirmed by a permutation test ($p = 0.017$) and tobit regressions (Table A14).

Table 5
Hypotheses and summary of findings.

Hypothesis		Finding	Section	Table
Goals and effort				
H1	Total goal(*Daily*)>total goal(*Weekly*)	✓	4.1	3
H1b	Total effort(*Daily*)>total effort(*Weekly*)	✓	4.2	3
	Controlling for total goal, total effort(*Daily*)=total effort(*Weekly*)	✓	4.2	3
H1b'	Total effort(*Daily(R)*)=total effort(*Aggregated*), as total goal equal	✓	5.1	A12
Goal achievement				
H2a	Goal achievement rate(*Daily*)=goal achievement rate(*Weekly*)	✓	4.2	A5
	Goal achievement rate(*Daily(R)*)=goal achievement rate(*Aggregated*)	✓	5.1	A5
H2b	Goal non-achievers are closer to total goal in *Daily* than in *Weekly*	✓	4.2	A6
	Goal non-achievers are closer to total goal in *Daily(R)* than in *Aggregated*	✗	5.1	A6
Effort substitution				
H3a	Monday effort(*Daily*)>Monday effort(*Weekly*)	✗	4.4	A13
	Monday effort(*Daily(R)*)>Monday effort(*Aggregated*)	✓	5.1	A14
H3b	Friday effort(*Daily*)<Friday effort(*Weekly*)	✗	4.4	A14
	Friday effort(*Daily(R)*)<Friday effort(*Aggregated*)	✓	5.1	A14

Note: Tables prefaced with 'A' are in the online appendix.

Our treatments offer an additional way to test Hypothesis H3b. All subjects stated a daily goal for Friday. Taking this goal as a benchmark, we indeed find that subjects in *Aggregated* were more likely to work harder than that benchmark on Friday in *Aggregated* than subjects in *Daily(R)* (Fisher's exact test, $p = 0.009$; logit marginal effect 17-20 percentage points, $p \leq 0.020$). We summarize our findings from this and the previous section in Table 5.

5.2. *The impact of minimum work requirements*

5.2.1. *Is it about getting started?*

Our model assumes that people 'just' choose their effort. In practice, the effort decision is more complex. Subjects need to follow the emailed link, start solving tables ('get started'), and then continue to work towards their goal ('get finished'). When we examine these two dimensions, subjects with daily goals were more likely to get started. On average there was one extra login during the week for *Daily* vs. *Weekly* (6.7 vs. 5.5, permutation test, $p = 0.108$) and subjects in *Daily* log in on more weekdays than subjects in *Weekly* (3.2 vs. 2.6, $p = 0.014$). Once logged on, subjects with daily goals completed 131 tables per login in *Daily* vs. 104 in *Weekly* ($p = 0.106$).

To explore further whether getting started helps subjects to get finished, we conducted treatments *DailyRequirement* and *WeeklyRequirement*. They differed from *Daily* and *Weekly*, respectively, only in that we introduced a minimal work requirement. Ariely and Wertenbroch (2002) argue that externally imposed commitment takes away flexibility and thus might be harmful – in particular for people without a self-control problem. We designed the work requirement to limit this problem. It required a subject to spend less than a minute to start working every day, but otherwise allowed for flexibility of whether and when to work. Specifically, subjects were informed that they needed to click the link in their email and complete at least one table per day to qualify for payments. If they failed to do so, they would lose all earnings for the week.[20]

Predictions (without a formal theory) The minimum work requirement does not commit subjects to fulfill a certain workload, but it commits them to 'get started' each day. The idea is that, once a subject clicked on the link and solved one table, he might continue to work. If subjects anticipate that the work requirement helps them to get started, thereby alleviating the problem of effort substitution, we expect them to set the same total goals in *WeeklyRequirement* and *DailyRequirement* and to provide the same total effort.

Results Subjects in *DailyRequirement* aimed to complete a total of 858 tables on average, whereas those in *WeeklyRequirement* aimed for 750 (permutation test, $p = 0.067$, Table A3). The treatment difference is not significant in tobit regressions, but has a non-negligible effect size (Table A15). Thus, the evidence is inconclusive. Further, we observe no significant difference in effort (permutation test, $p = 0.448$; Table A15), and there is no treatment difference in the number of logins ($p = 0.278$), weekdays logged on ($p = 0.195$), or tables completed per login ($p = 0.717$). The results suggest that the motivational power of a daily goal comes from helping people to get started working regularly and that the work requirement does the same. The fact that only few subjects completed just the one table required (14 single-table-for-a-day logins out of 460 subject-day combinations) indeed suggests that the minimum work requirement did get subjects started working. But our analysis in the next subsection shows that there is a downside to having a work requirement.

5.2.2. *Do work requirements complement internal commitment?*

To examine whether an externally enforced minimum work requirement complements internal commitment through goals, we now compare *Daily* and *DailyRequirement* as well as *Weekly* and *WeeklyRequirement*.

Predictions (without a formal theory) The work requirement allows flexible distribution of effort over the five workdays. Only subjects who anticipate being away from a computer with internet access for a whole day – a situation that is quite unlikely during term time – have a reason to drop out. So the work requirement should only lead to a small increase in drop out. Anticipating that the work requirement gets them to work more regularly, subjects should set a higher total goal and then provide a higher effort in *DailyRequirement* (*WeeklyRequirement*) than in *Daily* (*Weekly*).

Results Goals did not differ between *Daily* and *DailyRequirement* or between *Weekly* and *WeeklyRequirement* (permutation tests, $p = 0.172$ and $p = 0.262$; Table 6), but the effect sizes are non-negligible. We observe *lower* effort in *DailyRequirement* than in *Daily*, and no significant effort difference between *WeeklyRequirement* and *Weekly* (permutation tests, $p = 0.008$ and $p = 0.616$; Table 6).

To understand the mechanisms behind the surprising result that the work requirement harmed performance, we perform an exploratory analysis of the dropout behavior in the different treatments. The data contradict our premise that the work requirement does not affect drop-out.

[20] The only exception was if a person completed 1,000 tables before Friday – then no further actions were required (though this was not made explicit in the instructions in order to avoid an incentive to finish early).

Table 7 shows that the dropout probability increased from around 4 percent in *Daily* and *Weekly* to over 30 percent in *DailyRequirement* and *WeeklyRequirement*. Dropout occurred mostly right from the start and was not caused by subjects starting with the task to then stop later during the week.

Further, there was no significant effect of the requirement on the likelihood of working on Monday (bottom of Table 7). Thus, the work requirement appears to have pushed subjects to drop out who otherwise would have started working later than Monday – instead of getting them to start earlier.

In a next step, we examine whether the work requirement had any benefits for the subjects who did participate. Conditional on participation, the requirement did not have a significant effect on the total goal (*DailyRequirement* vs. *Daily* tobit coefficient 113.43, $se = 116.63$, $p = 0.332$; *WeeklyRequirement* vs. *Weekly* 157.31, $se = 131.72$, $p = 0.234$; $N = 209$). For effort, Table 6 ('No dropout' columns) reveals no significant difference between *DailyRequirement* and *Daily*, but significantly higher effort in *WeeklyRequirement* than in *Weekly*. If we consider each workday separately, subjects in *WeeklyRequirement* completed significantly more tables on any given day than subjects in *Weekly*.[21] That is, conditional on participation, subjects worked more steadily in *WeeklyRequirement* than in *Weekly*.

Taken together, these results suggest that a daily goal without an additional, externally enforced work requirement leads to the highest performance. The narrow goal bracket already motivates individuals to 'get started'. At the same time, it avoids the problem of dropout that the work requirement causes. When facing a weak internal commitment device (a weekly goal), the work requirement however does seem to benefit individuals who do not drop out, helping them to 'get started'.

6. Discussion

In this section, we discuss (alternative) mechanisms and extensions of our experiment.

6.1. Alternative mechanisms and explanations

Goal non-achievement Almost half of the subjects failed to reach their goal (Table 2). This is hard to explain with naïveté about the present bias alone, because our model predicts non-achievement only for a relatively severe present bias (cf. Fig. 2). Mistakes in updating beliefs might be one cause for not reaching a weekly goal. In our model, partially naïve individuals hold wrong beliefs about their effort. Adjusting for past observed effort, they keep updating their beliefs about future effort and goal achievement. Yet, such updating is not necessary under daily goals. Thus, the observed non-difference in goal achievement between daily and weekly goals speaks against mistakes in updating being a major source of goal non-achievement.

Further, it seems unlikely that non-achievement is due to a lack of time. 97 percent of subjects in *Daily*, *Weekly*, *Daily(R)*, and *Aggregated* set a goal that they could achieve in the maximum number of hours they could devote to the task (reported before setting goals), when taking into account their baseline productivity. Subjects also reported the maximum number of tables they thought they could solve in the time they had available. By this measure, 67 percent of goals were

[21] Mon: tobit coefficient 96.76**, $se = 48.44$; Tue: 71.06**, $se = 34.42$; Wed: 186.32***, $se = 43.40$; Thu: 109.35**, $se = 43.16$; Fr: 108.84**, $se = 44.28$; $N = 104$. Results are similar without controlling for total goal.

Table 6

Impact of work requirement on total goal and on total effort.

Dependent variable	Total goal	Total effort (tables counted)			
Sample[a]	All	All		No dropout	
	(1)	(2)	(3)	(4)	(5)
DailyRequirement[b]	118.91	−323.35**	−371.75**	151.83	80.33
vs. Daily	(100.53)	(154.97)	(150.40)	(142.91)	(130.22)
Margin.effect(requirement)[c]	51.34	−160.07**	−193.16**	74.03	44.09
Effect size[d]	0.17	−0.43	−0.52	0.21	0.13
WeeklyRequirement[b]	144.70	13.30	−43.27	496.37***	410.70***
vs. Weekly	(112.58)	(158.89)	(153.12)	(138.57)	(125.79)
Margin.effect[c]	77.12	6.79	−22.75	257.27***	220.77***
Effect size[d]	0.23	0.02	−0.06	0.67	0.57
Control for total goal	–	no	yes	no	yes
N	247	247	247	209	209

Notes. Each cell reports the tobit coefficient (with robust standard error in parenthesis) from a separate regression with the omitted category given by "vs.". Specifications include baseline productivity, a gender dummy, and a constant. Results are similar without controls or with the full set of controls. $^{*}p < 0.10$. $^{**}p < 0.05$. $^{***}p < 0.01$.

[a] All: all subjects in *Daily, Weekly, DailyRequirement, WeeklyRequirement*. No dropout: subsample with total effort> 0/satisfying daily login requirement.

[b] Dummy for *DailyRequirement* or *WeeklyRequirement*, respectively.

[c] Tobit marginal effect on the censored latent variable of the minimum work requirement vs. no requirement (at the means of control variables).

[d] $\frac{\text{Margin.effect}}{\text{Standard deviation of outcome in } \textit{Daily} \text{ or } \textit{Weekly}\text{, respectively}}$.

Table 7
Dropout.

Treatment	*N*	Drop out/zero effort[a]		Login Monday		Total effort (mean)	
		N	Percent	N	Percent	Full sample	Not dropped out
Daily	78	3	3.8	62	79.5	690.0	717.6
Weekly	77	3	3.9	51	66.2	520.8	541.9
DailyRequirement	47	17	36.2	35	74.5	487.1	744.9
WeeklyRequirement	45	15	33.3	32	71.1	558.3	811.1

Marginal effect of minimum work requirement vs. no requirement (percentage points).[b]

	Drop out/zero effort[a]	Login Monday
DailyRequirement vs. Daily[c]	30.13*** (7.53)	−5.71 (8.12)
WeeklyRequirement vs. Weekly[c]	29.10*** (7.58)	4.26 (8.67)

Notes.
[a] Dropout: counted zero tables/failed to satisfy the daily login requirement.
[b] Each cell reports the logit marginal effect (with robust standard error in parenthesis) from a separate regression with the omitted category given by "vs.". Specifications include total goal, baseline productivity, a gender dummy, and a constant. Results are similar without controls or with the full set of controls. Sample: *Daily, Weekly, DailyRequirement, WeeklyRequirement*, $N = 247$. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
[c] Dummy for *DailyRequirement* or *WeeklyRequirement*, respectively.

realistic. There is mixed evidence whether those who set such a 'subjectively realistic goal' were more likely to achieve their goal. There is no significant effect (Fisher's exact test, $p = 0.145$) but the logit marginal effect of 9.44 percentage points is non-negligible ($p = 0.134$, $N = 305$). Even so, still only 58 percent of subjects with a 'subjectively realistic goal' achieved their goal.

The descriptive evidence in Section 4.4 suggests that many subjects may have failed to correctly predict changes in circumstances later in the week. Irrespective of their original plans, subjects in *Daily* tended to provide lower effort at the end of the week than at the beginning of the week (Table 4). Some subjects might have underestimated how annoying the task was and decreased their effort after learning this. Or they might have failed to predict an increase in marginal costs of effort over the course of the week (see Section 4.4). For example, we observe that goal achievers had a lower total goal than non-achievers (permutation test, $p = 0.089$, but not significant in tobit regressions, $p \geq 0.133$). That is, achievers possibly better predicted the increase in marginal costs. Theoretically, this would have an effect similar to that of naïveté about the present bias.[22] Another reason could simply be that attention to the study decreased, the longer the study carried on. In line with this explanation, the share of subjects with at least one login on a given day decreased in both main treatments from Monday to Friday (among those who had not yet reached their total goal before Friday): from 75 to 59 percent in *Daily* and from 51 to 37 percent in *Weekly*.

Mistakes in goal setting Mistakes could explain our findings if they had led subjects to set a higher total goal when they picked five daily goals instead of one weekly goal. First, subjects might have felt reluctant to choose a large number, which would have lead to a lower total goal in *Weekly* compared to *Daily*. If reluctance of picking a large number was a driver, the proportion of subjects with *total goal*= 1000 should be lower in *Weekly* than in *Daily*. This is not the case (Fisher's exact test, $p = 0.631$; logit regressions $p \geq 0.445$). Second, subjects may have made small, upward biased mistakes when setting goals. As subjects had to make five choices when setting daily goals compared to one when setting a weekly goal, the sum of small mistakes would have been larger when setting daily goals. This explanation would predict a higher likelihood of the total goal being 'too high' in *Daily* compared to *Weekly*. A proxy is the estimated maximum number of tables that subjects thought they would be able to complete if they used all the spare time they had available to work on the tasks – a question they answered right before learning that they should set goals. Comparing the proportion of subjects for whom the total goal exceeds this number (30 percent in *Weekly* vs. 29 percent in *Daily*), we reject a treatment difference in the proportion of 'too high' goals (Fisher's exact test, $p = 1.000$; logit regressions $p \geq 0.913$).

Experience with the task A random effects panel regression yields no meaningful effect of experience on productivity (0.6 seconds *longer* completion time for the $1000th$ table compared to the first table). To test for possible effects of experience on goal setting and achievement, we exploit variation in how we recruited subjects (see Appendix E). We compare 71 subjects in *Daily* (out of 78), who had 2-3 hours experience with the real-effort task from a prior online experiment (Epper et al., 2018), with subjects in *Daily(R)*, who had no experience before. We find no significant difference in total goal (permutation test $p =0.547$; tobit regressions $p \geq 0.228$) or in the likelihood of achieving the goal (Fisher's exact test, $p = 0.382$; logit regressions, $p \geq 0.242$).

[22] One can easily see this in our parametric example with quadratic effort costs ($c_t(e_t) = (c_t\, e_t^2)/2$), where c_t has an inversely proportional effect to β on $e_{max,t}(\beta) = (2\beta)/c_t$.

Wrong beliefs about effort costs Suppose self 0 thinks that effort costs are $\underline{c}(e)$, while on the first working day they turn out to be $\bar{c}(e) > \underline{c}(e)$. With daily goals, self 0 sets goals $g_t = e_0^*$, where $b'(e_0^*) = \underline{c}'(e_0^*)$. This goal however might not be implementable (if $\bar{e}_{max} < e_0^*$, where $\beta(1 + b'(\bar{e}_{max})) = \bar{c}'(\bar{e}_{max})$). In this case, both with daily and weekly goals, all selves would provide $\bar{e}_{max}$ each period and fail to achieve the total goal. If $\bar{e}_{max} > e_0^*$, then all selves will stick to their daily goals. Yet, under a weekly goal it can happen that the originally planned $\hat{e}_{T,0} > \bar{e}_{max} > e_0^*$, so that self T will deviate and provide $\bar{e}_{max}$. Anticipating this, previous selves will increase their effort so that effort substitution will be less pronounced. The individual may or may not achieve the goal. Thus, wrong beliefs about effort costs could explain some of the observed goal non-achievement.

Uncertainty about effort costs Uncertainty about effort costs translates into uncertainty about effort. In a theoretical model, Koch and Nafziger (2016) demonstrate the effect of uncertainty on narrowly or broadly bracketed goals. A broad bracket allows pooling of risks across tasks, so that the individual suffers less often a loss due to goal non-achievement. Holding effort fixed, this increases the utility of the individual. Yet, exactly this risk pooling effect might dampen incentives because it is the fear of not achieving the goal that makes individuals strive for their goal. If there is little uncertainty, the ability to implement a better decision under narrow bracketing trumps the benefits from risk pooling under broad bracketing. Overall, this suggests that uncertainty will tend to strengthen the prediction of a lower total goal and lower total effort in *Weekly* compared to *Daily*.

Now consider learning about effort costs. Learning strategies (like “start small to learn your costs”) can take place in the same way under daily and weekly goals and thus would not lead to treatment differences in the absence of a self-control problem. Subjects may revise their goals after revelation of uncertainty, failing to meet expected effort (due to partial naïveté), or learning about effort costs might. Realizing that goals cannot be reached or are very expensive to reach, the subject revises them downwoard. As Koch and Nafziger (2016) demonstrate, goals are still effective in this case, but potentially have a lower motivational power so that the initial goal is not achieved. Yet, effort substitution also occurs when allowing for goal revision. Thus, our main predictions regarding treatment differences between *Daily* and *Weekly* carry over. In any case, our data provide no evidence of experience affecting goal setting or goal achievement (see above).

Time horizon The time frame of the project does not matter for the theoretical predictions. Our model predicts for any $T > 1$ that a broad goal leads to effort substitution and lower effort compared to narrow goals. This prediction however relies on the assumption that the next period is in ‘the future’, i.e., that self t discounts by the present bias β the payoffs occurring in $t + 1$ and beyond. Studies on time discounting typically consider periods of one day (e.g. Laury et al., 2012), but one study shows evidence of present bias even over a 5-minute interval (McClure et al., 2007).

In our experiment, we opted for a one-week time horizon and chose to remind people about their goals. Our empirical findings thus seem most applicable to settings were individuals face salient goals over a relatively short time horizon. Examples are project work or studying for an exam. With longer time horizons, goals may be less salient and reminders may be more important. An advantage of daily goals could be that they act as reminders by themselves. In addition, the bracketing of goals may affect how easy it is to monitor goal progress. Take for example a cab driver who can choose his working hours. It may be easier to monitor a fixed daily goal (such as "work 8 hours" or "earn $150") than to monitor progress toward a longer term goal because the latter requires keeping an account for accumulated effort.

Dropout The literature on *hidden costs of control* (Falk and Kosfeld, 2006) suggests aversion to being controlled as a possible explanation for dropout in the treatments with the work requirement.[23] Given the evidence that externally imposed goals increase performance (see, among others, Goerg and Kube, 2012), dropout may have been a response specifically to having been forced to work every day and not so much a response to having been forced to set goals. In particular, subjects might have reacted to having been *forced* to count *one* table. This very unambitious target was perhaps perceived as a nuisance rather than as an encouragement. Other external requirements or incentive schemes may work better. The optimal design of such schemes is an avenue for future research.

Power We performed power calculations only ex-post. Such ex-post calculations should be taken with care. We provide them to give some rough feeling for the sample size. Figure A4 in the online appendix shows the power analysis for the main Hypothesis H1b that we test. Based on the observed treatment difference in effort of 169 tables for *Daily* vs. *Weekly* and the actual number of participants in each treatment, the ex-post calculated power of a two-sided test would be 0.54, 0.77, and 0.85, respectively, for a significance level of 0.01, 0.05, and 0.1. That is, depending on the significance level, there would be a 54 to 85 percent chance that we fail to reject the null hypothesis of no treatment difference even if, in fact, there was a treatment difference. Thus, our main study would be underpowered for $\alpha \leq 0.05$, while for $\alpha = 0.1$ it would be correctly powered. For *Daily* vs. *Aggregated* we have a similar sample size. Comparisons involving treatments *WeeklyRequirement* and *DailyRequirement* have lower sample sizes and hence are most likely underpowered for conventional significance levels. On account of the limited power of tests, as remarked in the text, some care should be taken when interpreting statistically insignificant findings that however show non-negligible effect sizes (reported in the tables).

6.2. *Possible extensions*

Endogenous goal bracket We exogenously assigned the goal bracket (daily or weekly) to identify causal effects of the goal bracket. A direction for future research would be to let subjects choose their goal bracket. While such a treatment does not identify causal effects, it may reveal interesting correlations between the choice of the bracket, effort choices, and individual characteristics. Koch and Nafziger (2019) report some survey evidence in this direction.

Goal revision As discussed above, one reason for goal non-achievement may have been that subjects revised their goals. Future research could provide subjects with an explicit opportunity to revise their goals. This opens-up for new questions on the design of goals. Is the possibility of goal revision good or bad for overcoming self-control problems? While goal revision weakens the motivational power of goals (Koch and Nafziger, 2016), it allows subjects to react to resolution of uncertainty and thereby make better choices. For example, van Lent (2018) gave students (either expectedly or unexpectedly) the opportunity to revise their goals. He observes no difference in

[23] Another explanation builds on Bénabou and Tirole (2004). Here external control crowds out intrinsic motivation because it is a bad signal about effort costs. We cannot test this explanation. However, we believe that intrinsic motivation for the counting task is likely to have been low from the start.

grades between treatments with goal revision and without. Future research could control the type of uncertainty that subjects face, study the exact effort patterns (for daily vs. weekly goals), or study whether goal revision can explain goal non-achievement.

Day-by-day goal setting A further question is how the front-end delay in goal setting matters. In our study, subjects in *Daily* had to think about an entire profile of goals for the next week. Yet, some people might rather focus on their goals one day at a time, i.e., set their goal for Tuesday after having concluded work on Monday, and so on. Theoretically, as long as no uncertainty is resolved, it does not matter whether the individual sets the entire goal profile in advance or sets a goal day-by-day. Allowing subjects to set goals each day in *Daily* however would imply that we do not only vary the goal bracket in comparison to *Weekly*, but also allow subjects to react to possible resolution of uncertainty or to a better understanding of their present bias. It thus seems likely that subjects in *Daily* would perform better than in *Weekly*.

Disaggregation of weekly goals Another potential effect of goal bracketing is that achieving a weekly goal is cognitively more challenging than achieving a daily goal. Individuals have to understand how much effort to put in each day to achieve the weekly goal. Errors could lead both to under- or overshooting of the goal. To examine such channels, one could let subjects set a weekly goal and then disaggregate the goal into daily goals for them. Designing such a treatment however poses some challenges as to how exactly to disaggregate the weekly goal.

7. Conclusion

We provide a theoretical framework and experimental evidence for the motivational benefits of daily goals. In an online experiment, we exogenously assigned the goal bracket (daily or weekly) and let subjects choose their goals. Subjects worked harder under daily goals than under a weekly goal. The increase in effort was primarily related to the higher total goal level with daily goals compared to a weekly goal. In additional treatments, we exogenously shifted the framing (but not the level) of goals. Here subjects were less likely to procrastinate effort to the end of the week when we reminded them about their daily goals than when we reminded them, with the same frequency, about their aggregated weekly goal.

We conclude by discussing some broader implications of our findings. Many organizations struggle with a suboptimal effort allocation of their employees over time, often manifested in spikes of effort immediately prior to a bonus deadline (e.g. Asch, 1990). The effort substitution over time predicted by our model and found in the experiment is consistent with these spikes. Employees facing monthly performance targets may naturally adopt a monthly goal bracket for themselves. Our findings suggest that reframing a bonus threshold in terms of smaller and more frequent narrow goals can have beneficial effects.

In line with this, evidence from a field experiment that Cadena et al. (2011) ran at a Columbian bank shows how more narrowly bracketed goals led to better effort allocation over time. The bank faced the problem that its loan officers concentrated their effort for sourcing new clients and credit collection at the end of each month, just before monthly bonuses were calculated. This increased the costs of cash flow management for the bank and it contributed to loan officers feeling stressed during the second half of the month. In their intervention, Cadena et al. (2011) increased the frequency with which employees received reminders about goals from monthly to weekly periodicity without substantially altering the bonus structure. This lead to an 18 percent (10 percent) increase in new loans (renewal of loans) on average in the first two weeks of each month, without significantly affecting the overall level of loans per month or the quality of loan portfolios. In addition, workers reported lower stress levels.

Our findings tie in, more generally, with evidence on motivational benefits from externally inducing subjects to bracket narrowly and suggests ways to design interventions that can help individuals who struggle with self-control problems. For example, Soman and Cheema (2011) assigned a savings goal to Indian laborers. In the baseline condition, workers received their weekly wage in one envelope. In the treatment condition, the wage payments were split across two envelopes, with one containing the amount that the worker expressed as a savings goal. Actual savings in the treatment condition were higher than those in the baseline condition. Soman and Gourville (2001) found that single-performance ticket holders were more likely to carry through with the ex-ante desirable choice of seeing a theatre performance than multi-performance ticket holders. Gourville and Soman (2002) observe that members who self-selected into paying their gym membership fee each month attended the gym more regularly than those who paid the same overall fee annually, semi-annually, or quarterly. Relatedly, daily repayments of microcredit loans are often observed in developing countries (e.g. Afzal et al., 2017). Bauer et al. (2012) argue that such daily repayments act as external commitment devices for individuals with a self-control problem.

Acknowledgments

We thank Dan Ariely, Steffen Davidson, Kfir Eliaz, Dirk Engelmann, David Laibson, Gerd Mühlheusser, Daniel Schunk, Marco Schwarz, two anonymous referees, the associate editor, and participants at numerous conferences and seminars for helpful comments and discussions. Annette Mortensen, Mette Vad Andersen, Thomas Stephansen, Katrine Poulsgaard, and Heidi Christina Thysen provided excellent research assistance. Funding: This work was supported by the Aarhus University Research Foundation (grant AUFF-E-201-FLS-1-17) and the Danish Council for Independent Research (grant 12-124835).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jet.2019.104949.

References

Abeler, J., Falk, A., Götte, L., Huffman, D., 2011. Reference points and effort provision. Am. Econ. Rev. 101, 470–492.

Afzal, U., d'Adda, G., Fafchamps, M., Quinn, S., Said, F., 2017. Two sides of the same rupee? Comparing demand for microcredit and microsaving in a framed field experiment in rural Pakistan. Econ. J. 128, 161–2190.

Akina, Z., Karagozoglub, E., 2017. The role of goals and feedback in incentivizing performance. Manag. Decis. Econ. 38, 193–211.

Allen, E.J., Dechow, P.M., Pope, D.G., Wu, G., 2016. Reference-dependent preferences: evidence from marathon runners. Manag. Sci. 63, 1657–1672.

Ariely, D., Wertenbroch, K., 2002. Procrastination, deadlines, and performance: self-control by precommitment. Psychol. Sci. 13, 219–224.

Asch, B.J., 1990. Do incentives matter? The case of navy recruiters. Ind. Labor Relat. Rev. 43, 89–106.

Augenblick, N., Niederle, M., Sprenger, C., 2015. Working over time: dynamic inconsistency in real effort tasks. Q. J. Econ., 1067–1115.

Augenblick, N., Rabin, M., 2018. An experiment on time preference and misprediction in unpleasant tasks. Rev. Econ. Stud. 86, 941–975.

Bandura, A., Schunk, D.H., 1981. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. J. Pers. Soc. Psychol. 41, 586–598.
Bandura, A., Simon, K.M., 1977. The role of proximal intentions in self-regulation of refractory behavior. Cogn. Therapy Res. 1, 177–193.
Bauer, M., Chytilová, J., Morduch, J., 2012. Behavioral foundations of microcredit: experimental and survey evidence from rural India. Am. Econ. Rev. 102, 1118–1139.
Bénabou, R., Tirole, J., 2004. Willpower and personal rules. J. Polit. Econ. 112, 848–886.
Bisin, A., Hyndman, K., 2014. Present-Bias, Procrastination and Deadlines in a Field Experiment. National Bureau of Economic Research Working Paper No. w19874.
Brookins, P., Goerg, S.J., Kube, S., 2017. Self-chosen goals, incentives, and effort. Max Planck Institute for Research on Collective Goods. mimeo.
Bryan, G., Karlan, D.S., Nelson, S., 2010. Commitment devices. Annu. Rev. Econ. 2, 671–698.
Burger, N., Charness, G., Lynham, J., 2011. Field and online experiments on self-control. J. Econ. Behav. Organ. 77, 393–404.
Cadena, X., Schoar, A., Cristea, A., Delgado-Medrano, H.M., 2011. Fighting Procrastination in the Workplace: An Experiment. NBER Working Paper No. 16944.
Camerer, C., Babcock, L., Loewenstein, G., Thaler, R.H., 1997. Labor supply of New York city cabdrivers: one day at a time. Q. J. Econ. 112, 407–441.
Clark, D., Gill, D., Prowse, V.L., Rush, M., 2019. Using goals to motivate college students: theory and evidence from field experiments. Rev. Econ. Stat.. forthcoming.
Corgnet, B., Gómez-Miñambres, J., Hernán-Gonzalez, R., 2015. Goal setting and monetary incentives: when large stakes are not enough. Manag. Sci. 61, 2926–2944.
Corgnet, B., Gomez-Minambres, J., Hernan-Gonzalez, R., 2018. Goal setting in the principal-agent model: weak incentives for strong performance. Games Econ. Behav. 109, 311–326.
Dalton, P.S., Gonzalez, V., Noussair, C.N., 2015. Paying with Self-Chosen Goals: Incentives and Gender Differences. CentER Discussion Paper Series No. 2016-036.
Dupas, P., Robinson, J., 2016. The daily grind: cash needs, labor supply and self-control. Stanford University. mimeo.
Epper, T., Koch, A., Nafziger, J., 2018. Aarhus Survey. SSRN Working Paper No. 3123666.
Falk, A., Kosfeld, M., 2006. The hidden costs of control. Am. Econ. Rev. 96, 1611–1630.
Fischer, G., Ghatak, M., 2016. Repayment frequency in lending contracts. London School of Economics. mimeo.
Fudenberg, D., Levine, D.K., 2006. A dual-self model of impulse control. Am. Econ. Rev. 96, 1449–1476.
Goerg, S.J., Kube, S., 2012. Goals (th)at work – goals, monetary incentives, and workers' performance. Max Planck Institute for Research on Collective Goods. mimeo.
Gourville, J., Soman, D., 2002. Pricing and the psychology of consumption. Harv. Bus. Rev. 80, 90–96.
Harding, M., Hsiaw, A., 2014. Goal setting and energy conservation. J. Econ. Behav. Organ. 107, 209–227.
Heath, C., Larrick, R.P., Wu, G., 1999. Goals as reference points. Cogn. Psychol. 38, 79–109.
Himmler, O., Jäckle, R., Weinschenk, P., 2019. Soft commitments, reminders, and academic performance. Am. Econ. J. Appl. Econ. 11, 114–142.
Hsiaw, A., 2013. Goal-setting and self-control. J. Econ. Theory 148, 601–626.
Hsiaw, A., 2018. Goal bracketing and self-control. Games Econ. Behav. 111, 100–121.
Jain, S., 2009. Self-control and optimal goals: a theoretical analysis. Mark. Sci. 28, 1027–1045.
Kaiser, J., 2007. An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples. Stata J. 7, 402–412.
Kaur, S., Kremer, M., Mullainathan, S., 2015. Self-control at work. J. Polit. Econ. 123, 1227–1277.
Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. 121, 1133–1165.
Kirschenbaum, D.S., 1985. Proximity and specificity of planning: a position paper. Cogn. Therapy Res. 9, 489–506.
Koch, A.K., Nafziger, J., 2011. Self-regulation through goal setting. Scand. J. Econ. 113, 212–227.
Koch, A.K., Nafziger, J., 2016. Goals and bracketing under mental accounting. J. Econ. Theory 162, 305–351.
Koch, A.K., Nafziger, J., 2019. Correlates of narrow bracketing. Scand. J. Econ.. forthcoming.
Laibson, D., 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112, 443–477.
Laury, S.K., McInnes, M.M., Swarthout, J.T., 2012. Avoiding the curves: direct elicitation of time preferences. J. Risk Uncertain. 44, 181–217.
Locke, E.A., Latham, G.P., 2002. Building a practically useful theory of goal setting and task motivation. Am. Psychol. 57, 705–717.
Locke, E.A., Latham, G.P., 2015. Breaking the Rules: A Historical Overview of Goal-Setting Theory. Advances in Motivation Science, vol. 2. Academic Press, Elsevier, Oxford, UK, pp. 99–126.

McClure, S.M., Ericson, K.M., Laibson, D.I., Loewenstein, G., Cohen, J.D., 2007. Time discounting for primary rewards. J. Neurosci. 27, 5796–5804.

Neuhaus, G., 1993. Conditional rank tests for the two-sample problem under random censorship. Ann. Stat. 21, 1760–1779.

O'Donoghue, T., Rabin, M., 1999. Doing it now or later. Am. Econ. Rev. 89, 103–124.

O'Donoghue, T., Rabin, M., 2001. Choice and procrastination. Q. J. Econ. 116, 121–160.

Pope, D.G., Schweitzer, M.E., 2011. Is tiger woods loss averse? Persistent bias in the face of experience, competition, and high stakes. Am. Econ. Rev. 101, 129–157.

Rabin, M., Weizsäcker, G., 2009. Narrow bracketing and dominated choices. Am. Econ. Rev. 99, 1508–1543.

Read, D., Loewenstein, G., Rabin, M., 1999. Choice bracketing. J. Risk Uncertain. 19, 171–197.

Shefrin, H.M., Thaler, R.H., 1988. The behavioral life-cycle hypothesis. Econ. Inq. 26, 609–643.

Smithers, S., 2015. Goals, motivation and gender. Econ. Lett. 131, 75–77.

Soman, D., Cheema, A., 2011. Earmarking and partitioning: increasing saving by low-income households. J. Mark. Res. 48, S14–S22.

Soman, D., Gourville, J.T., 2001. Transaction decoupling: how price bundling affects the decision to consume. J. Mark. Res. 38, 30–44.

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

Toussaert, S., 2016. Connecting commitment to self-control problems: evidence from a weight loss challenge. mimeo, NYU.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science (N.S.) 211, 453–458.

Uetake, K., Yang, N., 2018. Harnessing the Small Victories: Goal Design Strategies for a Mobile Calorie and Weight Loss Tracking Application. SSRN Discussion Paper No. 2928441.

van Lent, M., 2018. Goal setting under uncertainty: a field experiment on rigid and flexible goals. University of Leiden. mimeo.

van Lent, M., Souverijn, M., 2017. Goal setting and raising the bar: a field experiment. University of Leiden. mimeo.

Wu, G., Heath, C., Larrick, R., 2008. A prospect theory model of goal behavior. University of Chicago. mimeo.