Self-chosen goals, incentives, and effort*

Philip Brookins† Sebastian Goerg‡ Sebastian Kube§

Preliminary Version

May 21, 2017

Abstract

In a series of randomized field experiments, we investigate the interplay between self-chosen work goals, monetary incentives, and work performance. Employees are observed in a natural work environment where they have to do a simple but effort-intense task. Output is perfectly observable and in most of our treatments workers are paid for performance. A regular piecerate contract serves as a benchmark, while in some treatments workers are paid the same piecerate but asked in addition to choose a non-binding work goal. We observe that the use of personal work goals leads to a significant output increase. Strikingly, the positive effect of self-chosen goals can persist even without any additional monetary incentives, i.e., without the piecerate. However, then the impact of self-chosen goals depends on the exact size of the goals and the difficulty of the task. Our results suggest that work contracts where workers themselves set goals and expectations can help to improve performances; even in the absence of monetary incentives.

Keywords: Goal setting, self-chosen goals, pay-for-performance contracts, workplace behavior, field experiment
JEL classification codes: A12, C93, D01, D03, D24, J24, J33, M52

*We thank the team of the MPI for Research on Collective Goods for making this study possible: Regina Goldschmitt, Sonja Hendriks, Inge Ljungström, Charlotte Mainzer, Martin Pollet, and Daniel Zimmel. We are grateful to Steffen Altmann, Brian Cooper, Christoph Engel and Armin Falk, Marco Kleine, Marta Serra-Garcia, Dana Sisak, as well as audiences at the University of Bonn, Center for Economics and Neuroscience, EALE Meeting in Cyprus, ESA Chicago, ESA Tucson, Florida State University, University of Frankfurt, University of Karlsruhe, University of Michigan, and the MPI for Research on Collective Goods in Bonn, for very helpful comments.

†Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113, Bonn, Germany; e-mail: philbrookins@gmail.com.

‡Department of Economics, Florida State University, Tallahassee, FL 32306-2180, USA; sqoerg@fsu.edu.

§Department of Economics, University of Bonn, Adenauerallee 24–42, 53113, Bonn, Germany; kube@uni-bonn.de.

1 Introduction

> *"The greatest danger for most of us is not that our aim is too high and we miss it, but that it is too low and we reach it."*
>
> attributed to Michelangelo (1475–1564)

> *"[A Goal] is the desired end state the individual reaches for; it is the ultimate aim of one's adopted action, the very cause of the action; it is the purpose toward which one is striving; it is the reason for doing and thinking."*
>
> Gordon B. Moskowitz and Heidi Grant (2009)

Goals are involved in many everyday situations. For example, people set a target weight when on a diet; learning goals are formulated and practiced in schooling; savings plans are used in order to purchase expensive consumption goods; countries agree on emission targets to protect the environment; quota systems are introduced to promote gender equality; firms make use of management styles that involve milestones improve upon production, sales or consumers' satisfaction; etc. Despite their omnipresence is the investigation of goals a relatively recent development within economics (e.g., Koch and Nafziger, 2011, 2016; Corgnet, Gómez-Miñambres and Hernán-Gonzalez, 2015; Dalton, Gonzalez and Noussair, 2016). Yet, goals directly relate to the existing studies on income targets (e.g., Fehr and Götte, 2007; Huffman and Götte, 2007; Crawford and Meng, 2011) and reference-dependent preferences in general (Kőszegi and Rabin, 2006, 2007; Abeler et al., 2011). In these studies, reference points against which (psychological) gains and losses are evaluated are implicitly set through expectations, in contrast choosing a goal might be interpreted as explicitly picking a reference point.

To provide empirical evidence on the endogenous selection of goals, i.e., references points, we explore *self-chosen non-binding goals* with the help of natural field setting, and the interplay between monetary incentives and individuals' goal setting. We do this in a natural work environment, in which workers' outputs are perfectly observable while exogenously varying the difficulty of the task and the accompanying monetary incentives.

In our natural field experiments we hired workers to help in the process of restructuring a large library. Their task was to search the library shelves for specific books from a given list. It was a one-time job and each worker was only employed for a single working day. Wage payments always included a fixed wage, but additional compensation depended on the experimental treatment. In the baseline treatment Piecerate, €0.10 was paid for each book found by the worker. The same piecerate was used in treatment Piecerate+Goal. The only difference was that, before workers started searching for the books, we asked them for their expectation about the number of books they thought they would find during the working day. Their stated belief was then referred to as their personal

work goal. Missing the self-chosen work goal, however, yielded no monetary consequences. The difference in behavior between PIECERATE and PIECERATE+GOAL thus identifies the effect of a self-chosen goal, holding monetary incentives constant between treatments. Our final experimental manipulation, treatment GOAL, consisted of the same goal setting process described in PIECERATE+GOAL, but workers did not received payments from a piecerate component. This treatment helps us to identify the effectiveness of goals on their own. To determine the robustnesses of our results we manipulated the task difficulty and implemented all treatments in a HIGH and a LOW DIFFICULTY environment, thereby yielding a 2×3 factorial design.

In line with a simple model of goal setting, we find that work goals significantly affect workers' output. Under both HIGH and LOW task difficulty is the average productivity in treatment PIECERATE+GOAL almost 15-17% higher than in PIECERATE; showing that self-chosen goals in the presence of monetary incentives increase performance. For GOAL the comparison with PIECERATE depends on the task difficulty: with a HIGH DIFFICULTY we observe a 17% higher output while with a LOW DIFFICULTY the difference is neglectable 2%. Our data indicate that this is due to workers failing to correctly adjust their goals to the work environment: goals chosen in the respective goal treatment do not differ significantly between HIGH DIFFICULTY and LOW DIFFICULTY. Thus, goals chosen under LOW DIFFICULTY induce effort only up to a certain point. However in PIECERATE+GOAL the additional monetary incentives are mitigating the small goal size. In addition, male workers tend to choose higher goals than female workers, which results in more women attaining their goals. Taken together, our results suggest that self-chosen personal work goals can increase performance and their impact on performance is more robust in the presence of (goal-independent) monetary incentives. Nevertheless, our study also shows that GOALS in the absence of any monetary incentives can induce the same or even higher performance as a PIECERATE contract.

Related Literature

The study of goals within economics has lead to a rapidly growing number of publications on diverse applications. Hence, goals have successfully been applied to topics such as social comparisons Falk and Knell (2004), energy consumption (Harding and Hsiaw, 2014), workplace behavior (Corgnet, Gómez-Miñambres and Hernán-Gonzalez, 2015), and educational performances of college students (van Lent and Souverijnz, 2016; Clark et al., 2016). In table A.1 of Appendix A we report more details and classifications of the models used in the economic goal-setting literature.

Broadly speaking, the literature on goals can be divided into two categories: *ex-*

ogenously imposed goals by a principal and *self-chosen goals* picked directly by an agent. *Exogenously imposed goals* can take the form of binding goals with monetary consequences (i.e., a bonus) or as non-binding goals where attainment has no additional consequences. Gómez-Miñambres (2012) and Corgnet, Gómez-Miñambres and Hernán-Gonzalez (2015) provide models for the latter situation in which a manager assigns a non-binding goal that functions as reference point for workers' intrinsic motivation. Using a lab experiment, Corgnet, Gómez-Miñambres and Hernán-Gonzalez (2015) test their model in a principal-agent setting. In their experiment managers indeed decide to use challenging but attainable goals and workers respond to these goals with increased output and reduced leisure activities.

The paper at hand fits into the category of *self-chosen goals* which have been shown to be a source of internal motivation to attenuate the self-control problem of hyperbolic discounters. Hsiaw (2010, 2013) models this in the context of an optimal stopping problem over an infinite time horizon, where uncertainty about outcomes generates an option value of waiting. In this model, by providing a reference point, the goal-setting attenuates the tendency to undervalue the option and to stop too early. A similar approach is presented in Koch and Nafziger (2011) who, too, study self-control problems. In their two-period model, goals are set in the first period and later used to evaluate the outcome in the second period. Assuming that agents care about reaching the reference standard set by the goal, goals become binding and help to overcome the self-control problem. If agents face minor uncertainty about their productivity they should bracket their goals narrowly (e.g., daily work goals) and they should bracket broadly (e.g., monthly work goals) if they face significant uncertainty (Koch and Nafziger, 2016).

Corresponding empirical evidence of goals being used as an efficient self-disciplining device is provided by Kaur, Kremer and Mullainathan (2010, 2015), suggesting that individuals indeed tend to use self-disciplining devices. They find that a fraction of workers in an Indian data-entry firm voluntarily agrees to incur a monetary loss when falling short of a self-chosen production target. Dalton, Gonzalez and Noussair (2016) provide further evidence with a lab-experiment. They investigate a self-chosen goal contract in which reaching the goal is backed up with monetary incentives, i.e., yielding a bonus when the self-chosen goal is reached. There the self-chosen goal contract is more cost effective than a piece rate contract and workers choose goals which they tend to outperform. We complement this literature by providing causal evidence from the field demonstrating that self-imposed goals can be effective if attainment does not result in additional payments and even further in the total absence of any monetary incentives.

Our findings are also of interest to the recent work on labor supply and income targets,

which implicitly deals with self-chosen goals as well (e.g., Camerer et al., 1997; Fehr and Götte, 2007; Huffman and Götte, 2007; Farber, 2008; Crawford and Meng, 2011). The main message from these studies is that temporary wage increases (e.g., windfall gains or increased productivity) can lead to a reduction in labor supply, arguing that such patterns could be a result of workers trying to achieve a certain daily income target. Our study differs insofar as our workers explicitly state their goals, and as these goals are on production rather than on income. Our findings suggest that implicit income targets can be replaced with an explicit production goal. This could potentially even mitigate the detrimental effects of income targets on efforts, though future studies would need to have treatments in which, for example, the piecerate or the production technology is varied to provide conclusive answers.

Outside of Economics, research on goals has a long tradition and is still flourishing (for comprehensive reviews, see, e.g., Locke and Latham, 2002, 2006, 2013; Moskowitz and Grant, 2009). In particular, literature in Psychology primarily focuses on the motivational and cognitive aspects of goals and the basic processes which translate goals into actions – distinguishing goals from (and relating them to) other motivational constructs. It is also exploring differences in goal contents, stressing that goals should be feasible and S.M.A.R.T. (Specific, Measurable, Attainable, Relevant and Timed). Furthermore, it is investigated how goals are regulated and pursued with the corresponding consequences. Almost all of these studies have consistently demonstrated that individuals' behavior is indeed affected by goals.[1] Our paper ties nicely with this literature by exploring the interaction between motivation, self-chosen goals, and incentives.[2]

The rest of this paper is organized as follows. In Section 2, we present a simple model and its predictions. We describe the experimental design of our field study in Section 3. In Section 4, we present detailed results on the effect of goals and work contracts on work behavior, as well as on the interaction between goal-setting and monetary incentives. Section 5 concludes the paper.

[1] Sometimes negatively, see Ordóñez et al. (2009) or for a concise survey Goerg (2015).

[2] For example, the recent comprehensive reference work for "The Psychology of Goals" edited by Moskowitz and Grant (2009) does not include a single instance of the terms "monetary incentive" or "contract". We are not aware of any study in this area that has focused on the interplay between self-chosen work goals and economic incentive contracts. What has been studied are different contracts with exogenously imposed (binding or non-binding) goals, e.g., Lee, Locke and Phan (1997) or Corgnet, Gómez-Miñambres and Hernán-Gonzalez (2015), or the influence of the general goal "to make money" on behavior, e.g., Aarts, Gollwitzer and Hassin (2004).

2 A Simple Model of Goalsetting

In this section, we present a simple model of motivation through goal-setting from which we generate testable predictions. We consider a production environment in which a single worker is incentivized to exert high effort by either monetary incentives (PIECERATE) or self-chosen goals (GOALS), or both (PIECERATE+GOALS). Our model shares similarities with recent theoretical contributions, but we differ in several important ways. First, we focus on self-chosen goals, and therefore, agents in our model have an additional control variable on top of choosing how much effort exerted on the task, i.e., the size of the goal. To capture the psychological or intrinsic utility an agent gains (loses) when they reach (do not reach) their goal, we follow the growing number of studies (Wu, Heath and Larrick, 2008; Gómez-Miñambres, 2012; Corgnet, Gómez-Miñambres and Hernán-Gonzalez, 2015; Hsiaw, 2010, 2013) who assume goals act as reference points that separate intrinsic utility into the domains of gains and losses.[3] Second, we modify the intrinsic utility function to capture the intensity to which the individual strives towards meeting her goal, commonly referred to as "goal commitment", which acts as a positive (negative) weight on the agent's intrinsic utility in the event that she exceeds (does not exceed) her goal (Gómez-Miñambres, 2012). Under this specification, we capture the realistic relationship that goals are positively correlated to personal standards, and thus, avoid circumstances where agents choose trivial goals.[4]

Our model is most closely related to Wu, Heath and Larrick (2008), who in addition to reference-dependence, assume that the intrinsic utility exhibits the properties of the value function from prospect theory (Kahneman and Tversky, 1979), namely, loss aversion and diminishing sensitivity. However, we allow the agent to have full control over the size of their goal and assume individuals only set non-trivial goals, i.e., goals in excess of their own personal standard. Dalton, Gonzalez and Noussair (2016) also extend the model of Wu, Heath and Larrick (2008) to handle endogenous goals as we do, they differ in that subjects are given a monetary bonus contingent on reaching their goal which is increasing in the size of their goal. In this way, even under the assumption that workers do not derive intrinsic utility from reaching their goal, workers who are given the bonus pay in addition to piecerate pay exert higher effort than the setting without self-chosen goals.

In addition to a fixed wage payment $w_F \geq 0$, a single risk-neutral agent is incentivized

[3] Again, refer to table A.1 of Appendix A for more details and classifications of the common models used in the economic goal-setting literature.

[4] Without the assumption that workers take their standards into account when choosing goals, it is always optimal for the worker to choose the lowest possible goal, i.e., a production goal of zero, and to carry out the maximization with the sole choice variable effort. This situation is unrealistic, and we therefore appropriately structure intrinsic utiltiy to rule this out.

to exert high effort $e \geq 0$ by either a pay-for-performance contract, or asked to choose a non-binding production goal g, or both. The pay-for-performance contract offers a piece-rate of $w_p > 0$ for each unit of output $y = ae$ produced, where $a \in (0, 1]$ represents the ability of the individual. We assume effort costs are strictly convex and given by $c(e) = de^2/2$, where $d \in \{d_L, d_H\}$ is the task difficulty parameter and $d_H > d_L > 0$.[5] Reaching or failing to reach the self-chosen goal does not affect the agent's utility through monetary compensation, but it does affect the agent's intrinsic utility $v(y, g, s)$, which is given by

$$v(y, g, s) = \begin{cases} \psi(g, s)(y - g)^{1/2} & \text{if } y > g \\ -\lambda\psi(g, s)(-(y - g))^{1/2} & \text{if } y \leq g \end{cases}, \tag{1}$$

where $\lambda > 1$ represents the coefficient of loss-aversion (Kahneman and Tversky, 1979), and $\psi(g, s)$ is the same goal commitment function used by Gómez-Miñambres (2012) to capture the intensity of goal commitment, which depends on the agents own personal standard $s > 0$ in the following way

$$\psi(g, s) = \begin{cases} sg & \text{if } g > s \\ 0 & \text{if } g \leq s \end{cases}. \tag{2}$$

There are several important features of this particular form of intrinsic utility. First, note that $v(\cdot)$ exhibits the main properties of the value function from prospect theory (Kahneman and Tversky, 1979): (i) the function is piecewise-continuous about the self-chosen goal, i.e., utility is separated into the domains of gains and losses with respect to the reference point g, (ii) there are diminishing gains, or losses, as $|y - g|$ increases, and (iii) agent's are assumed to be loss-averse, i.e., failing to reach the goal by a small amount $\epsilon > 0$ is more painful than surpassing it by the same amount. This functional form is beneficially not just for it's simplicity, but because empirical evidence supports the existence of these properties in the goal setting process (Heath, Larrick and Wu, 1999).

An additional feature is the goal commitment function $\psi(g, s)$ (Gómez-Miñambres, 2012). In the event that the goal selected is lower than or equal to their personal standard, the agent feels no commitment and receives no additional benefit from reaching her goal. However, for goals deemed challenging (i.e., $g > s$), the goal commitment function positively (negatively) impacts intrinsic utility provided she meet (does not meet) her goal. Thus, an agent with high standards sets high goals, and therefore has a larger commitment to her goal. Throughout the rest of the paper we assume the agent *always*

[5]For the purpose of the experiment, there is no exact way to estimate the precise form of the participants cost function. For simplicity, and to be able to obtain predictions, we assume quadratic costs without loss of generality.

sets challenging goals, which implies the goal commitment function is always given by $\psi(g, s) = sg$. It is natural to assume that the standard depends on the difficulty of the task at hand and the agent's ability. For simplicity, we will assume the following functional form $s(a, d) = \frac{a}{d}$. Thus, the standard is decreasing in task difficulty and increasing in ability.

The agent operates in only one of three possible incentive structures: PIECERATE, GOALS, or PIECERATE+GOALS. In the event the agent is only offered a piece-rate for each unit of output produced, the agent chooses an effort level $e = \frac{y}{a} > 0$ to maximize her total utility, which in this case is given by

$$u = w_F + w_p y - \frac{d}{2}\left(\frac{y}{a}\right)^2. \tag{3}$$

Equating marginal costs to marginal benefits, it is easy to see that the optimal level of output is $y^{PR} = a^2 w_P$.

In the other two cases of GOALS and PIECERATE+GOALS, the agent faces a slightly more complex optimization problem. In addition to output, the agent chooses the size of her production goal $g > s > 0$. Thus, the agent now chooses an output-goal pair (y, g) to maximize her total utility subject to the constraint $g > s$. If the agent is incentivized by both pay-for-performance compensation and goals (PIECERATE+GOALS), her total utility is given by

$$u = \begin{cases} w_F + w_P y + \frac{ag}{d}(y-g)^{1/2} - \frac{d}{2}\left(\frac{y}{a}\right)^2 & \text{if } y > g \\ w_F + w_P y - \frac{\lambda ag}{d}(-(y-g))^{1/2} - \frac{d}{2}\left(\frac{y}{a}\right)^2 & \text{if } y \leq g \end{cases}, \tag{4}$$

Total utility is similar for the case when the agent is only motivated by goal-setting (GOALS), but the piece-rate pay $w_P y$ is removed.

In either setting, three possibilities arise: (i) $y < g$, (ii) $y = g$, and (iii) $y > g$. First, note that it is never optimal for the agent to stop producing output at the moment she meets her goal (i.e., $y = g$). This is due to the discontinuity at $y = g$, in which marginal intrinsic utility tends to infinity as $y \to g^-$ and $y \to g^+$ for any $g \geq s$, implying the agent would be better off with either $y < g$ or $y > g$.[6]

Suppose the agent chooses the pair (y, g) such that $y < g$. Then the agent fails to meet her goal and receives the following total utility

$$u = w_F + w_p y - \frac{\lambda ag}{d}(-(y-g))^{1/2} - \frac{d}{2}\left(\frac{y}{a}\right)^2. \tag{5}$$

Thus, the agent's maximizes (5) subject to $g > s$ and $g > y$. The first-order conditions

[6]Holding output fixed, the same is true as the goal approaches output.

are:

$$w_p + \frac{\lambda a g}{2d}(g-y)^{-1/2} - \frac{d}{a^2}y = 0 \tag{6}$$

$$-\lambda\frac{a}{d}(g-y)^{1/2} - \frac{\lambda a g}{2d}(g-y)^{-1/2} = 0, \tag{7}$$

Note that Eq. (7) implies that $\frac{3}{2}g = y$, which contradicts the assumption that $g > y$. Therefore, it is never optimal for the agent to produce output below her goal.

Finally, suppose $y > g$. Then the agent maximizes

$$u = w_F + w_p y + \frac{ag}{d}(y-g)^{1/2} - \frac{d}{2}\left(\frac{y}{a}\right)^2 \tag{8}$$

subject to $y > g$ and $g > s$. The first-order conditions are:

$$w_p + \frac{ag}{2d}(y-g)^{-1/2} - \frac{d}{a^2}y = 0 \tag{9}$$

$$\frac{a}{d}(y-g)^{1/2} + \frac{ag}{2d}(y-g)^{-1/2} = 0, \tag{10}$$

From Eq. (10) we find that $\frac{3}{2}g = y$, consistent with our assumption that $y > g$, which can be substituted into (9) to find the optimal pair (y^*, g^*).[7]

Based on the above model we expect individuals facing a Low Difficulty to produce more output than the ones facing a High Difficulty. In addition we expect individuals' abilities to be correlated with the produced output. Figures B.1 and B.2 in Appendix B demonstrate these intuitive results graphically. In addition, do our goal and incentive manipulations yield the following testable hypothesis on the observed output :

Hypothesis 1 *Holding task difficulty and ability constant, the incentive structure ranking, from highest output to lowest, is:* Piecerate+Goals $\geq$ Goals $>$ Piecerate.

Hypothesis 1 results from examining the output while keeping levels of difficulty and ability fixed. Based on the above model Piecerate+Goals should outperform Goals, albeit by such a small amount that it might be empirically impossible to identify

[7] The second-order sufficient conditions for the optimality of (y^*, g^*) require that $(-1)^k|H_k(y^*, g^*)| > 0$ for $k = 1, 2$, where H_k is the k^{th} leading principal minor of Hessian matrix $H = \begin{bmatrix} \frac{\partial^2 u}{\partial y^2} & \frac{\partial^2 u}{\partial y \partial g} \\ \frac{\partial^2 u}{\partial y \partial g} & \frac{\partial^2 u}{\partial g^2} \end{bmatrix}$. It is easy to verify that (y^*, g^*) is a maximum provided $g^* > \frac{1}{2}\left(\frac{a^2}{d(2+a)}\right)^2$.

(therefore, we do not expect a strict inequality). However, both treatments, Piecerate+Goals and Goals, should significantly outperform Piecerate.

The above ranking is a direct result of self-chosen goals and we will now focus on the goal setting itself. Analogously to the previous discussion on outputs, we expect individuals facing a Low Difficulty to choose higher goals than the ones facing a High Difficulty and individuals' abilities to be correlated with the chosen goal. Again, figures B.1 and B.2 of Appendix B demonstrate these intuitive results graphically. With regard to our incentive manipulation we expect the following ranking:

Hypothesis 2 *Holding task difficulty and ability constant, we expect (marginally) higher goals in* Piecerate+Goals *than in* Goals.

Based on the above analyses we should observe the following relationship between output and goals:

Hypothesis 3 *Higher goals should lead to higher output.*

3 Experimental Design

To empirically test the effectiveness of self-chosen goals on workers' performance, we conducted a series of field experiments. The experiments took place at the library of a German research institute and participants had to find books according to a list given to them. In the first set of the experiments, we benefited from the fact that the books in the library had to be rearranged. For this every book had to be located in the library shelves and then relocated to a new place in a different shelf.[8] The second set of the experiments was conducted in the same library after the rearrangement of the books. The rearrangement of the books was done to make it easier to find books and, therefore, the task is intentionally easier in the second set of the experiments. In the following we will refer to the experiments with the original arrangement of books as High Difficulty experiments and to the experiments with the new arrangement as Low Difficulty experiments. All other elements, treatments, and procedures, which we will describe in the following, were identical under High Difficulty and Low Difficulty

[8] In this section, we focus on those elements of the work environment that are of central importance for our study. Comprehensive descriptions of the background of the library's restructuring, procedural details, work task and payment instructions used in the different treatments are provided in the appendix.

3.1 Treatments

We implemented three incentive schemes and all three were implemented in both HIGH DIFFICULTY and LOW DIFFICULTY yielding a 2×3 factorial design. In all schemes, subjects received a fixed payment of €22, but we manipulated the use of performance-related payments and self-chosen work goals. In PIECERATE workers receive a regular piecerate of €0.1 per produced output, i.e, each book relocated. In GOAL workers receive no additional payments beyond the fixed payment of €22. Instead we asked them for their expectation about the number of books they thought they would find during the working day. Their stated belief was then referred to as their personal work goal. Missing the self-chosen work goal yielded no monetary consequences. The third scheme PIECERATE+GOAL combines the features of the previous two schemes; workers receive a regular piecerate of €0.1 per produced output and are asked to specify their personal work goal.

Table 1: TREATMENTS

	HIGH DIFFICULTY			LOW DIFFICULTY		
	# Workers	Mean Age	Male	# Workers	Mean Age	Male
PIECERATE	25	23	52%	20	24.3	80%
GOALS	20	23.3	60%	20	23.6	55%
PIECERATE+GOALS	25	23.7	56%	20	22.8	45%
Total	70	23.3	55.7%	60	23.6	60%

In total 130 subjects participated in this study: 75 subjects in HIGH DIFFICULTY and 60 subjects in LOW DIFFICULTY. Table 1 gives the number of workers per treatment as well as gender and age composition per treatment.

3.2 Procedures

The job to search the library for books and relocate them was advertised online and via posters as a one-time job opportunity. Potential workers were informed that the job would last 3.5 hours and pay a minimum of 22 €. Workers applied online and were then randomly selected from the pool. They were then invited via email and asked to confirm the time slot we had allocated them to.

The experiments were carried out by a librarian, strictly following a fixed protocol. Upon arrival, the subject received a short manual describing the exact work task. After having read the written instructions, the subject got an extensive list of books to be searched. The subject's task was to work through the list sequentially: i) finding the

corresponding book in the shelves, ii) scanning its ID at a workstation and then iii) placing it in a book trolley. The exact order of books as given by the list had to be kept, because books would later be relabeled and placed into the shelves according to this order by the librarians.[9] The books on the list were not ordered alphabetically, so that the probability of two successive books on the list being close to each other in the shelves was virtually zero. Note that the simple work task is well suited for a field experiment. It is easy to understand and output is sensitive to the worker's effort due to the physical component involved in moving mobile shelves and walking around in the library. Moreover, due to the digital timestamp that is created in the database whenever a book's ID is scanned at the workstation, we also have a precise electronic measurement of workers' performance over time.

After instructions were completed, workers had to do two supervised trials, i.e., each subject had to search for two books and scan them at the workstation while a librarian was watching. This served three purposes. First, it ensured that each worker had understood the work task. Second, it provided workers with a rough estimate of how long they approximately need to find a book — which is important information for workers when they have to set themselves a goal. Third, the average time needed to find a book can be used to approximate subjects' general ability for this kind of task; allowing us to control for potential heterogeneity among individuals.[10] After subjects had left, the two trial books were put back in their original place, so that the ability measure was always based on the same two books.[11]

To avoid influences of the treatments on the ability measure, the exact payment scheme was announced only after the two trials. Like the task description, the scheme was handed out in written form and was additionally explained by a librarian. Subjects then could ask clarifying questions. Afterwards, if the treatment featured a self-chosen goal, subjects had to announce their personal goal and a post-it with it was attached to the display of the workstation. The librarian then started the timer before leaving the workplace, and the worker proceeded with the task for three hours. The workstation

[9]Consequently, afterwards the librarians had to control the order of the books and thereby checked the quality of the students' work. No complaints occurred; there was not a single reported incidence of quality problems.

[10]Throughout this paper, higher ability is defined as a lower search time for the second test-book. We are missing the ability measures for 10 subjects.

[11]This was done to ensure the comparability of the measure between subjects. Moreover, it allows us to check whether the work task becomes easier over time. This could happen as shelves are successively cleared when more and more books are removed as the study proceeds. Yet, this does not seem to be the case. We find no general time trend in ability measures over the course of our study; neither in High Difficulty ($r = -0.09$ and $p = 0.475$, Spearman rank correlation between day of experiment and ability) nor Low Difficulty($r = 0.004$ and $p = 0.974$).

always displayed the current number of scanned books. Subjects were allowed to take a break whenever necessary. After exactly three hours, the librarian returned, checked the total amount of scanned books, calculated the total payoff accordingly and paid the subject. In the end, the subject had to complete a short employee questionnaire on work satisfaction and left.

4 Results

Table 2 provides summary statistics on the time per book, i.e., speed of the worker, the workers' output, and the chosen goal size if applicable. The table reveals that, as expected, subjects worked faster and produced higher outputs in the second study with Low Difficulty than in the first study with High Difficulty. The comparisons of speed as well as outputs between the two studies are significant for each treatment (all $p < 0.001$, two-sided Fisher-Pitman permutation test for independent samples).[12]

Table 2: Summary Statistics

a. High Difficulty

	Time per Output (sec.)			Output			Goals		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Piecerate	113.5	111.1	26.9	102	101	20.6			
Goal	94	93.2	19.5	118.7	117.5	22.4	116.5	105	44.3
Piecerate+Goal	99.9	96.2	26.4	117.2	115	27.5	144	150	49.8

b. Low Difficulty

	Time per Output (sec.)			Output			Goals		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Piecerate	71	68.5	12.5	164.1	172.5	27.2			
Goal	68.8	70.3	13.6	167	158	29.6	115	100	55.7
Piecerate+Goal	60.3	59.8	10.2	192.9	189.5	25.9	150.5	155	47.3

In the following we present our results with respect to the implemented treatments. We will start with a look at workers' performances and investigate how fast they work and how much output they generate in the different treatments. Thereafter, we will focus more on the two treatments with self-chosen goals. There, we will first investigate

[12]The Fisher-Pitman permutation test is a more powerful non-parametric alternative to the Wilcoxon-Mann-Whitney rank-sum test. For more details refer to Kaiser (2009). For the remainder of this paper and if not otherwise reported p-values are given for the two-sided Fisher-Pitman permutation test for independent samples.

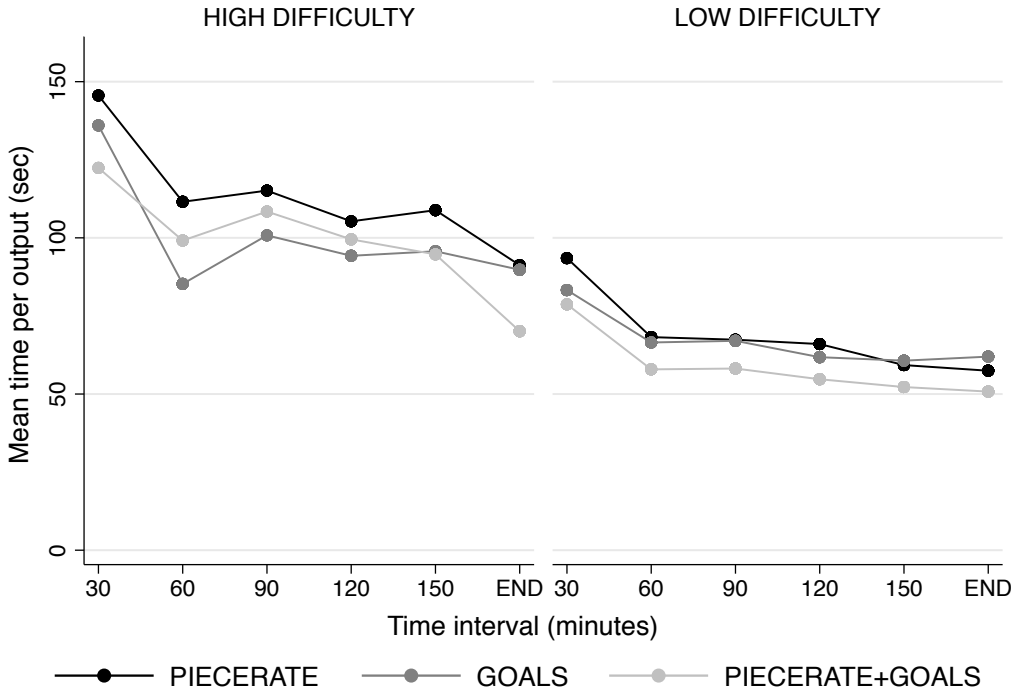


Figure 1: Mean time per output over the course of the experiment.

the determinants of the self-chosen goals, and then the influence of goals and monetary incentives on workers' output.

Figure 1 gives the workers' average speed, i.e., the time needed to produce one unit of output, through the course of the experiment. In the treatments with High Difficulty subjects work faster if they work with self-chosen goals. Subjects need significantly more time in Piecerate to produce one output than in Goal ($p = 0.006$) and Piecerate+Goal ($p = 0.078$). The slower working translates into a significantly smaller mean output of 102 in Piecerate compared to 118.7 in Goal ($p = 0.013$) and 117.2 in Piecerate+Goal ($p = 0.031$). Neither the speed nor the outputs differ significantly between Goal and Piecerate+Goal (speed $p > 0.414$ and output $p = 0.852$).

With Low Difficulty subjects work faster in Piecerate+Goal than in Goal ($p = 0.031$) and Piecerate ($p = 0.004$). Thus, workers' mean output of 192.9 in Piecerate+Goal is significantly larger than the mean output of 164.1 in Piecerate ($p = 0.001$) and the mean output of 167 in Goal ($p = 0.006$). While workers in Goal were faster and produced higher outputs than in Piecerate in High Difficulty it no longer holds in Low Difficulty; no significant differences in speed ($p = 0.590$) and output ($p = 0.749$) are observed. Yet, while worker produced a similar output in Piecerate and Goal they produced it much cheaper in the latter one.

From the employer side not only the produced output, but also the cost per produced output matters. In Piecerate and Piecerate+Goal workers earned 22€ plus 10€-Cent per output, while in Goal they only received the fixed wage of 22€. Thus, the cost per output is with both High and Low Difficulty significantly smaller in Goal than in Piecerate and Piecerate+Goal (for all comparisons $p < 0.001$).[13] Because of the higher output is the cost per book in Piecerate+Goal significantly smaller than in Piecerate ($p = .076$ in High Difficulty and $p = .002$ in Low Difficulty). If one would focus only on the marginal costs per book there would be no differences between Piecerate and Piecerate+Goal as we pay the exact same piecerate in both treatments. However, the marginal cost in Goal would still be lower, i.e., zero, as no piecerate is paid.

Before taking a closer look at the chosen goals, we complement our non-parametric analysis with a regression analysis, adding controls for other potential influences, in particular individuals' ability as measured by the average time needed in the ability stage.[14] Table 3 gives the results of the estimated models using ordinary least squares (OLS).[15] Models 1, 3, 5, and 7 replicate our non-parametric results comparing speed and output in the treatments separately for High and Low Difficulty. For the regressions comparing the time per output we utilize the full dataset which includes the time needed for each single unit of output; to control for within-subject correlation of times we cluster on the subject level. In Models 2, 4, 6, and 8 we add the individuals' time from the ability task, age, gender, dummies for day of the week as additional controls. In addition, in the models utilizing the time data we control for learning the task by including the current output number and its square.[16]

As expected is the time needed at the ability stage positively correlated with the time per output and negatively with the total output. Thus, subjects who are faster in the ability stage work faster on the task and produce a higher output. Neither gender nor age have a significant impact on speed or output. The rest of the table shows that our previously reported non-parametric results on speed and output are robust to the additional controls. Self-chosen goals in combination with monetary payoffs are a powerful way to increase output. But even without additional monetary incentives they can lead

[13] Cost per output with High Difficulty/Low Difficulty is .19/.24 Cents in Goal, .33/.24 Cents in Piecerate, and .30/.23 Cents in Piecerate+Goal.

[14] Ability neither differs significantly between treatments in High Difficulty ($p = 0.57$) nor in Low Difficulty ($p = 0.27$, both Kruskal-Wallis test).

[15] Table C.1 in the Appendix gives all coefficents.

[16] Subjects learn in this task and become better over time. In both High and Low Difficulty are the coefficients for the current output number negative and its square positive. This implies that subjects become faster over time, but at a decreasing rate.

Table 3: OLS explaining Time per Output and Total Output

	High Difficulty				Low Difficulty			
	Time Per Output		Total Output		Time Per Output		Total Output	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Goal	-17.72**	-12.15***	16.70**	16.07**	-2.646	-2.769	2.950	1.506
	(8.744)	(4.038)	(6.464)	(6.516)	(4.583)	(3.532)	(8.984)	(8.781)
Piecerate+Goal	-14.29**	-9.465**	15.24**	13.09*	-10.10**	-11.34***	28.80***	29.60***
	(7.085)	(4.531)	(6.880)	(7.276)	(4.418)	(3.677)	(8.390)	(9.321)
Time Ability		0.0576**		-0.0931**		0.0476**		-0.159***
		(0.0239)		(0.0376)		(0.0224)		(0.0408)
Constant	108.5***	175.5***	102***	116.6***	69.51***	94.37***	164.0***	181.3***
	(4.348)	(23.44)	(4.119)	(33.20)	(3.433)	(19.72)	(6.080)	(39.26)
Controls	-	age, gender, #book, #book^2 day of week	-	age, gender day of week	-	age, gender, #book, #book^2 day of week	-	age, gender, day of week
Observations	7,853	7,147	70	64	10,479	9,867	60	56
Subjects	70	64	70	64	60	56	60	56
R-squared	0.001	0.014	0.097	0.234	0.001	0.005	0.188	0.453

Robust standard errors in parentheses, Clustering on subject level for time per output
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

to higher output than a piecerate or at least to a similar output for a lower cost. We thus conclude our first results:

Result 1 *Our results under* High Difficulty *are in line with hypothesis 1. Under both* Piecerate+Goals *and* Goals *outputs are on similar levels and roughly 15% higher than under* Piecerate.

Result 2 *Our results under* Low Difficulty *are partially in line with hypothesis 1. Output under* Piecerate+Goals *is again roughly 15% higher than under* Piecerate. *However,* Goals *results in a similar output level as* Piecerate*, albeit at a lower cots.*

The previous part demonstrated that self-chosen goals can increase productivity. However, while that average outputs in Goal and Piecerate+Goal do not differ significantly under High Difficulty they do significantly under Low Difficulty. Yet, the incentives in Goal and Piecerate+Goal differ substantially: in Goal, no monetary incentives for additional outputs are provided while in Piecerate+Goal, the paid piecerate provides incentives for additional outputs. In the following we will take a closer look how the size of self-chosen goals is influenced by the two treatments Goal and Piecerate+Goal.

Table 4: OLS explaining Goal Size

	Goal	Goal	Goal	Goal
	(1)	(2)	(3)	(4)
Piecerate+Goal	31.29***	29.40***	32.41***	30.68***
	(10.64)	(10.79)	(10.41)	(10.50)
Low Difficulty	2.711	0.381	3.612	1.063
	(10.68)	(11.14)	(10.60)	(10.80)
Time Ability		-0.204**		-0.198**
		(0.0876)		(0.0834)
Male			22.90**	22.36**
			(10.62)	(10.73)
Age			-2.824	-3.294
			(2.682)	(2.608)
Constant	114.4***	148.8***	167.0**	211.9***
	(8.810)	(17.60)	(66.24)	(66.29)
Observations	85	81	85	81
R-squared	0.094	0.134	0.157	0.202

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Our data reveal that the size of the chosen goal is affected by the underlying incentive scheme. In line with hypothesis 2, goals tend to be higher when they lead to higher payoffs. With HIGH DIFFICULTY is the average goal of 144 books in PIECERATE+GOAL significantly higher than the average goal of 116 books in GOAL ($p = 0.064$).[17] We observe the same pattern with LOW DIFFICULTY: on average subjects chose a goal of 150.5 in PIECERATE+GOAL which is significantly higher than the average chosen goal of 115 in GOAL ($p = 0.039$). Interestingly, the goal size is influenced by the different incentive schemes, but not by the difference in difficulty. The chosen goals in GOAL and PIECERATE+GOAL do not differ significantly between HIGH and LOW DIFFICULTY (for both $p > .68$). Besides the incentive scheme is the goal size influenced by initial ability and gender. As expected, are goal size and the time needed in the initial ability test significantly correlated. Subjects who worked faster in the ability test tend to chose higher personal goals ($r = -0.24$ and $p = 0.0306$, Spearman rank correlation). In addition, male workers tend to choose goals that are roughly 22 units higher than the ones chosen by female workers ($p = 0.064$).[18]

Table 4 demonstrate that these results can be confirmed with OLS regressions. Even when controlling for additional covariates (difficulty, ability, gender, and age) is the estimate for the goal size roughly 30 books higher in PIECERATE+GOAL than in GOAL. The difficulty has no significant impact on the chosen goal size. This suggests that subjects fail to use the information form the ability stage to correctly update their expectations.[19] Gender and the initial ability influence the chosen goal size and as models 2, 3, and 4 demonstrate they influence the chosen goal size independently of each other. We thus conclude our second result,

Result 3 *In line with hypothesis 2, subjects choose higher goals in* PIECERATE+GOAL *than in* GOAL. *In general, workers with higher ability tend to chose higher goals. In addition, male workers tend to chose higher goals than female workers, independent of ability.*

As we have shown, workers chose within GOAL andPIECERATE+GOAL similar goal sizes regardless whether they work in the HIGH or LOW DIFFICULTY environment. Given the big output differences between HIGH and LOW DIFFICULTY this results in significantly different rates of workers achieving their self-chosen goal. In GOAL 55% of workers

[17]Note that the incentive schemes discussed in this paper have no negative monetary consequence if the goal is not reached.

[18]We observe similar differences in both GOAL (103 vs.125) and PIECERATE+GOAL (135 vs. 158).

[19]Subjects need on average 37 seconds less in the ability stage under LOW DIFFICULTY ($p = 0.0014$). However, the coefficient for LOW DIFFICULTY is not significant even if we do not control for ability.

reach their goal under HIGH DIFFICULTY and 85% under LOW DIFFICULTY ($p = 0.038$, χ^2-test). In PIECERATE+GOAL only 32% of workers reach their goal under HIGH DIFFICULTY and 85% under LOW DIFFICULTY ($p < 0.001$, χ^2-test). With HIGH DIFFICULTY is the failure rate higher in PIECERATE+GOAL than in GOAL, but it does not reach conventional levels of significance ($p = 0.121$, χ^2-test). With LOW DIFFICULTY we observe identical failure rates under both incentive schemes.

Obviously, reaching a goal is highly correlated with the size of the goal ($r = -0.5458$ with $p < 0.001$, point-biserial correlation) — the smaller the goal, the higher the likelihood to reach it. As previously demonstrated male workers tend to choose significantly higher goals than female workers and this, therefore, translates into significantly different failure rates across gender. Overall, 52% of male workers reach their goal, which is a significantly lower success rate than among female workers with 74% ($p = 0.035$,χ^2-test). This difference is mostly driven by the behavior in the LOW DIFFICULTY environment, in which it is relatively easy to attain a high goal. All female workers achieve their goal in LOW DIFFICULTY, while on average only 70% of males do ($p = 0.089$ in GOAL and $p = 0.038$ in PIECERATE+GOAL, both χ^2-test). We summarize these observations as follows:

Result 4 *Workers are more likely to attain their goals under* LOW DIFFICULTY. *On average, male workers fail more frequently to attain their self-chosen goals than female workers.*

Overall, and in line with hypothesis 3, is the chosen goal size significantly correlated with workers' output; the higher the goal the higher the output ($r = 0.273$ and $p = 0.012$, Spearman rank correlation). Figure 2 plots the relationship of goals and outputs for each incentive scheme in HIGH and LOW DIFFICULTY environments.

The positive correlation between goals and output is significant for GOAL with HIGH and LOW DIFFICULTY as well as for PIECERATE+GOAL with LOW DIFFICULTY (for all three treatments $r > 0.45$ and $p < 0.049$, Spearman rank correlation). However, they are not correlated in PIECERATE+GOAL with HIGH DIFFICULTY ($r = 0.07$ and $p = 0.739$, Spearman rank correlation). Table C.2 in the appendix shows OLS results regressing output and speed on goal size, ability, gender, and age for each treatment. The results confirm the positive effect of goal size on output in three of our four treatments and show that it is independent of ability.

While goal size is not significantly correlated with output in PIECERATE+GOAL under HIGH DIFFICULTY we have previously demonstrated that workers produce higher outputs than in PIECERATE without goals. In addition, we have demonstrated that workers

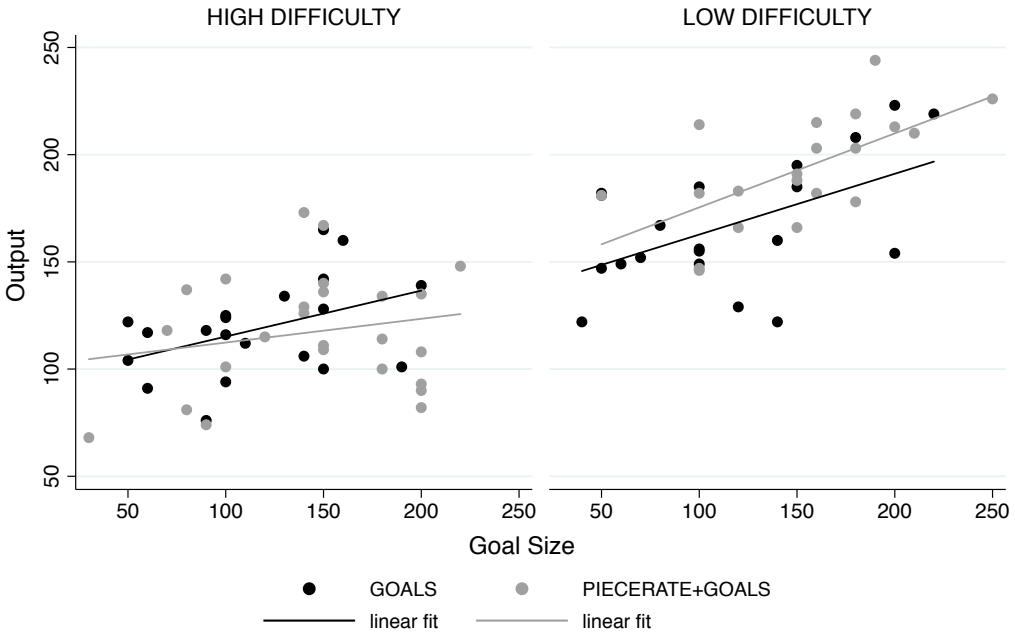


Figure 2: Relationship Goal Size and Output

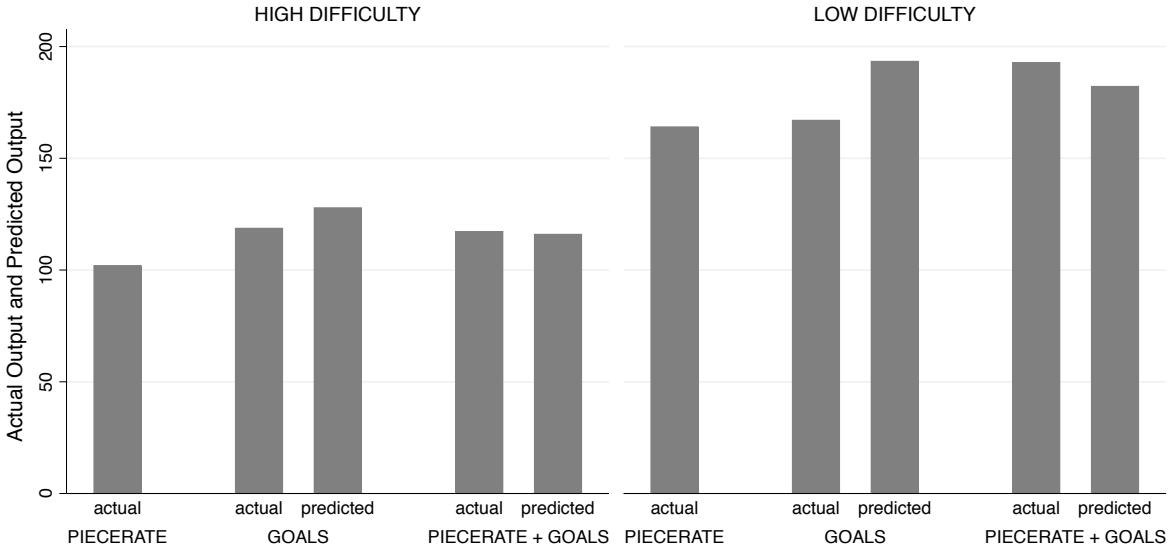
work faster with goals than without goals, however, only 32% of workers reach their goal in Piecerate+Goal under High Difficulty. These results combined suggest that in this treatment goals are set to high and once workers realize this goals loose their motivational power while the monetary incentives of the piecerate keep their motivational power. On the other hand, we observe a significant relationship between goals and output in Goal with Low Difficulty but no significant difference in output compared to Piecerate. At the same time goals are significantly higher in Piecerate+Goal than in Goal with the same attainment rate suggesting that workers could have chosen higher goals.

In a last step we will generate counterfactuals and analyze potential outcomes if workers would have chosen different goals. To produce counterfactuals we first separately estimate for both difficulty levels the influences on individuals' outputs Y_i with the following equation:

$$Y_i = \beta_0 + \beta_1 GoalSize_i + \beta_2 d_i^{\text{Goal}} + \beta_3 GoalSize_i \times d_i^{\text{Goal}} + \gamma \mathbf{X} + \epsilon_\mathbf{i}. \quad (11)$$

The output Y_i is determined by the chosen goal ($GoalSize_i$) and the goal scheme for which we control with an indicator (d_i^{Goals}) that is 1 in Goals and 0 in Piecerate + Goals. The interaction between the treatment indicator and the goal size controls for

differences between treatments with respect to the impact of the goal size. The vector $\mathbf{X}$ controls for the additional individual specific characteristics of gender, age and time needed in the ability stage. In a second step, we predict the output $\hat{Y}_i$ of worker i if a different goal would have been chosen. For this the actual goal is replaced with the average goal of the same ability quantile in the other goal treatment.[20] Figure 3 shows the results of this exercise and displays them next to the original outputs.



The predicted outputs $\hat{Y}_i$ are based on the estimation $Y_i = \beta_0 + \beta_1 GoalSize_i + \beta_2 d_i^{\text{Goal}} + \beta_3 GoalSize_i \times d_i^{\text{Goal}} + \gamma \mathbf{X} + \epsilon_\mathbf{i}$ done for each of the two difficulties. The prediction $\hat{Y}_i$ uses the mean goal of the same ability quantile of the other treatment.

Figure 3: Actual and Predicted Outputs in Both Studies

These counterfactuals do not change the comparisons under High Difficulty. Workers in Goal would have produced slightly higher outputs if they had chosen the average goal from Piecerate+Goal, yet the output would still not differ from the output in Piecerate+Goal ($p = 0.138$) and still be significantly higher than in Piecerate ($p < 0.001$). If workers in Piecerate+Goal would have chosen the average goal from Goal it would have had only a negligible impact on the output which would still be significantly higher than the one Piecerate ($p = 0.004$) and insignificant compare to the one in Goal ($p = 0.598$).

In Low Difficulty comparisons with the counterfactuals look different for Goal.

[20] Using the average goal per ability quantile leeds to higher goals for high ability workers and lower goals for low ability workers. Qualitatively similar results are obtained if only the overall average goals of the treatments are used.

If in GOAL workers would have chosen the higher goals of PIECERATE+GOAL it would have lead to significantly higher outputs than in PIECERATE ($p < 0.001$). Average output would then be very similar to the one in PIECERATE+GOAL ($p = 0.936$). The comparisons in PIECERATE+GOAL are not different with the counterfactuals. While output would have decreased with the average goal from GOAL it would still have been higher than the outputs in PIECERATE ($p = 0.007$) and GOAL ($p = 0.023$).

Result 5 *In line with hypothesis 3 accurately sized goals lead to higher output. The output in* GOAL *is very sensitive to the goal size and suffers under badly chosen goals. However, in* PIECERATE+GOAL *monetary incentives guarantee that workers work hard even when the chosen goals turn out to be unattainable or to low.*

5 Discussion and conclusions

In this paper, we explored how behavior is affected by self-chosen goals when attaining the goal yields no additional monetary consequences. Our benchmark was a regular piece-rate contract without any goals. Compared to the benchmark, we observed significantly higher outputs with self-chosen goals in the presence of a regular piece-rate contract. The performance of self-chosen goals in the absence of monetary incentives depended on the task difficulty; in a high difficulty environment we observed significantly higher output and in a low difficulty environment we observed similar outputs as in the benchmark. Our data indicate that workers failed to correctly adjust their goals to the difficulty of the work environment resulting in goals that were to low in the work environment with low difficulty. This failure impeded the effectiveness of goals in the absence of monetary incentives. However, the impact of goals on performance was more robust in the presence of (goal-independent) monetary incentives. This evidence that self-chosen goals in the presence of larger incentives yield more robust outcomes is consistent with evidence from exogenously assigned goals in principal-agent settings (Corgnet, Gómez-Miñambres and Hernán-Gonzalez, 2015). In addition and in line with Dalton, Gonzalez and Noussair (2016), male workers tended to choose higher goals than female workers and consequently a substantial fraction of male workers failed to attain their goals.

Taken together, our results suggest that self-chosen personal work goals can increase performance and their impact on performance is more robust in the presence of monetary incentives. Nevertheless, our study also shows that goals in the complete absence of monetary incentives can still induce the same or even higher performances as a regular piece-rate contract.

References

Aarts, Henk, Peter M Gollwitzer, and Ran R Hassin. 2004. "Goal contagion: perceiving is for pursuing." *Journal of personality and social psychology*, 87(1): 23.

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference points and effort provision." *American Economic Review*, 101(2): 470–492.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor supply of New York City cabdrivers: One day at a time." *Quarterly Journal of Economics*, 407–441.

Clark, Michiel, David Gill, Victoria Prowse, and Mark Rush. 2016. "Using Goals to Motivate College Students: Theory and Evidence from Field Experiments." *IZA Discussionpaper*, , (10283).

Corgnet, Brice, Joaquín Gómez-Miñambres, and Roberto Hernán-Gonzalez. 2015. "Goal setting and monetary incentives: When large stakes are not enough." *Management Science*, 61(12): 2926–2944.

Crawford, Vincent P, and Juanjuan Meng. 2011. "New york city cab drivers' labor supply revisited: Reference-dependent preferences with rationalexpectations targets for hours and income." *American Economic Review*, 101(5): 1912–1932.

Dalton, Patricio S, Victor Gonzalez, and Charles N Noussair. 2016. "Self-Chosen Goals: Incentives and Gender Differences." *working paper*.

Falk, Armin, and Markus Knell. 2004. "Choosing the Joneses: Endogenous goals and reference standards." *The Scandinavian Journal of Economics*, 106(3): 417–435.

Farber, Henry S. 2008. "Reference-dependent preferences and labor supply: The case of New York City taxi drivers." *American Economic Review*, 98(3): 1069–1082.

Fehr, Ernst, and Lorenz Götte. 2007. "Do workers work more if wages are high? Evidence from a randomized field experiment." *American Economic Review*, 97(1): 298–317.

Goerg, Sebastian J. 2015. "Goal setting and worker motivation." *IZA World of Labor*.

Gómez-Miñambres, Joaquín. 2012. "Motivation through goal setting." *Journal of Economic Psychology*, 33(6): 1223–1239.

Harding, M., and A. Hsiaw. 2014. "Goal setting and energy conservation." *Journal of Economic Behavior & Organization*, 107(209-227).

Heath, Chip, Richard P Larrick, and George Wu. 1999. "Goals as reference points." *Cognitive psychology*, 38(1): 79–109.

Hsiaw, Alice. 2010. "Goal-setting, social comparison, and self-control."

Hsiaw, Alice. 2013. "Goal-setting and self-control." *Journal of Economic Theory*, 148(2): 601–626.

Huffman, D., and L. G. Götte. 2007. "Affect as a source of motivation in the workplace." In *Emotion and Decision-Making.* , ed. Roy Baumeister and George Loewenstein. New York:Russell Sage.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the Econometric Society*, 263–291.

Kaiser, Johannes. 2009. "An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples." *The Stata Journal*, 7(3): 402–412.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2010. "Self-control and the development of work arrangements." *American Economic Review*, 100(2): 624–628.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-control at work." *Journal of Political Economy*, 123(6): 1227–1277.

Koch, Alexander, and Julia Nafziger. 2016. "Goals and bracketing under mental accounting." *Journal of Economic Theory*, 162: 305–351.

Koch, Alexander K, and Julia Nafziger. 2011. "Self-regulation through Goal Setting*." *Scandinavian Journal of Economics*, 113(1): 212–227.

Kőszegi, Botond, and Matthew Rabin. 2006. "A model of reference-dependent preferences." *Quarterly Journal of Economics*, 1133–1165.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-dependent risk attitudes." *American Economic Review*, 97(4): 1047–1073.

Lee, Thomas W, Edwin A Locke, and Soo H Phan. 1997. "Explaining the assigned goal-incentive interaction: The role of self-efficacy and personal goals." *Journal of Management*, 23(4): 541–559.

Locke, Edwin A, and Gary P Latham. 2002. "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey." *American psychologist*, 57(9): 705.

Locke, Edwin A, and Gary P Latham. 2006. "New directions in goal-setting theory." *Current directions in psychological science*, 15(5): 265–268.

Locke, Edwin A, and Gary P Latham. 2013. *New developments in goal setting and task performance.* Routledge.

Moskowitz, Gordon B, and Heidi Grant. 2009. *The psychology of goals.* Guilford Press.

Ordóñez, Lisa D., Maurice E. Schweitzer, Adam D. Galinsky, and Max H. Bazerman. 2009. "Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting." *Academy of Management Perspectives*, 23(1): 6–16.

van Lent, Max, and Michiel Souverijnz. 2016. "Goal Setting and Raising the Bar: A Field Experiment." *Working Paper.*

Wu, George, Chip Heath, and Richard Larrick. 2008. "A prospect theory model of goal behavior." *mimeo, University of Chicago.*

A Summary of goal theory papers

Table A.1 presents a summary of the existing theoretical models on goal setting. For each study, we provide details regarding model type, goal type (e.g., exogenously or endogenously set), state and choice variables, and main comparative statics on worker productivity. The following categories are included:

Goal type

There are two types of goals that have been considered in the literature: Exogenous goals, which are taken as given by the individual, and endogenous goals, which the agent sets for himself. A third type of goal exists not touched on by the studies in Table A, known as participatory goals. Participatory goals are goals which are determined by two agents simultaneously, such as between an employee and her employer.

Reference-dependent preferences

An agent is said to have reference-dependent preferences provided her utility depend on her performance relative to a standard, which in the present paper, is the agent's goal or a function of the goal.

Present-biased individuals

A present-biased individual is typically captured by a "many selves" model in which the agent at different points in time value payoffs differently. For example, a time $t = 0$ an individual places equal weights on the benefit and cost of their future efforts, whereas the same individual at time $t = 1$ places higher weights on the costs of current efforts as compared to current benefits.

Loss aversion

Losses loom larger than benefits.

Diminishing sensitivity

As the absolute difference between output and the reference point increases, so does the associated payoff in absolute utility. Put differently, the marginal benefit of exceeding the reference point is smaller the further away from the reference point. This also holds in the domain of losses.

Consumption utility

By consumption utility we refer to agent's who derive utility directly from their output, e.g., pecuniary incentives, letter grades, etc. This is different from intrinsic or psychological utility which agents derive in the presence of goal-setting.

Study	Goal Type: Endogenous	Goal Type: Exogeneous	Reference dependence	Present-biased individuals	Loss aversion	Diminishing sensitivity†	Consumption utility	Empirical analysis	Setting‡
Falk and Knell (2004)	✓		✓			+	✓	✓	2
Wu et al. (2008)		✓	✓		✓	+/−			4
Jain (2009)	✓		✓	✓			✓		4
Suvorov and van de Ven (2009)	✓	✓	✓	✓	✓		✓		4
Hsiaw (2010)	✓		✓	✓			✓		4
Koch and Nafziger (2011)	✓		✓	✓	✓		✓		4
Gómez-Miñambres (2012)		✓	✓		✓	+	✓		1
Hsiaw (2013)	✓		✓	✓			✓		4
Harding and Hsiaw (2014)	✓		✓	✓		+/−		✓	3
Corgnet et al. (2015)		✓	✓		✓	+/−	✓	✓	1
Clark et al. (2016)	✓		✓	✓	✓		✓	✓	2
Dalton et al. (2016)	✓		✓		✓	+/−	✓	✓	1
Koch and Nafziger (2016)	✓		✓	✓	✓		✓		4
Hsiaw (2016)	✓		✓	✓	✓		✓		4
Lent and Souverijn (2016)	✓	✓	✓					✓	2
This paper	✓		✓		✓	+/−	✓	✓	1

† + denotes denotes diminishing sensitivity in the domains of gains and − in the domains of losses.
‡ Each study can be broken down into one of the following settings:
1 - Workplace incentives, 2 - Educational improvement, 3 - Energy conservation, 4 - General environment

Table A.1: Summary of goal-setting models.

B Figures

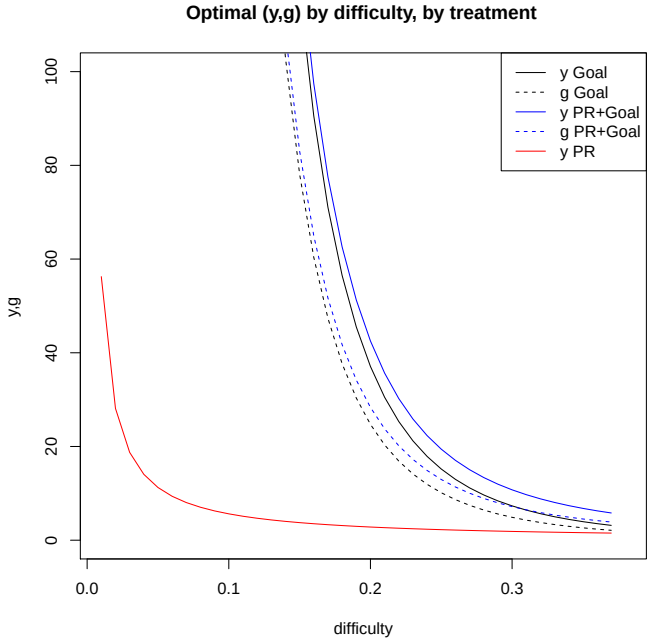


Figure B.1: Optimal output y and goal g (if present) as a function of difficulty, by treatment.

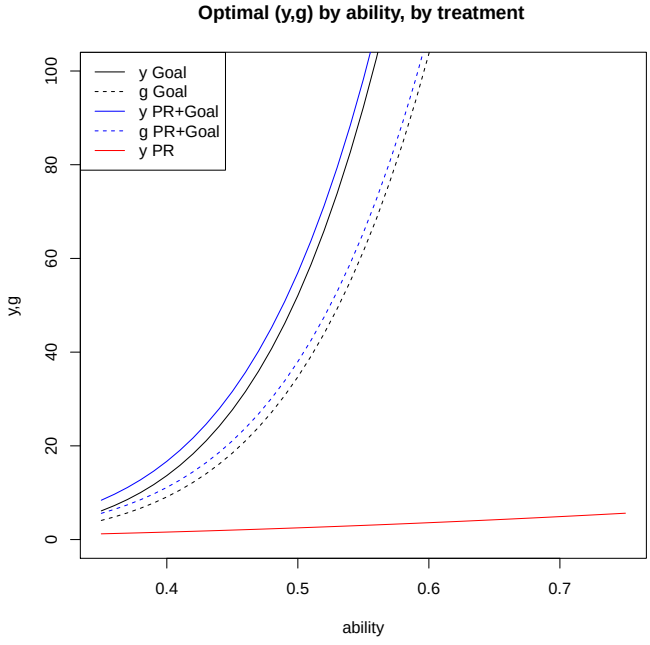


Figure B.2: Optimal output y and goal g (if present) as a function of ability, by treatment.

C Tables

Table C.1: OLS explaining Time per Output and Total Output

	High Difficulty				Low Difficulty			
	Time Per Output		Total Output		Time Per Output		Total Output	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Goal	-17.72**	-12.15***	16.70**	16.07**	-2.646	-2.769	2.950	1.506
	(8.744)	(4.038)	(6.464)	(6.516)	(4.583)	(3.532)	(8.984)	(8.781)
Piecerate+Goal	-14.29**	-9.465**	15.24**	13.09*	-10.10**	-11.34***	28.80***	29.60***
	(7.085)	(4.531)	(6.880)	(7.276)	(4.418)	(3.677)	(8.390)	(9.321)
Time Ability		0.0576**		-0.0931**		0.0476**		-0.159***
		(0.0239)		(0.0376)		(0.0224)		(0.0408)
Age		0.192		-0.208		0.0838		-0.563
		(0.779)		(1.259)		(0.704)		(1.486)
Male		1.396		-5.027		-1.243		7.927
		(3.748)		(6.391)		(4.080)		(8.435)
BookNumber		-2.190***				-0.722***		
		(0.423)				(0.156)		
BookNumber2		0.0106***				0.00304***		
		(0.00272)				(0.000946)		
Tuesday		-8.586		12.56		1.790		15.21
		(5.681)		(8.372)		(5.111)		(12.32)
Wednesday		-7.998		10.76		-4.524		28.93**
		(5.476)		(8.459)		(4.709)		(12.25)
Thursday		-7.050		10.96		-3.231		19.33*
		(6.208)		(9.484)		(3.753)		(10.22)
Friday		-8.118		12.33		3.325		4.196
		(5.502)		(9.122)		(5.020)		(11.69)
Constant	108.5***	175.5***	102***	116.6***	69.51***	94.37***	164.0***	181.3***
	(4.348)	(23.44)	(4.119)	(33.20)	(3.433)	(19.72)	(6.080)	(39.26)
Controls	-	age, gender, #book, #book^2 day of week	-	age, gender day of week	-	age, gender, #book, #book^2 day of week	-	age, gender, day of week
Observations	7,853	7,147	70	64	10,479	9,867	60	56
Subjects	70	64	70	64	60	56	60	56
R-squared	0.001	0.014	0.097	0.234	0.001	0.005	0.188	0.453

Robust standard errors in parentheses, Clustering on subject level for time per output
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table C.2: OLS explaining Time per Output and Total Output

	High Difficulty				Low Difficulty			
	Goal		Piecerate+Goal		Goal		Piecerate+Goal	
	Total Output	Time p. Output	Total Output	Time p. Output	Total Output	Time p. Output	Total Output	Time p. Output
Goal Size	0.216**	-0.153**	0.00937	-0.0364	0.263**	-0.0975**	0.336***	-0.0982**
	(0.0855)	(0.0677)	(0.132)	(0.138)	(0.119)	(0.0405)	(0.111)	(0.0355)
Ability Time	-0.158*	0.122	-0.0957	0.0720	-0.223	0.123**	-0.172**	0.00944
	(0.0846)	(0.0706)	(0.0929)	(0.0717)	(0.137)	(0.0568)	(0.0705)	(0.0655)
Male	-16.36**	10.28*	-7.846	5.787	4.590	-6.100	-4.819	5.731
	(6.938)	(5.620)	(11.28)	(9.587)	(13.47)	(5.658)	(6.892)	(4.890)
Age	0.324	0.230	-2.427	1.824	-4.038	2.345	-1.301	-0.497
	(2.141)	(1.981)	(2.317)	(1.965)	(5.341)	(1.918)	(3.815)	(2.148)
Constant	121.3*	79.72	192.7***	42.43	263.4*	8.751	199.2*	81.70
	(63.43)	(58.89)	(58.48)	(52.83)	(125.7)	(46.79)	(97.33)	(56.18)
Observations	19	2,255	24	2,783	19	3,185	19	3,710
Subjects	19	19	24	24	19	19	19	19
R-squared	0.469	0.001	0.115	0.001	0.448	0.003	0.534	0.001

Robust standard errors in parentheses, Clustering on subject level for time per output
$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

D Procedural Details

In 2010, the library of the Max Planck Institute for Research on Collective Goods in Bonn had to be restructured from an alphabetical order to an order by topic. Each single book out of roughly 35,000 books had to be searched and found in the library shelves and had then to be relocated to a different shelf. The library uses mobile aisle shelving allowing only one subject at a time to work in this area (see also Figure D.1). Consequently, there was always only a single worker in the library at a time; usually one in the morning and one in the afternoon (allocation of treatments to the day of the week and the time of the day was randomized).

Figure D.1: Picture of the library showing the mobile aisle shelving

Job offers were advertised at the University of Bonn. The announcement read that it was an one-time job opportunity lasting for 3.5 hours and paying a minimum amount of €22. The announcement contained information about the background (re-organizing a library), as well as a short description of the work task. Everyone who saw the poster was allowed to apply for the job. The only restriction was that, given the height of the bookshelves, participants needed to be taller than 1.70 meters. Job candidates had to apply online, stating their preferred working times. Among all applicants, a student research assistant randomly selected participants for our study and offered them an open slot. The time slots (and thus also the subjects) were randomly assigned to the different treatments. Subjects were informed via email about their time slot. Two days before

the actual day of work, each subject received an email reminding him or her again of the time slot, location, duration, and minimum wage payment. Workers were not aware that they were participating in a field experiment. Indeed, due to the authenticity of the work task and work environment, we had no report of any subject asking whether their job was part of an experiment. For the second wave with low difficulty the whole procedure was repeated in 2014.

Upon arrival, the subject received a short written manual for the work task and was additionally instructed by one of the librarians who strictly followed a fixed protocol (see Appendix E for an English translation of the German instructions). Then the subject received a list of books to be searched. While the books in the shelves were alphabetically ordered, the books on the list were ordered by topic. This implied that the probability of two successive books on the list being close to each other in the shelves was rather low. The list was so long that it was obvious to the subject that he or she would not have been able to find all the books on list within three hours, which is important because otherwise finishing the entire list might have served as a goal as well – potentially overriding our treatment manipulation.

The exact task for the subject consisted of the following steps:

1. Pick the book from the top of the list and search for it in the library shelves.
2. Scan the book's ID at a computer workstation and mark the book on the list. If the book is borrowed, instead of the book a placeholder will be at that position in the shelf. In that case, a barcode label for borrowed books should be scanned and the book should be marked in the list correspondingly.
3. Place the book (or the book's placeholder) in a book trolley. Stick to the exact order as it is given on the list of books.
4. Pick the next book from the list and start over.

The librarian emphasized that the subject should work sequentially through the list, keeping the same order on the book trolley as on the list since the books would later be re-labeled and placed into the shelves in this order.[21] After the work task had been explained, each subject had to search for two test books, scan them at the workstation and place them on a book trolley. This procedure served three purposes. First, it ensured that each worker had understood the work task. Second, it provided workers with a rough estimate of their own ability – i.e., how long they approximately need to find a single book - which is important information for workers when they have to set themselves a goal.

[21] According to the librarians who applied the new labels, the order was always kept but for a few exceptions.

Third, because scanning a book provides us with a timestamp for each book, we can approximate subjects' initial ability by using the time difference between the test-books' first and second timestamp.

The treatment manipulation, i.e., the exact payment scheme, was introduced only after the subject had found the two test books and scanned them at the workstation. Like the task description, the payment scheme was also handed out in written form and was additionally explained by a librarian. Subjects then could pose clarifying questions. Afterwards, if the treatment featured a self-chosen goal, subjects had to announce their personal goal. The goal was noted on a post-it that was attached to the display of the workstation. The librarian started a timer and left the workplace. The subject started working for three hours. The workstation always showed the current number of scanned books. Thus subjects were informed about their current earnings and their distance to the goal at any time. Subjects were allowed to take a break whenever necessary. After exactly three hours, the librarian returned, checked the total amount of scanned books and calculated the total payoff accordingly. In case the subject was found to be in the process of scanning books at the computer terminal, he or she was allowed to finish the scanning and stop working afterwards. In the end, a short questionnaire was handed out, eliciting the difficulty of the task, subjects' satisfaction with the personal performance, and their general well-being.

In the first run under high difficulty, subjects searched for a total of 11,461 books during the course of the present study, which is roughly one third of the library's holding of books. After approximately 5,000 books had been handled, the shelves were compressed and filled up with books from other parts of the library. This ensured that the amount of books in the shelves was similar for each subject; and in fact, we do not observe that subjects become faster as the holding of books declines over the course of the experiment.[22] In the second run, books were direclty put back after the experiment. Thus, the inventory of the shelves is identical for all subjects and again no significant correlation between date and initial ability can be found ($r = 0.004$ and $p = 0.974$ Spearman rank correlation). After all sessions of the field experiment had been completed, subjects were debriefed via email.

[22]We tested for a correlation between date and initial ability, which turned out to be not significant ($r = 0.09$ and $p = 0.475$, Spearman rank correlation). This means that the initial ability test which is (by design) not influenced by the treatments does not become easier or more difficult over time.

E Experimental instructions

E.1 Instructions for book search

Your task is to find the books on the list you have just been handed out. Please work your way from the top of the list to the bottom. Once you have found a book, scan it on your PC and place it in the designated book trolley.
In a next step, our librarians will provide the books with new labels and place them in a different location. Because the sorting of the labels corresponds to the order of the books on your list, it is extremely important to us that you locate, scan, and place the books on the book trolley in the same order.
It might happen that you are not able to find a book. If the book has been borrowed, there should be a plastic tab in its place instead. This tab should contain this information and the corresponding book details. In this case, please mark the corresponding book on your list with an "A" and place the plastic tab in the designated storage container as a substitute for the missing book. Also, make sure to scan the barcode found on the green card ("LOAN").
Another possibility might be that you cannot find a book because it has been misplaced by a previous user. In this case, please mark the corresponding book on your list with a "00" and place a white sheet of paper in the designated book trolley instead. If you are not sure whether the book you found in the shelf is the same book as on your list, please point this out by putting a white sheet of paper inside the book and placing it in the designated book trolley.
It might also happen that you find more than one copy of a given book. In this case, please mark the corresponding book on your list with a "2" (or "3", "4",...) and then scan all of the copies.
Please approach us if there is a problem, which you are not able to solve on your own.

E.2 PIECERATE instructions

We ask you to execute your task carefully. However, try to find as many books as possible during the next 3 hours. The more books you find, the higher the cash payment will be which you will get from us immediately after the time has expired. The following applies:

- You will receive a base salary of €22. This means that you will get at least €22 for the 3 hours you are here.
- In addition to your base salary, you will receive a bonus payment. The amount of the bonus payment will depend on the amount of books that you have successfully searched for and found, as you will get an extra 10 cents for every book you find.
- A book counts as "found" only if you have either scanned the book (also the extra copies if there are any) or if you have scanned the plastic tab that belongs to a borrowed book. Missing books (barcode on the red card "MISSING") do NOT count as found.

E.3 PIECERATE + GOAL instructions

We ask you to execute your task carefully. However, try to find as many books as possible during the next 3 hours. The more books you find, the higher the cash payment will be which you will get from us immediately after the time has expired. The following applies:

- You will receive a base salary of €22. This means that you will get at least €22 for the 3 hours you are here.
- In addition to your base salary, you will receive a bonus payment. The amount of the bonus payment will depend on the amount of books that you have successfully searched for and found, as you will get an extra 10 cents for every book you find.
- A book counts as "found" only if you have either scanned the book (also the extra copies if there are any) or if you have scanned the plastic tab that belongs to a borrowed book. Missing books (barcode on the red card "MISSING") do NOT count as found.

Apart from that, you have to estimate the amount of books that you believe you will find. This estimate represents your personal goal.

E.4 GOAL instructions

We ask you to execute your task carefully. However, try to find as many books as possible during the next 3 hours. The following applies:

- You will receive a base salary of €22. This means that you will get €22 for the 3 hours you are here.
- A book counts as "found" only if you have either scanned the book (also the extra copies if there are any) or if you have scanned the plastic tab that belongs to a borrowed book. Missing books (barcode on the red card "MISSING") do NOT count as found.

Apart from that, you have to estimate the amount of books that you believe you will find. This estimate represents your personal goal.