(i) title: **Self-regulation through goal setting**

(ii) authors: Alexander K. Koch (School of Economics and Management, Aarhus University, Building 1322, 8000 Aarhus C, Denmark, Email: akoch@econ.au.dk, Tel.: +45 8942 2137.)

Julia Nafziger (School of Economics and Management, Aarhus University, Building 1322, 8000 Aarhus C, Denmark, Email: jnafziger@econ.au.dk, Tel.: +45 8942 2048.)

(iii) abstract:

Goals are an important motivator. But little is known about why and how people set them. We address these questions in a model based on two stylized facts. i) Goals serve as reference points for performance. ii) Present-biased preferences create self-control problems. We show the power and limits of self-regulation through goals. Goals increase an individual’s motivation – but only up to a certain point. Furthermore, they are painful self-disciplining devices. Thus, greater self-control problems may result in tougher goals; but for a severe present bias goals either lack motivating force, or are too painful to be accepted.

(iv) Keywords: Goals; Self-control; Motivation; Time inconsistency; Psychology.

(iv) JEL Classification: A12, C70, D91
1 Introduction

Goals are a cornerstone of the human motivational system. While the role of goals has been largely ignored by economists, psychologists have studied extensively the way in which goals affect task performance. According to Bandura, this research has shown that “the regulation of motivation by goal setting is a remarkably robust phenomenon” (p.xii of his foreword to Locke and Latham’s (1990) survey on goals). The typical study takes goals as exogenously assigned and measures how such given goals affect performance. However, little is known about why and how people set goals for themselves. We address these questions in a model based on two stylized facts from psychology and behavioral economics. i) Goals serve as reference points for performance. ii) Present-biased preferences create self-control problems. Our contribution is to show how the choice of goals can serve as an instrument for regulating own behavior, and how individuals optimally choose such goals. We thereby also highlight the limits of self-regulation through goal setting.

People often exhibit time-inconsistent preferences because they overemphasize immediate costs or benefits relative to more distant ones (e.g., Ainslie and Haslam 1992, Strotz 1955, Phelps and Pollak 1968, Laibson 1997, O’Donoghue and Rabin 1999a). Such a present

1Bargh (1990), Carver and Scheier (1999), and Oettingen et al. (2001) all point to the issue succinctly summarized by Fuhrmann and Kuhl (1998, p.653): “Goals are typically considered to be ‘just there’ and the interest of research mostly remains restricted to the process of how the organism manages to reach goals. We believe that the process by which goals are formed and become an object of self-regulation deserves much more attention in self-regulation research.”
bias can create self-control problems. For instance, a person finds that working hard on a task scheduled in a week’s time is the optimal thing to do. But once the date of the task arrives, the person may decide to shirk, because the costs of working hard become more salient when they are immediate and may now loom larger than the gains, which only accrue in the future. A sophisticated individual, who anticipates his self-control problem (see e.g., O’Donoghue and Rabin 1999a), can attempt to regulate his future behavior. In line with Baumeister and Heatherton’s (1996) notion that “without clear and consistent standards self-regulation will be hampered”, we show how goals help an individual exert control over future actions.

A key property of such goals is that they affect individuals’ perceptions of future outcomes. Locke and Latham (2002, p.709-710) state in their survey of the empirical goal research in psychology that “goals serve as the inflection point or reference standard for satisfaction versus dissatisfaction […] For any given trial, exceeding the goal provides increasing satisfaction as the positive discrepancy grows, and not reaching the goal creates increasing dissatisfaction as the negative discrepancy grows.” Building on these stylized facts, we model goals as reference points against which actual task outcomes are measured, assuming that people display loss aversion. Hence, goal setting leads to an endogenous reference point for task performance.

We show that setting a challenging goal for future performance makes future selves strive

---

2Heath, Larrick, and Wu (1999) point out the similarity to the value function in Kahneman and Tversky’s (1979) Prospect Theory, and present evidence supporting this view. Fehr and Götte (2007) link individual-level measures of loss aversion for bicycle messengers to effort changes in response to an exogenous variation in the piece rate. Their results support a reference dependent preference model, where messengers set a daily income goal.
harder: they become motivated by the fear of falling short of the goal. An increase in the goal level set today allows an individual to raise his motivation to work hard in the future. This result is in line with Locke and Latham’s (1990) summary of evidence concerning the relationship between motivation and (exogenously assigned) goal difficulty: “people with high goals produce more because they are dissatisfied with less. […] This is why they are motivated to do more than those with easy goals.”

This brings us then to our main question: How do people set goals for themselves? An individual with a more severe present bias is more tempted to shirk. Thus, the greater the present bias, the higher the goal required to motivate effort. However, there are limits to self-regulation, because a higher goal increases the individual’s motivation only up to a certain point. As a consequence, for a severely present-biased individual self-regulation with goals is not feasible. But our model shows that this feasibility constraint need not be what prevents the individual from self-regulating. A more severe limit may arise because goals are painful self-disciplining devices. While a higher goal raises the motivation of a future self, it also increases the chances of falling short of it and then suffering a psychological loss. Consistent with evidence from psychology (e.g., Locke 1968, Hollenbeck and Klein 1987, Hollenbeck et al. 1989), we show that very difficult goals might not be “accepted” – even if self-regulation with a very difficult goal is feasible. As a result, the optimal goal is non-monotonically related to the severity of the present bias. For intermediate ranges, the individual compensates for a stronger bias with a higher goal. If however the individual faces a severe present bias, goals may not provide sufficient motivation to

---

3 Similarly, Abeler et al. (2009) show in a controlled lab experiment that higher reference levels lead to higher effort.
sidestep the self-control problem; or they are just too painful. Such a person gives up on self-regulation.

Our paper contributes to the small economics literature on goal choice. In parallel and independent work to our paper, Suvorov and van de Ven (2008) and Hsiaw (2009b) also ask how goals can help overcome self-control problems, but adopt a very different interpretation of goals by modeling them as self-fulfilling rational expectations. In Hsiaw (2009b), this implies that the individual cannot actively choose his goal – an important ingredient of our model. Suvorov and van de Ven (2008) interpret goals as past expectations regarding future effort, in the spirit of Köszegi and Rabin (2006). In this framework, goals can be stochastic, and bounds on goal setting arise because the individual can only choose among those reference points that are consistent with one of the multiple personal equilibria. In our setting, the individual can continuously adjust goals, e.g., setting a higher aspiration level than the expected outcome, or choosing to be less ambitious. Motivational bounds on goal setting hence arise endogenously in our model, and not from a limited choice set. Falk and Knell’s (2004) social comparison model highlights other motivations for goal setting than self-regulation: *self-enhancement* (comparisons with others make people feel better) and *self-improvement* (comparisons with others can help improve own performance).

Our paper links the literature on goals with the behavioral economics literature that deals with the question of how present-biased individuals cope with self-control problems (for

---

4Koch and Nafziger (2009) and Hsiaw (2009a) provide extensions to multiple tasks, to study goal bracketing.
an overview see e.g., Brocas et al. 2004). Much of this literature focuses on the role of external commitment technologies for achieving *precommitment* (Elster 2000). For example, people may invest in illiquid assets, or sign binding contracts in order to overcome self-control problems in savings, consumption and retirement decisions (Laibson 1997, Diamond and Köszegi 2003, Carrillo 2005) or to overcome low effort provision and procrastination (O’Donoghue and Rabin 1999b, DellaVigna and Malmendier 2004, Carrillo and Dewatripont 2008). Most closely related to our approach is the paper of Carrillo and Dewatripont (2008), who study how promises to other parties help overcome self-control problems. Breaking a promise leads to costs from a loss in reputation or sanctions enforced by explicit contracts. As the model has no uncertainty about the performance measure, making promises is costless if the individual takes the ‘right’ actions. Our setting highlights that performance standards (whether internally or externally enforced) are costly devices for disciplining a future self, because of the risk of falling short of the standard. A few contributions deal with intrapersonal strategies to overcome self-control problems, as does our paper. In a model where an individual has imperfect recall about past motives, Bénabou and Tirole (2004) explain why internal commitment devices actually work. In contrast, and similar to our approach, Benhabib and Bisin (2005) and Herweg and Müller (2008) assume the presence of such an internal commitment device and ask how an individual can use it to regulate behavior. Benhabib and Bisin study a consumption saving model, in which an individual may invoke control processes to inhibit automatic processes that are prone to temptation. Herweg and Müller show how self-imposed deadlines help overcome procrastination.
2 A model of goal setting

We consider the behavior of an individual with a present bias who faces a self-control problem. At date 1, the individual chooses his effort for a task. The costs of effort are immediate, whereas the outcome and the related utility realize one period later, at date 2. Due to a present bias, a self-control problem arises. At date 0, the individual thinks that working hard is optimal. But then he shirks once he makes the effort choice at date 1. To regulate his behavior at date 1, the individual can set, at date 0, a goal that serves as a reference point for the future outcome to be achieved.

The task. At date 1, the individual has to complete a task which requires him to put in some level of effort \( e \in \{e, \bar{e}\} \). While high effort causes an immediate utility cost \((c(\bar{e}) \equiv c)\), low effort causes no disutility \((c(e) = 0)\). One period later, at date 2, the task outcome \( y \in [y, \bar{y}], \ y \in [-\infty,0], \ \bar{y} \in (0, \infty] \) and any benefits arising from it realize. The task outcome depends stochastically on the individual’s effort according to the distribution \( F(y|e) \), with density \( f(y|e) > 0 \ \forall y \in [y, \bar{y}] \) and mean \( 0 < E(y|e) < \infty \). High effort raises the outcome in the sense of strict first-order stochastic dominance:

\[ F(y|\bar{e}) < F(y|e) \ \forall y \in (y, \bar{y}). \]

Goal setting. At date 0, the individual chooses a goal \( a \geq y \), which becomes anchored in his head as a standard for the future utility related to the outcome \( y \). A standard

\[ \text{The assumption } y \leq 0 \text{ can be dropped at the cost of additional notation in the proof of Lemma 2.} \]

\[ \text{For simplicity we refer to the output support as } y \in [y, \bar{y}], \text{ which should of course be read as } y \in [y, \tilde{y}] \text{ for } \tilde{y} = \infty \text{ (and analogously for } y = -\infty). \]

\[ \text{The lower bound on the goal captures the idea that the individual cannot set a goal “I want to} \]
that will be met for sure \((a = y)\) is called no goal. The assumption that an individual has the capacity to set goals for himself that remain meaningful over time is grounded in the psychology literature. For example, Gollwitzer (1999) writes: “By forming goal intentions, people translate their noncommittal desires into binding goals. The consequence of having formed a goal intention is a sense of commitment that obligates the individual to realize the goal.”

**Timing and utilities.** At date 0, the individual chooses his goal level and no payoff-relevant events occur: \(u_0 = 0\). Date-1 utility reflects the immediate cost of effort exerted on the task: \(u_1 = -c(e)\). At date 2, the individual experiences utility related to the realized task outcome relative to the goal. It consists of two components. The first component is the individual’s consumption utility or instrumental utility, \(v(y)\), that corresponds to the traditional outcome-based utility in economic models. The second component is the individual’s psychological utility, which takes the form of a Kahneman and Tversky (1979) value function. That is, psychological utility depends on the difference between the outcome realization \(y\) and the goal \(a\) (measured in consumption utility units). The individual achieve \(-\infty\) points in the exam” if the lowest possible outcome is zero points. The assumption hence excludes that the individual can mentally make himself infinitely happy by lowering his reference point, in which case no economic decision would be of any importance to him. Note however that if \(y = -\infty\), we do allow for the possibility of \(a = -\infty\), as utility then will be bounded.

For terminology and further discussion see Köszegi and Rabin (2006). As deviations from the goal are measured in consumption utility units, missing a goal on a task that is not very important hurts only a little (e.g., not hitting the waste paper basket when throwing a ball of crumpled paper), whereas missing a goal on a task that is important hurts a lot (e.g., failing the last try on a professional qualifying examination). Furthermore, psychological utility is defined over outcomes and not over actions, analogous
experiences a psychological gain $\mu^+(v(y) - v(a))$ from satisfying the goal, and a psychological loss $\mu^-(v(a) - v(y))$ from falling short of the goal, where losses count more than gains of equal size. We assume that the individual exhibits linear loss aversion (Tversky and Kahneman 1991), which is used in many applications (see e.g., DellaVigna 2009 for examples). Moreover, to avoid notational clutter, we also assume linear consumption utility. Summarizing, we write date-2 utility as:

$$u_2(y, a) = y + \eta \left( \max\{y - a, 0\} - \mu \times \max\{a - y, 0\} \right),$$

with $\mu > 1$ being the coefficient of loss-aversion\footnote{Gächter et al. (2007) provide evidence of considerable heterogeneity in loss aversion across individuals and tasks: the vast majority of individuals are loss averse, with the average coefficient of loss aversion (measured as the willingness-to-accept/willingness-to-purchase ratio) being around 2 – similar to previous studies.} and $\eta > 0$ the weight attached to gain-loss utility. In the absence of a goal, simply set $a = y$. It will be convenient to write the expected value of $u_2(y, a)$ given effort level $e$ in terms of a value function of goal setting:

$$g(e, a) = \int_y^a y f(y|e) dy + \left( \mu \int_y^a (y-a) f(y|e) dy + \int_y^a (a-y) f(y|e) dy \right)$$

$$= (1 + \eta) E(y|e) - \eta a - \eta (\mu - 1) \int_y^a F(y|e) dy,$$

where (1) uses integration by parts.

**Present bias and self-control problem.** The individual has time-inconsistent preferences, modeled as $(\beta, \delta)$-preferences (e.g., Phelps and Pollak 1968, Laibson 1997, O’Donoghue to the traditional outcome-based utility setting. Such a formulation also arises if there is uncertainty about the ‘right’ type of action to take, because not all possible future circumstances are known or can be described ex ante (an approach, e.g., taken by Rayo and Becker 2007).
and Rabin 1999a). The parameter $\delta$ corresponds to the standard exponential discount factor (for simplicity, $\delta = 1$). The second parameter, $\beta \in [0, 1)$, captures the extent to which a present bias causes current payoffs to be more salient than future payoffs. The expected utility of the individual at date $t \in \{0, 1\}$, $U_t = u_t + \beta \sum_{\tau > t} E(u_\tau)$, is given by:

$$U_0 = \beta \{-c(e) + g(e, a)\},$$
$$U_1 = -c(e) + \beta g(e, a).$$

While the incarnation of the individual at date 0 (self 0) weighs equally the effort costs and the potential future benefits related to the task outcome, his date-1 incarnation (self 1) places a larger relative weight on the effort costs due to his present bias. We assume that the individual is sophisticated, i.e., he knows about his present-biased preferences.

To illustrate how goals can serve as self-discipling devices, we assume that a conflict of interest between self 0 and self 1 arises in the absence of a goal (i.e., if $a = y$). High effort is optimal from the perspective of self 0 if the expected utility it yields ($\beta \{g(\bar{e}, y) - c\}$) exceeds that from low effort ($\beta g(\underline{e}, y)$). That is, he wants his future self to work hard if and only if

$$\beta \{g(\bar{e}, y) - g(\underline{e}, y) - c\} \geq 0,$$

(2)

or – written out using (1) – if and only if the gain from high effort $(1+\eta) \{E(y|\bar{e}) - E(y|\underline{e})\}$ (strictly positive by Assumption 1) outweighs the effort costs $c$. As immediate costs hurt self 1 more, he will actually provide low effort if and only if

$$\beta \{g(\bar{e}, y) - g(\underline{e}, y)\} - c < 0.$$

(3)

Thus, a conflict of interest between self 0 and self 1 arises whenever both (2) and (3) hold,
referred to as a *self-control problem* in the behavioral economics literature.\[10\]

**Assumption 2** The individual faces a self-control problem: in the absence of a goal \((a = y)\), self 0 strictly prefers high effort (i.e., \(g(\bar{e}, y) - g(e, y) - c > 0\)) but self 1 will shirk, i.e.,

\[
\beta_{SC} \equiv \frac{c}{g(\bar{e}, y) - g(e, y)} > \beta. \tag{SC}
\]

3 Self-regulation through goal setting

3.1 Goals as a source of motivation

We now move to the goal setting problem of the individual. Starting with the effort choice made by self 1, we ask how he responds to a goal \(a\). We then determine by backward induction the optimal goal that self 0 will set.

For a given goal level \(a\), self 1 provides high effort if and only if his expected utility from hard work exceeds that from shirking: \(\beta g(\bar{e}, a) - c \geq \beta g(e, a)\). By Condition \(\text{[SC]}\), this incentive constraint fails to hold at \(a = y\). So a higher goal can possibly get around the self-control problem if \(g(\bar{e}, a) - g(e, a)\) is increasing in the goal level – which is the case in our model. For any effort level, a higher goal raises the chance that the individual ends up falling short of the aspired level \(a\). Motivation increases with a higher goal because the individual is *even more* likely to fall short of \(a\) if self 1 shirks. To see this formally, \[10\]

\[\text{Evidence from the psychology literature shows that ‘do-your-best’ goals often fail to motivate people (e.g., Locke and Latham 1990). Within our model, it can easily be seen why: if self 0 told self 1 to ‘do his best’, without specifying a concrete goal level \(a\), self 1 indeed does what is best for him by maximizing utility from his date-1 perspective. So whenever Assumption \(\text{[2]}\) holds, self 1 will shirk rather than pick the best effort choice from the perspective of self 0.}\]
write out the incentive constraint using (1):

$$\beta \eta (\mu - 1) \int_a^\infty [F(y|\bar{e}) - F(y|\bar{e})] \, dy \geq c - \beta \{g(\bar{e}, y) - g(e, y)\}. \tag{4}$$

The right-hand side of Equation (4) is the net-utility that self 1 obtains from shirking in the absence of a goal – a positive number, as the benefit from the task does not compensate for the cost of working hard because of the individual’s present bias (Assumption 2). The left-hand side of Equation (4) measures the fear of suffering more often a psychological loss from not reaching the goal if self 1 shirks: it is non-negative and increasing in the goal level, because the probability of output below $a$ is higher under low than under high effort (by Assumption 1).

However, there is a limit to the motivating force of higher goals. To understand this intuitively, suppose output is bounded from above. Raising the goal above the maximum level of output that can possibly be achieved means that it is impossible to reach the goal, regardless of how hard the individual works. So an increase in the goal level above $\bar{y}$ has no additional motivating force, i.e., only achievable goals provide incremental incentives. As a result, the left-hand side of (4) is bounded. In the proof of Lemma 1, we then show that this bound applies more generally, i.e., also in case $\bar{y} = \infty$ the individual cannot push his motivation to infinity.

**Lemma 1**

1. **Goal setting increases the motivation for self 1 to exert effort, relative to having no goal.** High (‘difficult’) goal levels motivate more than low (‘easy’) ones:

$$\frac{\partial}{\partial a} (\beta \{g(\bar{e}, a) - g(e, a)\} - c) \begin{cases} > 0 & \text{for all } a \in (y, \bar{y}), \\ = 0 & \text{otherwise.} \end{cases}$$
2. There is a bound on the motivating force of goals:

\[
\lim_{a \to \infty} \{g(\bar{e}, a) - g(e, a)\} = \{g(\bar{e}, \bar{y}) - g(e, \bar{y})\} < \infty \quad \forall \bar{y} \in (0, \infty].
\]

In numerous studies of how exogenously given goals affect task performance, psychologists indeed document that ‘difficult’ goals induce better outcomes than ‘easy’ goals, or the absence of any goal (see e.g., the survey by Locke and Latham 1990). Our result matches this pattern by formalizing the psychological loss-aversion mechanism proposed by psychologists.

Can goal setting always help an individual get around his self-control problem? That is, does there exist for every possible present bias \(\beta \in [0, 1]\) a goal \(\hat{a} \geq y\) that makes the incentive constraint of self 1 bind: \(\beta (g(\bar{e}, \hat{a}) - g(e, \hat{a})) = c\)? All else equal, a more severe present bias reduces the incentives for self 1 to provide effort. Self 0 can counteract the resulting lack in motivation with a more difficult goal. But, as Lemma 1 shows, self 0 cannot push effort incentives arbitrarily high. This in turn implies that even a very large goal cannot motivate an individual with a severe present bias.

**Lemma 2** If the present bias is not too strong (\(\beta \geq \beta_{IC}\)), there exists a goal \(\hat{a} > y\) that motivates self 1, where \(\hat{a}\) is defined by \(\beta (g(\bar{e}, \hat{a}) - g(e, \hat{a})) = c\), and

\[
\beta_{IC} \equiv \frac{c}{g(\bar{e}, \bar{y}) - g(e, \bar{y})} < \beta_{SC}.
\]

The more severe the present bias (i.e., the lower \(\beta\)), the higher the goal level \(\hat{a}\) that is required to motivate self 1 to exert high effort: \(\frac{\partial \hat{a}}{\partial \beta} < 0\) for \(\hat{a} \in (y, \bar{y})\). If the present bias is too strong (\(\beta < \beta_{IC}\)), self-regulation with a goal is not feasible, i.e., there exists no goal \(a \geq y\) that can solve the self-control problem.
3.2 Goals are painful self-disciplining devices

Holding effort fixed, a higher goal increases the chance of falling short of the aspired outcome. This weighs more heavily than the chance of exceeding $a$, because losses loom larger than gains of equal size. An increase in the goal level therefore has a direct negative effect on the expected utility for self 0, i.e., it is “painful”. As Locke and Latham (1990, p.79) put it: “raising the goal level simply shifts the valence function to a higher plane; the individual must do more for less.”

Lemma 3 Goals are painful self-disciplining devices. Holding fixed the effort level, the expected utility of self 0 is decreasing in the goal level $a$: $\frac{\partial \{g(e,a)-c(e)\}}{\partial a} < 0$.

Clearly, the only purpose of setting a painful goal is to discipline self 1 to put in more effort than he would exert in the absence of a goal. An individual with no self-control problem, or an individual whose present bias is so severe that goals have no motivating force ($\beta < \beta_{IC}$) will never specify a painful goal. So in terms of goal setting, individuals with a severe present bias behave like people without a self-control problem. But while people without a self-control problem work hard, people with a severe present bias shirk.

What about an individual with a self-control problem, for whom self-regulation is in principle feasible ($\beta_{IC} \leq \beta < \beta_{SC}$)? Does he always want to set a goal, i.e., is he, in the words of the psychologists, willing to “accept” the goal $\hat{a} > y$ necessary to motivate self 1? Self 0 weighs the expected gains from higher effort by self 1 against the direct negative impact on self 0’s expected utility from goal $\hat{a}$:

$$\beta \{g(\bar{e}, \hat{a}) - c\} \geq \beta g(e, y).$$

(5)
Lemma 2 showed that the required goal for self-regulation increases with the severity of the present bias. But a higher goal is also more painful, making it less likely that Condition (5) is satisfied. Clearly, self 0 would always be willing to specify the goal required to motivate self 1 if Condition (5) held at the highest goal level that has incremental motivating force \((a = \bar{y})\), i.e., if \(g(\bar{e}, \bar{y}) - c \geq g(e, y)\). As we argue in the appendix, this holds if the gain in instrumental utility \(E(y|\bar{e}) - E(y|e)\) is large. Then the feasibility condition \(\beta \geq \beta_{IC}\) determines whether self-regulation will occur or not. If however the gain in instrumental utility is small, i.e., \(g(\bar{e}, \bar{y}) - c < g(e, y)\), limits on self-regulation may also arise because goals are painful self-disciplining devices. Hence, the feasibility constraint need not be what keeps the individual from engaging in self-regulation. The following result provides the formal conditions.

**Proposition 1**

1. If the gain in instrumental utility is large, i.e., \(g(\bar{e}, \bar{y}) - c \geq g(e, y)\), and \(\beta \geq \beta_{IC}\), then self 0 sets a painful goal \((\hat{a} > y)\) and self 1 provides high effort. If \(\beta < \beta_{IC}\), self 0 sets \(a = y\) and self 1 provides low effort.

2. If the gain in instrumental utility is small, i.e., \(g(\bar{e}, \bar{y}) - c < g(e, y)\), then there exists a cutoff \(\beta_0\) satisfying \(\beta_{IC} < \beta_0 < \beta_{SC}\), such that

   (a) if \(\beta \in [\beta_0, \beta_{SC})\), self 0 sets a painful goal \((\hat{a} > y)\) and self 1 provides high effort.

   (b) if \(\beta < \beta_0\), self 0 sets \(a = y\) and self 1 provides low effort, even though there exist goal levels that would motivate self 1 to exert high effort.

Part 2 illustrates that if the gain in instrumental utility is small, goals that motivate self 1 are only accepted if the present bias is not too severe \((\beta \geq \beta_0)\). Then the required goal is
not too difficult so that the expected gains from working hard and achieving the goal not only outweigh the expected losses from falling short of it, but also make the self-set goal more attractive than the alternative of having no goal and accepting that self 1 shirks. In contrast, individuals with a severe present bias ($\beta \in [\beta_{IC}, \beta_0]$) rationally choose not to set tough goals for themselves, even though self-regulation with a goal is feasible. The required goal is too painful to be accepted. So, consistent with the view in psychology (e.g., Hollenbeck and Klein 1987), our model predicts that goal acceptance moderates the goal-difficulty-motivation relation.

Part 1 shows that self-regulation is less painful and goals are more likely to be accepted if the task outcome provides a large instrumental utility gain. Still self-regulation with goals has its limits. Because the incentive effects of even very large goals are bounded, individuals with a very severe present bias ($\beta < \beta_{IC}$) cannot motivate themselves with a goal and thus abstain from goal setting.

Based on the comparison of parts 1 and 2, and because $\beta_{IC} < \beta_0$, our model predicts that additional instrumental utility gains can play an important role for people with a severe present bias. In the absence of these gains, people may not accept the goal that is necessary to motivate them and give up on self-regulation. Indeed, the psychology literature discusses that monetary incentives or public announcements of goals enhance goal acceptance (e.g., Locke 1968, Hollenbeck and Klein 1987, Hollenbeck et al. 1989). For example, the success of schemes such as Weight Watchers relies on concerns for reputation that arise from peer pressure, when goals are publicly announced. Commitment contracts are another example of shaping the instrumental utility related to task outcomes with the help of monetary incentives. One common form of such agreements are bets that a certain
outcome will be achieved. To provide a tool for engaging in such commitment contracts, the economists Ian Ayres, Dean Karlan and Jonathan Goldberg recently designed the website StickK.com. It allows an individual to specify a goal (e.g., lose 5kg of weight), a penalty for not reaching it (e.g., donate EUR 1,000 to charity X), and to designate a person who both verifies whether this goal was achieved and who enforces the penalty. Our instrumental utility component may encompass such monetary incentives, or concerns for reputation next to the direct benefits from a good performance (like improved health from a weight loss, better chances for a job with good grades) and can thereby explain why such instruments enhance goal acceptance.

4 Conclusion

This paper analyzes the process of self-regulation through self-set goals. We model an individual with a present bias that causes a self-control problem and show that he can enhance his motivation to work hard in the future by raising the goal against which performance will be measured. A challenging goal increases the psychological cost of shirking because the individual suffers a loss from falling short of the goal. What our model shows is that this can increase the motivation of the individual up to a certain point. Thus, goals can help some people engage in self-regulation. However, there are limits to self-regulation. Because goals are painful self-disciplining devices, the individual may rationally choose not to set a tough goal for himself and rather give up on self-regulation.
Appendix

Proof of Lemma \[1\]

The left-hand side of (4) is the part of the incentive constraint \(\beta \{g(\bar{e}, a) - g(e, a)\} - c \geq 0\) that depends on the goal level \(a\). Differentiating with respect to \(a\), we get

\[
\frac{\partial}{\partial a} (\beta \{g(\bar{e}, a) - g(e, a)\} - c) = \beta \eta (\mu - 1) \left[ F(a|\underline{e}) - F(a|\bar{e}) \right], \tag{6}
\]

which is strictly positive for \(a \in (\underline{y}, \bar{y})\) by Assumption \[1\] and equal to zero, otherwise. Specifically, if \(\bar{y}\) is bounded, increasing \(a\) beyond \(\bar{y}\) does not increase incentives, because \(F(a|\underline{e}) = F(a|\bar{e}) = 1\) for \(a \geq \bar{y}\). This yields the first part of the lemma.

We next show that \(\int_{\underline{y}}^{\bar{y}} [F(y|\underline{e}) - F(y|\bar{e})] \, dy\) is bounded for any \(\underline{y} \in [-\infty, 0]\), or \(\bar{y} \in (0, \infty]\).

\[
\int_{\underline{y}}^{\bar{y}} [F(y|\underline{e}) - F(y|\bar{e})] \, dy = \int_{0}^{\bar{y}} [1 - F(y|\bar{e})] \, dy - \int_{\underline{y}}^{0} F(y|\bar{e}) \, dy
\]

\[
= \left( \int_{0}^{\bar{y}} [1 - F(y|\bar{e})] \, dy - \int_{\underline{y}}^{0} F(y|\bar{e}) \, dy \right)
\]

\[
= E(y|\bar{e}) - E(y|\underline{e}) < \infty.
\]

where the third line follows from a standard rule to express expectations, which is applicable also for \(\bar{y} = \infty\) or \(\underline{y} = -\infty\) (see e.g., p.341 of Wolfstetter 2002). The final inequality follows from \(0 < E(y|e) < \infty\) and Assumption \[1\]. Using this bound in the incentive constraint, and the fact that for \(\bar{y} < \infty\)

\[
\lim_{a \to \infty} \int_{\underline{y}}^{a} [F(y|\underline{e}) - F(y|\bar{e})] \, dy = \int_{\underline{y}}^{\bar{y}} [F(y|\underline{e}) - F(y|\bar{e})] \, dy,
\]

we obtain the second part of the lemma:

\[
\lim_{a \to \infty} \{g(\bar{e}, a) - g(e, a)\} = \lim_{a \to \infty} \eta (\mu - 1) \int_{\underline{y}}^{a} [F(y|\underline{e}) - F(y|\bar{e})] \, dy + (\eta + 1) \{g(\bar{e}, \bar{y}) - g(e, \bar{y})\}
\]

\[
= (\eta \mu + 1) \{E(y|\bar{e}) - E(y|\underline{e})\} = g(\bar{e}, \bar{y}) - g(e, \bar{y}).
\]
Proof of Lemma 2

From Condition (SC) we know that $\beta \{ g(\bar{e}, y) - g(e, y) \} - c < 0$. Hence, $\hat{a} > y$. Lemma 1 shows that $\lim_{a \to \infty} \{ g(\bar{e}, a) - g(e, a) \} = \{ g(\bar{e}, \bar{y}) - g(e, \bar{y}) \}$ and that $\beta \{ g(\bar{e}, a) - g(e, a) \} - c$ is increasing in $a$ for $a \in (y, \bar{y})$. So if, for a given $\beta$, $\beta \{ g(\bar{e}, \bar{y}) - g(e, \bar{y}) \} - c > 0$, we know that there exists a unique $\hat{a} \in (y, \bar{y})$, such that $\beta \{ g(\bar{e}, \hat{a}) - g(e, \hat{a}) \} - c = 0$. Applying the implicit function theorem, we get

$$\frac{d \hat{a}}{d \beta} = -\frac{g(\bar{e}, y) - g(e, y) + \eta (\mu - 1) \int_y^{\hat{a}} [F(y|e) - F(y|\bar{e})] \, dy}{\eta (\mu - 1) [F(\hat{a}|e) - F(\hat{a}|\bar{e})]} < 0,$$

for $\hat{a} \in (y, \bar{y})$.

As shown above, the goal with maximum motivating force is $\bar{y}$. So the binding incentive constraint at $\hat{a} = \bar{y}$ defines implicitly $\beta_{IC}$: $\beta_{IC} (g(\bar{e}, \bar{y}) - g(e, \bar{y})) = c$. Because $\frac{d \hat{a}}{d \beta} < 0$, $\beta_{IC}$ is the minimum value of $\beta$ for which self-regulation with a goal is feasible.

Proof of Lemma 3

The expected utility of self 0, given goal $a$ and fixed effort $e$, is $\beta \{ g(e, a) - c(e) \}$. Differentiating (1) with respect to $a$ shows that

$$\frac{\partial \beta \{ g(e, a) - c(e) \}}{\partial a} = -\eta [1 + (\mu - 1) F(a|e)] < 0.$$

Proof of Proposition 1

Self 0 is ready to accept goal $a$ if it motivates high effort by self 1 and yields a higher expected utility relative to setting no goal and accepting that self 1 shirks, i.e., if $\Phi(a) \equiv$
\( g(\bar{e}, a) - c - g(e, y) \geq 0. \) From Lemma 3 it follows that \( \Phi(a) \) is a continuous and strictly decreasing function. Together with Lemma 1 this implies that self 0 will never set \( a > \bar{y} \) (the goal for which the maximum motivating force is achieved). Now, if \( \Phi(\bar{y}) \geq 0 \), the utility of self 0 from any goal \( a \in [y, \bar{y}] \) will be positive; so it follows from Lemma 2 and 3 that self 0 will set \( \hat{a} \) whenever \( \beta \geq \beta_{IC} \) (i.e., whenever self-regulation is feasible). Written out, one sees that the condition

\[
g(\bar{e}, \bar{y}) - c - g(e, y) \equiv \Phi(\bar{y}) = (1 + \eta \mu) E(y|\bar{e}) - (1 + \eta) E(y|e) - \eta (\mu \bar{y} - y) - c \geq 0, \tag{7}
\]

can be satisfied, for example, if \( E(y|\bar{e}) \) is sufficiently high relative to \( E(y|e) \). This shows the first part of Proposition 1.

The second part of Proposition 1 deals with the case where \( \Phi(\bar{y}) < 0 \). Then, as \( \Phi(a) \) is a continuous, strictly decreasing function and \( \Phi(y) > 0 \) (by Assumption 2), there exists a unique value \( a_0 \in (y, \bar{y}) \) such that \( \Phi(a_0) = 0 \). This is the maximum goal level that self 0 would ever choose to get self 1 to exert high effort.

The next argument shows that there exists a \( \beta_0 \in (\beta_{IC}, \beta_{SC}) \) such that self 1’s incentive constraint binds at \( a_0 \). Define \( \Delta(a, \beta) \equiv \beta \{g(\bar{e}, a) - g(e, a)\} - c \). Condition (SC) tells us that self 1’s incentive constraint binds if \( \beta = \beta_{SC} \) and \( a = y \): \( \Delta(y, \beta_{SC}) = 0 \). So, because \( \Delta(y, \beta_{SC}) \) is continuous and using the result in (6) that \( \frac{\partial \Delta(a_0, \beta)}{\partial a} > 0 \) for \( a \in (y, \bar{y}) \), we obtain that \( \Delta(a_0, \beta_{SC}) > 0 \). Similarly, from Lemma 2 we know that \( \Delta(\bar{y}, \beta_{IC}) = 0 \), so \( \Delta(a_0, \beta_{IC}) < 0 \). Using \( \frac{\partial \Delta(a_0, \beta)}{\partial \beta} = g(\bar{e}, a_0) - g(e, a_0) > 0 \) (implied by \( g(\bar{e}, \bar{y}) - g(e, y) > 0 \) together with Lemma 1), there exists a unique \( \beta_0 \in (\beta_{IC}, \beta_{SC}) \) such that \( \Delta(a_0, \beta_0) = 0 \). Thus, \( \beta_0 \) is the minimum value of the present-bias parameter \( \beta \) for which self 0 would set a goal to motivate self 1.

\[\text{\[11\]}\text{E.g., for } \bar{y} = \infty \text{ or } y = -\infty, \text{ we have } \Phi(\bar{y}) < 0, \text{ as substituting into (7) shows.}\]
References


