

IZA DP No. 3893

Self-Regulation through Goal Setting

Alexander K. Koch
Julia Nafziger

December 2008

Self-Regulation through Goal Setting

Alexander K. Koch

*University of Aarhus
and IZA*

Julia Nafziger

University of Aarhus

Discussion Paper No. 3893
December 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Self-Regulation through Goal Setting^{*}

Goals are an important source of motivation. But little is known about why and how people set them. 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. We show how goals permit self-regulation, but also that they are painful self-disciplining devices. Greater self-control problems therefore lead to stronger self-regulation through goals only up to a certain point. For severely present-biased preferences, the required goal for self-regulation is too painful and the individual rather gives up.

JEL Classification: A12, C70, D91

Keywords: goals, self-control, motivation, time inconsistency, psychology

Corresponding author:

Alexander K. Koch
School of Economics and Management
University of Aarhus
Building 1322
8000 Aarhus C
Denmark
E-mail: akoch@econ.au.dk

^{*} We thank Johan Lagerlöf for many helpful discussions, and are grateful to Mathias Dewatripont, Guido Friebel, Michael Naef and Luis Rayo for their comments.

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 (1990a) survey of the psychology literature on goals). The typical study takes goals as exogenously assigned and measures how such given goals affect performance in various tasks. However, little is known about why and how people set goals for themselves (see e.g. Bargh 1990, Carver and Scheier 1999, Oettingen, Pak, and Schnetter 2001). 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. The contribution of our paper is to show how the choice of goals can serve as an instrument for regulating own behavior, and how individuals will optimally choose such goals. We thereby also highlight the limits of self-regulation through goal setting: because goals are painful self-disciplining devices, individuals with a severe present bias may rationally choose not to set tough goals for themselves and rather give up on self-regulation.

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 bias* means that an individual cannot trust the way he will behave in the future, and thus creates a need for self-regulation. Suppose, for instance, a person finds working hard on a task scheduled in a week’s time is the optimal thing to do, when weighing today the future cost of effort against the future benefits from a higher task outcome. But once the date of the task arrives, the person may then actually decide to shirk rather than work hard. Such a preference reversal occurs because the costs of working hard become more salient when they are immediate and now loom larger than the gains, which only accrue in the future. In other words, the present bias creates an intrapersonal conflict of interest.¹

If the person attempts to influence his future behavior by simply telling himself ‘to do his best’ in a week’s time, then the future self will indeed choose the ‘best’ effort level – namely, to shirk. The problem is that what is best from the future self’s perspective is not what today’s self finds optimal. Hence, setting a naïve ‘do-your-best’ goal is not a successful self-regulation strategy. Consistent with this, Locke and Latham (1990a) conclude from their review of 183 studies on goal setting that ‘do-your-best’ goals lead to low performance.

¹There are other ways to model interpersonal conflicts and self-control problems: e.g., the “doer-planner model” (Thaler and Shefrin 1981), the “dual self model” (Fudenberg and Levine 2006), or the “temptation utility model” (Gul and Pesendorfer 2001).

People may nevertheless specify such goals, as they are not fully aware of their self-control problem – believing that their future selves have the same preferences over choices as they have today.

In contrast, 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 specific goals help such sophisticated individuals exert control over future actions.

A key property of such goals is that they affect individuals’ perceptions of future outcomes. In their surveys of the large body of empirical research carried out by psychologists, Locke and Latham state that “having a goal means to use it as a standard for evaluating one’s performance” (1990a, p.78), and 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” (Locke and Latham 2002, p.709-710). We take these stylized facts as starting points and model goals as reference points against which actual task outcomes are measured, assuming that people display *loss aversion* regarding goal achievement.² Hence, the process of goal setting leads to an endogenous reference point for task performance.³

We then show that setting a challenging goal for future performance makes future selves strive harder: they become motivated by the fear of falling short of the goal. Thus, an increase in the goal level set today allows an individual to raise his motivation to work hard in the future, by increasing the psychological cost of shirking. This result is in line with Locke and Latham’s (1990a) summary of evidence concerning the relationship between motivation and (exogenously assigned) goal difficulty: “subjects with specific, hard goals exerted more effort than those with less-difficult goals.” Our model derives formally their explanation that “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 that would be needed to motivate the future self – suggesting a monotone relationship between the self-set goal and the present bias. However, as our model shows, there are limits to self-regulation. The reason is that goals are painful

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.

³See Köszegi and Rabin (2006) for a different approach to modeling endogenous reference points. In their setting, the individual’s prior expectation about an outcome serves as the reference point. The idea that goals serve as reference points, which are in part actively chosen by the individual, can also be found in Falk and Knell (2004) (see below).

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 despite high effort. The process of goal setting hence trades-off anticipated psychological costs against benefits from self-regulation. The comparative statics of our model show that the optimal goal is non-monotonically related to the severity of the individual's present bias. For intermediate ranges of the present bias the individual compensates for a stronger bias with a higher goal. If however the individual faces a severe present bias, the goal required to sidestep the self-control problem becomes too painful. Such a person rationally gives up on self-regulation.

The paper is organized as follows. We next discuss the related literature. Section 2 introduces the model, which we analyze in Section 3. Section 4 concludes the paper. All proofs are in the appendix.

Related Literature

Goals have mainly been studied by psychologists (see references above), and we are aware of only one paper in the economics literature: Falk and Knell (2004) investigate in a social comparison model how the motives of *self-enhancement* (people use comparisons with others to make themselves feel better) and *self-improvement* (people use comparisons with others to improve their own performance) affect goal choice. The individual's reference standard for his personal outcome depends both on his chosen goal and the average outcome of others in the comparison group. In contrast to our setting, the individual does not display loss aversion regarding goal achievement. Instead, a higher goal lowers the costs of providing effort. Moreover, effort and the ability of an individual are complements in the production function. Falk and Knell show that the optimal goal is increasing in the ability of an individual – implying that people compare themselves to similar others. Our study is complementary to theirs as it highlights another important motivation for goal setting: the regulation of own behavior when people face self-control problems.

We therefore link the literature on goals with the behavioral economics literature on time-inconsistent preferences. Most closely related are theoretical contributions that deal with the question how present-biased individuals cope with self-control problems (for an overview see e.g. Brocas, Carrillo, and Dewatripont 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, Della Vigna and Malmendier 2004, Carrillo and Dewatripont

2008).⁴ From this literature Carrillo and Dewatripont (2008) is most closely related to our approach: they study how promises to other parties help overcome self-control problems. In their setting, breaking a promise leads to costs from a loss in reputation or other sanctions enforced by explicit contracts.⁵ 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.

While most of the literature focuses on external commitment strategies, a few contributions deal with intrapersonal strategies, as our paper. In a model where an individual has imperfect recall about past motives, and hence draws inference about these motives based on past actions (like living up to a *personal rule*), Bénabou and Tirole (2004) can explain why people do not abandon unpleasant personal rules, i.e. why internal commitment devices actually work. In contrast, 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. These papers thus are similar to our approach. 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. Carrillo and Mariotti (2000) and Bénabou and Tirole (2002) show that a hyperbolic discounting agent may rationally avoid information, because extra knowledge may move beliefs to a range where inefficient procrastination occurs.

2 Model

We consider the behavior of an individual with present-biased preferences who faces a self-control problem. At date 1, the individual chooses the effort to put into completing a task. The costs of effort are immediate, whereas the outcome (and the utility related to the outcome) realize only one period later, at date 2. When making the effort decision at date 1, the individual overemphasizes the immediate costs relative to the more distant benefit of a good task performance because he has a present bias. This bias can create an intrapersonal conflict of interest. At date 0, the individual may think that working hard on the task is

⁴There is also a growing empirical literature on this topic. For example, Gine, Karlan, and Zinman (2008) examine the effects of a commitment contract for smokers. Ashraf, Karlan, and Yin (2006) study the effects of a commitment savings product in the Philippines. Thaler and Benartzi (2004) design a savings product allowing costumers to commit to higher future savings. Newton *et al.* (1998) show that consumers strategically ration their purchase quantities of goods that are likely to be consumed on impulse. Ariely and Wertenbroch (2002) study the effects of deadlines on procrastination.

⁵The economists Ian Ayres, Dean Karlan and Jonathan Goldberg turned this idea into the website stickK.com. It allows people to make promises to relatives or friends, who are supposed to enforce these *commitment contracts* which put reputation or money at stake.

optimal. But then he might shirk 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.

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

Assumption 1 $F(y|\bar{e}) < F(y|\underline{e}) \quad \forall y \in (0, \bar{y})$.

Goal setting. At date 0, the individual chooses a goal a , which then becomes anchored his head as a standard for the future outcome y to be achieved. 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.” One of the findings in the psychology literature however is that such goals cannot be arbitrarily set, but must be “realistic” (see e.g. Locke and Latham 1990b). To capture bounds on goal setting imposed by Nature, we require that $a \in [0, \bar{y}]$. That is, the goal level must correspond to a possible outcome of the task.

There is extensive psychological evidence that people experience a loss if they fall short of a goal and a gain if they exceed it. The losses loom larger than gains of equal size. For example, Bandura (1989, p.1180) summarizes that “people seek self-satisfactions from fulfilling valued goals” and experience “discontent with substandard performances” (see also the introduction). We capture this with a value function in the spirit of Kahneman and Tversky (1979): depending on the difference between the outcome realization y and the goal a , the individual experiences a psychological gain $\mu^+(y - a)$ from satisfying the goal if $y \geq a$, and a psychological loss $\mu^-(a - y)$ from falling short of the goal if $y < a$. Following Köszegi and Rabin’s (2006) recent application of Kahneman and Tversky’s (1979) gain-loss utility framework to psychological utility, we assume that the individual exhibits linear loss aversion: $\mu^+(y-a) = \max\{y-a, 0\}$ and $\mu^-(y-a) = \mu \times \max\{a-y, 0\}$, with $\mu > 1$.

Present bias. The individual has time-inconsistent preferences in the sense of Strotz (1955). A present bias causes current payoffs to be more salient than future payoffs. Following the literature (e.g. Phelps and Pollak 1968, Laibson 1997, O’Donoghue and Rabin 1999a), we model this using (β, δ) -preferences (or *quasi-hyperbolic discounting*). At date t , the

individual discounts a future payoff at date $t + s$ ($s \geq 1$) at rate $\beta \delta^s$. The parameter δ corresponds to the standard exponential discount factor. For $\beta = 1$, the individual thus behaves as in a standard time-consistent discounting framework. For $\beta < 1$ the individual has a present bias, because the current period is overweighted relative to future periods. As β is the parameter of interest, without loss of generality, we will assume that $\delta = 1$. Hence, the expected utility of the individual at date $t \in \{0, 1\}$, U_t , is given by:

$$\begin{aligned} U_0 &= u_0 + \beta E[u_1 + u_2], \\ U_1 &= u_1 + \beta E[u_2], \end{aligned}$$

where u_t is the *instantaneous* utility in period t . While the incarnation of the individual at date 0 (*self 0*) weighs future utilities u_1 and u_2 equally, his date-1 incarnation (*self 1*) puts a larger relative weight on u_1 if $\beta < 1$.

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:

$$u_2(y, a) = v(y) + [\max\{y - a, 0\} - \mu \times \max\{a - y, 0\}].$$

It consists of two parts:⁶ First, a *consumption utility* or *instrumental utility* component, $v(y)$ ($v(y)' > 0$ and $v(y)'' < 0$). This corresponds to the traditional outcome-based utility in economic models. Second, the *psychological utility* component introduced above, which depends on the difference between the outcome realization y and the goal a .⁷

3 Analysis

3.1 Intrapersonal conflicts of interest

To understand why and how goals may act as a positive motivator, we show first when a conflict of interest between self 0 (the date-0 incarnation of the individual) and self 1 (the date-1 incarnation of the individual) arises in the absence of a positive goal level. For self 0 the choice of effort still lies in the future, and he therefore weighs equally the effort costs and the potential future benefits related to the task outcome. High effort is optimal from his perspective if the expected utility it yields ($\beta \{E[v(y)|\bar{e}] + E[y|\bar{e}] - c\}$) exceeds that from

⁶For terminology see Köszegi and Rabin (2006).

⁷Psychological utility is here defined over outcomes and not over actions (a similar approach is e.g. taken by Rayo and Becker 2007). This is analogous to the traditional outcome-based utility setting, in which people also care about outcomes rather than actions. Such a formulation arises naturally 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*. However, similar arguments as outlined in the following would apply if we assumed that the psychological utility was defined over actions and people set goals for actions.

low effort ($\beta \{E[v(y)|\underline{e}] + E[y|\underline{e}]\}$). That is, he wants his future self to work hard if and only if

$$\beta \{ \Delta E[v(y)] + \Delta E[y] - c \} \geq 0, \quad (1)$$

where $\Delta E[v(y)] \equiv E[v(y)|\bar{e}] - E[v(y)|\underline{e}]$ and $\Delta E[y] \equiv E[y|\bar{e}] - E[y|\underline{e}]$ denote the expected gains in instrumental utility and psychological utility, respectively.

Only in the case with no present bias ($\beta = 1$) does self 1 always choose exactly as self 0 would. Self 1 however overemphasizes immediate costs and rewards because $\beta < 1$. As a result, self 1 may prefer to shirk, while self 0 wants self 1 to work hard. To see this formally, note that for self 1 it is optimal to provide low effort if and only if

$$\beta \{ \Delta E[v(y)] + \Delta E[y] \} - c < 0. \quad (2)$$

Thus, whenever both (1) and (2) hold, there is a conflict of interest between self 0 and self 1 (this is often referred to as a *self-control problem* in the behavioral economics literature). This leads to the following definition:

Definition 1 *There is an intrapersonal conflict of interest, i.e. the individual faces a self-control problem, if and only if:*

$$\frac{c}{\beta} > \Delta E[v(y)] + \Delta E[y] > c. \quad (\text{CoI})$$

The condition (CoI) implies the existence of a threshold $\beta_1 = \frac{c}{\Delta E[v(y)] + \Delta E[y]}$ for a conflict of interest to arise: for a mild present bias, $\beta \geq \beta_1$ there is no intrapersonal conflict of interest. However, with a more severe bias, $\beta < \beta_1$ there is a self-control problem.

3.2 ‘Do-your-best’ goals and naïveté about own present bias

Before discussing how people set specific goals to motivate themselves, we briefly want to ask whether a much simpler ‘do-your-best’ goal can also serve as a source of motivation. Here self 0 simply tells self 1 to do his best, without specifying a concrete positive goal level a . In response, self 1 indeed does what is best for him and maximizes utility from his date-1 perspective. But whenever Condition (CoI) holds, the best effort choice of self 1 does not match the best effort choice from the perspective of self 0. Thus, such a goal is as good as no goal, and fails to address intrapersonal conflicts of interest. While the argument is simple, it helps understand the evidence from the psychology literature that ‘do-your-best’ goals often fail to motivate people and tend to lead to low performance (e.g. Locke and Latham 1990a).

Lemma 1 *‘Do-your-best’ goals fail to motivate high effort if there is an intrapersonal conflict of interest, i.e. if Condition (CoI) holds.*

Why do people nevertheless set ‘do-your-best’ goals that do not help overcome self-control problems? Setting such goals comes at no costs (they are not painful, in contrast to the specific goals that we consider in the next section). Also, individuals often are not fully aware of their time-inconsistent preferences, i.e. are *naïve* in the sense of O’Donoghue and Rabin (1999a). Self 0 may think that self 1 attaches the same relative weights to future outcomes as he does today. In such cases, self 0 believes that the optimal effort choice from his perspective corresponds to the best effort choice that self 1 will make.

3.3 Specific goals as a source of motivation

We now move to the goal setting problem for an individual who is aware of his present bias and correctly anticipates his future behavior, i.e. who is *sophisticated* in the sense of O’Donoghue and Rabin (1999a). To analyze this problem we start with the effort choice made by the incarnation of the individual at date 1, and ask how self 1 responds to a specific goal a . We then determine by backward induction the optimal goal that self 0 will set.

Given the goal a , self 1 decides on his effort $e \in \{e, \bar{e}\}$, maximizing his expected utility

$$U_1(e|a; \beta) \equiv -c(e) + \beta \left\{ E[v(y)|\bar{e}] + \mu \int_0^a (y - a) f(y|\bar{e}) dy + \int_a^{\bar{y}} (y - a) f(y|\bar{e}) dy \right\}. \quad (3)$$

High effort ($e = \bar{e}$) is optimal for self 1 if only if the expected gain in date-2 utility (discounted by the present-bias parameter β) outweighs the immediate cost of effort. In other words, the following ‘incentive constraint’ must be satisfied:

$$\Phi(a; \beta) \equiv U_1(\bar{e}|a; \beta) - U_1(e|a; \beta) \geq 0. \quad (4)$$

As we show in the appendix, the left hand side is increasing in the goal level a , i.e. $\frac{\partial \Phi(a; \beta)}{\partial a} > 0$. Thus, a ‘difficult goal’ is a better motivator than an ‘easy goal’. This might seem counterintuitive at first, because a higher goal also raises the chance that the individual ends up falling short of the aspired level a , even though he worked hard. However, the incentive constraint tells us that the individual is *even more* likely to fall short of the goal a if self 1 shirks. This is the channel through which a higher (more difficult) goal increases the incentives to work hard. We summarize this in the following lemma:

Lemma 2 *Goals that specify a specific aspiration level for the outcome increase the motivation for self 1 to exert effort. High (‘difficult’) goal levels more so than low (‘easy’) ones or no goals, i.e. $\frac{\partial \Phi(a; \beta)}{\partial a} > 0$.*

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 1990a). Our above result matches this pattern by formalizing the psychological loss-aversion mechanism proposed by

psychologists for the functioning of goals.

Because a higher goal increases the motivation of self 1 to provide effort (Lemma 2), it can serve as a tool for self-regulation and potentially help get around intrapersonal conflicts of interest. How does the goal required for self-regulation relate to the extent of the individual's present bias? All else equal, a lower β , i.e. a more severe present bias, reduces the incentives for self 1 to provide effort. Self 0 can counteract a lack in motivation arising from this with a more difficult goal. Thus, we expect a positive relationship between the goal level that is necessary to motivate self 1 in the presence of a conflict of interest, and the severity of the present bias. To see this formally, denote by \hat{a} the minimum goal level that makes the incentive constraint (4) hold with equality: $\Phi(\hat{a}; \beta) = 0$. The following lemma shows how \hat{a} varies with the severity of the present bias, captured by the parameter β :

Lemma 3 *Suppose an individual faces an intrapersonal conflict of interest, i.e. Condition (CoI) holds. Then, the more severe the present bias (i.e. the lower β), the higher the goal level \hat{a} that is required to motivate self 1 to exert high effort, i.e. $\frac{\partial \hat{a}}{\partial \beta} > 0$. If the present bias is too strong ($\beta < \underline{\beta}$) no 'realistic' goal ($a \in [0, \bar{y}]$) can help get around the intrapersonal conflict of interest, where*

$$\underline{\beta} \equiv \frac{c}{\Delta E[v(y)] + \mu \Delta E[y]}.$$

The result shows that an individual needs to respond to a more severe self-control problem with a more ambitious goal to motivate his future self. The scope for this however has its limits, because according to the evidence from the psychology literature only 'realistic' goals lead to the psychological gain-loss effects at the heart of our model. But these limits on goal setting are not the constraints that prevent an individual from self-regulating. Instead, limits arise endogenously from goal being painful self-disciplining devices as we show in the next section, where we ask about possible costs from goal setting.

3.4 Goals are painful self-disciplining devices

Faced with a self-control problem, self 0 may attempt to motivate his future self to choose high effort by setting himself a goal a . To see the possible costs from setting such a goal, suppose that the future effort level e was fixed (i.e. independent of the chosen goal) and consider the expected utility of self 0 with goal level a . With slight abuse of notation, this expected utility is:

$$U_0(a|e) = \beta \left\{ \int_0^{\bar{y}} v(y) f(y|e) dy + \mu \int_0^a (y - a) f(y|e) dy + \int_a^{\bar{y}} (y - a) f(y|e) dy - c(e) \right\}. \quad (5)$$

Holding the effort level e fixed, a higher goal increases the chance of falling short of the aspired outcome a . This weighs more heavily than the chance of exceeding a , because losses loom larger than gains of equal size ($\mu > 1$). An increase in the goal level therefore has

a direct negative effect on the expected utility for self 0. As the psychologists Locke and Latham (1990a, p.79) put it: “raising the goal level simply shifts the valence function to a higher plane; the individual must do more for less.” A higher goal level thus is painful because it reduces the expected utility resulting from a given effort level.

Lemma 4 *Goals are painful self-disciplining devices: $\frac{\partial U_0(a|e)}{\partial a} < 0$.*

So the purpose of setting $a > 0$ is to discipline self 1 to put in more effort than he would exert in the absence of a positive goal. An individual faced with an intrapersonal conflict of interest therefore needs to determine in a first step the minimum goal level \hat{a} required to motivate self 1 to work hard. In a second step, self 0 then weighs the expected gains from higher effort by self 1 against the direct negative impact on self 0’s expected utility from the goal \hat{a} . In other words, self 0’s expected utility with goal \hat{a} and high effort in the future has to exceed the utility from low effort in the future and $a = 0$:

$$U_0(\hat{a}|\bar{e}) \geq U_0(0|\underline{e}). \quad (6)$$

Lemma 3 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 (6) is satisfied. So will an individual with an intrapersonal conflict of interest ($\beta < \beta_1$) ever set a positive goal for himself? And does the constraint that goals have to be realistic ever bind and prevent self-regulation? Clearly, the more important the gain in instrumental utility $\Delta E[v(y)]$ is, the more likely it is that (6) holds, and the lower $\underline{\beta}$ from Lemma 3. It is thus illuminating to study the above questions for the case where $\Delta E[v(y)] = 0$, because this reduces to a minimum the scope and desirability of self-regulation through goal setting. The next proposition summarizes our findings.

Proposition 1 *Suppose that the task offers potential psychological rewards exceeding the cost of effort ($\Delta E[y] > c$), but no additional instrumental utility ($v(y) = 0$). Then,*

- i) if the individual has no intrapersonal conflict of interest ($\beta \geq \beta_1$), self 0 sets no goal ($a = 0$) or a ‘do-your-best’ goal, and self 1 exerts high effort, where $\beta_1 \equiv c/\Delta E[y]$.*
- ii) there exists a cutoff β_0 satisfying $\frac{1}{\mu}\beta_1 < \beta_0 < \beta_1$, such that for $\beta \in (\beta_0, \beta_1)$ self 0 sets a specific goal ($a > 0$) and self 1 provides high effort.*
- iii) if $\frac{1}{\mu}\beta_1 \leq \beta < \beta_0$, self 0 sets $a = 0$ and self 1 provides low effort, even though there exist ‘realistic goal levels $a \in (0, \bar{y}]$ that would motivate self 1 to exert high effort.*
- iv) if $\beta < \frac{1}{\mu}\beta_1$, self-regulation is not possible because there exists no ‘realistic goal level that motivates self 1 to exert high effort; this however is never a binding constraint.*

The first part of the proposition is straightforward: there is no reason to set a painful goal if there is no self-control problem to solve. If however faced with an intrapersonal conflict of interest, self 0 weighs the benefits of self-regulation against the expected psychological costs. Goals that motivate self 1 to work hard pay off if the present bias is not too severe ($\beta \geq \beta_0$). For this parameter range, the required goal is sufficiently low 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 living with self 1 shirking.

For a stronger present bias, which is not so severe that self-regulation no longer would be feasible ($\beta \in [\beta_1/\mu, \beta_0]$), motivating self 1 to work hard requires a very ambitious goal. This leaves only a small chance of obtaining psychological satisfaction from exceeding the goal, but a high chance of not meeting the aspired outcome level and suffering a psychological loss, which also hurts self 0. Hence, self 0 prefers not to impose these psychological costs on himself, even though they would allow him to get around his self-control problem and motivate self 1 to work hard.

For $\beta < \beta_1/\mu$, no admissible goal would motivate self 1 to work hard. However, the constraint that goals have to be ‘realistic’ is not what keeps the individual from engaging in self-regulation, as the preceding argument showed. Limits on self-regulation through goal setting arise because goals are painful self-disciplining devices. Individuals with a severe present bias therefore rationally choose not to set tough goals for themselves, even if this were feasible, and rather give up on self-regulation.

Thus, in terms of goal setting, sophisticated individuals with a severe present bias behave like people without a self-control problem (i.e. people with a mild or no present bias) or naïve individuals: they choose no goal. But while people without a self-control problem will work hard once faced with the task, people with a severe present bias will shirk – as will those who abstained from goal setting out of naïveté.

For the case where the task outcome also provides an instrumental utility gain $\Delta E[v(y)]$ there are two effects. First, the range of parameters β for which an intrapersonal conflict of interest exists is smaller (i.e. β_1 is lower). Second, if a self-control problem exists, self-regulation is less painful, so a positive goal is more likely to be optimal (pushing β_0 down). Thus, our model predicts that additional instrumental utility gains $\Delta E[v(y)]$ related to the task outcome can play a particularly important role for people with severe time-inconsistency problems. In the absence of these gains they would otherwise not set goals for themselves and give up on self-regulation.

Next to direct benefits from a good performance (like improved health from a weight loss, better chances for a job with good grades), instrumental utility gains might also encompass monetary incentives, or concerns for reputation. For example, the success of schemes such

as e.g. Weight Watchers relies on concerns for reputation that arise from peer pressure when goals are publicly announced. Regarding monetary incentives, Locke and Latham (1990b, p.240) observe that “[m]onetary incentives strengthen goal commitment, providing people value money, the amount of money is sufficiently large and the incentives are not tied to goals perceived as impossible”. 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 enforces the penalty.

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. A premise of our model is that people have the capacity to set goals for themselves that remain meaningful over time – a fact that is well documented in the psychology literature. What our model shows is that such goals help some people to engage in self-regulation. However, there are limits to self-regulation even if an individual can commit to his goal: 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 Lemma 2.

The incentive constraint for self 1 (4) can be rewritten as

$$\beta \left\{ \Delta E[v(y)] + \Delta E[y] + (\mu - 1) \left(\int_0^a y [f(y|\bar{e}) - f(y|\underline{e})] dy + a [F(a|\underline{e}) - F(a|\bar{e})] \right) \right\} \geq c. \quad (7)$$

The higher the goal a , the more likely it is that the incentive constraint is satisfied:

$$\frac{\partial \Phi(a; \beta)}{\partial a} = \beta (\mu - 1) [F(a|\underline{e}) - F(a|\bar{e})] > 0 \quad \forall a \in (0, \bar{y}), \quad (8)$$

by Assumption 1. ■

Proof of Lemma 3.

As $\beta < 1$ it follows from Condition (CoI) and (8) that $\hat{a} > 0$. The binding incentive constraint (7) defines \hat{a} for a given β . Applying the implicit function theorem we get

$$\frac{d\hat{a}}{d\beta} = - \frac{\Delta E[v(y)] + \Delta E[y] + (\mu - 1) \left(\int_0^{\hat{a}} y [f(y|\bar{e}) - f(y|\underline{e})] dy + \hat{a} [F(\hat{a}|\underline{e}) - F(\hat{a}|\bar{e})] \right)}{(\mu - 1) [F(\hat{a}|\underline{e}) - F(\hat{a}|\bar{e})]} < 0,$$

for $\hat{a} \in (0, \bar{y})$. For the second part, evaluate the binding incentive constraint (7) at the maximum admissible goal level \bar{y} and rearrange terms. ■

Proof of Lemma 4.

Holding effort constant and differentiating Equation (5) with respect to a gives

$$\frac{\partial U_0(a|e)}{\partial a} = - [1 + (\mu - 1) F(a|e)] < 0.$$

Combining this insight with Lemma 2 yields the result. ■

Proof of Proposition 1.

i) From Lemma 4 we know that self 0 never sets a in excess of the level \hat{a} required to incentivize his future self. Denote by β_1 the value of the present-bias parameter β for which, given $a = 0$, self 1 would be just indifferent between exerting effort or not. Condition (CoI) and $\Delta E[v(y)] = 0$ imply that $\beta_1 \equiv c/\Delta E[y]$. The result follows because \hat{a} decreases with β (Lemma 3).

iv) Follows from Lemma 3, as $\Delta E[v(y)] = 0$ implies that $\underline{\beta} = \frac{1}{\mu} \beta_1$.

ii) and iii) We have $U_0(0|\bar{e}) - U_0(0|\underline{e}) > 0$ because $\Delta E[y] > c$. Moreover,

$$U_0(\bar{y}|\bar{e}) - U_0(0|\underline{e}) = \beta \{ \mu [E(y|\bar{e}) - \bar{y}] - E(y|\underline{e}) - c \} < 0.$$

From the proof of Lemma 4 it follows that $U_0(a|\bar{e}) - U_0(0|\underline{e})$ is a continuous and strictly decreasing function. Thus, by the intermediate value theorem, there exists a unique value

$a_0 \in (0, \bar{y})$ such that $U_0(a_0|\bar{e}) - U_0(0|\underline{e}) = 0$. 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 $\beta_0 \in (\underline{\beta}, \beta_1)$ such that self 1's incentive constraint holds at a_0 . We make use of $\Phi_a > 0$ (see Equation (8)). From i) we know that self 1's incentive constraint (4) binds at β_1 and $a = 0$: $\Phi(0; \beta_1, 0) = 0$. So, $\Phi(a_0; \beta_1, 0) > 0$. Similarly, from Lemma 3 we know that $\Phi(\bar{y}; \underline{\beta}, 0) = 0$, so $\Phi(a_0; \underline{\beta}, 0) < 0$. Using $\Phi_\beta > 0$, by the intermediate value theorem, there exists a unique $\beta_0 \in (\underline{\beta}, \beta_1)$ such that $\Phi(a_0; \beta_0, 0) = 0$. Thus, β_0 is the *minimum* value of the present-bias parameter β for which self 0 would set a goal to incentivize self 1. ■

References

- AINSLIE, G., AND N. HASLAM (1992): “Hyperbolic Discounting,” in *Choice over Time*, ed. by G. Loewenstein, and J. Elster, pp. 57–92. Russell Sage Foundation, New York.
- ARIELY, D., AND K. WERTENBROCH (2002): “Procrastination, Deadlines, and Performance: Self-control by Precommitment,” *Psychological Science*, 13, 219–224.
- ASHRAF, N., D. KARLAN, AND W. YIN (2006): “Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines,” *Quarterly Journal of Economics*, 121(2), 635–672.
- BANDURA, A. (1989): “Human Agency in Social Cognitive Theory,” *American Psychologist*, 44(9), 1175–1184.
- BARGH, J. A. (1990): “Auto-motives: Preconscious Determinants of Social Interaction,” in *Handbook of Motivation and Cognition: Foundations of Social Behavior*, ed. by E. T. Higgins, and R. M. Sorrentino, vol. 2, pp. 93–130. Guilford Press, New York.
- BAUMEISTER, R. F., AND T. F. HEATHERTON (1996): “Self-regulation Failure: An Overview,” *Psychological Inquiry*, 7(1), 1–15.
- BÉNABOU, R., AND J. TIROLE (2002): “Self-confidence and Personal Motivation,” *Quarterly Journal of Economics*, 117(3), 871–915.
- (2004): “Willpower and Personal Rules,” *Journal of Political Economy*, 112(4), 848–886.
- BENHABIB, J., AND A. BISIN (2005): “Modeling Internal Commitment Mechanisms and Self-control: A Neuroeconomics Approach to Consumption-Saving Decisions,” *Games and Economic Behavior*, 52(2), 460–492.
- BROCAS, I., J. D. CARRILLO, AND M. DEWATRIPONT (2004): “Commitment Devices under Self-control Problems: An Overview,” in *Reasons and Choices*, ed. by I. Brocas, and J. D. Carrillo, vol. 2 of *The Psychology of Economic Decisions*, chap. 4, pp. 49–66. Oxford University Press, Oxford.
- CARRILLO, J. D. (2005): “To be Consumed with Moderation,” *European Economic Review*, 49(1), 99–111.
- CARRILLO, J. D., AND M. DEWATRIPONT (2008): “Promises, Promises, ...,” *Economic Journal*, 118, 1453–1473.
- CARRILLO, J. D., AND T. MARIOTTI (2000): “Strategic Ignorance as a Self-disciplining Device,” *Review of Economic Studies*, 67(3), 529–544.

- CARVER, C. S., AND M. F. SCHEIER (1999): “Themes and Issues in the Self-regulation of Behavior,” in *Advances in Social Cognition*, ed. by R. S. Wyer, vol. 12, pp. 1–105. Erlbaum, Mahwah, NJ.
- DELLA VIGNA, S., AND U. MALMENDIER (2004): “Contract Design and Self-control: Theory and Evidence,” *Quarterly Journal of Economics*, 119(2), 353–402.
- DIAMOND, P., AND B. KÖSZEGI (2003): “Quasi-hyperbolic Discounting and Retirement,” *Journal of Public Economics*, 87(9-10), 1839–1872.
- ELSTER, J. (2000): *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge University Press, Cambridge, MA.
- FALK, A., AND M. KNELL (2004): “Choosing the Joneses: Endogenous Goals and Reference Standards,” *Scandinavian Journal of Economics*, 106(3), 417–435.
- FUDENBERG, D., AND D. K. LEVINE (2006): “A Dual-self Model of Impulse Control,” *American Economic Review*, 96(5), 1449–1476.
- GINE, X., D. KARLAN, AND J. ZINMAN (2008): “Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation,” mimeo, Yale University.
- GOLLWITZER, P. M. (1999): “Implementation Intentions: Strong Effects of Simple Plans,” *American Psychologist*, 54(7), 493–503.
- GUL, F., AND W. PESENDORFER (2001): “Temptation and Self-control,” *Econometrica*, 69(6), 1403–1435.
- HEATH, C., R. P. LARRICK, AND G. WU (1999): “Goals as Reference Points,” *Cognitive Psychology*, 38, 79–109.
- HERWEG, F., AND D. MÜLLER (2008): “Performance of Procrastinators: On the Value of Deadlines,” mimeo, University of Bonn.
- KAHNEMAN, D., AND A. TVERSKY (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47(2), 263–292.
- KÖSZEGI, B., AND M. RABIN (2006): “A Model of Reference-dependent Preferences,” *Quarterly Journal of Economics*, 121(4), 1133–1165.
- LAIBSON, D. (1997): “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics*, 112(2), 443–477.
- LOCKE, E. A., AND G. P. LATHAM (1990a): *A Theory of Goal Setting & Task Performance*. Prentice Hall, Englewood Cliffs, NJ.

- (1990b): “Work Motivation and Satisfaction: Light at the End of the Tunnel,” *Psychological Science*, 1(4), 240–246.
- (2002): “Building a Practically Useful Theory of Goal Setting and Task Motivation,” *American Psychologist*, 57(9), 705–717.
- NEWTON, H. N., C. BARNHART, P. H. VANCE, AND K. WERTENBROCH (1998): “Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice,” *Marketing Science*, 17(4), 317–337.
- O’DONOGHUE, T., AND M. RABIN (1999a): “Doing It Now or Later,” *American Economic Review*, 89(1), 103–124.
- (1999b): “Incentives for Procrastinators,” *Quarterly Journal of Economics*, 114(3), 769–816.
- OETTINGEN, G., H.-J. PAK, AND K. SCHNETTER (2001): “Self-regulation of Goal Setting: Turning Free Fantasies about the Future into Binding Goals,” *Journal of Personality and Social Psychology*, 80(5), 736–753.
- PHELPS, E. S., AND R. A. POLLAK (1968): “On Second-best National Saving and Game-equilibrium Growth,” *Review of Economic Studies*, 35(2), 185–199.
- RAYO, L., AND G. S. BECKER (2007): “Evolutionary Efficiency and Happiness,” *Journal of Political Economy*, 115(2), 302–337.
- STROTZ, R. H. (1955): “Myopia and Inconsistency in Dynamic Utility Maximization,” *Review of Economic Studies*, 23(3), 165–180.
- THALER, R. H., AND S. BENARTZI (2004): “Save More Tomorrow (TM): Using Behavioral Economics to Increase Employee Saving,” *Journal of Political Economy*, 112(1), 164–187.
- THALER, R. H., AND H. M. SHEFRIN (1981): “An Economic Theory of Self-control,” *Journal of Political Economy*, 89(2), 392–406.