DISCUSSION PAPER SERIES

IZA DP No. 4049

Commitment to Self-Rewards

Alexander K. Koch
Julia Nafziger

February 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Commitment to Self-Rewards

Alexander K. Koch
Aarhus University
and IZA

Julia Nafziger
Aarhus University

Discussion Paper No. 4049
February 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 4049
February 2009

ABSTRACT

Commitment to Self-Rewards*

Self-administered rewards are ubiquitous. They serve as incentives for personal accomplishments and are widely recommended as tools for overcoming self-control problems. However, it seems puzzling why self-rewards can work: the prospect of a reward has a motivating force only if the threat of self-denial of the reward after low performance is credible. We explain how a rational forward-looking individual may achieve commitment to self-rewards, by applying Köszegi and Rabin's (2006) model of endogenous reference point formation to a self-regulation problem. Our results show why and when self-regulation built on self-rewards can be successful and thus illustrate the power, but also the limits, of self-rewards.

JEL Classification: A12, C70, D86

Keywords: self-control, motivation, self-reinforcement, time inconsistency, reference-dependent preferences

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 are grateful to Mathias Dewatripont, Ott Toomet, and Frederic Warzynski for helpful comments and discussions.

1 Introduction

The pleasure of the moment often seduces us to act against our long run interests: we are tempted to shirk on unpleasant tasks – such as studying for an exam, writing a report, dieting, or saving money. To increase our motivation we frequently promise ourselves a reward if we persist and do accomplish a particular task. We might, for instance, tell ourselves "If I write three pages for this report, I'll buy a coffee". But it seems puzzling why such a promise can work without external enforcement: the prospect of a contingent reward can only have a motivating force if the threat of withholding it is credible. Attempting to motivate myself with a coffee is futile if I know that, should I fail to write anything, I will still buy the coffee.

Nevertheless, self-administered rewards (*self-rewards*) are ubiquitous.[1] Moreover, they are recommended in almost any related self-help guide and figure prominently in the professional treatment of problem behaviors (see e.g. Bandura 1971, 1976, 1986, 2005, Locke, Latham and Erez (1988), Febbraro and Clum 1998, Faber and Vohs 2004, Clum and Watkins 2007).[2] This suggests that self-reward strategies must have some power. So why and when is self-regulation built on self-rewards successful? And what determines whether commitment to self-rewards is possible? The aim of our paper is to shed light on these enduring questions by applying Köszegi and Rabin's (2006) model of endogenous reference point formation to a self-regulation problem.

The need for self-regulation arises from a present-bias (Strotz 1955, Laibson 1997) that creates a self-control problem.[3] From today's perspective the individual finds it optimal to complete a task scheduled in the future. But once the date of the task arrives, she will be tempted to shirk: all else equal, the current cost of effort looms larger (due to the present bias) than the future benefit from task completion. Anticipating her self-control problem, the individual attempts to regulate her future behavior by making a noncommittal promise to herself of a state-contingent self-reward, like buying something nice if (and only if) the task is completed.

Given this promise, buying the good after task completion is something the individual expected to happen; whereas buying it when not completing the task comes unexpectedly. And

[1] The consumer research literature documents the wide-spread "therapeutic" use of *self-gifts* as incentives for personal accomplishments (e.g. Mick and DeMoss 1990, Mick and Faure 1998, Kivetz and Simonson 2002). Such self-gifts "entail fulfilling an explicit or implicit self-contract in order to approve consumption indulgence" (Mick and DeMoss 1990, p.326).

[2] Psychologists refer to such self-regulation strategies as *self-reinforcement* (Bandura 1971, 1976).

[3] There are alternative ways to model intrapersonal conflicts and self-control problems: e.g., Thaler and Shefrin's (1981) *doer-planner model*, Fudenberg and Levine's (2006) *dual self model*, or Gul and Pesendorfer's (2001) *temptation utility model*.

whether expectations are met or not matters: according to Köszegi and Rabin (2006, 2007) past expectations become reference points. When evaluating an outcome relative to such a reference point, people often display loss-aversion in the sense of Kahneman and Tversky (1979).[4] However, the reference point of an individual is not arbitrary: for an individual who is rational and forward-looking it must indeed be optimal (not) to buy the good if she expected before (not) to buy it. That is, actions and expectations must constitute a *personal equilibrium* (Köszegi and Rabin 2006, Köszegi 2007).

To illustrate the consequences of these ideas for self-regulation that this paper develops, consider a self-reward strategy of the form "If I complete the task, I'll buy that nice pair of shoes I saw in the shop window; but if I shirk I will deny myself the shoes." If the price of the shoes is very low, the individual will always buy them – no matter what her past expectations were: the increase in consumption utility outweighs even a psychological loss from spending an unexpected amount of money. Hence, the part of the self-promise "...if I shirk I will deny myself the shoes" is not credible. But then the self-regulation strategy unravels in a personal equilibrium: the individual rationally expects that she will buy the shoes, irrespective of the task outcome. She thus looks forward to the same continuation utility, whether she works on the task or not. This leaves the self-control problem of the individual unresolved, and she will therefore not complete the task. Conversely, if the price of the shoes is very high, she knows that she will never buy them – so the part of the self-promise "If I complete the task, I'll buy that nice pair of shoes..." is not credible. Again this makes it impossible for the individual to rationally expect anything else for the future. In both of these cases, self-rewards are not credible and self-regulation is doomed to fail.

Are promises of self-rewards thus a futile attempt to regulate own behavior? The answer is no, or more precisely, it depends. For an intermediate price range, buying as well as not buying can be part of a personal equilibrium – similar to what Köszegi and Rabin (2006) show: if the individual expects to buy the shoes she will buy them; if she expects not to buy the shoes she will not buy them. The individual can exploit these multiple equilibria for the purposes of self-regulation, because they allow her to believe in the promise that she will buy the shoes upon task completion, but would deny herself the shoes if she failed to work on the task.[5] Thus, the self-reward is credible if its price lies in this intermediate range, i.e. if the good involved is neither too extravagant nor too much of a bargain or a necessity.

[4]Köszegi and Rabin (2007, p.1048) illustrate this idea as follows: "An employee who had expected a $50,000 salary will assess a salary of $40,000 as a loss, and a taxpayer who had expected to pay $30,000 in taxes will treat a $20,000 tax bill as a gain." For a discussion of evidence that expectations-based counterfactuals influence how people react to outcomes see Köszegi and Rabin (2006, 2007).

[5]A distinguishing feature from other models where an individual chooses *optimal beliefs* or has *motivated beliefs* (e.g. Akerlof and Dickens 1982, Landier 2000, Brunnermeier and Parker 2005, Gollier and Muermann 2006, Bénabou and Tirole 2007, Brunnermeier, Gollier and Parker 2007) is that beliefs in a personal equilibrium are rational expectations of future outcomes.

Credibility however is not enough. To help overcome the self-control problem the reward must also provide appropriate incentives: completing the task must increase the continuation utility enough to offset the temptation to shirk. This places an additional constraint on the price of the good. It must be sufficiently low, so that buying the good upon task completion is enough of a reward to motivate the individual to put in the required effort.
The more difficult the task is, or the stronger the present bias is, the lower the price of the (given) reward good must be in order to satisfy the incentive constraint, as the comparative statics of our model show. But then this constraint may conflict with the requirement that the self-reward must be credible: namely, if the price of the good is very low, the individual will always buy it, irrespective of her past expectations. In other words, it might happen that whenever the reward is sufficiently powerful to provide incentives, the threat of self-denial is not credible – and whenever self-denial is credible, the reward is not powerful enough. In these cases, self-regulation based on the promise to consume the available reward good only upon task-completion is doomed to fail.
In contrast, self-regulation works if both the credibility and incentive constraints are met. In such a case, the *preferred personal equilibrium* involves a self-reward coupled with rational expectations of the individual that she will complete the task and buy the good upon task completion, but would deny herself the good if she failed to work on the task. Our results therefore not only offer an explanation for the wide-spread use and success of self-reward strategies, but they also help understand possible limits to self-regulation.

Related literature

Our main contribution is to the literature that deals with the question of how present-biased individuals cope with self-control problems (for overviews see e.g. Elster 2000, Brocas, Carrillo and Dewatripont 2004). A large body of work focuses on the role of external commitment technologies. It explains why people incur costs – for example by investing in illiquid assets, signing binding contracts, or making binding promises to other parties – in order to overcome self-control problems in savings, consumption and retirement decisions (e.g. Laibson 1997, Diamond and Köszegi 2003, Carrillo 2005), or to overcome low effort provision and procrastination (e.g. O'Donoghue and Rabin 1999b, DellaVigna and Malmendier 2004, Carrillo and Dewatripont 2008). The growing strand of empirical contributions on this topic shows, for example, that consumers strategically ration their purchase quantities of goods that are likely to be consumed on impulse (Newton *et al.* 1998), and examines the effects of commitment contracts for smokers (Gine, Karlan and Zinman 2008), commitment savings products (Thaler and Benartzi 2004, Ashraf, Karlan and Yin 2006), as well as deadlines (Ariely and Wertenbroch 2002).
Only a few papers deal with intrapersonal strategies, as we do here. Most of them assume the presence of an internal commitment device – such as a goal, a deadline, or some control

process – and ask how people design and use such devices optimally to regulate behavior (Benhabib and Bisin 2005, Herweg and Müller 2008, Koch and Nafziger 2008). Our paper endogenously derives conditions under which individuals can achieve internal commitment. It is thus most closely related to Bénabou and Tirole (2004), who study personal rules – another important motivator mentioned in the psychology literature next to self-rewards. In their model, individuals have imperfect recall about past motives, and hence draw inference about these motives based on their past actions (like living up to a personal rule in a situation that puts their willpower to a test). The gain from maintaining a high self-reputation is what can motivate an individual to stick to a personal rule even though the task turned out to be unexpectedly unpleasant (e.g. the individual underestimated how strong the craving for a cigarette or a glass of alcohol is in the morning).[6] Our results complement Bénabou and Tirole (2004) by showing how internal commitment may arise also in routine tasks, for which an individual knows her (in)ability to resist temptation (e.g. writing three pages for a report or studying for an exam).

2 Model

Overview and timing. We consider the behavior of an individual with present-biased preferences who faces a task. At date 1, the individual chooses whether or not to exert effort to complete the task. The cost of effort is 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 cost relative to the more distant benefit of task completion, because she has a present bias. This bias creates an intrapersonal conflict of interest: at date 0, the individual thinks that working on the task is optimal. But once she makes the effort choice at date 1, all else equal, she will shirk. Anticipating her self-control problem, the individual can attempt to regulate her behavior by making a noncommittal self-promise, e.g. of rewarding herself at date 3 if the task was completed.

The task. Completing the task requires effort ($e = \bar{e}$) at date 1 and leads to a delayed benefit $\bar{y}$ at date 2. Should the individual however shirk ($e = \underline{e}$), the task remains unfinished and yields only $\underline{y} < \bar{y}$ at date 2. Effort causes an immediate disutility $c(\bar{e}) \equiv c > 0$, whereas $c(\underline{e}) = 0$.

Self-reward. There is an opportunity at date 3 to buy one unit of a good at price p. At

[6] According to Bénabou and Tirole (2004, p.858), a natural interpretation is that in "normal times" the ability to resist impulses "is a known" but it may be hard to predict for "times of stress – whether caused by abstinence, the proximity of temptation, or cues that intensify 'visceral' cravings."

date 0, the individual can make a noncommittal promise to herself of a consumption plan (*self-reward*) for date 3 that conditions on the task outcome $y \in \{\underline{y}, \bar{y}\}$. The state-contingent consumption plan $x(y) \in \{0, 1\}$ can for example take the form: "buy the good if the task was completed" ($x(\bar{y}) = 1$) and "don't buy the good if the task remains unfinished" ($x(\underline{y}) = 0$). While for concreteness we interpret the self-reward as a consumption good that can be purchased at price p, other interpretations are possible: the reward might be any activity that brings pleasure (v) and causes pain or has an opportunity cost (p).

Preferences. When making her effort decision at date 1, the individual overemphasizes the immediate cost relative to the more distant benefit of a completed task. That is, the individual has present-biased preferences (Strotz 1955). 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*). The first parameter, δ, corresponds to the standard exponential discount factor (for simplicity we assume $\delta = 1$). The second parameter, $\beta \in [0, 1)$, captures the extent of the present bias, and is the parameter of interest in our model. The utility of the individual at date $t \in \{0, .., 3\}$ is given by:

$$U_t \quad = \quad u_t + \beta \left[\sum_{\tau=t+1}^{3} u_\tau \right], \tag{1}$$

where u_t is the (expected) *instantaneous utility* in period t (to be specified below). For instance, at date 0 the individual (*self 0*) weighs future utilities u_1, u_2, and u_3 equally; but at date 1 the individual (*self 1*) puts a larger relative weight on u_1 by discounting u_2 and u_3 with $\beta < 1$, reflecting her present bias. We assume that the individual is *sophisticated* in the sense of O'Donoghue and Rabin (1999a), i.e. self 0 knows about the present-biased preferences of her future selves.

The instantaneous utility in each period has the properties of Köszegi and Rabin's (2006) reference-dependent utility. It is composed of two components: intrinsic *consumption utility* (which corresponds to the outcome-based utility traditionally studied in economics) and *gain-loss* utility. The latter is related to the deviation of the consumption utility from its reference point (measured in consumption-utility units): losing something that the individual values (e.g. a gold ring) is more painful than losing something that she does not value (e.g. a paper clip). Both consumption utility and gain-loss utility are assumed to be additively separable across the dimensions of the period outcome.[7] We capture these features with a linear loss aversion framework, as in the applications of Köszegi and Rabin (2006) and Heidhues and Köszegi (2008): if an outcome differs along dimension k from its reference point r_k by z, the corresponding gain-loss utility is $\mu_k(z) = \eta_k\, z$ for $z \geq 0$ and $\mu_k(z) = \eta_k\, \lambda\, z$ for $z < 0$. The parameter $\eta_k \geq 0$ measures the weight of gain-loss utility regarding dimension k in the utility function, and $\lambda > 1$ is the *coefficient of loss aversion*.

[7] See Köszegi and Rabin (2004, 2006) for a discussion of this assumption.

At date 0, no payoff-relevant outcomes occur, and thus $u_0 = 0$. At date 1, the individual makes her effort choice and incurs the cost of effort, which result in a consumption utility of $-c(e)$ and a gain-loss utility of[8] $\mu_e(-c(e) - r_e)$, i.e.

$$u_1 = -c(e) + \mu_e(-c(e) - r_e). \tag{2}$$

At date 2, the individual receives utility from the task outcome y, and hence

$$u_2 = y + \mu_y(y - r_y). \tag{3}$$

At date 3, the individual has the opportunity to purchase one unit of a good at price $p \geq 0$. Normalizing endowments to (0,0), she derives utility $v(x)$ from $x \in \{0, 1\}$ units of the good and money:[9]

$$u_3 = v(x) - p \cdot x + \mu_x(v(x) - r_x) + \mu_p(-p \cdot x - r_p), \tag{4}$$

with $v(1) \equiv v > v(0) = 0$.

Personal equilibrium. The reference points regarding the outcomes for dates $t > 0$ are determined by the expectations that the individual forms at date 0. These expectations are assumed to be rational in the sense of a *personal equilibrium* defined by Köszegi and Rabin (2006), i.e. continuation play must be consistent with optimal behavior at each date given expectations. If, for example, the individual expects to buy the good at date 3, then her reference point in the good dimension is $r_x = v$ and in the money dimension it is $r_p = -p$. In a personal equilibrium, it must then indeed be optimal for the individual to buy the good when date 3 is reached, given that her reference point is (r_x, r_p).

A crucial feature of this setup is that multiple personal equilibria may exist for some parameter range. For example, in one personal equilibrium the individual expects that she will buy the good, and it is optimal for her to buy the good; in the other personal equilibrium, she expects that she will not buy the good, and then indeed she does not buy the good. When such multiple personal equilibria exist, we assume that self 0 chooses her favorite plan according to Köszegi and Rabin's (2006) notion of a *preferred personal equilibrium*: she implements the personal equilibrium which provides her with the highest expected utility. Such a plan can include state-contingent consumption, where the promise of self 0 to consume the good only in certain states translates into corresponding expectations.

Self-control problem. We will consider the case where the present bias of the individual is sufficiently strong, so that an intrapersonal conflict of interest arises (Assumption 1 below

[8]For ease of exposition, we abuse notation and index the dimension for gain-loss utility and the reference point with the respective choice or outcome variable.

[9]We follow Köszegi and Rabin (2006, 2007) and Heidhues and Köszegi (2008) in assuming that preferences are also over monetary wealth in case the reward is interpreted as buying a good. For a discussion of this assumption see Köszegi and Rabin (2007).

provides the formal condition): at date 0, the individual thinks that working hard to complete the task at date 1 is optimal. But, in the absence of a contingent reward, self 1 will prefer to shirk once she faces the task at date 1.

3 Analysis

Our analysis proceeds as follows. First, we consider the decision of self 3 whether or not to buy the good – given the expectations formed by self 0 (this part of the analysis is based on Köszegi and Rabin 2006). Next, we turn to the effort decision of self 1 at the task stage. Finally, we consider the design of the self-reward scheme by self 0: What expectations can arise in a personal equilibrium given the behavior of self 1 and self 3? And when and how can these be exploited to design a credible and incentive compatible reward?

To show more clearly our main insights about designing and committing to self-administered rewards, we make the simplifying assumption that the individual perceives no gain-loss utility related to the task-outcome or effort dimensions, i.e. $\eta_y = \eta_e = 0$. Furthermore, let $\eta_p = \eta_x = 1$. In the appendix, we allow for the more general case and show that our results are robust.

The decision of self 3: to buy or not to buy?

The reference point for self 3 is the individual's past expectation regarding the date-3 outcome. We need to distinguish two cases: did the individual expect to buy the good, or not (in a given state of the world)? Suppose first she did, and her reference points regarding the good and money dimensions thus are $r_x = v$ and $r_p = -p$, respectively. Following through with this expectation provides self 3 with consumption utility $v-p$ from purchasing the good. As her reference points for each outcome dimension are met, she perceives no gain or loss. If, instead, she abstains from buying the good, her consumption utility is 0. So she perceives a loss in the goods dimension and a gain in the money dimension. This results in gain-loss utility $-\lambda\, v + p$. Hence, it is indeed optimal for self 3 to buy the good if $v - p \geq -\lambda\, v + p$, i.e. if

$$p \leq \frac{v\,(1+\lambda)}{2} \equiv p_{max}.$$

Should the price exceed p_{max}, then self 3 will never consume the good – no matter what expectations the individual held in the past.

Suppose next that the individual did not expect to buy the good, i.e. $r_x = 0$ and $r_p = 0$. Not buying the good at date 3 then provides self 3 with consumption and gain-loss utility of zero. If she buys the good, however, her consumption utility is $v - p$ and she perceives gain-loss utility of $v - \lambda\, p$. Thus, it is optimal for self 3 not to purchase the good if $0 \geq 2\, v - p\,(1+\lambda)$, i.e. if

$$p \geq \frac{2\, v}{1+\lambda} \equiv p_{min}.$$

Should the price be less than p_{min}, then self 3 will always consume the good – no matter what expectations the individual held in the past.
In the intermediate price range, $p \in [p_{min}, p_{max}]$, two possibilities arise: if self 0 forms the expectation of buying the good, self 3 will buy the good; and if self 0 forms the expectation of not buying the good, self 3 will not buy the good.

Lemma 1 (Köszegi and Rabin 2006)

1. *Suppose $p < p_{min}$. Then self 3 always buys the good, irrespective of past expectations.*
2. *Suppose $p > p_{max}$. Then self 3 never buys the good, irrespective of past expectations.*
3. *Suppose $p \in [p_{min}, p_{max}]$. Then two possibilities exist: self 3 will buy the good if self 0 forms the expectation of buying the good; and self 3 will not buy good if self 0 forms the expectation of not buying the good.*

The decision of self 1: work or shirk?

To understand why and how a contingent reward may act as a positive motivator, we show first that a conflict of interest between self 0 and self 1 may arise in the absence of such a self-regulation strategy. That is, we consider the case where the date-2 task outcome does not affect the continuation utility at date 3. What is the effort decision of self 1 in this case? She will work on the task if and only if the gain in consumption utility from task completion at date 2, discounted by her present-bias parameter β, exceeds the immediate disutility from effort:

$$\beta \{\bar{y} - \underline{y}\} - c \geq 0. \tag{5}$$

Self 0, in contrast, does not distort date-1 costs relative to date-2 benefits and thus wants self 1 to complete the task if and only if:

$$\{\bar{y} - \underline{y}\} - c \geq 0. \tag{6}$$

As $\beta < 1$, inequality (6) may hold while inequality (5) does not. Then an intrapersonal conflict of interest arises: self 1 strictly prefers to shirk, but self 0 would prefer her to put in effort to complete the task.

Assumption 1 (Self-control problem) *There is a conflict of interest between self 0 and self 1, i.e. the individual faces a self-control problem, if and only if:*

$$\frac{c}{\beta} > \bar{y} - \underline{y} \geq c. \tag{SCP}$$

As self-regulation is only an issue if the individual faces a self-control problem, we will henceforth consider the case where Condition (SCP) holds. In an attempt to resolve this intrapersonal conflict of interest, self 0 can make a noncommittal promise of a state-contingent

consumption plan. If the individual will indeed stick to this plan at date 3, then the anticipated continuation utility following task completion, $\bar{u}_3$, is distinct from that if the task remains unfinished, $\underline{u}_3$. And these distinct continuation utilities can provide incentives for self 1: given that self 0 made a credible promise of a self-administered reward, self 1 completes the task if and only if:

$$\beta\{\bar{y} + \bar{u}_3\} - c \geq \beta\{\underline{y} + \underline{u}_3\} \quad \Leftrightarrow \quad \beta\{\bar{y} - \underline{y} + \bar{u}_3 - \underline{u}_3\} - c \geq 0. \tag{7}$$

For clarity of exposition we will assume that the good available at date 3 is a *net reward* from the perspective of self 0, i.e. $v > p$. Thus, a self-reward of the form "If I complete the task I will reward myself and buy the good, if I shirk I will deny myself the good" ($x(\bar{y}) = 1$, $x(\underline{y}) = 0$) will – if the plan is credible – lead to $\bar{u}_3 = v - p > \underline{u}_3 = 0$.[10] The 'incentive constraint' for self 1, inequality (7), can then be rewritten as:

$$\tilde{p} \equiv [\bar{y} - \underline{y} + v] - \frac{c}{\beta} \geq p. \tag{8}$$

Put differently, whenever $p \leq \tilde{p}$ a self-reward has the power to provide incentives. The more difficult the task is (i.e. the higher c), or the more severe the present bias is (i.e. the lower β), the stronger the incentives must be, i.e. the lower the price ceiling $\tilde{p}$ for the good to still be attractive enough to help overcome the self-control problem.

Self 0: self-regulation or resignation?

To examine whether or not a state-contingent consumption plan is not only incentive compatible, but also credible, we need to ask whether it can be part of a personal equilibrium. Three different cases arise as a consequence of Lemma 1. First, if the price of the good is less than p_{min}, self 3 will always consume the good – no matter what expectations the individual had. The part of the self-promise "...if I shirk I will deny myself the good" therefore is not credible. In a personal equilibrium, self 0 will always expect her future self to consume the good and self 3 will indeed do so. Hence, $\bar{u}_3 = \underline{u}_3$ and the noncommittal promise of a self-administered reward cannot provide the additional incentives necessary for bridging the conflict of interest between self 0 and self 1. A similar issue arises in the second case, where the price of the good exceeds p_{max}. Self 3 will never consume the good – no matter what expectations the individual had. This means that the part of the self-promise "If I complete the task I will reward myself and buy the good..." is not credible. In a personal equilibrium, self 0 will expect not to buy the good and self 3 will indeed not purchase it. Again, $\bar{u}_3 = \underline{u}_3$ and there are no additional incentives for self 1 to exert effort.

[10] Self-reinforcement based on a *net punishment* is analogous. In such a case, the individual specifies $x(\bar{y}) = 0$, $x(\underline{y}) = 1$, as the consumption benefit v is lower than the disutility p caused along another dimension. While with a self-administered reward self-denial of the reward has to be a credible threat, with a self-punishment plan the expectation of carrying through with the unpleasant plan in case of shirking has to be credible.

In the third case, where $p \in [p_{min}, p_{max}]$, two personal equilibria arise: when self 0 forms the expectation of buying the good, then it is indeed optimal for self 3 to buy the good; when self 0 forms the expectation of not buying the good, then not buying is indeed optimal for self 3. This multiplicity of personal equilibria provides self 0 with scope to pick her *preferred personal equilibrium*, and thereby motivate her future self with a credible self-administered reward scheme of the form "If I complete the task I will reward myself and buy the good, if I shirk I will deny myself the good." Stated differently, the noncommittal promise is a self-fulfilling plan: given expectations, it is optimal for self 3 to buy the good if the task was completed and not to buy it otherwise. This plan however can only help overcome the self-control problem if the self-reward is sufficiently powerful to motivate self 1, i.e. if $p \leq \tilde{p}$ so that the incentive constraint of self 1 holds.

In sum, whenever $p \leq \tilde{p}$ in addition to $p \in [p_{min}, p_{max}]$ the preferred personal equilibrium involves self 0 promising herself a contingent reward plan that motivates self 1 to work on the task. While self 1 is disciplined by the prospect of losing out on the consumption utility from the reward if she does not complete the task, self-denial of the reward does not occur on the equilibrium path and hence does not harm the individual.[11] Hence, forming the expectation not to buy after shirking increases the expected utility of self 0 relative to the alternative without a self-reward plan, where the self-control problem persists and the task does not get completed. We summarize our findings in the following proposition:

Proposition 1

Suppose that self 0 faces a self-control problem, i.e. Condition (SCP) holds. Then for

1. *$p < p_{min}$ or $p > p_{max}$, self-regulation based on a self-administered reward is not credible: in the unique personal equilibrium, self 3 never ($p > p_{max}$) or always ($p < p_{min}$) consumes the good and the task does not get completed.*

2. *$p_{max} \geq p \geq p_{min}$ and $p > \tilde{p}$, self-regulation based on a self-administered reward is credible. However, the reward is not large enough to provide self 1 with sufficient incentives to overcome the self-control problem. In the preferred personal equilibrium, self 0 expects to always buy the good (or to buy it if the task is not completed) and the task does not get completed.*

3. *$v > p \geq p_{min}$ and $p \leq \tilde{p}$, self-regulation based on a self-administered reward is credible and the reward can motivate self 1 to work on the task. In the preferred personal equilibrium, self 0 makes a promise of buying the good only after task completion and the task gets completed.*

[11] This holds true also for stochastic production: in such a case the individual can still condition the self-reward on her (non-stochastic) effort.

The result sheds light on what on the surface looks like a puzzling phenomenon: self-rewards are widely used and tend to be rather successful when employed in the right circumstances. Proposition 1 shows that, under certain conditions, a rational forward-looking individual can indeed overcome a self-control problem with the promise to self-administer a reward after achieving a specific target, and the threat to deny herself the reward otherwise.

Two forces may however constrain self-regulation, depending on the characteristics of the available reward good. First, commitment to the self-reward strategy has to be possible, captured by the credibility constraint $p \in [p_{min}, p_{max}]$. In other words, the good must be neither something that the individual will consume anyway – no matter what her expectations and reference point are (e.g. because it is a necessity or a bargain) – nor so extravagant or expensive that she will never consume it. Second, the reward has to be large enough to motivate the individual to act despite her present bias, captured by the incentive constraint $p \leq \tilde{p}$.

These two forces may be in irreconcilable conflict with each other: in order for the good to be a powerful incentive tool it must come at a low enough price ($p \leq \tilde{p}$); but then the opportunity to purchase the good may be so attractive that the individual will always end up consuming the good (because $p < p_{min}$). The following corollary captures this:

Corollary 1 (Comparative Statics)

The higher the cost of effort c for completing the task is, or the more severe the present bias is (i.e. the lower β), the lower the price ceiling $\tilde{p}$ for the incentive constraint $p \leq \tilde{p}$. If c is high or β is low relative to the consumption value v of the good, $\tilde{p}$ falls below p_{min}:

$$\tilde{p} < p_{min} \quad \Leftrightarrow \quad v_{min} \equiv \frac{1+\lambda}{\lambda - 1}\left[\frac{c}{\beta} - (\bar{y} - \underline{y})\right] > v,$$

i.e. the available good cannot at the same time serve as a reward that provides sufficiently strong incentives for effort and be a credible self-reward.

While credibility of the self-reward does not depend on the task difficulty or the present bias of the individual (c and β do not affect p_{min} and p_{max}), these factors influence $\tilde{p}$. If the task is very burdensome or the present bias is severe relative to the utility from consuming the good available as a reward, so that $v_{min} > v$, then self-regulation is doomed to fail: if the reward is sufficiently powerful, the threat of self-denial is not credible – and if self-denial is credible, the reward is not powerful enough.

The result thus helps understand the limits to self-regulation based on self-rewards. The available reward good may come at the 'wrong' price, so that either the incentive constraint $p < \tilde{p}$ or the credibility constraint $p \in [p_{min}, p_{max}]$ is not met. Or the severity of the self-control problem makes it impossible to satisfy both constraints, regardless of the price at which the available reward good can be purchased, because $v_{min} > v$. This explains why

there still is a large demand for (often costly) external commitment devices.[12] These only need to satisfy the individual's incentive constraint to help overcome a self-control problem, i.e. $p < \tilde{p}$ – credibility is not an issue as the reward is externally enforced.[13]

Our analysis also helps explain the advice that self-help guides give about how to use self-rewards. The comparative statics result above illustrates that it becomes more difficult to find an 'appropriate' reward good (if one can choose among several ones) in the face of arduous tasks or when the individual has a strong present bias. Correspondingly, self-help guides typically recommend to start with thinking about suitable contingent rewards: you should use 'small rewards' that are sufficiently enticing to help motivate yourself, but that are not so important to you that you would go ahead and get them even if you failed to achieve the prescribed targets (e.g. 'play a game of pinball' rather than 'buy the car that I need anyway'). What these targets should be is then considered in a second step: if possible, you should divide tasks into relatively unambitious 'milestones' that make the cost of reaching a reward small enough relative to the benefit, so that the promised reward for each step has bite (e.g. 'write three pages' rather than 'write 30 pages').[14]

4 Conclusion

Three decades ago Bandura raised the question why self-administered rewards can work:

> *"One of the significant, but insufficiently explored, issues in self-reinforcement is why people adhere to contingencies requiring difficult performances, thereby temporarily denying themselves [until requisite performances are attained] rewards over which they exercise full control."* (Bandura 1976, p.140)

We provide a model based on Köszegi and Rabin (2006) that explains how commitment to such self-rewards can be achieved, and what factors determine whether self-regulation is possible. Our results show that self-reward strategies are credible only if the good involved

[12] If self-regulation is not possible, self 0 has a willingness to pay for access to an external commitment device of $\beta\left[\bar{y}-\underline{y}-c\right]$.

[13] External commitment devices often rely on making it costly to deviate from the preferred action of self 0, i.e. they punish deviations. The threatened punishment $v-p<0$ works if the incentive constraint of self 1 is satisfied, i.e. $\left[\bar{y}-\underline{y}+p-v\right]-\frac{c}{\beta}\geq 0$. Schelling (1992) lists examples, such as depositing self-incriminating letters to be published if the individual fails the target. In contrast, a self-punishment strategy would – analogously to the self-reward strategy we discussed – require additionally that the threat is credible, i.e. that the individual believes that she would stick to punishing herself (e.g. by engaging in an unpleasurable activity) if she deviated from the prescribed course of action.

[14] For example, if the costs of drafting a report are convex then writing 30 pages in one period is more than ten times as painful as writing three pages in one period. In this case, splitting the task makes the self-control problem for each step less severe in relative terms than the one for the overall task: $\frac{c(3)}{\beta}-\frac{\bar{y}-\underline{y}}{10}<\frac{1}{10}\left[\frac{c(30)}{\beta}-(\bar{y}-\underline{y})\right]$.

is neither too extravagant nor too much of a bargain or a necessity. But the reward also has to be sufficiently large to provide incentives for the individual to exert effort on the task. Successful self-regulation based on self-administered rewards is possible if the available reward good satisfies these credibility and incentive constraints.

Thus, our results offer an explanation for the wide-spread use and success of self-rewards. But they also help understand the possible limitations of self-rewards: according to the comparative statics of our model, the more difficult the task is, or the stronger the present bias of the individual is, the harder it becomes to jointly satisfy the credibility and incentive constraints. That is, these constraints may conflict with each other: to motivate the individual the reward might have to be so attractive that it becomes impossible for her to believe that she will actually carry through with her threat of self-denial when shirking. For a tough self-control problem it may therefore be difficult to find an 'appropriate' reward good.

The limits to self-regulation that our analysis identifies show that self-rewards are not a panacea. They work only when adopted in the right circumstances, explaining why there still is demand for (often costly) external commitment devices: self-administered rewards can help overcome relatively easy motivational problems, but the really hard self-control problems require external commitment.

Appendix

To bring out the main driving forces, we assumed so far that the individual does not perceive gain-loss utility related to the date-1 effort cost and the date-2 task outcome ($\eta_e = \eta_y = 0$), and we normalized the weight of gain-loss utility at date 3 ($\eta_x = \eta_p = 1$). Here we consider the more general case where, as in Köszegi and Rabin (2006), the individual places equal weight on gain-loss utility across *all* outcome dimensions: $\eta_k = \eta > 0$ for $k = \{e, y, x, p\}$. As before, we assume that self 0 wants her future self to complete the task, i.e. inequality (6) holds.

Now, because self 0 holds expectations regarding the effort of self 1 and the task outcome at date 2, deviations from these reference points trigger gain-loss utility. For instance, with the expectation that the task will be completed, the individual's reference level for the effort cost is $r_e = -c$, and for the task outcome it is $r_y = \bar{y}$. From the perspective of self 1, shirking thus would lead to an immediate gain in the effort-cost dimension of $\eta\, c$ and a discounted loss in the task dimension at date 2 of $-\beta\, \eta\, \lambda\, (\bar{y} - \underline{y})$.

These gain-loss utility components may provide an additional source of motivation for self 1 to complete the task, and therefore change the conditions under which a conflict of interest between self 0 and self 1 arises (in the absence of a self-reward). To see this *task-anticipation effect*, consider the incentives of self 1 in the absence of a self-administered reward. Given the expectation that self 1 will complete the task, it is indeed optimal for self 1 to do so if and only if:

$$\bar{y} - \underline{y} \geq \frac{1+\eta}{1+\eta\,\lambda}\,\frac{c}{\beta} \equiv \Delta\, y'. \tag{9}$$

Effort incentives for self 1 are stronger because the individual suffers a loss if she does not reach the anticipated task outcome.[15] If, on the other hand, self 0 expects self 1 not to provide effort, self 1 will indeed shirk if and only if:

$$\bar{y} - \underline{y} \leq \frac{1+\eta\,\lambda}{1+\eta}\,\frac{c}{\beta} \equiv \Delta\, y''. \tag{10}$$

Because $\Delta\, y'' > \Delta\, y'$, we are left with three cases:

1. $\bar{y} - \underline{y} > \Delta\, y''$: Self 1 will complete the task – no matter what past expectations were.
2. $\bar{y} - \underline{y} < \Delta\, y'$: Self 1 will shirk – no matter what past expectations were.
3. $(\bar{y} - \underline{y}) \in [\Delta\, y', \Delta\, y'']$: Two personal equilibria exist – one with the self-fulfilling expectation of task completion and one with the self-fulfilling expectation of the task remaining unfinished. In the preferred personal equilibrium, self 0 forms the expectation that the task will be completed.

[15] The requirement for the case where $\eta_y = \eta_e = 0$ that $\bar{y} - \underline{y} \geq c/\beta$ is stronger because $\lambda > 1$ implies that $c/\beta > \Delta\, y'$.

Overall, the condition for a self-control problem to exist hence becomes

$$\Delta\, y' > \bar{y} - \underline{y} > c. \tag{SCP'}$$

Again, self 0 can attempt to use a self-reward to overcome the self-control problem that arises when Condition (SCP') holds. Analogously to the analysis in the main part, for this to be successful, the self-reward must be credible (i.e. $p \in [p_{min}, p_{max}]$) and the anticipated date-3 continuation utilities $\bar{u}_3$ and $\underline{u}_3$ must be sufficient to motivate self 1. The latter means that self 0's expectation that the task will be completed has to be self-fulfilling. This can be part of a personal equilibrium only if the following incentive constraint for self 1 is satisfied:

$$\beta\,\{(1+\eta\,\lambda)(\bar{y}-\underline{y}) + \bar{u}_3 - \underline{u}_3\} \geq (1+\eta)\,c. \tag{11}$$

For the case where the good provides net consumption utility $v - p > 0$ we thus obtain a characterization of the preferred personal equilibria that parallels Proposition 1, replacing Condition (SCP) with Condition (SCP') and the threshold $\tilde{p}$ with

$$\tilde{p}' \equiv (1+\eta\,\lambda)(\bar{y}-\underline{y}) + v - \frac{1+\eta}{\beta}\,c. \tag{12}$$

In sum, the qualitative features of the self-regulation problem are similar to the simplified setting in the main text. What the new conditions show in addition is i) self-control problems are less likely to arise (but can still arise) if the individual perceives gain-loss utility related to the task outcome and effort cost dimension; ii) when a self-control problem exists, a higher weight on gain-loss utility η makes it more likely that the individual can overcome it using a self-reward.

References

Akerlof, George A. and William T. Dickens, "The Economic Consequences of Cognitive Dissonance," *American Economic Review*, June 1982, *72* (3), 307–319.

Ariely, Dan and Klaus Wertenbroch, "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science*, May 2002, *13* (3), 219–224.

Ashraf, Nava, Dean Karlan, and Wesley Yin, "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, May 2006, *121* (2), 635–672.

Bandura, Albert, "Vicarious and Self-Reinforcement Processes," in Robert Glaser, ed., *The Nature of Reinforcement*, New York: Academic Press, 1971, pp. 228–278.

___ , "Self-Reinforcement: Theoretical and Methodological Considerations," *Behaviorism*, 1976, *4* (2), 135–155.

___ , *Social Foundations of Thought and Action: A Social Cognitive Theory*, New Jersey: Prentice Hall, 1986.

___ , "The Evolution of Social Cognitive Theory," in Ken G. Smith and Michael A. Hitt, eds., *Great Minds in Management: The Process of Theory Development*, Oxford: Oxford University Press, 2005, pp. 9–35.

Bénabou, Roland and Jean Tirole, "Willpower and Personal Rules," *Journal of Political Economy*, August 2004, *112* (4), 848–886.

___ **and** ___ , "Identity, Dignity and Taboos: Beliefs as Assets," CEPR Discussion Paper no. 6123 2007.

Benhabib, Jess and Alberto Bisin, "Modeling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions," *Games and Economic Behavior*, August 2005, *52* (2), 460–492.

Brocas, Isabelle, Juan D. Carrillo, and Mathias Dewatripont, *Commitment Devices under Self-Control Problems: An Overview*, The Psychology of Economic Decisions. Vol.2: Reasons and Choices, Oxford University Press, 2004.

Brunnermeier, Markus K. and Jonathan A. Parker, "Optimal Expectations," *American Economic Review*, September 2005, *95* (4), 1092–1118.

___ **, Christian Gollier, and Jonathan A. Parker**, "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns," *American Economic Review*, May 2007, *97* (2), 159–165.

Carrillo, Juan D., "To be Consumed with Moderation," *European Economic Review*, January 2005, *49* (1), 99–111.

___ **and Mathias Dewatripont**, "Promises, Promises, ...," *Economic Journal*, August 2008, *118* (531), 1453–1473.

Clum, George A. and Patti Lou Watkins, *Handbook of Self-Help Therapies*, New York: Routledge, 2007.

DellaVigna, Stefano and Ulrike Malmendier, "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, May 2004, *119* (2), 353–402.

Diamond, Peter and Botond Köszegi, "Quasi-Hyperbolic Discounting and Retirement," *Journal of Public Economics*, September 2003, *87* (9-10), 1839–1872.

Elster, Jon, *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*, Cambridge, MA: Cambridge University Press, 2000.

Faber, Ronald and Kathleen D. Vohs, "To Buy or Not to Buy? Self-Control and Self-Regulatory Failure in Purchase Behavior," in Roy F. Baumeister and Kathleen D. Vohs, eds., *Handbook of Self-Regulation: Research, Theory, and Applications*, New York, NY: Guilford, 2004, pp. 509–524.

Febbraro, Greg A. R. and George A. Clum, "Meta-Analytic Investigation of the Effectiveness of Self-Regulatory Components in the Treatment of Adult Problem Behaviors," *Clinical Psychology Review*, March 1998, *18* (2), 143–161.

Fudenberg, Drew and David K. Levine, "A Dual-Self Model of Impulse Control," *American Economic Review*, December 2006, *96* (5), 1449–1476.

Gine, Xavier, Dean Karlan, and Jonathan Zinman, "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation," mimeo, Yale University 2008.

Gollier, Christian and Alexander Muermann, "Optimal Choice and Beliefs with Ex Ante Savoring and Ex Post Disappointment," IDEI Working Paper No. 462 2006.

Gul, Faruk and Wolfgang Pesendorfer, "Temptation and Self-Control," *Econometrica*, November 2001, *69* (6), 1403–1435.

Heidhues, Paul and Botond Köszegi, "Competition and Price Variation When Consumers Are Loss Averse," *American Economic Review*, September 2008, *98* (4), 1245–1268.

Herweg, Fabian and Daniel Müller, "Performance of Procrastinators: On the Value of Deadlines," Bonn Econ Discussion Papers no. 3-2008 January 2008.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, March 1979, *47* (2), 263–291.

Kivetz, Ran and Itamar Simonson, "Earning the Right to Indulge: Effort as a Determinant of Customer Preferences Toward Frequency Program Rewards," *Journal of Marketing Research*, May 2002, *39* (2), 155–170.

Koch, Alexander K. and Julia Nafziger, "Self-Regulation Through Goal Setting," IZA Discussion Paper no. 3893 December 2008.

Köszegi, Botond, "Utility from Anticipation and Personal Equilibrium," mimeo, University of California, Berkeley 2007.

___ **and Matthew Rabin**, "A Model of Reference-Dependent Preferences," Economics Department, University of California, Berkeley, Working Paper E04-337 April 2004.

___ **and** ___ , "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, November 2006, *121* (4), 1133–1165.

___ **and** ___ , "Reference-Dependent Risk Attitudes," *American Economic Review*, September 2007, *97* (4), 1047–1073.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, May 1997, *112* (2), 443–477.

Landier, Augustin, "Wishful Thinking: A Model of Optimal Reality Denial," mimeo, MIT 2000.

Locke, Edwin, Gary P. Latham, and Miriam Erez, "The Determinants of Goal Commitment," *Academy of Management Review*, January 1988, *13* (1), 23–39.

Mick, David Glen and Corinne Faure, "Consumer Self-Gifts in Achievement Contexts: The Role of Outcomes, Attributions, Emotions, and Deservingness," *International Journal of Research in Marketing*, October 1998, *15* (4), 293–307.

___ **and Michelle DeMoss**, "Self-Gifts: Phenomenological Insights from Four Contexts," *Journal of Consumer Research*, December 1990, *17* (3), 322–32.

Newton, Harry N., Cynthia Barnhart, Pamela H. Vance, and Klaus Wertenbroch, "Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice," *Marketing Science*, Fall 1998, *17* (4), 317–337.

O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *American Economic Review*, March 1999, *89* (1), 103–124.

___ and ___ , "Incentives For Procrastinators," *Quarterly Journal of Economics*, August 1999, *114* (3), 769–816.

Phelps, E. S. and R. A. Pollak, "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, April 1968, *35* (2), 185–199.

Schelling, Thomas C., "Self-Command: A New Discipline," in George Loewenstein and Jon Elster, eds., *Choice Over Time*, New York: Russell Sage Foundation, 1992, chapter 7, pp. 167–176.

Strotz, R. H., "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 1955, *23* (3), 165–180.

Thaler, Richard H and H M Shefrin, "An Economic Theory of Self-Control," *Journal of Political Economy*, April 1981, *89* (2), 392–406.

Thaler, Richard H. and Shlomo Benartzi, "Save More Tomorrow (TM): Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, February 2004, *112* (1), S164–S187.