



IZA DP No. 5802

## Goals and Psychological Accounting

Alexander K. Koch  
Julia Nafziger

June 2011

# Goals and Psychological Accounting

**Alexander K. Koch**

*Aarhus University  
and IZA*

**Julia Nafziger**

*Aarhus University*

Discussion Paper No. 5802

June 2011

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto: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 Foundation. 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

### Goals and Psychological Accounting<sup>\*</sup>

We model how people formulate and evaluate goals to overcome self-control problems. People often attempt to regulate their behavior by evaluating goal-related outcomes separately (in narrow psychological accounts) rather than jointly (in a broad account). To explain this evidence, our theory of endogenous narrow or broad psychological accounts combines insights from the literatures on goals and mental accounting with models of expectations-based reference-dependent preferences. By formulating goals the individual creates expectations that induce reference points for task outcomes. These goal-induced reference points make substandard performance psychologically painful and motivate the individual to stick to his goals. How strong the commitment to goals is depends on the type of psychological account. We provide conditions when it is optimal to evaluate goals in narrow accounts. The key intuition is that broad accounts make decisions or risks in different tasks substitutes and thereby create incentives to deviate from goals. Model extensions explore the robustness of our results to different timing assumptions and goal and account revision.

JEL Classification: A12, C70, D81, D91

Keywords: quasi-hyperbolic discounting, reference-dependent preferences, loss aversion, self-control, mental accounting, goals

Corresponding author:

Alexander K. Koch  
Institut for Økonomi  
Aarhus University  
Building 1322  
8000 Aarhus C  
Denmark  
E-mail: [akoch@econ.au.dk](mailto:akoch@econ.au.dk)

---

<sup>\*</sup> We thank Johannes Abeler, Kathrin Breuer, Stefano DellaVigna, Dirk Engelmann, Guido Friebel, Botond Köszegi, Georg Nöldeke, Wendelin Schnedler, Heiner Schumacher, Yoram Weiss, seminar participants at the Universities of Aarhus, East Anglia, Humboldt Berlin, Bern, Heidelberg, Frankfurt, Zürich, ZEW Mannheim, and audiences at the ESEM meetings in Barcelona and Glasgow, the ESA in Copenhagen, and M-BEES in Maastricht for helpful comments and discussions.

# 1 Introduction

Many people find it difficult to resist tempting choices that go against their long-run interest. Failure to overcome such *self-control problems* leads, for example, to overindulgence, or to underprovision of useful but unpleasant activities. Much of the literature focuses on external commitment devices as tools for overcoming self-control problems, such as investing in illiquid assets to protect savings (for surveys see Elster 2000, Bryan *et al.* 2010). Yet most instances of self-control in everyday life do not rely on external commitment (cf. Rachlin 1995). We study how people achieve internal commitment by formulating and evaluating goals. The central question we ask is whether an individual should evaluate goals from different tasks together, in one broad psychological account, or if he should evaluate them separately in narrow psychological accounts.

Many examples suggest that people adopt narrow accounts, or using popular language, “narrow goals”. People often track expenditures in mental accounts for narrowly defined categories, such as entertainment, clothing, or food (e.g., Heath and Soll 1996). Many diet programs recommend daily nutrition goals. For instance, the Weight Watchers PointsPlus™ system assigns points to food based on their nutritional content and sets a personalized, daily goal. At the same time, not all goals are narrow. People do not have an account for every item they buy, or for every possible consumption category. And diet programs typically combine daily nutrition goals with the recommendation to weigh yourself only at weekly intervals.<sup>1</sup>

There has been some discussion that narrow accounts help to overcome self-control problems. Shefrin and Thaler (1988) model how assigning wealth to distinct, narrow accounts allows consumers to control their short-run urge to overspend. But they assume that consumers have a low marginal propensity to spend out of some exogenously given accounts.<sup>2</sup> Camerer *et al.* (1997) and Read *et al.* (1999) informally discuss that narrowly evaluated goals (such as daily work goals) may provide better self-control. A broad goal would allow the individual to slack off and tell himself that he will make up for today’s shortfall by working harder tomorrow. But the generality and the robustness of this mechanism are not clear.

In sum, there is a need to formalize the driving forces of how people structure their accounts and explore the robustness of narrow accounting. Our contribution is to develop a theory of endogenous formation of narrow or broad accounts in a self-control setting. We build on and formalize ideas from the mental accounting literature and combine them with insights from the literature on goal-setting. Mental accounting is typically associated with how people organize, evaluate, and keep track of their financial activities (Thaler 1980, 1985, 1990, 1999). We apply the logic of a mental account to goals and to motivational problems that go beyond financial activities. To make this distinction clear, we adopt the concept of a *psychological account* from Tversky and Kahneman (1981, p.456), defined as “*an outcome frame which specifies (i) the set of elementary outcomes that*

---

<sup>1</sup>For weekly weighing and daily calorie targets see, e.g., the recommendations of the British NHS: <http://www.nhs.uk/Livewell/Newyear/Pages/NYdiettips.aspx> (last accessed May 2011).

<sup>2</sup>Indeed a large body of evidence in the mental accounting literature suggests that people do not treat money as fully fungible when a label is attached to a part of their budget (e.g., Heath and Soll 1996). However, the underlying theoretical mechanisms are not well understood.

are evaluated jointly and the manner in which they are combined and (ii) a reference outcome that is considered neutral or normal.”<sup>3</sup>

What distinguishes our approach from the previous literature is that we explicitly model the processes through which psychological accounts impose constraints on future behavior, and how these constraints change with the breadth of accounts. By asking what goals are self-enforcing under a certain type of account, we derive boundary conditions for self-regulation. These show that narrowly evaluated goals allow people to fully overcome mild self-control problems and help alleviate more severe self-control problems. By asking whether and when the individual can achieve the same, better, or worse outcomes with a broad account, we obtain conditions for the optimality of narrow or broad accounts in various settings (such as when a person faces risky or riskless choices, sequential or simultaneous choices). To address the question why people do not borrow from other accounts (i.e., transform their narrow accounts into a broad account), we consider in an extension of our model an individual who has the opportunity to revise his goals or accounts.

Our paper also contributes to the choice bracketing literature that asks why people often focus on a single choice in isolation from other related choices. Read *et al.* (1999) discuss that narrow bracketing may arise because of cognitive capacity limitations, cognitive inertia, pre-existing heuristics, or for motivational reasons. We provide a model for the last aspect – showing that evaluating outcomes from a decision in isolation from other related decisions can be a commitment device for a time-inconsistent agent. However, our aim is not to provide a general theory of narrow bracketing. In particular, our model does not explain the kind of bracketing in lottery choices famously illustrated by the ABCD example of Tversky and Kahneman (1981). The example illustrates a tendency to make each choice in isolation when people face several choices over pairs of lotteries. Such narrow bracketing leads to dominated choices (Tversky and Kahneman 1981, Thaler *et al.* 1997, Rabin and Weizsäcker 2009). Moreover, in contrast to, for example, Rabin and Weizsäcker (2009) we assume that the individual is able to consider all decisions. In our model, he may deliberately choose narrow accounts. Our findings thus suggest that instances of narrow bracketing need not always reflect choice errors. Similar in spirit, Kőszegi and Rabin (2009) show that a person may rationally behave “as if” he brackets narrowly. Their focus however is not on self-control problems. They consider a person who faces sequential risky choices and derives anticipatory gain-loss utility from changes in beliefs.

We consider an individual who at some future date(s) makes two decisions. These decisions may be, for instance, how hard to work in two different tasks or how much to consume of two different goods. The individual has a present bias which we model using *quasi-hyperbolic discounting* (Strotz 1955, Phelps and Pollak 1968, Laibson 1997). The present bias creates a self-control problem. Before facing the choice problem the individual prefers a different decision than he does, all else equal, at the time when he actually makes his choice.

To motivate his future self, the individual sets goals. A goal is a plan, such as “I want to study 8 hours on Monday for the exam”. Goals give rise to expectations about outcomes, such as “I will get

---

<sup>3</sup>Furthermore, Thaler (1999, p.186) uses “the term ‘mental accounting’ to describe the entire process of coding, categorizing, and evaluating events”, while a psychological account just describes a frame for evaluation.

a good grade”. Because the individual has loss-averse, reference-dependent preferences (Kahneman and Tversky 1979) such expectations matter. Following Kőszegi and Rabin (2006, 2007), past expectations serve as reference points for future selves (for experimental evidence see Abeler *et al.* 2011). If expectations are exceeded, the individual feels a gain (getting a B feels good if one expected to fail the exam). If expectations fail to materialize, the individual feels a loss (getting a B is disappointing if one expected an A). The justification for reference-dependent preferences in this setting stems from two sources. The first is Tversky and Kahneman’s (1981) above definition of a psychological account. The second is the psychology literature on goal setting, which emphasizes the idea that goals induce reference points and that people are loss averse regarding goal achievement (for a survey see Locke and Latham 1990). Heath *et al.* (1999) explicitly point out the similarity to the value function in Kahneman and Tversky’s (1979) Prospect Theory.

For goals to induce expectations, the individual must believe that they can be accomplished. And indeed one of the findings in the psychology literature is that goals must be “realistic” and “attainable” (e.g., Hollenbeck *et al.* 1989). Similarly, popular self-help guides stress that goals should be “SMART” – specific, measurable, attainable, realistic, and timely.<sup>4</sup> We capture these ideas by assuming that goals and the expectations they induce are rational in the sense of a *personal equilibrium* (Kőszegi and Rabin 2006, Kőszegi 2010). If the individual expects to work 8 hours, it must indeed be optimal to do so given the induced reference points for related costs and benefits.

In section 3, we consider a model where task outcomes are deterministic. Our analysis starts with asking how and when goals in combination with narrow accounts can help the individual to overcome his self-control problem. The goal to make a certain decision creates an “attachment” to the associated outcomes. This attachment increases the individual’s willingness to actually make the desired decision. Specifically, if the individual deviates from his goal, say he works less than desired, he will face a psychological loss. And the fear of a loss from falling short of the goal pushes the individual to stick to the goal. As we show, a narrowly evaluated goal can help people overcome self-control problems that are not too severe, and alleviate more severe self-control problems. This result confirms Loewenstein’s (1999) conjecture that goal setting can explain why many people with time-inconsistent preferences often behave in a time-consistent fashion.

Next, we endogenize the choice of accounts. With deterministic task outcomes, broad and narrow accounts yield the same utility for the same decision levels. However, broad accounts create incentives to deviate from goals by substituting decision levels between tasks (“*decision substitution*”). An individual who adopts narrow accounts focuses on the health benefits and costs of a particular meal, or of a work-out activity on a particular day. In contrast, an individual with broad accounts evaluates outcomes from various meals, or various activities together. Thereby, broad accounts allow the individual to find excuses to substitute away from unpleasant activities such as “if I stop studying today, I can make up with a better grade on the next exam, for which I’ll study harder”. Narrow accounts preclude such excuses: “if I stop studying today, I’ll be unhappy because I’ll get a bad grade in the exam”. This ability to substitute between decisions decreases the set of imple-

---

<sup>4</sup>Typing “SMART goals” into a search engine brings up hundreds of hits. The acronym’s first appearance in print is in Doran (1981), but it seems to have been in use earlier.

mentable goals relative to what can be achieved with narrow accounts. As a consequence, optimal self-regulation may call for narrowly evaluated goals. We give conditions when broad accounts are strictly dominated.

First, we clarify that the mere presence of a self-control problem is not sufficient to cause decision substitution. What matters is whether there is a relative bias between tasks. Such a bias arises, for example, if the consumption set involves not only “mundane” goods, but also so-called virtue goods (goods that provide benefits in the future) or vice goods (goods that have future costs). This result helps understand the categories that people adopt when forming mental budgets, such as “entertainment” versus “household expenses”. A relative bias also arises in sequential settings, because the present bias makes costs in a future task feel less painful than costs incurred now. With broad accounts, the decision-making self has an incentive to slack off a bit now and rely on his future self to make up for the short fall. This helps to understand why many diet programs advocate narrow, daily goals for calorie intake. Or why people are more prone to choose vice goods if tasks are repeated than in an isolated choice situation (for experimental evidence see Khan and Dhar 2007). Our analysis however also highlights that decision substitution only arises in sequential task settings if the individual is prone to considering sunk costs. Only if the individual takes into account past outcomes, i.e., only if he is able to leave his broad account open over time, is he willing to compensate for yesterday’s shortfall.

Second, we show that the not so severe self-control problems are the ones for which decision substitution undermines self-regulation, because in these cases there is motivational slack for the future self. In contrast, goals set to deal with a severe self-control problem effectively commit the future self not to go “over target” to compensate for a short fall in the other task. For that reason decision substitution does not arise.

In section 4, we extend the deterministic model and show that our results are robust to goal and account revision. To model how reference points adjust if the individual changes his mind about goals we apply the anticipatory utility approach of Köszegi and Rabin (2009). In further extensions, we explore different assumptions regarding the point in time when the individual evaluates accounts, and we discuss different forms of accounting. For example, in relation to mental accounting, our qualitative results are unchanged or even strengthened if the individual only has accounts for expenditures.

In section 5, we introduce stochastic task outcomes and stochastic reference points. This helps complete the picture. Our above results tell us that narrow psychological accounts often facilitate self-regulation, but also that a person does not always need to make psychological accounts as narrow as possible. And indeed, people do not seem to set accounts for every single item they buy or for every minute they work. Outcome uncertainty provides a counterbalancing rationale for evaluating decisions together. The choice of psychological accounts now has a *direct* impact on the individual’s utility – not just an indirect effect through the decision level as with deterministic outcomes. If one holds decision levels constant, broad accounts provide strictly higher utility than narrow accounts. The reason is that broad accounts allow to pool risks (Gneezy and Potters 1997,

Thaler *et al.* 1997).<sup>5</sup> Think of a student who, despite much studying, does poorly on an exam because of a “bad” day. This causes a psychological loss under narrow accounts. Under broad accounts bad luck in one exam does not always cause a loss, because good luck in another exam can sometimes compensate for this. This *risk-pooling effect* extends to stochastic reference points, as shown by Kőszegi and Rabin (2007).

In the absence of time-inconsistent preferences, the individual strictly prefers broad accounts because of the risk-pooling effect. However, in the presence of motivational problems, narrow accounts can be strictly optimal because accounting affects incentives. As we show, broad accounts make risks in different tasks substitutes, which may weaken incentives (*risk-incentive effects*). To establish these effects, we turn off the potential driving forces for the inferiority of broad accounts considered before: there is no relative bias and our main result is driven by unilateral deviations (so decision substitution, which stems from joint deviations that pay off under a relative bias, cannot play a role).<sup>6</sup> We show that if there is not too much uncertainty about the impact that a decision will have on outcomes, narrow accounts are optimal because they have greater motivating force than broad accounts. Thus, we predict that people facing tasks with little uncertainty, such as routine work, should adopt narrow accounts (e.g., specify a daily work goal). People facing considerable uncertainty (researchers, writers) should instead adopt broad accounts (e.g., a monthly target). Similarly, the result can explain the advice that dieters should not weigh daily: day-to-day weight is subject to considerable fluctuations that are outside the control of the individual.

In section 6, we bring together the ideas of sections 3, 4, and 5. We consider a setting with sequential tasks (i.e., a relative bias arises), outcome uncertainty, and adaptation of reference points. Specifically, we allow the reference distribution to adapt not only to changes of mind but also to the arrival of new information about the state of the world. For example, an individual may learn before he provides effort, how productive his effort will be. This can capture in a stylized way the problem of a cab driver, for whom the effective wage varies with demand shifters such as weather and conventions. He learns at the start of the day whether it is a busy or a not so busy day and then decides how many hours to drive – inducing reference points for income and effort costs.

We show that risk pooling opportunities, the key advantage of broad accounts, disappear. In contrast, the key disadvantage of broad accounts, decision substitution can arise. Decision substitution is central to the argument put forward by Camerer *et al.* (1997) why individuals who can choose their working hours, such as cab drivers, often appear to adopt narrow, daily income targets.<sup>7</sup> Read

---

<sup>5</sup>Thaler *et al.* (1997) illustrate the benefits of risk pooling with a gamble that Paul Samuelson proposed to one of his colleagues:  $(0.5 \circ -\$200; 0.5 \circ \$100)$ . The colleague rejected the proposal, saying that the \$100 loss would hurt him more than he would enjoy the \$200. But he was willing to accept several independent repetitions of the gamble. Assuming a reference point of zero and a coefficient of loss aversion  $\lambda = 2.5$ , the utility from the one-shot gamble is  $1/2(200) + 1/2(-250) < 0$ . In contrast, it pays to accept two independent gambles, because the utility is  $1/4(400) + 1/2(100) + 1/4(-500) > 0$ . Risk pooling allows a gain of 200 in one gamble to offset a loss of 100 in the other gamble, reducing the chances of falling into the loss domain.

<sup>6</sup>Adding a relative bias and considering decision substitution would only strengthen our result that broad accounts are sometimes worse than narrow accounts.

<sup>7</sup>Camerer *et al.* (1997) triggered a lively debate about the estimation of wage elasticities of labor supply (Götte



*et al.* (1999, p.189) explain: “If [cab drivers] had, for example, picked a weekly target they might have been tempted to quit early on any given day, while assuring themselves that they could make up the deficiency later in the week.”

Without such an explanation daily income targets seem puzzling because they lead to negative income elasticities. A broader goal would allow a time-consistent agent to increase earnings and leisure by working fewer hours on days with low wages and compensating with extra hours on days with high wages. By endogenizing the reference point, Kőszegi and Rabin (2006) are able to reconcile several puzzling empirical findings from labor supply data on cab drivers, stadium vendors and bicycle messengers – explaining why there can be both a negative relationship between effort and wages and a positive relationship between the decision to show up for work and expected wages. Yet all of this literature fixes the evaluation horizon to derive empirical implications (see DellaVigna 2009, p.326). Our model endogenizes the evaluation horizon and shows that adopting a narrow account, such as considering a single day, is not an error, but indeed can be optimal.

## **Related literature**

Our paper contributes to the research on goals which has been carried out mainly by psychologists (for an overview see Locke and Latham 1990). Most of these studies investigate the effect of *given* goals (exogenously assigned by an experimenter) on performance in various tasks. We analyze how individuals design and evaluate goals, contributing to the small economics literature on self-regulation through goal setting. The narrow accounting part of our model relates to previous work on goals or promises as reference points in single-task settings by Carrillo and Dewatripont (2008), Suvorov and van de Ven (2008), and Koch and Nafziger (2011). It extends that work, in particular, by analyzing continuous decisions in stochastic environments and allowing for stochastic reference points. Moreover, we are the first to consider goal and account revision. Our main contribution is to develop a theory of endogenous formation of narrow or broad accounts. This goes beyond the previous literature by explaining the aforementioned phenomena of how people psychologically structure their goals in different accounts. Most closely related is Hsiaw (2010b), who – in parallel and independent work – extends her single-task model (Hsiaw 2010a) to study goal bracketing with two sequential, continuous-time optimal stopping problems. The environment and intuitions developed are different from those in our paper. The self-control problem arises from the tension between terminating a project at its current value versus the option value of waiting for a potentially higher payoff. Goals help counter the tendency of the present-biased individual to stop projects too early. Because the decision whether or not to stop a project does not directly influence costs and benefits, decision substitution plays no role in Hsiaw’s setting. Moreover, she assumes that a narrow goal is evaluated as soon as a project stops, whereas broad goals postpone the evaluation until the second project is stopped. Because of discounting, postponing the evaluation makes deviations from the goal in the first task less painful and thus weakens incentives under broad goals. Her comparative static results relating to risk arise from these differences in evaluation horizons. The driving forces in our analysis are different, as we deliberately turn off this timing effect (in section

---

*et al.* 2004, Farber 2005, 2008, Fehr and Gőtte 2007, Crawford and Meng 2011).

4, we discuss what happens if we add it in our model).

Our paper also contributes to the theoretical literature dealing with the question of how present-biased individuals cope with self-control problems (for an overview see Brocas *et al.* 2004). Much of this literature focuses on the role of external commitment technologies for achieving *pre-commitment* (Elster 2000). Only a few contributions deal with intra-personal strategies. Related to the aforementioned papers on goal setting, Suvorov and van de Ven (2008) and Koch and Nafziger (2009b) model the use of self-rewards. Most assume the presence of an internal commitment device. Bisin and Hyndman (2009) and Herweg and Müller (2011) show how the ability to set deadlines helps individuals with time-inconsistent preferences to regulate behavior. Benhabib and Bisin (2005) model the use of neural control processes to inhibit automatic processes that are prone to temptation. Bénabou and Tirole (2004) explain why internal commitment devices can actually work if an individual has imperfect knowledge about his willpower. Our approach applies to different informational environments (perfect vs. imperfect self-knowledge) and relies on different mechanisms (expectations-based reference points vs. self-signaling).

## 2 The Model

**The decisions.** The individual faces two decisions  $x_i \in [0, \infty)$ ,  $i = 1, 2$ , at some date(s)  $t \geq 1$ . Each decision  $i$  involves immediate costs  $-c_i(x_i)$  and benefits  $b_i(x_i)$ , as well as delayed costs or benefits  $h_i(x_i)$ . For instance, a decision may be about consumption of a *vice* good, such as eating potato chips. The individual decides what quantity  $x_i$  of chips to consume, for which he pays  $c_i(x_i) = p_i x_i$ , letting  $p$  be the per-unit price of chips. Chips are tasty when consumed ( $b_i(x_i) > 0$ ) but unhealthy in the long run ( $h_i(x_i) < 0$ ). Or a decision may be about a *virtue* activity that requires immediate costs ( $c_i(x_i) > 0$ ) to achieve long-run benefits ( $h_i(x_i) > 0$ ). Examples are taking an unpleasant medicine (such as self-injecting insulin) and tasks with costly effort (such as studying for an exam, dieting, and working out in the gym). All outcomes are deterministic (we consider stochastic task outcomes in section 5). We allow for the possibility that either  $b_i(x_i)$ ,  $c_i(x_i)$ , or  $h_i(x_i)$  is always zero. But to guarantee a meaningful problem, we require that each decision involves some costs and benefits. For those we assume:

**Assumption 1** *Costs and benefits are strictly increasing and continuously differentiable, with  $b_i(0) = c_i(0) = h_i(0) = 0$ ,  $b'_i(x_i) > 0$ ,  $c'_i(x_i) > 0$  for  $x_i > 0$ ,  $b''_i(x_i) \leq 0$  and  $c''_i(x_i) \geq 0$  (at least one strict inequality); and corresponding conditions on  $h_i(x_i)$ , depending on whether these are delayed costs or benefits. Furthermore,  $b'_i(0) + h'_i(0) - c'_i(0) > 0$  and  $\lim_{x_i \rightarrow \infty} [b'_i(x_i) + h'_i(x_i) - c'_i(x_i)] < 0$ .*

**Goal design.** Before the individual makes the actual decisions at date(s)  $t \geq 1$ , self 0 (the date-0 incarnation of the individual) sets goals for these decisions and plans how their outcomes shall be evaluated: in a broad psychological account, or in narrow psychological accounts.

The *goals*  $\hat{x}_1$  and  $\hat{x}_2$  induce expectations about future costs and benefits. The individual expects that he will experience the cost  $c_i(\hat{x}_i)$ , the benefit  $b_i(\hat{x}_i)$  and, at a later date, the benefit or cost  $h_i(\hat{x}_i)$ . For example, the individual may have the goal of studying 8 hours on Monday for an exam. With this goal he expects that his future self on Monday will not go out, watch TV, ... (the

opportunity costs of studying); and he expects to obtain a good grade on the exam (the delayed benefits). These goal-induced expectations become reference points against which the individual evaluates actual task outcomes.

The type of *psychological account* determines how the individual evaluates outcomes from the two decisions. If self 0 designs narrow accounts, the individual will compare the actual cost of a decision  $c_i(x_i)$  with the expected cost  $c_i(\hat{x}_i)$ ; the actual benefit  $b_i(x_i)$  with the expected benefit  $b_i(\hat{x}_i)$ ; and the actual cost or benefit  $h_i(x_i)$  with the expected outcome  $h_i(\hat{x}_i)$ . On the other hand, if the individual sets a broad account, he will compare the joint costs of the two decisions  $c_1(x_1) + c_2(x_2)$  with the expected joint costs  $c_1(\hat{x}_2) + c_2(\hat{x}_2)$ ; the actual benefits  $b_1(x_1) + b_2(x_2)$  with the expected joint benefit  $b_1(\hat{x}_2) + b_2(\hat{x}_2)$ ; and  $h_1(x_1) + h_2(x_2)$  with the expected joint outcome  $h_1(\hat{x}_2) + h_2(\hat{x}_2)$ . Note that psychological accounts allow the individual to evaluate narrowly or broadly outcomes from different tasks that fall within the same category. Across outcome dimensions we assume separability, following Kőszegi and Rabin (2006). That is, there are separate accounts for costs and benefits. For example, having to give up more leisure time than expected for studying feels like a loss – even though it may yield a better grade than expected, which feels like a separate gain. We discuss this assumption in section 4.2.

**Reference-dependent instantaneous utility.** Following Kőszegi and Rabin (2006), the instantaneous utility is composed of two components. The first component is *consumption* utility which comprises costs and benefits that accrue during the period. It corresponds to the outcome-based utility traditionally studied in economics. The second component is *psychological gain-loss* utility. It arises from the comparison of the realized outcome in a psychological account with its reference point. The individual experiences a psychological gain if an outcome exceeds the goal-induced reference point and a psychological loss if it falls short of the reference point. Gain-loss utility takes the form of Kahneman and Tversky’s (1979) value function. Losses loom larger than gains. For tractability we assume linear loss aversion, as is common in applications (Tversky and Kahneman 1991; see DellaVigna 2009 for applications). Gain-loss utility is defined over outcomes and not over actions, analogous to the traditional outcome-based utility. If an outcome differs from its reference point (measured in consumption-utility units) by  $z$ , the corresponding gain-loss utility is  $\mu(z) = \eta z$  for  $z \geq 0$  and  $\mu(z) = \eta \lambda z$  for  $z < 0$ .<sup>8</sup> The parameter  $\eta \geq 0$  measures the weight of gain-loss utility in the utility function, and  $\lambda > 1$  is the coefficient of loss aversion.

**Present bias.** The individual has a present bias which we model using  $(\beta, \delta)$ -preferences. The first parameter,  $\delta$ , corresponds to the standard exponential discount factor (for simplicity,  $\delta \equiv 1$ ). The second parameter,  $\beta \in (0, 1)$ , captures the extent to which the individual overemphasizes immediate utility flows relative to more distant utility flows. The utility of the individual at date

---

<sup>8</sup>This ensures that the individual feels more pain if he falls short of the reference point in a dimension that he values (e.g., failing the last try on an examination) than if he falls short of the reference point in a dimension that he does not value (e.g., not hitting the waste paper basket when throwing a ball of crumpled paper).

$t \in \{0, 1, \dots, T\}$ ,  $T \geq 2$ , is given by

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

where  $u_t$  is the *instantaneous utility* (specified above). For instance, self 0 weighs future utility flows  $u_1$  and  $u_2$  equally; but self 1 puts a larger relative weight on  $u_1$  because he discounts  $u_2$  with  $\beta < 1$ , reflecting his present bias. As we are interested in modeling deliberate self-regulation through goal design, we assume that the individual knows about his present-biased preferences, i.e., he is *sophisticated* in the sense of O'Donoghue and Rabin (1999).

**Timing.** Self 0 designs goals by setting a goal  $\hat{x}_i$  for each decision and by deciding how to evaluate outcomes (in narrow or broad accounts). No payoff-relevant events occur otherwise. Self 1 then chooses  $x_i$ ,  $i \in \{1, 2\}$ , and experiences immediate benefits  $b_i(x_i)$  as well as immediate costs  $c_i(x_i)$ . The delayed benefits or costs  $h_i(x_i)$  realize at date 2. We also study a sequential decision scenario where self 1 chooses  $x_1$ , self 2 chooses  $x_2$ , and delayed outcomes realize at date 3. Gain-loss utility accrues at date 2, i.e., after all the decisions are completed (at date 3 with sequential decisions). That is, we focus on the effects that goal evaluation has if the point at which the individual perceives gain-loss utility does not change with the type of psychological account. We discuss the implications of alternative timing assumptions in section 4.2.

**Intra-personal conflicts of interest and absolute bias.** While self 0 designs goals, the actual decisions are made by a future self. If self 0 had the choice he would prefer the decision to equate marginal costs and benefits from his perspective. That is, the *preferred decision of self 0 for task  $i$* ,  $x_{0,i}^*$ , satisfies

$$b'_i(x_{0,i}^*) - c'_i(x_{0,i}^*) + h'_i(x_{0,i}^*) = 0. \quad (1)$$

On the other hand, if he was given the choice, the self who carries out decision  $i$  would prefer as goal the decision which equates marginal benefits to marginal costs from his perspective. That is, the *preferred goal of the decision-making self for task  $i$* ,  $x_i^*$ , satisfies

$$b'_i(x_i^*) - c'_i(x_i^*) + \beta h'_i(x_i^*) = 0. \quad (2)$$

The decision-making self discounts delayed outcomes with  $\beta < 1$  relative to the other outcomes, whereas self 0 weights all outcomes equally. As a consequence,  $x_{0,i}^* > x_i^*$  with delayed benefits ( $h_i(x_i) > 0$  for  $x_i > 0$ ) and  $x_{0,i}^* < x_i^*$  with delayed costs ( $h_i(x_i) < 0$  for  $x_i > 0$ ). Hence, an intra-personal conflict of interest arises whenever a decision involves delayed outcomes, and we say the decision-making self has an absolute bias.

**Definition 1 (Absolute bias)** *The decision-making self has an absolute bias in decision  $i$  if the decision involves delayed outcomes, i.e., if  $h_i(x_i) \neq 0$  for  $x_i > 0$ .*

**Implementable decisions and goals.** We assume that the individual has the capacity to set goals and accounts that remain meaningful over time. The goals of self 0 induce expectations. These

expectations serve as reference points for future incarnations of the individual. That is, reference points do not adjust instantaneously if the individual deviates from his goals or revises his goals before a decision. This captures the evidence from the psychology literature that people translate their noncommittal desires into goals which then give a sense of commitment (cf. Gollwitzer and Sheeran 2006). In section 4.1, we extend our model to analyze adjustment of goals and psychological accounts.

Goals, the expectations that they induce, and the individual's decisions must constitute a *personal equilibrium*. In each continuation game, given the expectations induced by the goal  $\hat{x}_i$ , the actual decision  $x_i$  must be consistent with the goal. That is,  $x_i = \hat{x}_i$  in equilibrium. For example, if self 0 sets the goal of studying 8 hours for an exam, then he expects his future self to incur the costs  $c_i(8)$  for 8 hours of studying and looks forward to the benefits  $h_i(8)$  resulting from the exam grade. Given these reference points for study costs and study success, it must indeed be optimal for the future self to study 8 hours. The equilibrium concept captures the idea from the psychology literature that goals have to be realistic.<sup>9</sup>

Any decision or goal that can arise in a personal equilibrium under a narrow account is said to be *narrow-accounting implementable*. Similarly, for broad accounts we call decisions and goals that can arise in a personal equilibrium *broad-accounting implementable*. Among these implementable goals, self 0 picks his *preferred personal equilibrium*, i.e., the combination of goals and accounts that maximizes his utility.<sup>10</sup>

**Notation.** From the perspective of self 0, decision  $i \in \{1, 2\}$  with delayed benefits  $h_i(x_i) \geq 0$  involves *overall* benefits  $B_{0,i}(x_i) = \beta (b_i(x_i) + h_i(x_i))$  and *overall* costs  $C_{0,i}(x_i) = \beta c_i(x_i)$ . Overall benefits and costs from the perspective of self  $t > 0$  are denoted in a similar fashion. In the case of simultaneous decisions, self 1 makes decisions  $i \in \{1, 2\}$  and receives  $B_{1,i}(x_i) = b_i(x_i) + \beta h_i(x_i)$  and  $C_{1,i}(x_i) = c_i(x_i)$  if there are delayed benefits. Analogous definitions apply for delayed costs and for sequential decisions.

### 3 Analysis

Faced with his intra-personal conflicts of interest, self 0 attempts to influence the decisions of his future self through the goals that he designs. For the problem to be interesting, we assume that the decision-making self has an absolute bias in at least one of the two decisions. The cases of delayed benefits and delayed costs are symmetric. Hence, we focus our discussion, propositions, and proofs on the case of delayed benefits ( $h_i(x_i) \geq 0$  for all  $x_i > 0$ , with strict inequality for at least one

---

<sup>9</sup>It also rules out that the individual can make himself arbitrarily happy by lowering his reference point. Being unable to fool himself about his future behavior, the individual cannot systematically generate pleasant surprises. For example, a student who is well prepared for an exam cannot tell himself that he will fail the exam to get a bigger kick out of a good grade.

<sup>10</sup>Note the distinction from the preferred decision of self 0 determined above. The latter solves for a *choice-acclimating personal equilibrium* (CPE, Köszegi and Rabin 2007) where the expectations adapt to a committed decision. In the absence of a present-bias and with deterministic outcomes, self 0 can always implement his preferred decision in an (unacclimating) personal equilibrium.

task). Where relevant we address the mixed case where one task involves delayed benefits and the other involves delayed costs.

### 3.1 Goal setting with narrow psychological accounts

Which decisions can self 0 implement if he sets narrowly evaluated goals? A goal  $\hat{x}_i$  for task  $i$  induces a cost reference point  $c_i(\hat{x}_i)$ , output reference point  $b_i(\hat{x}_i)$ , and a reference point for the delayed benefits  $h_i(\hat{x}_i)$ . Solving backward, we ask when the decision-making self  $t$  has an incentive to deviate from the goal  $\hat{x}_i$  given these reference points. Decision  $x_i$  does not affect the utility from the other task and we suppress it.

If the decision-making self  $t$  sticks to the goal ( $x_i = \hat{x}_i$ ), he meets expectations in that task. So there will be no psychological gain or loss when outcomes are evaluated. That is, the utility for self  $t$  from task  $i$  is  $B_{t,i}(\hat{x}_i) - C_{t,i}(\hat{x}_i)$ . In contrast, falling short of the goal ( $x_i < \hat{x}_i$ ) creates a psychological gain because costs are lower than expected. But it also leads to a psychological loss from falling short of the expected benefits. Gain-loss utility is discounted by  $\beta$ , because outcomes are evaluated in the last period. Hence, the utility for self  $t$  in task  $i$  if he deviates to  $x_i < \hat{x}_i$  is:

$$B_{t,i}(x_i) - C_{t,i}(x_i) + \eta\beta(c_i(\hat{x}_i) - c_i(x_i)) - \eta\beta\lambda(b_i(\hat{x}_i) - b_i(x_i)) - \eta\beta\lambda(h_i(\hat{x}_i) - h_i(x_i)).$$

For goal  $\hat{x}_i$  to be implementable, self  $t$  should have no incentive to lower his decision. The utility from sticking to the goal has to exceed the utility from falling short of it, or equivalently,

$$\begin{aligned} & (1 + \eta\beta\lambda)b_i(\hat{x}_i) + \beta(1 + \eta\lambda)h_i(\hat{x}_i) - (1 + \eta\beta)c_i(\hat{x}_i) \\ \geq & (1 + \eta\beta\lambda)b_i(x_i) + \beta(1 + \eta\lambda)h_i(x_i) - (1 + \eta\beta)c_i(x_i). \end{aligned} \quad (3)$$

Inequality (3) holds for any goal  $\hat{x}_i$  that is not “too high”:  $\hat{x}_i \leq x_{max,i}^N$ , where  $x_{max,i}^N$  is defined by

$$(1 + \eta\beta\lambda)b'_i(x_{max,i}^N) + \beta(1 + \eta\lambda)h'_i(x_{max,i}^N) = (1 + \eta\beta)c'_i(x_{max,i}^N). \quad (4)$$

What motivates self  $t$  is the fear of facing a loss in the benefit dimension should he fall short of the goal. Because of loss aversion, these losses in the benefit dimensions count more than gains in the cost dimensions – as long as the goal does not exceed  $x_{max,i}^N$ . Similarly, the decision-making self has no incentive to deviate and surpass the goal if it is not “too low”:  $\hat{x}_i \geq x_{min,i}^N$ , where  $x_{min,i}^N$  is defined by

$$(1 + \eta\beta)b'_i(x_{min,i}^N) + \beta(1 + \eta)h'_i(x_{min,i}^N) = (1 + \eta\beta\lambda)c'_i(x_{min,i}^N). \quad (5)$$

Hence, there exists a continuum of narrow-accounting implementable goals  $[x_{min,i}^N, x_{max,i}^N]$ . Among these, self 0 picks as his goal the narrow-accounting implementable goal that maximizes his utility:

$$\hat{x}_i = \operatorname{argmax}_{x_i \in [x_{min,i}^N, x_{max,i}^N]} B_{0,i}(x_i) - C_{0,i}(x_i). \quad (6)$$

For a task without an absolute bias, there is no intra-personal conflict of interest. So self 0 sets the goal equal to his preferred decision  $x_{0,i}^*$ . For a task with an absolute bias, self 0 can implement his preferred decision  $x_{0,i}^*$  whenever  $x_{0,i}^* \in [x_{min,i}^N, x_{max,i}^N]$ . When does this arise? Note that  $x_{min,i}^N < x_{0,i}^*$

always holds. So self 0 can fully overcome his self-control problem whenever  $x_{0,i}^* \leq x_{max,i}^N$ . The larger  $\beta$ , the more the selves agree on perceived delayed costs and benefits. And the more they agree, the more likely it is that self 0 can implement his preferred decision. This is best illustrated with the special case where all benefits are delayed. While the preferred decision of self 0 solves  $h'_i(x_{i0}^*) = c'_i(x_{i0}^*)$ , the maximal implementable goal solves  $\beta \frac{1+\eta\lambda}{1+\eta\beta} h'_i(x_{max,i}^N) = c'_i(x_{max,i}^N)$ . Hence, for values of  $\beta$  that satisfy  $\eta(\lambda - 1) \geq (1 - \beta)/\beta$ , we have  $x_{max,i}^N \geq x_{0,i}^*$ . Then self-regulation through goal setting is successful.

Self-regulation is constrained if the individual faces a more severe self-control problem – because  $\beta$  is so small that  $x_{max,i}^N < x_{0,i}^*$ . In this case, the best self 0 can do is to set goal  $\hat{x}_i = x_{max,i}^N$ . Nevertheless, self 0 can nudge his future self to a more ambitious decision than self 1 would want on his own. The maximal implementable goal exceeds the preferred goal of the decision-making self,  $x_i^*$ , defined by equation (2).<sup>11</sup> The reason is that having a goal creates an “attachment” to the outcomes that the goal implies. This attachment can push self 1 to choose a decision that exceeds  $x_i^*$ .

### Proposition 1

1. *There exists a range of narrow-accounting implementable goals  $[x_{min,i}^N, x_{max,i}^N]$ .*
2. *The preferred goal of the decision-making self for task  $i$  satisfies  $x_i^* < x_{max,i}^N$ . The preferred decision of self 0 for task  $i$  satisfies  $x_{0,i}^* > x_{min,i}^N$  and there exists a cut-off  $\tilde{\beta} \in (0, 1)$  such that  $x_{0,i}^* \leq x_{max,i}^N$  for  $\beta \geq \tilde{\beta}$ . Self 0 chooses as his goal  $\hat{x}_i = \min\{x_{0,i}^*, x_{max,i}^N\}$ .*

## 3.2 Broad versus narrow accounts

This section develops one of the main themes of the paper, namely that the possibility of decision substitution is a key disadvantage of goal setting with broad accounts. First, we illustrate this in a simple, repeated work task setting in which effort costs in the future task feel less painful than effort costs now. As we show next, more generally, problems of decision substitution arise whenever a *relative bias* exists between the decisions.

### 3.2.1 Sequential effort decisions: The problem of effort substitution

Working life is full of tasks that require costly effort on different dates to achieve future benefits. Examples are studying for exams, working on long-term projects, or working as a freelancer (as authors, journalists, and cabdrivers often do). Similarly, it takes repeated investments in the form of physical exercise and good dietary choices to maintain a healthy life-style.

We capture these aspects in a stylized setting, where the individual faces a twice repeated task that requires costly effort to produce some future output. Self 1 works at date 1 and incurs immediate effort costs  $c(x_1)$ . Self 2 works at date 2 and incurs immediate effort costs  $c(x_2)$ . There are no

---

<sup>11</sup>Note that  $x_{min,i}^N > x_i^*$  can arise for  $\lambda\beta < 1$ . Redefining the the preferred goal of self 1 as  $\max\{x_i^*, x_{min,i}^N\}$  would not affect the presence of an intra-personal conflict of interest, and would not change our analysis.

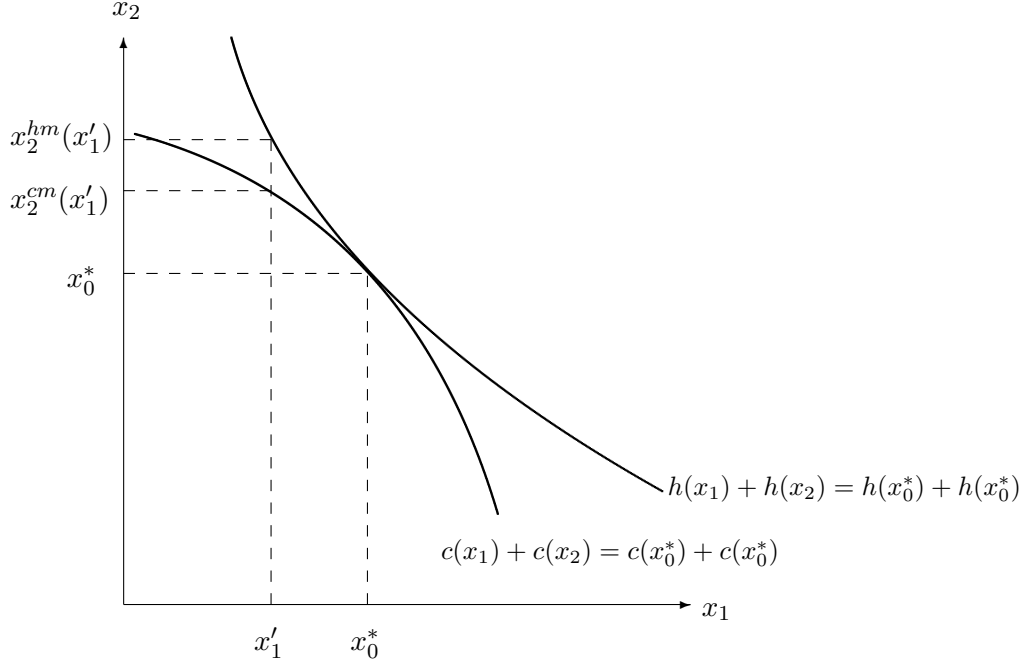


Figure 1: Illustration of effort substitution

immediate benefits. At date 3, delayed benefits  $h(x_1)$  and  $h(x_2)$  realize, and the individual experiences gain-loss utility from the evaluation of outcomes. Because tasks are symmetric, we drop subscripts whenever possible.

The range of narrow-accounting implementable goals is the same as in section 3.1. Because outcomes for different efforts are evaluated in isolation from each other, it does not matter whether effort decisions are made sequentially or simultaneously. Can the same goals be implemented with broad accounts? To analyze this question, we solve backward – starting with self 2 who faces the second task.

**Self 2.** Suppose the individual has the goal of exerting  $x_{min}^N < x_0^* < x_{max}^N$  in each task. Consider first how self 2 chooses after self 1 sticks to the goal for task 1. The actual costs and benefits from task 1 match their reference levels and cancel out from gain-loss utility. So the problem of self 2 looks exactly like that under narrow accounting. Thus, self 2 sticks to his goal for task 2.

Suppose that self 1 works less hard than he was supposed to and deviates to  $x_1' < x_0^*$ . If self 2 sticks to his goal by exerting  $x_2 = x_0^*$ , he suffers a psychological loss from falling short of the reference point in the benefit dimension; and he perceives a gain in the cost dimension. His utility is:  $\beta [h(x_1) + b(x_0^*)] - c(x_0^*) + \eta \beta [c(x_0^*) - c(x_1)] + \eta \beta \lambda [h(x_1) - b(x_0^*)]$ . Because the loss in the benefit dimension weighs more heavily on the margin than the psychological gain in the cost dimension, it pays off for self 2 to substitute at least partially for the shortfall created by self 1 and provide  $x_2 > x_0^*$ .



However, self 2 does not make up fully for the shortfall in benefits. Compensating for the loss in the benefit dimension requires that self 2 puts in such an amount of effort that overall costs exceed expected costs. The reason is that the *benefit-matching effort decision*  $x^{hm}(x'_1)$  – defined by  $h(x'_1) + h(x^{hm}(x'_1)) = h(x_0^*) + h(x_0^*)$  – exceeds the *cost-matching effort decision*  $x^{cm}(x'_1)$  – defined by  $c(x'_1) + c(x^{cm}(x'_1)) = c(x_0^*) + c(x_0^*)$ . This follows because benefits are concave and costs are convex in effort (see figure 1). Increasing effort beyond the cost-matching level does not pay off because self 2 would fight a marginal loss in the benefit dimension at the expense of a bigger marginal loss in the cost dimension.

One way to avoid effort substitution is to set a goal that pushes self 2 as much as possible. It follows directly from our single-task analysis that self 2 will never provide more effort than the maximal narrow-accounting implementable goal  $x_{max}^N$ . If self 2 faces goal  $x_{max}^N$  under broad accounts, he is credibly committed not to exceed this goal – even if self 1 deviates.

**Lemma 1** *Consider a twice repeated task with immediate costs and delayed benefits. Suppose self 1 deviates from his goal and provides  $x_1 < \hat{x}_1$ .*

1. *Suppose  $x_{max}^N \leq x_0^*$ . Faced with goal  $x_{max}^N$  under broad accounts, self 2 does not respond to the deviation of self 1.*
2. *Suppose  $x_{max}^N > x_0^*$ . Faced with goal  $x_0^*$  under broad accounts, self 2 responds to the deviation of self 1 with  $x_2 = \min\{x^{cm}(x'_1), x_{max}^N\} > x_0^*$ . Effort substitution is partial:  $x_2$  never fully compensates for the lost benefits from task 1.*

**Self 1.** Does it pay off for self 1 to deviate from the goal  $x_0^* \in (x_{min}^N, x_{max}^N)$  and rely on self 2 to partially make up for the shortfall? The answer is yes. To provide some intuition, consider a marginal downward deviation  $x'_1 = x_0^* - \epsilon$ , to which self 2 responds by matching costs. Because of this response, the deviation causes no gain-loss utility in the cost dimension. However, self 1 can reduce the costs that he incurs. He can save  $c'(x_0^*)$  on the margin by “borrowing” effort from his future self at “price”  $\beta c'(x_0^*)$ . The fact that self 2 only partially compensates for the shortfall in benefits does not stop self 1 from deviating. The loss in the benefit dimension is only of second order. Therefore, self 1 will deviate at least a little bit from  $x_0^*$  to achieve first-order savings on his effort costs.

**Proposition 2** *Consider a twice repeated task with immediate costs and delayed benefits, and suppose that  $x_{min}^N < x_0^* < x_{max}^N$ . Because  $x_0^*$  is narrow-accounting implementable but not broad-accounting implementable for both tasks, goal setting with narrow accounts is strictly optimal.*

### 3.2.2 Simultaneous decisions: Relative bias as cause for decision substitution

The sequential-decision setting illustrates how self 1’s biased view of the relative benefits and costs in the two tasks may create opportunities for decision substitution with broad accounts. We now show that a relative bias between tasks is the key driving force for broad accounts to do strictly worse than narrow accounts.

## Broad accounts cannot do better than narrow accounts

The marginal incentives for deviations from the goal in a single task (*unilateral* deviations), are the same under broad and narrow accounts. But with broad accounts, self 0 additionally has to make sure that the decision-making self does not have a profitable *joint* deviation. For instance, our sequential-decision setting shows that it may pay off to lower one decision relative to the goal and compensate by increasing the other decision. The additional constraint to prevent joint deviations weakly reduces the range of implementable goals. Broad accounts thus cannot improve self-regulation in settings with deterministic outcomes.

**Lemma 2** *A necessary condition for goals  $(x_1, x_2)$  to be broad-accounting implementable is that  $x_1$  and  $x_2$  are narrow-accounting implementable.*

In the following, we ask when broadly evaluated goals can replicate successful self-regulation with narrow accounts, and under what circumstances broad accounts perform strictly worse. That is, we focus on the case  $x_{min,i}^N < x_{0,i}^* < x_{max,i}^N$ . If  $x_{0,i}^*$  is not implementable, self 0 sets goal  $x_{max,i}^N$ . For sequential tasks, narrow and broad accounts then do equally well. For simultaneous tasks, matters are more complicated. In appendix A.5, we provide a sufficient condition for broad accounts to perform equally well as narrow accounts in settings where  $x_{0,i}^* > x_{max,i}^N$ .

### When do narrow and broad accounts perform equally well?

Narrowly evaluated goals  $x_{0,i}^* \in (x_{min,i}^N, x_{max,i}^N)$  allow the individual to overcome his self-control problem. Self 0 can implement the same decisions under broad accounts if self 1 and self 0 perceive in the same way the trade-off between overall costs and overall benefits in the two decisions. That is, if

$$\frac{B'_{0,1}(x_1)}{B'_{0,2}(x_2)} = \frac{B'_{1,1}(x_1)}{B'_{1,2}(x_2)} \quad \text{and} \quad \frac{C'_{0,1}(x_1)}{C'_{0,2}(x_2)} = \frac{C'_{1,1}(x_1)}{C'_{1,2}(x_2)} \quad \text{for all } (x_1, x_2). \quad (7)$$

In geometric terms, condition 7 means that the (overall) iso-benefit and iso-cost curves for self 0 and self 1 overlap. This happens, for example, when both decisions involve only delayed benefits and immediate costs. We use this case to illustrate why there is no profitable joint deviation under broad accounts if condition 7 holds.

Self 1 has no incentive to change both decisions in the same direction. Certainly, jointly reducing both decisions would increase the consumption utility of self 1 – because of the intra-personal conflict of interest. But the effect on gain-loss utility is the same as that of two unilateral deviations, because the individual is not compensating for the shortfall in benefits in one task through higher benefits in the other task. Such a joint deviation does not pay off, because a goal  $x_{0,i}^* < x_{max,i}^N$  deters unilateral deviations.

Self 1 has no incentive to increase one decision and decrease the other decision. At  $(x_{0,1}^*, x_{0,2}^*)$ , the iso-cost and iso-benefit curves of both self 0 and self 1 are tangent. This implies that any joint deviation globally lowers consumption utility of self 1. Either it lowers benefits by more than the saved costs, or it increases costs by more than the benefits. Why any joint deviation also globally lowers gain-loss utility is not straightforward. To build intuition consider one particular deviation,

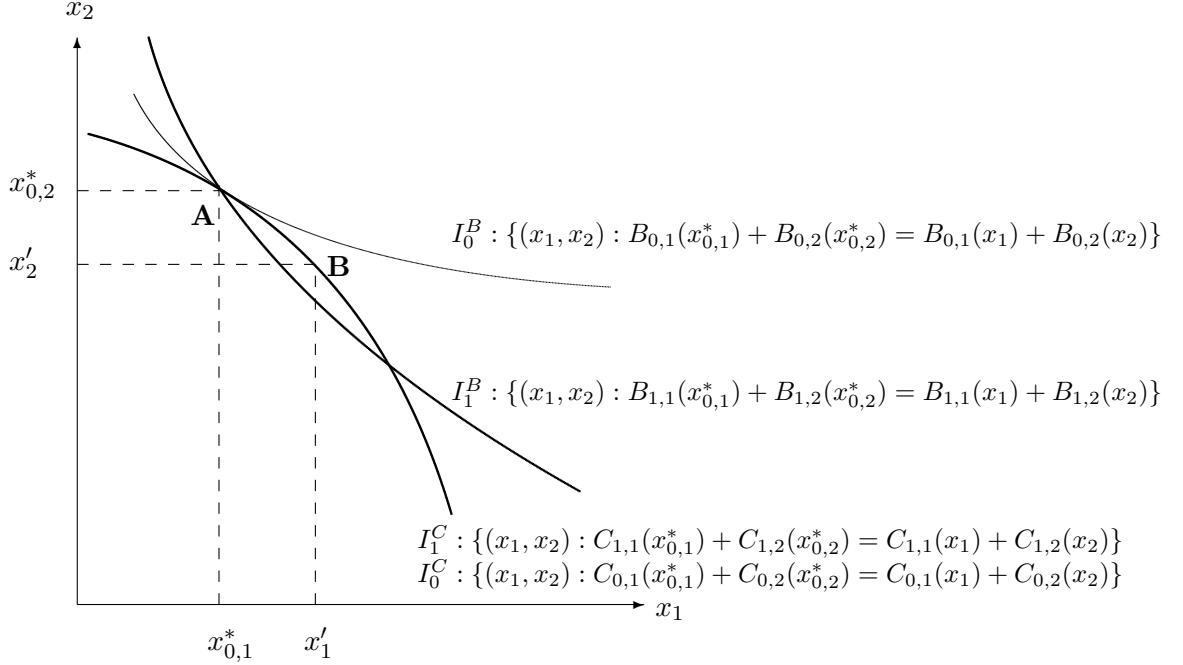


Figure 2: Decisions with a relative bias: Deviations can increase self 1's consumption utility

namely a move along the iso-cost curve in figure 1, to  $x'_1 < x_{0,1}^*$ . While such a move ensures that goal-induced expectations in the cost dimension are met, it leads to lower benefits. To match the shortfall decision 1 causes in the benefit dimension, self 2 has to provide  $x_2 = x^{hm}(x'_1)$ . That is, self 2 has to increase costs beyond the cost-matching level, reached at  $x^{cm}(x'_1)$ .

### Successful self-regulation cannot always be replicated with broad accounts

While narrowly evaluated goals  $x_{0,i}^* \in (x_{min,i}^N, x_{max,i}^N)$  allow for successful self-regulation, it may be impossible to replicate these goals with broad accounts if condition (7) fails to hold at  $(x_{0,1}^*, x_{0,2}^*)$ .

**Definition 2 (Relative bias)** *There is a relative bias between decisions if*

$$\frac{B'_{1,1}(x_{0,1}^*)}{B'_{1,2}(x_{0,2}^*)} \neq \frac{C'_{1,1}(x_{0,1}^*)}{C'_{1,2}(x_{0,2}^*)}.$$

A relative bias means that, from the perspective of self 1, the marginal rates of substitution between costs and benefits are not equalized at the preferred decisions of self 0. A necessary condition for a relative bias to arise is that there are some delayed outcomes and that there is some asymmetry between the two tasks. Examples are situations where only one of the tasks involves delayed outcomes (either costs or benefits); or where task occur sequentially (such as in section 3.2.1).

Figure 2 illustrates how a relative bias can give rise to a profitable joint deviation. Point A, where the iso-cost ( $I_0^C$ ) and iso-benefit ( $I_0^B$ ) curves of self 0 are tangent, describes the preferred decisions of self 0. Self 0 and self 1 agree on the trade-off between the two decisions in the cost dimension: their iso-cost curves,  $I_0^C$  and  $I_1^C$ , overlap. However, the selves disagree on the trade-off in the

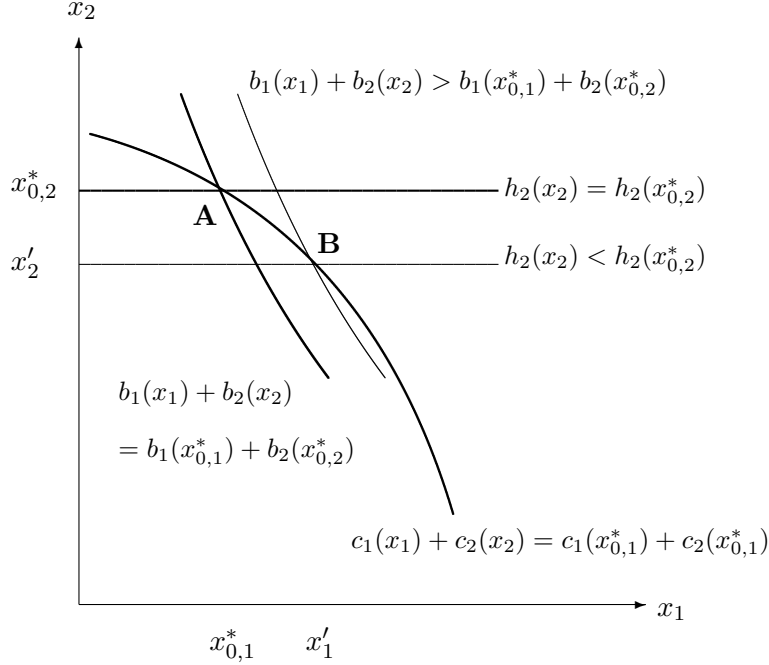


Figure 3: Decisions with a relative bias: Gains/losses in cost and benefit dimensions

benefit dimension. Because of the relative bias, the iso-benefit curve from the perspective of self 1 ( $I_1^B$ ) intersects with his iso-cost curve in point  $A$ . Self 1 thus can increase his consumption utility by favoring the decision toward which he is relatively biased. In geometric terms, this means moving south-east from  $(x_{0,1}^*, x_{0,2}^*)$  into the lens spanned by the iso-benefit and iso-cost curves. For example, self 1 can increase his total benefits while keeping costs constant by moving along the iso-cost curve to point  $B$ .

Gains in consumption utility however are accompanied by some psychological losses. Any increase in overall benefits results from a loss in one benefit dimension (say immediate benefits) and a gain in the other, separately evaluated benefit dimension (delayed benefits). This can most easily be seen for the case where only one of the decisions involves a delayed outcome, say a delayed benefit in decision 2. As figure 3 illustrates, moving along the iso-cost curve from point  $A$  to  $B$  allows to increase immediate benefits, but only at the cost of lower delayed benefits. More generally, starting from  $(x_{0,1}^*, x_{0,2}^*)$  it is impossible to increase one type of benefit without raising costs or decreasing the other type of benefit.

Overall, the positive effect a deviation has on consumption utility dominates if the loss in one of the benefit dimensions is not felt too heavily, i.e., if  $\lambda$  is not too high. In this case, a deviation pays off and  $(x_{0,1}^*, x_{0,2}^*)$  cannot be implemented under broad accounts. The next result provides a sufficient condition.

**Proposition 3** *Suppose  $x_{min,i}^N < x_{0,i}^* < x_{max,i}^N$ , i.e.,  $x_{0,i}^*$  is narrow-accounting implementable.*

1. *A sufficient condition for  $(x_{0,1}^*, x_{0,2}^*)$  to be broad-accounting implementable is that condition (7) holds, i.e., that selves 0 and 1 trade-off the two decisions in the same way. Then narrow*

and broad accounts are both optimal.

2. A sufficient condition for  $(x_{0,1}^*, x_{0,2}^*)$  not to be broad-accounting implementable is that decisions involve a relative bias and that  $(1 - \beta)/\beta > \eta(\lambda - 1)$ . Then narrow accounts are strictly optimal.

### 3.3 Discussion

Broad accounts make decision levels in different periods substitutes. This undermines self-regulation. The temptation not to work today is exacerbated by the consideration that not working today will induce the future self to work more and at least partially make up for the short fall. Experimental evidence by Khan and Dhar (2007) illustrates such substitution. They confront subjects with a virtue-vice consumption decision: subjects have to choose between an unhealthy and a healthy snack as “compensation” for filling in a survey. Subjects choose the vice good more often in a repeated choice treatment than in an isolated choice treatment (70 percent vs. 48 percent,  $p < 0.05$ ). In the former, they have the opportunity to make up for choosing the vice good today by choosing the virtue good next time. In treatments where the next reward is fixed, even more subjects choose vice in the virtue-next condition (90 percent,  $p < 0.05$ ), while the fraction of vice choices in the vice-next condition is not significantly different from that in the isolated choice treatment (43 percent). Khan and Dhar (2007) conclude that the repeated choice scenario – which one can think of as inducing a broad account – is detrimental to self-control (they examine different explanations such as variety seeking and show that these are not the driving forces). In a similar vein, Fishbach *et al.* (2006) observe that anticipated progress toward a fitness goal increases the likelihood of indulgence in unhealthy food – consistent with decision substitution. These observations cannot be explained with willpower depletion models, because the effect comes from thinking about future choices rather than from having resisted a current temptation.

The phenomenon of decision substitution – caused by a relative bias – helps to understand how people structure their accounts. When deciding on what decisions to evaluate jointly, proposition 3 suggests that a person should ask himself whether he is relatively biased toward one of the decisions. For example, he should not group vice and virtue goods together. This would lead to overconsumption of vice goods and underconsumption of virtue goods. But he might well group several vice goods, or several virtue goods together.

How narrowly defined the specific categories need to be depends on the consumption preferences of the person and the relative trade-offs that these preferences imply. For example, a generic psychological account for food might work well for people who are not tempted to overeat on unhealthy food (because, despite their present bias, they quickly reach a point of satiation). But it will be a disaster for a person struggling with a diet. Such a person is better off adopting narrower categories such as “fruits and vegetables” and “sweets”, or “main meals” and “snacks”.<sup>12</sup> The idea of preventing substitution toward some categories of food, or over time, is also reflected in many diet programs. For example, Weight Watchers’ PointsPlus<sup>TM</sup> system assigns points to food

---

<sup>12</sup>A goal could be to follow the American Cancer Society’s advice to eat 5 servings of fruits and vegetables a day.

based on their nutritional content. It aims to support a balanced diet through a personalized, daily points goal. A similar intuition applies for spending on food and drinks. At a restaurant dinner it is easier for most people to get carried away on drinks than on food. And indeed consumers appear to adopt distinct accounts for food and drinks and hence do not treat them as substitutes (see the field experiment by Abeler and Marklein 2010).

Whether decision substitution arises also depends on the characteristics of the decisions. It is quite natural that a person can substitute in his head “similar” outcomes (e.g., money spent on different items while grocery shopping). A series of experiments by Khan and Dhar (2007) confirms this and, at the same time, suggests that substitution does not work for distinct outcome categories. The separability assumption of Kőszegi and Rabin (2006) captures this idea. To illustrate the consequences of this assumption, suppose one decision involves a delayed health benefit only (say taking fluoride tablets to prevent future tooth decay) and the other decision involves an immediate benefit only (say going to see a movie). While immediate monetary outlays can be measured in the same account, the health benefit and the joy of watching the movie are perceived as distinct dimensions. So the individual effectively has narrow accounts in the two benefit dimensions. And these narrow accounts in the benefit dimension already prevent decision substitution as our results shows. To see this consider the sufficient condition for decision substitution in proposition 3. It requires, in addition to a relative bias, that

$$(1 - \beta)/\beta > \eta(\lambda - 1). \tag{8}$$

This condition always fails if one of the decisions involves no current benefit (i.e.,  $b'_i(x_i) = 0$  for one  $i$ ). To see this suppose  $x_{0,i}^* \in (x_{min,i}^N, x_{max,i}^N)$ , i.e., goals  $x_{0,i}^*$  are narrow-accounting implementable. Using  $h'_i(x_{0,i}^*) = c'_i(x_{0,i}^*)$ , equation (4) together with  $x_{0,i}^* < x_{max,i}^N$  implies  $(1 - \beta)/\beta \leq \eta(\lambda - 1)$ . In contrast, if both decisions involve immediate and delayed benefits, condition (8) holds for an individual with a strong present-bias (i.e., low  $\beta$ ). In this case, the goals are not broad-accounting implementable.

The above examples illustrate the role of the separability assumption for different kinds of benefits. If benefits are perceived as distinct categories, decision substitution does not arise even though the individual has a broad account for monetary outlays. A similar argument holds for benefits that occur at different points in time: Whether decision substitution arises over time depends on whether or not the individual can choose to integrate immediate and delayed outcomes. We assume integrability over time in section 3.2.1 and return to a discussion of this issues in section 4.2.

Finally, our analysis reveals that grouping decisions with a relative bias together does not automatically create a problem of decision substitution. The condition  $x_{0,i}^* < x_{max,i}$  shows that it is the not so severe self-control problem for which decision substitution undermines incentives, because in these cases there is motivational slack for the future self. In contrast, goals set to deal with a severe self-control problem effectively commit the future self not to go “over target” to compensate for a short fall in the other task. Hence, decision substitution does not arise.

## 4 Extensions of the no-uncertainty model and robustness

In this section, we discuss different forms of accounting, alternative assumptions on when gain-loss utility is felt, and how our results are affected by possibilities to adjust goal design.

### 4.1 Goal revision

A premise of our analysis is that people have the capacity to set goals for themselves which remain meaningful over time. To model this we assumed that reference points do not adjust if an individual revises his goals before a decision. We now explore the robustness of our results if goal revisions have an effect on reference points.<sup>13</sup> The following extension of our model, based on Kőszegi and Rabin (2009), considers simultaneous decisions. Section 6 covers sequential decisions and reference point adjustment in response to the resolution of uncertainty.

#### 4.1.1 Adjusting goal levels

To allow for goal adjustment, we split a decision-making period  $t$  into two subperiods, where the individual feels the same way about task benefits and costs.<sup>14</sup> Self 0 designs goals  $\hat{x}_{0,i}$ . At date  $t.I$ , the individual can adjust goals. Revised goals  $\hat{x}_{t,i}$  become anchored in the individual's head and induce new reference points in future evaluation of outcomes. At date  $t.II$ , the individual makes the decisions. All delayed outcomes occur at some future date  $\tau$ . Goals and goal revisions have to be rational, i.e., they must be consistent with personal equilibrium.

Changes in beliefs are carriers of utility according to the evidence reviewed by Kőszegi and Rabin (2009). Applying their modeling approach, we assume that the individual derives “anticipatory” gain-loss utility from changes in goals (in addition to consumption utility and “contemporaneous” gain-loss utility). Specifically, the individual compares future outcomes that arise under the new goal  $\hat{x}_{t,i}$  with the outcomes that would occur under the past goal  $\hat{x}_{0,i}$ . This causes anticipatory gain-loss utility  $\gamma_{t,t} \mu(b(\hat{x}_{t,i}) - b(\hat{x}_{0,i})) + \gamma_{\tau,t} \mu(h(\hat{x}_{t,i}) - h(\hat{x}_{0,i})) + \gamma_{t,t} \mu(c(\hat{x}_{0,i}) - c(\hat{x}_{t,i}))$ . The parameter  $\gamma_{t,t} \leq 1$  is the weight self  $t.I$  attaches to anticipatory gain-loss utility from outcomes that occur at date  $t.II$ ; and  $\gamma_{\tau,t} \leq 1$  is the weight attached to anticipatory gain-loss utility for outcomes that occur at date  $\tau > t$ .<sup>15</sup>

---

<sup>13</sup>How fast reference points adjust to revised goals is an open question. There is some mixed evidence on adjustment to new information. Bettors who accumulate losses tend to make riskier bets at the end of the day to erase these losses, suggesting that their reference points do not adapt to losses from previous bets (see Camerer 2003, p.296). With time, reference states seem to adjust, as reflected by the phenomenon of *hedonic adaptation* (see Frederick and Loewenstein 2003). Experimental evidence on short-term responses is inconclusive: Matthey and Dwenger's (2008) findings suggest no short-term adjustment of reference states, Gill and Prowse's (2011) suggest quick adjustment.

<sup>14</sup>The  $\beta - \delta$  model approximates the property of the hyperbolic discount function (e.g., Ainslie 2001, Loewenstein and Prelec 1992) that discount rates decline over time. But it does not capture the fact that close to the decision date the individual agrees more with the decision-making self's point of view than with that of self 0. The extension mimics the latter property.

<sup>15</sup>It seems plausible to assume that a worsening of expectations about a date far away in the future is less painful than a worsening of expectations about a closer date, i.e., that  $\gamma_{\tau,t}$  decreases in  $\tau - t$  (cf. Kőszegi and Rabin 2009). It is a matter of interpretation of  $\gamma_{t,t}$  whether any discounting occurs between dates  $t.I$  and  $t.II$ , i.e., whether the

Consider an individual who faces narrow accounts. For self  $t.II$  our results from section 3.1 on what (possibly revised) goals are implementable continue to hold. With similar arguments applied to the anticipatory gain-loss utility, self  $t.I$  has no incentive to revise the inherited goal for task  $i$  downward if it does not exceed  $\tilde{x}_{max,i}^N$ , defined by

$$(1 + \eta \gamma_{t,t} \lambda) b'_i(\tilde{x}_{max,i}^N) + (\beta + \eta \gamma_{\tau,t} \lambda) h'_i(\tilde{x}_{max,i}^N) = (1 + \eta \gamma_{t,t}) c'_i(\tilde{x}_{max,i}^N). \quad (9)$$

What changes compared to  $x_{max,i}^N$  defined in equation (4) are the factors by which different gain-loss utility components are weighted. Gain-loss utility from immediate outcomes receives weight  $\gamma_{t,t}$  and that from delayed outcomes receives weight  $\gamma_{\tau,t}$ . To derive  $\tilde{x}_{max,i}^N$  we assume that the individual evaluates all outcomes in the last period and hence discounts by  $\beta$  these future experiences of then contemporaneous gain-loss utility. Depending on whether  $\gamma_{t,t}$  and  $\gamma_{\tau,t}$  are smaller or larger than  $\beta$ ,  $\tilde{x}_{max,i}^N$  may be smaller or larger than  $x_{max,i}^N$ . To deter downward goal adjustments and decisions that deviate from the goal, self 0 therefore cannot set a goal that exceeds  $\min\{x_{max,i}^N, \tilde{x}_{max,i}^N\}$ . Similarly, the goal cannot fall below  $\max\{x_{min,i}^N, \tilde{x}_{min,i}^N\}$ .<sup>16</sup>

Overall, we see that the possibility of goal adjustment does not necessarily restrict goal setting by self 0. The larger  $\gamma_{t,t}$  and  $\gamma_{\tau,t}$ , the larger  $\tilde{x}_{max,i}^N$ . Intuitively, the worse the individual feels after revising his goals, the less likely is it that goal revision hampers self-regulation. In contrast, if  $\gamma_{t,t} = \gamma_{\tau,t} = 0$ , the individual has no anticipatory feelings from changing his goals and goal adjustment is without cost. In this case, goals are not effective in overcoming self-control problems and the choice of psychological accounts does not matter.

In the presence of anticipatory gain-loss utility, we still find that narrow accounts can be strictly optimal for simultaneous decisions. Only the sufficient condition in proposition 3, part 2 needs to be adjusted. If contemporaneous gain-loss utility is experienced at the time when the corresponding payoffs are experienced, the condition becomes

$$(1 - \beta) > \eta (\lambda \gamma_{\tau,t} - \gamma_{t,t}). \quad (10)$$

#### 4.1.2 Adjusting accounts

Suppose that the preferred decisions of self 0 are implementable under narrow accounts but not under a broad account. If he inherits narrow accounts, self  $t.I$  may want to shift to broad accounts so that these become anchored in the individual's head by date  $t.II$  and make decision substitution possible for self  $t.II$ . Anticipating these behavioral consequences self  $t.I$  must also revise goals.<sup>17</sup> Revision of goals and accounts by self  $t.I$  immediately triggers anticipatory feelings from comparing outcomes under his new goals with those that would occur under the past goals. Because the individual only just started thinking about his new accounts, he feels anticipatory gain-loss utility in the old, narrow accounts. The only way to avoid the (negative) anticipatory feelings is not to

---

feelings are anticipatory or contemporaneous.

<sup>16</sup>The cutoff  $\tilde{x}_{min,i}^N$  is defined by  $(1 + \eta \gamma_{t,t}) b'_i(\tilde{x}_{min,i}^N) + (\beta + \eta \gamma_{\tau,t}) h'_i(\tilde{x}_{min,i}^N) = (1 + \eta \gamma_{t,t} \lambda) c'_i(\tilde{x}_{min,i}^N)$ .

<sup>17</sup>Decisions and beliefs must constitute a personal equilibrium in every contingency, i.e., they must do so on and off the equilibrium path.



revise accounts. Specifically, because expectations will be met in the continuation game starting at date  $t.II$ , there is only anticipatory gain-loss utility. The account revision problem of self  $t.I$  therefore boils down to the goal adjustment problem under narrow accounts. Above we showed that the individual has no incentive to revise goals as long as these fall within a certain range.

**Corollary 1** *Suppose revisions of goals and accounts trigger anticipatory utility.*

1. *There exists a range of narrow-accounting implementable goals  $\left[ \max\{x_{min,i}^N, \tilde{x}_{min,i}^N\}, \min\{x_{max,i}^N, \tilde{x}_{max,i}^N\} \right]$ .*
2. *Suppose the preferred decisions of self 0 are narrow-accounting implementable. A sufficient condition for them not to be broad-accounting implementable is that decisions involve a relative bias and that inequality (10) holds. Then narrow accounts are strictly optimal.*

## 4.2 Evaluating and closing accounts

### 4.2.1 Timing of gain-loss utility

To facilitate the exposition and comparison across cases, we use as lead scenario that the individual closes all accounts at the last date, i.e., he experiences gain-loss utility once outcomes from all tasks are settled. Depending on the application, alternative assumptions may be more realistic. To model situations where the reference points adjust after deviations by the decision-making self, one can apply our extension from section 4.1 and assume that the individual feels anticipatory gain-loss utility from such adjustments. Similarly, the extension allows us to capture situations where the individual feels gain-loss utility at the time he experiences the relevant consumption-utility component. For this, set  $\gamma_{t,\tau} = \beta$  and  $\gamma_{t,t} = 1$  in equations (9) and (10). Compared with our main model, where  $\gamma_{t,\tau} = \gamma_{t,t} = \beta$ , these changes affect  $x_{max,i}^N$  and the sufficient condition in part 2 of proposition 3. Still there is a range of implementable goals, and decision substitution can arise. For sequential decisions two additional effects kick in. First, a broad account implicitly means that the individual is able to leave the account open until all information relevant for the account arrives: the earliest date when the broad account can be closed is when self 2 makes his decision. We discuss this assumption in the next subsection. Second, different timing assumptions affect the best response of self 1 and 2 under broad accounts if the respective other self sticks to his goal. While the maximal implementable goal for self 2 is equal to the one under narrow accounts ( $x_{max,2}^B = x_{max,2}^N$ ), it can be shown that for self 1  $x_{max,1}^B$  can be higher or lower than  $x_{max,1}^N$ . In the absence of decision substitution, and depending on the specific timing assumption, this can make narrow accounts strictly worse or strictly better than broad accounts for a certain parameter range. When broad accounts push the evaluation of outcomes further into the future, discounting makes deviations in the first task feel less painful than under narrow accounts. This incentive-decreasing effect of broadly evaluated goals is a key driving force in the analysis of Hsiaw (2010b). While interesting, the effects of timing do not invalidate our previous insights: decision substitution can arise, and once it arises, broad accounts do strictly worse than narrow accounts.

### 4.2.2 Leaving accounts open over time: sunk cost effects

Because of its intuitive appeal, decision substitution has been advanced as a plausible explanation for narrow bracketing in a range of settings where tasks occur sequentially (see Read *et al.* 1999). Our model shows that for such decision substitution to arise with sequential tasks and broad accounts, the individual must be either prone to sunk costs, or neglect the cost dimension entirely. Such tight links between mental accounting and sunk costs have previously been noted. Thaler (1999, p.202), for instance, writes “My own thinking about mental accounting began with an attempt to understand why people pay attention to sunk costs.” A common illustration of the connection between sunk costs and accounting due to Kahneman and Tversky is that of a family that drives through a blizzard to a basketball game, which they only do because they already paid for the tickets.

In our model, an individual with a broad account evaluates outcomes only after all relevant information for the account becomes available. Hence, when deciding on his effort in the second task, the individual considers whether the overall cost account for both tasks is balanced. That is, he pays attention to the sunk cost from the first task. With the alternative assumption that gain-loss utility is experienced as soon as a task-related payoff occurs, accounts would follow the exogenous timing of outcomes. In our sequential task setting, this would imply that the individual always narrowly evaluates costs, and that decision substitution does not arise.<sup>18</sup>

Is it reasonable to assume that people can leave psychological accounts open over time? Thaler (1999) discusses that the arrival of information often goes hand in hand with the closure of an account.<sup>19</sup> However, other evidence suggests that people are able to leave their accounts open even after payoff information arrives: people have mental accounts for categories of expenses accumulated over a week or a month rather than for individual purchases (e.g., Heath and Soll 1996); private investors tend to evaluate sales transactions relative to historic purchase prices rather than the current market value of their assets, as evidenced by the *disposition effect* (Odean 1998); and bettors take riskier gambles after losses (see footnote 13).

### 4.2.3 Other types of accounts

If the individual only had expenditure accounts, as is often assumed in the mental accounting literature, our results would be unchanged or even strengthened. Consider, for instance, two consumption decisions with immediate monetary costs. One involves only delayed benefits and the other involves only immediate benefits. In our main model, broad accounts need not be dominated. For example, the sufficient condition for decision substitution from proposition 3 fails to hold for goals  $x_{0,i}^* < x_{max,i}^N$  (see section 3.3). But if the individual only tracks expenditures in a

---

<sup>18</sup>Decision substitution still arises in a setting where the individual feels some gain-loss utility once a payoff occurs but also evaluates outcomes in a broad account that can be left open until all relevant payoffs occur.

<sup>19</sup>Referring to the “Samuelson lottery” (cf. footnote 5), Thaler writes: “Samuelson’s colleague should accept any number of trials of this bet strictly greater than one *as long as he does not have to watch*”. This suggests that watching the single tosses induces a narrow account for each lottery. For related evidence on how the frequency of feedback affects the evaluation of financial decisions, see Gneezy *et al.* (2003).

broad mental account, decision substitution always arises. As long as he keeps total expenditures fixed, the individual can freely substitute consumption. Substituting away from the decision with delayed benefits always increases the utility of self 1 because of the relative bias. This insight helps understand why having a mental account for food expenditures might be a good idea to control expenditures for food (vis-a-vis other categories) but a bad idea to control one’s diet.

Similarly, our results carry over if a person only considers benefits (e.g., a student who only cares about the grade and does not feel gain-loss utility regarding time spent studying); or if there is gain-loss utility over decisions and not over outcomes (e.g., a student feels anger or guilt if he works less than prescribed by his goal).

The concept of a *topical account* (Tversky and Kahneman 1981, Kahneman and Tversky 1984) is similar to a narrow psychological account because it relates to one specific activity or decision. But it differs from our approach by requiring the individual to integrate costs and benefits in one account. The separability assumption in our setting is crucial for reference dependence to have any impact. Indeed, as Kőszegi and Rabin (2006, p.1138) write, “In combination with loss aversion, this separability is at the crux of many implications of reference-dependent utility, including the endowment effect.” Note also that the individual would never choose to create a topical account – if he could design his gain-loss utility in this way – because integrating costs and benefits in one account makes goal setting ineffective.

## 5 Goals for tasks with stochastic outcomes

The following extension of our model allows for stochastic task outcomes. What changes is that the choice of psychological accounts has a *direct* impact on the individual’s gain-loss utility. If one holds decision levels constant, broad accounts provide strictly higher gain-loss utility than narrow accounts because broad accounts allow to pool risks. In the absence of time-inconsistent preferences, the individual therefore strictly prefers broad accounts. But, in the presence of motivational problems, narrowly evaluated goals may be optimal. To isolate the additional driving forces with stochastic outcomes, we consider a setting where selves 0 and 1 trade off the two decisions in the same way. That is, broad and narrow accounts would perform equally well if outcomes were deterministic (cf. proposition 3).

### 5.1 The model with outcome uncertainty

**The decisions.** We consider two symmetric, simultaneous decisions with immediate costs and delayed benefits and extend the setting from section 2 as follows. A decision may either have a strong or a weak impact on the delayed outcome. With probability  $\pi \in (0, 1]$  the (marginal) impact is high and the delayed benefit in task  $i$  is  $h_H(x_i)$ . With probability  $1 - \pi$  the impact is low and the delayed benefit is  $h_L(x_i)$ , where  $h_H(x_i) > h_L(x_i) \geq 0$ ,  $h'_H(x_i) > h'_L(x_i) \geq 0$ , and  $h''_H(x_i) < h''_L(x_i) \leq 0$  for all  $x_i$ . Note that the model nests the no-uncertainty case ( $\pi \in \{0, 1\}$ ). To facilitate exposition and make effects more transparent, we explain the intuition for the results with a special case. We set  $h_L(x_i) = 0$  for all  $x_i$  (“a failure”), and denote  $h(x_i) \equiv h_H(x_i) \geq 0$  (“a

success”). Details for the more general case and proofs are relegated to the appendix. As before, we assume that accounts are closed and evaluated after all outcomes realize (our results do not rely on this specific assumption).

**Stochastic reference points.** The modeling follows Kőszegi and Rabin (2006, 2007). We assume that the individual has stochastic reference points. With narrow accounts, a goal  $\hat{x}_i$  induces reference point  $c(\hat{x}_i)$  in the cost dimension and a reference lottery for the benefit of  $(\pi \circ h(\hat{x}_i); (1 - \pi) \circ 0)$ . With broad accounts, goals  $(\hat{x}_1, \hat{x}_2)$  induce reference point  $c(\hat{x}_1) + c(\hat{x}_2)$  in the cost dimension and a reference lottery for the joint benefit of  $(\pi^2 \circ [h(\hat{x}_1) + h(\hat{x}_2)]; \pi(1 - \pi) \circ h(\hat{x}_1); \pi(1 - \pi) \circ h(\hat{x}_2); (1 - \pi)^2 \circ 0)$ . The individual evaluates a stochastic outcome according to its expected utility. The utility of a particular outcome is the average of how this outcome feels relative to each possible realization of the reference lottery. This is best illustrated with an example. Suppose that the individual has goal  $\hat{x}_i$ , makes decision  $x_i = \hat{x}_i$ , and evaluates outcomes narrowly. The gain-loss utility in the benefit dimension for decision  $i$  is given by

$$\pi \{ \pi \mu(h(\hat{x}_i) - h(\hat{x}_i)) + (1 - \pi) \mu(h(\hat{x}_i) - 0) \} + (1 - \pi) \{ \pi \mu(0 - h(\hat{x}_i)) + (1 - \pi) \mu(0 - 0) \}.$$

The individual expects to achieve benefit  $h(\hat{x}_i)$  with probability  $\pi$ . This outcome gives rise to the “mixed feelings” captured within the first curly brackets. Partially it feels like meeting the reference state  $h(\hat{x}_i)$  (to which the reference lottery assigns weight  $\pi$ ). And partially it feels like a gain compared to the reference state 0 (to which the reference lottery assigns weight  $1 - \pi$ ). Analogously, the individual expects benefit 0 with probability  $1 - \pi$ . As captured within the second curly brackets, this outcome feels like a loss relative to the probability- $\pi$  reference state  $h(\hat{x}_i)$ . And it feels like meeting the probability- $(1 - \pi)$  reference state 0.

## 5.2 The preferred decisions of self 0 and the risk-pooling effect

The individual knows that – even if he sticks to his goals – he only meets his outcome targets on average. In expectation, the individual hence suffers a loss. How large the loss is differs for narrow and broad accounts and, therefore, the preferred decisions differ. Denote the preferred decisions of self 0 for the two symmetric tasks under narrow accounts by  $x_0^N$  and those under broad accounts by  $x_0^B$ . Self 0 takes into account that his expectations adapt to his committed decisions. Technically, the equilibrium concept for the preferred decisions is that of *choice-acclimating personal equilibrium* (Kőszegi and Rabin 2007). To study goal design, we ask in section 5.3 whether self 0 can implement his preferred decisions in an (unacclimating) personal equilibrium.<sup>20</sup>

---

<sup>20</sup>In the model with deterministic outcomes, the present bias is the only source of a conflict of interest. With stochastic outcomes, there is an externality of the decision on future beliefs. Even in the absence a present bias, this externality may cause a conflict of interest in the sense that the preferred decision cannot be implemented in a personal equilibrium (cf. Kőszegi and Rabin 2007). Because this effect is not the driving force of our main results, we postpone discussion of it to section 5.4.2.

### 5.2.1 Narrow psychological accounts

We start by deriving the utility of self 0 for a committed decision  $x_i > 0$ . The stochastic benefit for a decision  $i$  is compared to its reference lottery  $(\pi \circ h(x_i); (1 - \pi) \circ 0)$ . With probability  $1 - \pi$ , benefit 0 realizes. This partially feels like a loss, because with probability  $\pi$  the individual expected the benefit  $h(x_i)$ . Only with probability  $1 - \pi$  expectations are met. With probability  $\pi$ , benefit  $h(x_i)$  realizes. This partially feels like a gain, because the individual expected a benefit realization of 0 with probability  $1 - \pi$ . The probability of a gain hence is the same as that of a loss:  $\pi(1 - \pi)$ . As losses loom larger than gains ( $\lambda > 1$ ), the individual expects a net psychological loss. There is no gain-loss utility in the cost dimension, because cost expectations are always met. In sum, the utility of self 0 for a committed decision  $x_i$  in task  $i$  is:

$$\beta \{ \pi h(x_i) [1 - \eta(\lambda - 1)(1 - \pi)] - c(x_i) \}.$$

Maximizing yields the first-order condition for the preferred decision of self 0,

$$\pi h'(x_0^N) [1 - \eta(\lambda - 1)(1 - \pi)] \leq c'(x_0^N), \quad (11)$$

which holds with equality for  $x_0^N > 0$ . A choice-acclimating personal equilibrium implies a strong notion of risk aversion, in the sense that the individual may choose stochastically dominated options (Kőszegi and Rabin 2007). Decision  $x_i > 0$  causes positive expected consumption utility but it exposes the individual to the risk of falling short of his expectations. To avoid psychological losses, the individual can choose  $x_i = 0$ . That way he is sure that his expectations are always met – he expects to fail and will fail. To exclude the dominated choice  $x_i = 0$ , we assume that gain-loss utility does not swamp concerns for consumption utility (cf. Herweg *et al.* 2010). The following assumption implies that the preferred decision of self 0 is positive and increasing in  $\pi$ :

**Assumption 2 (No dominance of gain-loss utility)**  $1 - \eta(\lambda - 1)(1 - \pi) > 0$ .

### 5.2.2 Broad psychological accounts

With broad psychological accounts, decisions  $x_1$  and  $x_2$  induce reference lottery  $(\pi^2 \circ [h(x_1) + h(x_2)]; \pi(1 - \pi) \circ h(x_1); \pi(1 - \pi) \circ h(x_2); (1 - \pi)^2 \circ 0)$  in the benefit dimension. How does the gain-loss utility differ from that under narrow accounts? As above, self 0 expects a net loss. With narrow accounts a loss occurs in, say, task 1 when the individual expects to succeed but actually fails. This happens with probability  $\pi(1 - \pi)$ . The event “the individual expects to succeed in task 1, but fails” however need not result in a loss under broad accounts, because task 2 creates a buffer against a failure in task 1. With probability  $\pi(1 - \pi)$ , self 0 expects that he will fail in task 2 but actually succeeds. This success in task 2 makes up for the shortfall in task 1. Such *risk pooling* reduces the per-task probability of a net loss from  $\pi(1 - \pi)$  under narrow accounts to  $\pi(1 - \pi)[1 - \pi(1 - \pi)]$  under broad accounts. Hence, the utility of self 0 for symmetric decisions  $x_1$  and  $x_2$  under broad accounts is:<sup>21</sup>

$$\beta \{ (h(x_1) + h(x_2)) \pi [1 - \eta(\lambda - 1)(1 - \pi)(1 - \pi(1 - \pi))] - (c(x_1) + c(x_2)) \}. \quad (12)$$

---

<sup>21</sup>In appendix B.1, we show that the preferred decisions of self 0 are indeed symmetric.

Maximizing with respect to  $x_1$  and  $x_2$  yields the symmetric first-order conditions:

$$\pi h'(x_0^B) [1 - \eta(\lambda - 1)(1 - \pi)(1 - \pi(1 - \pi))] \leq c'(x_0^B), \quad (13)$$

holding with equality for  $x_0^B > 0$ . Assumption 2 ensures an interior solution. Comparison with equation (11) shows that self 0 is willing to “risk more” on each individual decision, i.e.,  $x_0^B > x_0^N$ .

**Proposition 4 (Kőszegi and Rabin 2007)**

1. *Holding decision levels constant, the utility of self 0 is higher with broad accounts than with narrow accounts for all  $\pi \in (0, 1)$ .*
2. *The preferred decision of self 0 is higher with broad accounts than with narrow accounts:  $x_0^B > x_0^N$  for all  $\pi \in (0, 1)$ .*

It is a well known property that loss averse individuals benefit from pooling risks. Furthermore, risk pooling makes such individuals less risk averse. Proposition 4 is a variant of that result for stochastic reference points. It goes back to Kőszegi and Rabin (2007, proposition 5).<sup>22</sup> For experimental evidence see, for example, Gneezy and Potters (1997, 2003).

**5.3 The goal design problem**

We turn to the goal design problem of self 0. Solving backward, we ask what conditions guarantee that self 1 will not lower his decision in one task relative to the goal set for that task under broad or narrow accounts, respectively. Considering such unilateral deviations provides us with the maximal implementable goal under narrow accounts  $x_{max}^N$ , and an upper bound on implementable goals under broad accounts, which, with slight abuse of notation, we denote by  $x_{max}^B$ . Considering, in addition, joint deviations would pinpoint exactly the maximal broad-accounting implementable goal. But as discussed below, for our main purpose it suffices to consider unilateral deviations. The first-order conditions defining  $x_{max}^N$  and  $x_{max}^B$ , respectively, have the same structure. At  $x_{max}^A$ ,  $A \in \{B, N\}$ , self 1 is just indifferent between sticking to this goal and deviating to a lower decision:

$$\beta [\pi + \eta \{ \pi + (\lambda - 1) \Pr(\text{loss}^A) \}] h'(x_{max}^A) = (1 + \beta \eta) c'(x_{max}^A), \quad (14)$$

where  $\Pr(\text{loss}^A)$  denotes the probability of a loss caused by a decision that falls short of the goal  $x_{max}^A$ . The higher  $\Pr(\text{loss}^A)$ , the higher is  $x_{max}^A$  – because self 1 has stronger incentives not to lower his effort.

**5.3.1 Narrow psychological accounts**

For narrow accounts,  $\Pr(\text{loss}^N) = \pi^2$ . A deviation to  $x_1 < \hat{x}$  only matters if it actually has an impact on the outcome, i.e., if the outcome is a success. This happens with probability  $\pi$ . The

---

<sup>22</sup>They show that  $U(F|F) + U(G|G) \leq U(F + G|F + G)$  for any distribution functions  $F$  and  $G$ . In our case, with symmetric decisions  $x_1 = x_2 = x$ , we have  $F = G = (\pi \circ h(x); (1 - \pi) \circ 0)$  and  $F + G = (\pi^2 \circ h(x) + h(x); \pi(1 - \pi) \circ h(x); \pi(1 - \pi) \circ h(x); (1 - \pi)^2 \circ 0)$ .

outcome  $h(x_1)$  then feels like a loss compared to reference standard  $h(\hat{x})$ . With stochastic reference points, the reference lottery assigns probability  $\pi$  to  $h(\hat{x})$ . Hence,  $x_{max}^N$  is defined by

$$\beta \pi \{1 + \eta [1 + (\lambda - 1) \pi]\} h'(x_{max}^N) = (1 + \beta \eta) c'(x_{max}^N). \quad (15)$$

Outcome uncertainty decreases incentives. The maximal narrow-accounting implementable goal is lower than it would be with deterministic outcomes. There are two reasons for this. The first is a standard consumption-utility effect. The less likely it is that the decision “matters”, i.e., the lower  $\pi$  is, the smaller the marginal impact of the decision on expected consumption utility,  $\pi h'(x_i)$ . The second reason is linked to gain-loss utility. A goal is successful if the fear of falling short of expectations prevents the individual from lowering his decision relative to the goal. With deterministic outcomes, a marginal deviation results in a loss with probability one. With stochastic outcomes, a marginal deviation is less painful because it only increases the probability of a loss by  $\pi^2$ .

In the absence of a present bias ( $\beta = 1$ ), the maximal implementable goal always exceeds the preferred decision of self 0 (compare equations 11 and 15). With a present bias, successful self-regulation is possible if  $\beta$  is sufficiently high, so that  $x_0^N \leq x_{max}^N$ . For lower  $\beta$ , we have  $x_{max}^N < x_0^N$  and self-regulation is constrained by the maximal implementable goal  $x_{max}^N$ .

### 5.3.2 Broad psychological accounts

What are the incentives of self 1 under broad accounts to marginally lower his decision in one task, say task 1, relative to the goal? To derive  $\Pr(\text{loss}^B)$  induced by such a unilateral deviation, suppose that self 1 faces symmetric goals (in appendix B.2.2, we show that setting symmetric goals is indeed optimal). The effects are easiest to explain by distinguishing the probability of success for the two decisions. Denote them by  $\pi_1$  and  $\pi_2$  (we still assume  $\pi_1 = \pi_2$ ).

Risk-pooling allows the individual to offset a lower than expected outcome in one task with a higher than expected outcome in the other. This has a positive and a negative effect on  $\Pr(\text{loss}^B)$ , and thereby on how incentives under a broad account compare with incentives under narrow accounts (*risk incentive effects*). Let us first consider the negative effect. Under narrow accounts, the probability of a shirking-induced loss in task 1 is  $\pi_1^2$ . With probability  $\pi_1^2$ , the individual expects to succeed in task 1 and the decision actually matters. In contrast, under broad accounts, the individual does not always suffer a loss in this case. With probability  $\pi_2(1 - \pi_2)$ , the individual expects to fail in task 2 but actually succeeds. So even if self 1 deviates in task 1, he can meet the reference state  $h(\hat{x}) + 0$  by succeeding in task 2. That is, the deviation does not always lead to a loss where it would have done so under narrow accounts. Thus, broad accounts additively reduce the probability of a shirking-induced loss by

$$\pi_1^2 \times \pi_2(1 - \pi_2). \quad (16)$$

In sum, the individual fears less to fall short of his reference standard if he deviates under broad accounts than under narrow accounts. This dampens incentives.

Let us next consider the positive effect on incentives. Decision  $x_1$  not only affects losses in task 1 but it can also provide a buffer against losses in task 2. With probability  $\pi_1(1 - \pi_1)$ , the individual

expects to fail in task 1 but the outcome actually is  $h(x_1)$ . In this case, a deviation to  $x_1 < \hat{x}$  does not cause a loss under narrow accounts. The reason is that  $h(x_1)$  beats the failure reference state of 0. But with broad accounts such a deviation causes a loss if, in addition, the individual expects to succeed in task 2 but is unlucky. That is, with probability  $\pi_2(1 - \pi_2)$ , the joint outcome  $h(x_1) + 0$  falls short of the joint reference state  $h(\hat{x}) + 0$  as soon as  $x_1 < \hat{x}$ . Compared to narrow accounts, the probability of a shirking-induced loss thus additively increases by

$$\pi_1(1 - \pi_1) \times \pi_2(1 - \pi_2). \quad (17)$$

The positive incentive effect reflects “raised expectations” from adopting broad accounts. The individual knows that if he sticks to the goals, risk pooling will eliminate some losses that would occur with narrow accounts. Faced with this tougher reference lottery, he has more to lose from shirking because a low outcome feels more painful with broad accounts than it does with narrow accounts.<sup>23</sup>

Overall, the negative effect in (16) exceeds the positive effect in (17) if and only if  $\pi \geq \frac{1}{2} \equiv \underline{\pi}$ . Then shirking hurts the individual more with narrow accounts than with broad accounts. In the more general setting, where the decision affects the failure-outcome (i.e.,  $h'_L(x) > 0$ ), the threshold  $\underline{\pi}$  can be lower than 1/2.

Adding (16) and (17) to  $\pi^2$  gives  $\Pr(\text{loss}^B)$ . Plugging this into equation (14) yields an upper bound on broad-accounting implementable goals,  $x_{max}^B$ :

$$\beta \pi [1 + \eta \{1 + \pi(\lambda - 1)(1 + (1 - \pi)(1 - 2\pi))\}] h'(x_{max}^B) = (1 + \beta \eta) c'(x_{max}^B). \quad (18)$$

As for narrow accounts,  $x_{max}^B$  can be smaller or larger than  $x_0^B$  depending on  $\beta$ .

**Lemma 3** *The maximal implementable goal under narrow accounts,  $x_{max}^N$ , is strictly greater than the upper bound for implementable goals under broad accounts  $x_{max}^B$  if and only if  $\pi > \underline{\pi}$ , where  $\underline{\pi} \in (0, \frac{1}{2}]$ . In the special case where  $h_L(x) = 0$  for all  $x$ ,  $\underline{\pi} = 1/2$ .*

The risk-incentive effect extends to more than two tasks as we show in appendix B.4.

### 5.3.3 Goal design by self 0: When are narrow accounts optimal?

The risk-pooling effect causes the utility of self 0 to be higher with a broad account than with narrow accounts, holding the decision levels constant (cf. proposition 4). Hence, there is no reason to adopt narrow accounts if successful self-regulation is feasible with broad accounts. If however the individual has a relatively strong present bias (i.e., low  $\beta$ ), such that  $x_{max}^B < x_0^B$ , the motivational force of broadly evaluated goals is too weak to overcome the self-control problems. Nevertheless, the individual may be able to implement his preferred decision – or at least a decision closer to

---

<sup>23</sup>In a model with deterministic reference points, risk pooling does not change the reference points under broad accounts. Hence, the positive incentive effect cannot arise. In contrast, the negative incentive effect arises whenever the reference point with broad accounts is smaller than the expected benefit in a single task. Because then, in expectation, part of the outcome in one task is available and can be used to make up for a shortfall in the other task. For a model along these lines, see Koch and Nafziger (2009a).



his preferred decision – with narrow accounts. Because of the risk-incentive effect,  $x_{max}^B < x_{max}^N$  if  $\pi > \underline{\pi}$ .

Does the risk-pooling or risk-incentive effect dominate for  $\pi > \underline{\pi}$ ? Broad and narrow accounts do equally well for  $\pi = 1$ , because then  $x_{max}^B = x_{max}^N$  and risk pooling plays no role. Introducing a bit of uncertainty, by marginally reducing  $\pi$ , leads to a larger drop in utility under broad accounts than under narrow accounts. That is, narrow accounts are optimal when there is little uncertainty that the decision has a strong impact on the task outcome. For large  $\pi$  the risk-pooling effect is of second order. Intuitively, if the individual lives up to his goals, there is only a small probability of suffering a loss with narrow accounts that could have been avoided with a broad account. The risk-incentive effect, however, is of first order:  $x_{max}^B$  drops faster than  $x_{max}^N$  does if one lowers  $\pi$ .

**Proposition 5** *Suppose  $x_{max}^B < x_0^B$ . There exists a cutoff  $\hat{\pi} \in (\underline{\pi}, 1)$ , such that for  $\hat{\pi} < \pi < 1$  narrow accounts are strictly optimal.*

The result shows that even if the individual has no relative bias between tasks and the decision substitution effect outlined in section 3 is not present, narrow accounts can be strictly optimal. Outcome uncertainty gives rise to additional risk-incentive effects. For high values of  $\pi$ , narrow accounts are optimal for a present-biased individual because their greater motivational force compensates for the lack of risk-pooling opportunities.

### 5.3.4 Discussion

Proposition 5 suggests that an individual should adopt narrow accounts if he faces little uncertainty that the decision will have a strong impact on the task outcome. If there is substantial risk that effort has no or only a low impact, broad accounts are better. Think of teaching preparation versus research. The former involves tasks where there is a close relationship between effort and outcomes. Hence, our model says that it is a good idea to set tight goals for teaching preparation, such as “every day, spend one hour preparing”, or “prepare  $x$  slides on a given day”. In comparison, research involves tasks where success on a given day might be uncertain – despite best efforts (think of writing up a proof). So our model suggests that researchers should not look at the quality of a days’ work but evaluate their research outcomes over a longer period.

Advice in popular guides on goal setting is consistent with our results. For example, a blog for freelance journalists notes the need to balance between those days when you are “in the spirit” and highly productive and those days when you “labor just to get one article going”, concluding:<sup>24</sup> “Setting weekly goals for your article writing could prove more useful than daily goals. The thing is that no matter how noble our business goals may be, life gets in the way sometimes [...]”.

Similarly, individuals on a diet face daily fluctuations that they cannot control. Weight loss programs, such as Jenny Craig<sup>®</sup> therefore often recommend to weigh oneself weekly rather than daily. On the other hand, as discussed before, such programs also recommend daily calorie targets or daily point goals for healthy and unhealthy food. That is, these programs recommend narrow accounts for immediate costs and benefits but a broad, weekly account for the uncertain weight outcome.

---

<sup>24</sup> <http://EzineArticles.com/4204813> (last accessed May 2011).

Our results suggests why. Narrow accounts help prevent decision substitution (if there is a relative bias), while the broad outcome account allows for risk pooling.<sup>25</sup>

## 5.4 Extensions and robustness

### 5.4.1 Joint deviations

To derive the sufficient conditions for optimality of narrow accounts, it is enough to determine the bounds  $x_{max}^B$  and  $x_{min}^B$  for which self 1 has no incentive to deviate from the goal in a *single* task. Allowing for the possibility of *joint* deviations may further shrink the set of broad-accounting implementable goals. Checking for joint deviations in the general case is extremely tedious and not very illuminating. This exercise can only strengthen the result on the optimality of narrow accounts in proposition 5.<sup>26</sup> For the special case where  $h_L(x) = 0$  for all  $x$ , it is however relatively easy to show that  $x_{min}^B$  and  $x_{max}^B$  are tight bounds for the broad-accounting implementable decisions. This yields the following corollary to proposition 5.

**Corollary 2** *Suppose  $h_L(x) = 0$  for all  $x$ .*

1. *The bounds  $x_{min}^B$  and  $x_{max}^B$  are tight.*
2. *Suppose  $x_{max}^B < x_0^B$ . There exists a cutoff  $\hat{\pi} \in (1/2, 1)$ , such that for  $\pi \in (\hat{\pi}, 1)$  narrow accounts are strictly optimal and for  $\pi \in (0, \hat{\pi}]$  broad accounts are strictly optimal.*

### 5.4.2 Time consistency in the absence of a present bias

The present-bias is the driving force for the suboptimality of a broad account in proposition 5 as it can cause  $x_0^B > x_{max}^B$ . In the absence of a present-bias, we have  $x_0^B < x_{max}^B$  and  $x_0^N < x_{max}^N$  (and corollary 2 tells us that the bound  $x_{max}^B$  is tight for  $h_L(x) = 0$  for all  $x$ ). That is, these bounds do not prevent self 0 from implementing his preferred decision. However, the minimal implementable goal may prevent self 0 from implementing his preferred decision because the individual's preferences may be time inconsistent even in the absence of a present bias.

As carefully laid out in Kőszegi and Rabin (2007), a loss averse individual with stochastic reference points is more risk averse when a decision is committed to well in advance than when the decision is made for given expectations. The same happens in our model. Self 1 takes expectations as given in the preferred personal equilibrium (PPE). In contrast, the preferred decision of self 0 accounts for the externality that a committed choice has on expectations:  $x_0^B$  is a choice-acclimating personal equilibrium (CPE). Because self 0 is more risk averse than self 1, it is possible that self 0 prefers a decision  $x_0^B$  that is lower than the minimal implementable goal  $x_{min}^B$ . That is, the set of CPE need not coincide with the set of PPE – even though there is no present bias. The choice of accounts

---

<sup>25</sup>Note that the risk-pooling and risk-incentive effects go though unchanged if a broad benefit account goes in hand with a narrow cost account. In contrast, with sequential tasks decision substitution occurs only if the individual has a broad cost account (see section 4.2.2).

<sup>26</sup>The only additional insight from such an exercise is whether the unilateral deviations alone pin down the precise cutoff  $\underline{\pi} \in (0, \frac{1}{2}]$  in lemma 3 or whether the cutoff actually is lower.

then may matter because  $x_{min}^N < x_{min}^B$ , as stated in the result below. Replacing assumption 2 with the following assumption excludes this form of time inconsistency for an individual with no present bias.<sup>27</sup>

**Assumption 2'**  $1 - \eta(\lambda - 1)(1 - \pi) \geq \frac{1+\eta}{1+\eta\lambda}$ .

The following corollary shows that the inferiority of broad accounting in proposition 5 is driven by the individual's present bias and that it is not an artefact of other types of time inconsistency that may arise with reference dependent preferences.

**Corollary 3**

1. *The lower bound on the minimal implementable goal under broad accounts exceeds the minimal implementable goal under narrow accounts:  $x_{min}^N < x_{min}^B$ .*
2. *If assumption 2' holds then  $x_{min}^N < x_0^N$  and  $x_{min}^B < x_0^B$  for all  $\beta \leq 1$ .*
3. *Suppose  $h_L(x) = 0$  for all  $x$  and assumption 2' holds. In the absence of a present-bias (i.e.,  $\beta = 1$ ), broad accounts are strictly optimal, because  $x_{min}^B < x_0^B < x_{max}^B$  and these bounds are tight.*

## 6 Adaptation of reference points to information

Sometimes people learn about the performance in a past task before they start with a new task. For example, a student may get to know his grade in one exam before he studies for the next exam. And sometimes people receive information before working on a task about how productive their effort will be. For example, a cab driver may discover from observing the weather at the start of his shift whether it will be a busy or a slow day.

The following extension of our model shows that the risk-pooling effect, the key advantage of broad accounts, disappears in settings with interim or pre-task information.<sup>28</sup> That is, for fixed decision levels, broad and narrow accounts do equally well. Nevertheless, in the presence of a relative bias, narrow accounts may help achieve better self-regulation. Decision substitution continues to be a problem and prevents the individual from implementing the same decisions under broad accounts than under narrow accounts.

Decision substitution is central to the (controversial) discussion, summarized in the introduction, that many cab drivers appear to adopt narrow, daily income targets. Our results contribute to the theoretical underpinning of that debate in two ways. First, we show that decision substitution is a robust phenomenon. Second, by endogenizing the evaluation horizon we show that a narrow account, such as considering a daily income target, need not be an error but can indeed be optimal.

---

<sup>27</sup>While assumption 2 already rules out that self 0 would like to implement stochastically dominated choices, it does not rule out that self 0 would prefer a lower decision level than the minimal implementable goal.

<sup>28</sup>This is related to the effects discussed in Kőszegi and Rabin (2009, section 3). They argue that under certain circumstances an individual may behave “as if” he is a narrow bracketer.

## 6.1 The model extension

**The decisions.** Two symmetric decisions occur sequentially. With probability  $\pi$ , a task is in a productive state where the delayed benefit is  $h(x_i) \equiv h_H(x_i) \geq 0$ . And with probability  $1 - \pi$ , it is in an unproductive state, yielding a delayed benefit of zero.

**Anticipatory utility.** We extend the approach of section 4.1 to stochastic outcomes and allow the reference distribution to adapt not only to changes of mind but also to new information or deviations from goals. Such adaptations trigger anticipatory utility, because the individual contrasts the old reference distribution with the new reference distribution. As in Kőszegi and Rabin (2009), the individual compares the worst percentile of outcomes under the new distribution with the worst percentile of outcomes under the old distribution, then the second worst percentile of outcomes, and so on (see below for specific examples of how one calculates anticipatory utility).<sup>29</sup>

**Timing.** We consider the following *pre-task information* scenario. Self 1.I learns the productivity in task 1, updates the reference distribution and possibly revises his goals for tasks 1 and 2. Given the (revised) goals and reference distribution, self 1.II makes the decision for task 1. In period 2, self 2.I learns the productivity in task 2, updates the reference distribution and possibly revises the goal for task 2. Self 2.II makes the decision for task 2. Costs occur when decisions are made, and the delayed benefits from both tasks realize in period  $\tau \geq 3$ .

Proposition 6, derived below, holds also for the case with *interim-information*. In this scenario, self 2.I learns about the performance in task 1, updates the reference distribution and possibly revises his goals for task 2. Self 2.II makes the decision for task 2. Because the main driving forces are essentially the same as in the pre-task information scenario, we state this without proof (derivations are available from the authors).

To focus on how updating of the reference distribution affects risk pooling and decision substitution, we assume that all outcomes are evaluated in the final period  $\tau$  in the chosen type of accounts. Anticipatory utility compares how an account will be evaluated under the revised reference distribution to how it would have been evaluated under the old reference distribution. Hence, anticipatory utility at date  $t$  assigns weight  $\gamma_{\tau,t} \leq 1$  to outcomes that are affected by a revision of reference distributions.<sup>30</sup> We assume that  $\gamma_{\tau,t}$  is weakly decreasing in  $\tau - t$  (cf. footnote 15). Denote the equilibrium decision that the individual makes in task  $i = t$  if it is in the productive state by  $\hat{x}_t$ . In the unproductive state, the individual chooses a zero decision.

---

<sup>29</sup>Note that in contrast to Kőszegi and Rabin (2009), a broad account allows anticipatory feelings to relate to outcomes from different periods. They assume that each account only comprises outcomes from a single period.

<sup>30</sup>The additional insights from allowing the choice of accounts to shift the timing of (anticipatory) gain-loss-utility are related to those discussed in section 4.2.1 (results are available from the authors). In the absence of decision substitution, broad accounting might bring a small advantage, because updating of expected costs occurs later than under narrow accounting.

## 6.2 The risk-pooling effect disappears

Under narrow accounts, self  $t.I$  inherits from self 0 the reference distribution  $[\pi \circ h(\hat{x}_t); (1 - \pi) \circ 0]$  for the delayed benefit and  $[\pi \circ c(\hat{x}_t); (1 - \pi) \circ 0]$  for the cost. He then observes the state and updates the reference distributions. If he observes that the state is productive, he knows that he will incur benefit  $h(\hat{x}_t)$  and cost  $c(\hat{x}_t)$  with probability 1. However, he previously expected that with probability  $1 - \pi$  the unproductive state occurs, i.e., costs and benefits of zero. The comparison of the new reference distribution with past expectations triggers anticipatory utility  $\eta(1 - \pi)[\gamma_{\tau,t} h(\hat{x}_t) - \gamma_{\tau,t} \lambda c(\hat{x}_t)]$ . Similarly, if self  $t.I$  observes that the state is unproductive, this triggers anticipatory utility  $-\eta\pi[\gamma_{\tau,t} \lambda h(\hat{x}_t) - \gamma_{\tau,t} c(\hat{x}_t)]$ . Taking expectations over the anticipatory utility in the possible states for the two tasks and adding consumption utility, the utility of self 0 under narrow accounts is

$$\beta \pi \sum_{t=1}^2 \{h(\hat{x}_t) - c(\hat{x}_t) - \eta(1 - \pi)(\lambda - 1)\gamma_{\tau,t}(h(\hat{x}_t) + c(\hat{x}_t))\}. \quad (19)$$

Under broad accounts, self  $1.I$  inherits from self 0 the reference distribution for the delayed benefits  $[\pi^2 \circ (h(\hat{x}_1) + h(\hat{x}_2)); \pi(1 - \pi) \circ h(\hat{x}_1); \pi(1 - \pi) \circ h(\hat{x}_2); (1 - \pi)^2 \circ 0]$ . He learns the productivity for task 1 and updates the reference distribution to  $[\pi \circ (h(x_1) + h(\hat{x}_2)); (1 - \pi) \circ (h(x_1) + 0)]$ , where  $x_1 \in \{0, \hat{x}_1\}$ , depending on the realized state. Self  $2.I$  learns the productivity for task 2 and updates the reference distribution to  $h(x_1) + h(x_2)$ ,  $x_2 \in \{0, \hat{x}_2\}$ . Similarly, for the costs. Each revision of expectations triggers anticipatory utility. For example, if self  $1.I$  learns that the state is unproductive, this triggers anticipatory utility

$$\begin{aligned} & -\eta\gamma_{\tau,1} \left[ \lambda \pi^2 (h(\hat{x}_1) + h(\hat{x}_2) - 0 - h(\hat{x}_2)) + \lambda \pi(1 - \pi) (h(\hat{x}_1) + 0 - 0 - 0) \right. \\ & \left. + \pi(1 - \pi) (0 + h(\hat{x}_2) - 0 - h(\hat{x}_2)) + (1 - \pi)^2 (0 + 0 - 0 - 0) \right] + \eta\gamma_{\tau,1} \pi c(\hat{x}_t) \\ & = -\eta\pi[\gamma_{\tau,1} \lambda h(\hat{x}_t) - \gamma_{\tau,t} c(\hat{x}_t)]. \end{aligned}$$

Proceeding in this fashion for the other states and dates, one obtains exactly the same utility of self 0 as under narrow accounts, given in equation (19). Thus, the risk-pooling effect vanishes. As a consequence, self 0 has the same preferred decisions under broad and narrow accounting. For the unproductive state it is zero. For the productive state in task  $t$  the preferred decision  $x_{0,t}^*$  maximizes the utility in equation (19) and is defined by

$$\frac{1 - \eta\gamma_{\tau,t}(1 - \pi)(\lambda - 1)}{1 + \eta\gamma_{\tau,t}(1 - \pi)(\lambda - 1)} h'(x_{0,t}^*) = c'(x_{0,t}^*). \quad (20)$$

Assumption 2 guarantees an interior solution. The assumption that  $\gamma_{\tau,t}$  is weakly decreasing in  $\tau - t$  implies that  $x_{0,1}^* \geq x_{0,2}^*$ .

## 6.3 Narrow-accounting implementable goals

Because goals can be revised, there are several constraints on implementable goals. Self  $t.I$  observes the state and may revise his goal for task  $t$ . If the state is unproductive, he has no incentive to

do so. So suppose he observes the productive state. This information, and the possible downward revision of his goal for task  $t$ , trigger anticipatory utility. Following similar steps as in our analysis above shows that self  $t$  has no incentive to revise his goal downward as long as it does not exceed some threshold  $\tilde{x}_{max,t}^N$  (all thresholds are derived in appendix C). An additional constraint is that self  $1.I$  does not revise the goal for the productive state of task 2, giving another upper bound on the goal for task 2,  $\tilde{x}_{max,2}^N$ .

Self  $t.II$  faces no uncertainty about the task outcome. Hence, he has no incentive to deviate from the (possibly already revised) goal, if it does not exceed  $x_{max,t}^N$ , as defined for  $b'_i(\cdot) = 0$  in equation (9).

Overall, the maximal implementable goal for task 1 hence is  $\min\{x_{max,1}^N, \tilde{x}_{max,1}^N\}$ , and for task 2 it is  $\min\{x_{max,2}^N, \tilde{x}_{max,2}^N, \tilde{x}_{max,2}^N\}$ . The additional bound for task 2,  $\tilde{x}_{max,2}^N$ , is not binding if current anticipatory utility weighs at least as strongly as discounted future anticipatory utility (i.e.,  $\gamma_{\tau,1} \geq \beta \gamma_{\tau,2}$ ).

## 6.4 Decision substitution under broad accounts

Suppose goals  $(x_{0,1}^*, x_{0,2}^*)$  are narrow-accounting implementable.<sup>31</sup> Are they broad-accounting implementable? Solving backwards, self  $2.I$  inherits a reference point that already takes into account the actual decision  $x_1$ . Thus, the outcomes related to task 1 cancel out from any anticipatory utility. So for self  $2.I$  the same maximal decision is implementable as under narrow accounts.

However, by arguments similar to those outlined in section 3, decision substitution at date 1 may prevent the narrow-accounting implementable decisions from being broad-accounting implementable. At date 1 the individual prefers to shift costs to the future. Therefore, self  $1.I$  may lower his goal for task 1 and revise upward his goal for task 2 to compensate.<sup>32</sup>

To see this consider a small deviation along the iso-cost curve for which  $c(x_1) + c(x_2) = c(x_{0,1}^*) + c(x_{0,2}^*)$  and  $h(x_{0,2}^*) \leq h(x_{0,1}^*) < h(x_1) + h(x_2) < h(x_{0,1}^*) + h(x_{0,2}^*)$ . What are the marginal incentives for such a deviation from goals  $x_{0,t}^*$ ? Upon observing the productive state and revising goals in this way, self  $1.I$  has utility (dropping arguments)

$$\begin{aligned} & \beta h_1 - c_1 + \beta \pi (h_2 - c_2) + \eta \gamma_{\tau,1} \times [\text{anticipatory utility: learn state in task 1 \& goal revision}] \\ & + \beta \eta \gamma_{\tau,2} \times [\text{future anticipatory utility: learn state in task 2}]. \end{aligned} \quad (21)$$

In appendix C, we show that such a deviation leads to a gain in utility for  $\pi$  near 1 and a drop in utility for  $\pi$  near 0. The intermediate value theorem implies that there exists a threshold  $\hat{\pi} \in (0, 1)$ , such that decision substitution pays off if  $\pi > \hat{\pi}$ . That is, for  $\pi > \hat{\pi}$  the preferred decisions of self 0 are not broad-accounting implementable and narrow accounting is optimal. With our findings, summarized in the following proposition, we establish the robustness of our earlier results on the optimality of narrow accounts.

---

<sup>31</sup>While the possibility of goal revision adds the above constraints on narrow-accounting implementable decisions, one can show that there exist parameter ranges where narrow goals allow to fully overcome the self-control problems in both tasks for all  $\pi \in (0, 1)$ .

<sup>32</sup>Preventing self  $1.I$  from only revising the goal for task 2 gives the already defined upper bound  $\tilde{x}_{max,2}^N$ .

**Proposition 6** *Suppose the individual receives either pre-task or interim information and that revisions of the reference distribution trigger anticipatory utility.*

1. *The risk-pooling effect disappears: Holding decision levels constant, the utility and preferred decisions of self 0 are equal under broad and narrow accounts.*
2. *Suppose the preferred decisions of self 0 are narrow-accounting implementable. There exists a cutoff  $\hat{\pi} \in (0, 1)$ , such that for all  $\pi > \hat{\pi}$  the preferred decisions of self 0 are not broad-accounting implementable and narrow accounts are strictly optimal.*

## 7 Conclusion

As pointed out by Brocas *et al.* (2004), one can transfer many of the insights from how people deal with *intra*-personal conflicts of interest (arising from time-inconsistent preferences) to principal-agent settings with *inter*-personal conflicts of interest (arising, e.g., from moral hazard and limits of contracting). For example, and corresponding to our setting, employers can use goal setting as an intrinsic motivator to alleviate possible moral hazard problems. A large number of experiments by psychologists show that goals set by another party can increase the motivation of an individual (for a survey, see Locke and Latham 1990). In a review of research on goal setting in organizations, Latham and Yukl (1975) conclude that many of the insights from those experiments carry over to firms.

What lessons can one draw for firms from our theory of psychological accounting? One way for an organization to influence its employees' psychological accounts is by framing performance appraisals appropriately. The firm can conduct appraisals more or less frequently, and it can provide general performance feedback or feedback tailored to a specific task. Another, more direct way is through the assignment of agents to tasks. A firm can either assign the same agent to perform two tasks, or a different agent to each task.

In settings with a close relationship between effort and output (where incentive problems might be caused by an incomplete contract), our model suggests that firms should tailor performance appraisals to the task, or employ different agents for each task if agents have a relative bias between the tasks. Such a relative bias can arise because agents are more intrinsically motivated to carry out one of the tasks. For instance, many nurses are attracted to the profession because they derive satisfaction from patient care. To them other duties, such as administration, supervision, and teaching, often are less attractive (e.g., Shields and Ward 2001; for further discussion and arguments related to biases for tasks and task specialization see Prendergast 2007).

In settings with a stochastic relationship between effort and output, our model suggests that an agent performing routine work should be evaluated more frequently against shorter-term targets than an agent performing more creative and more uncertain tasks such as in R&D. Similarly, our model suggests that agents in routine tasks specialize or have narrow performance goals, but that for tasks with more outcome uncertainty we should observe broader goals and more diversification. Another area to which ideas from our paper can be applied are labeling effects. Category or product labels chosen by policy makers and firms often influence decisions. By naming in-kind

benefits “housing benefits” or “child benefits” policy makers may be evoking narrow accounts for these categories. Such narrow accounts prevent substitution away from the targeted category and thereby increase consumption in the targeted category (for a recent discussion of such labeling effects and experimental evidence see Abeler and Marklein 2010). Bertrand *et al.* (2006) discuss that bank accounts could be designed specifically to conform to people’s mental-accounting schemes. They suggest that earmarks such as a “fridge account”, an “education account”, or a “car account” could help the poor manage their spending and savings. Our modeling approach can help understand in what circumstances imposing narrow categories can have a beneficial effect on decisions (from the perspective of the policy maker).

Similarly, our ideas can contribute to the analysis of advertising and pricing decisions by firms. An interesting application is how supermarkets bundle products together to evoke certain psychological accounts, and thereby influence consumers’ purchase decisions. Package size also plays a role for how purchases are bracketed. Wertenbroch (1998), for instance, documents that consumers often purchase “sinful” products like cigarettes in small packages at higher per-unit price, presumably to limit consumption. In line with our narrow accounting story, they find evidence that demand is rationed because consumers’ demand for vice goods increases relatively less in response to price reductions than that for virtue goods.



# Appendix

## A Goal setting with deterministic task outcomes

We repeatedly use the following concepts. For the pair of goals  $(\hat{x}_1, \hat{x}_2)$  and a deviation to  $x_1 < \hat{x}_1$  define the *cost-matching decision*  $x^{cm}(x_1)$  by  $c_1(\hat{x}_1) + c_2(\hat{x}_2) = c_1(x_1) + c_2(x^{cm}(x_1))$ , the *delayed-benefit-matching decision*  $x^{hm}(x_1)$  by  $h_1(\hat{x}_1) + h_2(\hat{x}_2) = h_1(x_1) + h_2(x^{hm}(x_1))$ , and the *immediate-benefit-matching decision*  $x^{bm}(x_1)$  by  $b_1(\hat{x}_1) + b_2(\hat{x}_2) = b_1(x_1) + b_2(x^{bm}(x_1))$ .

### A.1 Proof of Proposition 1

Part 1: see the main text. Part 2 repeatedly uses assumption 1 and  $\lambda > 1$ . Rearranging equation (4) gives

$$c'_i(x_{max,i}^N) = \frac{1 + \eta\beta\lambda}{1 + \eta\beta} b'_i(x_{max,i}^N) + \beta \frac{1 + \eta\lambda}{1 + \eta\beta} h'_i(x_{max,i}^N) > b'_i(x_{max,i}^N) + \beta h'_i(x_{max,i}^N). \quad (22)$$

Because  $b'_i(x_i^*) + \beta h'_i(x_i^*) = c'_i(x_i^*)$  this implies  $x_i^* < x_{max,i}^N$ . Similarly, because  $b'_i(x_{0,i}^*) + h'_i(x_{0,i}^*) = c'_i(x_{0,i}^*)$ , equation (5) implies  $x_{0,i}^* > x_{min,i}^N$ . For  $\beta = 1$  equation (22) implies  $x_{0,i}^* < x_{max,i}^N$ , and for  $\beta = 0$  it implies  $x_{0,i}^* > x_{max,i}^N$ . Because  $\frac{1 + \eta\beta\lambda}{1 + \eta\beta}$  and  $\beta \frac{1 + \eta\lambda}{1 + \eta\beta}$  are strictly increasing in  $\beta$ , the existence of a cut-off  $\tilde{\beta} \in (0, 1)$  follows from the intermediate value theorem. Finally, self 0 implements  $\hat{x}_i = \min\{x_{0,i}^*, x_{max,i}^N\}$  because his utility is increasing in  $x_i$  for  $x_{min,i}^N < x_i < x_{0,i}^*$  and decreasing for  $x_i > x_{0,i}^*$ .

### A.2 Proof of Lemma 1

In a preliminary step, we show that the cost-matching decision is lower than the benefit-matching one:  $x^{hm}(x_1) > x^{cm}(x_1) > \hat{x}_2$ . Note that the iso-cost and iso-benefit curves are tangent at  $\hat{x}_1 = \hat{x}_2$ . For joint deviations from the goals  $x_1 < \hat{x}_1$  and  $x_2 > \hat{x}_2$ , we have  $c'_1(x_1) \leq c'(\hat{x}_1)$  and  $c'(x_2) \geq c'(\hat{x}_2)$  as well as  $h'(x_1) \geq h'(\hat{x}_1)$  and  $h'(x_2) \leq h'(\hat{x}_2)$ . Hence

$$\left| \frac{dx_2}{dx_1} \right|_{\sum_i c(x_i) = \sum_i c(\hat{x}_i)} = \frac{c'(x_1)}{c'(x_2)} \leq \frac{c'(\hat{x}_1)}{c'(\hat{x}_2)} = \frac{h'(\hat{x}_1)}{h'(\hat{x}_2)} = 1 \leq \left| \frac{dx_2}{dx_1} \right|_{\sum_i h(x_i) = \sum_i h(\hat{x}_i)} = \frac{h'_1(x_1)}{h'_2(x_2)}.$$

At least one of the inequalities above is strict (assumption 1).

Next, we show that the utility of self 2 is strictly increasing for  $x_2 < \min\{x^{cm}(x_1), x_{max}^N\}$  and strictly decreasing for  $x_2 > \min\{x^{cm}(x_1), x_{max}^N\}$ . Following a deviation of self 1 to  $x_1 < \hat{x}_1$ , there is a loss in the benefit dimension and a gain in the cost dimension for responses of self 2  $\hat{x}_2 \leq x_2 < \min\{x^{cm}(x_1), x_{max}^N\}$ . For such values of  $x_2$  the derivative of the utility of self 2 is

$$U'_2 = \beta(1 + \eta\lambda)h'(x_2) - (1 + \eta\beta)c'(x_2) > \beta(1 + \eta\lambda)h'(x_{max}^N) - (1 + \eta\beta)c'(x_{max}^N) = 0.$$

For  $x_2 > x_{max}^N$  the utility of self 2 is always strictly decreasing. Part 1 follows because the goal is set to  $\hat{x}_2 = x_{max}^N$ . To show the remainder of part 2, suppose  $x^{cm}(x_1) \leq x_{max}^N$  and consider goal

$\hat{x}_2 = x_0^*$ . Taking the right-derivative of the utility of self 2 at  $x_2 = x^{cm}(x_1)$ :

$$\begin{aligned} U_2' \Big|_{x_2 \downarrow x^{cm}(x_1)} &= \beta(1 + \eta\lambda) h'(x^{cm}(x_1)) - (1 + \beta\eta\lambda) c'(x^{cm}(x_1)) \\ &< (1 + \eta\beta\lambda) [h'(x^{cm}(x_1)) - c'(x^{cm}(x_1))] < 0, \end{aligned}$$

because  $h_2'(x_2) - c_2'(x_2) < 0$  for  $x_2 > x_0^*$ . Thus, the utility of self 2 is strictly decreasing for  $x^{cm}(x_1) < x_2 < x^{hm}(x_1)$ . Similarly, at  $x^{hm}(x_1)$  the left-derivative of the utility of self 2 is  $\beta h_2'(x^{cm}(x_1)) - (1 + \beta\eta\lambda) c_2'(x^{cm}(x_1)) < 0$  and the right-derivative is  $\beta(1 + \eta) h_2'(x^{cm}(x_1)) - (1 + \beta\eta\lambda) c_2'(x^{cm}(x_1)) < 0$ , because  $x_2 > x_{min}^N$  (cf. equation 5). Hence, the utility of self 2 is maximized at  $x_2 = \min\{x^{cm}(x_1), x_{max}^N\}$ .

### A.3 Proof of Proposition 2

By way of contradiction, suppose that  $x_0^*$  is broad-accounting implementable in both periods. We first show that there exist small deviations  $x_1 = x_0^* - \epsilon$ ,  $\epsilon > 0$ , such that  $x^{cm}(x_1) < x_{max}^N$ . As  $x_{max}^N > x_0^*$ , we have  $c(x_{max}^N) + c(x_0^*) > c(x_0^*) + c(x_0^*)$ . By continuity, there exists an  $\epsilon > 0$  such that  $c(x_{max}^N) + c(x_0^* - \epsilon) > c(x_0^*) + c(x_0^*)$ . By construction,  $c(x^{cm}(x_1)) + c(x_0^* - \epsilon) = c(x_0^*) + c(x_0^*)$ . Hence,  $c(x_{max}^N) + c_1(x_0^* - \epsilon) > c(x^{cm}(x_1)) + c(x_0^* - \epsilon)$ , which implies  $x^{cm}(x_1) < x_{max}^N$ .

Self 2 responds with cost-matching to such a small deviation (cf. lemma 1). Furthermore, we know from the proof of lemma 1 that  $x^{hm}(x_1) > x^{cm}(x_1) > x_0^*$ , i.e., there is a loss in the benefit dimension. Hence, the left-derivative of the utility of self 1 at decision  $x_1 = x_0^*$  is

$$U_1' \Big|_{\epsilon \downarrow 0} = \beta(1 + \eta\lambda) h'(x_0^*) + \beta(1 + \eta\lambda) h'(x_0^*) \frac{dx_2}{dx_1} - \left[ c'(x_0^*) + \beta c'(x_0^*) \frac{dx_2}{dx_1} \right].$$

Substituting in for  $\frac{dx_2}{dx_1} = -\frac{c'(x_1)}{c'(x_2)}$  gives

$$\begin{aligned} U_1' \Big|_{\epsilon \downarrow 0} &= \beta(1 + \eta\lambda) \left[ h'(x_0^*) - h'(x_0^*) \frac{c'(x_0^*)}{c'(x_0^*)} \right] - \left[ c'(x_0^*) - \beta c'(x_0^*) \frac{c'(x_0^*)}{c'(x_0^*)} \right] \\ &= -(1 - \beta) c'(x_0^*) < 0. \end{aligned}$$

Hence, deviating from  $x_0^*$  increases the utility of self 1. Thus,  $x_0^*$  is not broad-accounting implementable.

### A.4 Proof of Proposition 3

For both parts note that  $x_{0,i}^*$ ,  $i = 1, 2$  being narrow-accounting implementable rules out unilateral deviations (cf. lemma 2). Furthermore, it rules out joint deviations in which self 1 either increases both decisions or decreases both decisions (changing both decisions in the same way results in the same gains and losses as a unilateral deviation). So we only need to consider joint deviations in which self 1 increases one decision and decreases the other.

#### Part 1

Without loss of generality, suppose that self 1 deviates to  $x_1' < x_{0,1}^*$  and  $x_2' > x_{0,2}^*$ . We are interested in the utility difference  $U_1(x_1', x_2' | x_{0,1}^*, x_{0,2}^*) - U_1(x_{0,1}^*, x_{0,2}^* | x_{0,1}^*, x_{0,2}^*)$ . By the gradient theorem, we

can express this change in utility as a line integral along a curve with endpoints  $(x_{0,1}^*, x_{0,2}^*)$  and  $(x'_1, x'_2)$ . That is, denoting  $U_1(\cdot) \equiv U_1(\cdot | x_{0,1}^*, x_{0,2}^*)$ ,

$$U_1(x'_1, z(x'_1)) - U_1(x_{0,1}^*, z(x_{0,1}^*)) = - \int_{x'_1}^{x_{0,1}^*} \underbrace{\left[ \frac{\partial}{\partial x_1} U_1(x_1, z(x_1)) + \frac{\partial}{\partial x_2} U_1(x_1, z(x_1)) \frac{dz(x_1)}{dx_1} \right]}_{\equiv \phi'(x_1)} dx_1, \quad (23)$$

where  $z : [x'_1, x_{0,1}^*] \rightarrow [x_{0,2}^*, x'_2]$  is an arbitrary bijective function with  $z(x'_1) = x'_2$  and  $z(x_{0,1}^*) = x_{0,2}^*$ . Further, the gradient theorem implies that the line integral does not depend on the path  $z(x_1)$  between the two endpoints, which will turn out to be useful for our purposes. We show that the integrand  $\phi'(x_1) > 0$  for any  $x_1 \in [x'_1, x_{0,1}^*]$ , which implies that the utility difference in (23) is negative. That is, any deviation  $(x'_1, x'_2)$  causes a drop in overall utility.

We start with a few useful observations on the relevant cases to consider. If there are both immediate and delayed benefits, condition (7) implies that the delayed and immediate iso-benefit curves lie on top of each other. Specifically, condition (7) holds if and only if

$$\frac{b'_1(x_1) + \beta h'_1(x_1)}{b'_2(x_2) + \beta h'_2(x_2)} - \frac{b'_1(x_1) + h'_1(x_1)}{b'_2(x_2) + h'_2(x_2)} \propto (1 - \beta) [b'_1(x_1) h'_2(x_2) - h'_1(x_1) b'_2(x_2)] = 0 \quad \forall (x_1, x_2).$$

There are five cases where this holds: (i)  $\beta = 1$ , (ii)  $h'_1(x_1) = h'_2(x_2) = 0 \forall (x_1, x_2)$ , (iii)  $b'_1(x_2) = b'_2(x_2) = 0 \forall (x_1, x_2)$ , (iv)  $h'_i(x_i) = b'_i(x_i) = 0 \forall x_i$  for one task  $i \in \{1, 2\}$ , and (v)  $\frac{h'_1(x_1)}{h'_2(x_2)} = \frac{b'_1(x_1)}{b'_2(x_2)} \forall (x_1, x_2)$ . Cases (i) and (ii) each imply that  $x_{0,i}^*$  globally maximizes the utility of self 1, and hence he has no incentive to deviate from the goals. We excluded case (iv) by assumption. This leaves cases (iii) and (v) to consider in the proof.

There are three possible types of deviations  $(x'_1, x'_2)$ : (a)  $x_{0,2}^* < z(x'_1) < x^{cm}(x_1)$ , (b)  $x^{cm}(x_1) \leq z(x'_1) < x^{bm}(x_1) = x^{hm}(x_1)$ , (c)  $x^{bm}(x_1) \leq z(x'_1)$ . Figure 4 illustrates these.

**(a)** Because  $(x'_1, x'_2)$  lies below the iso-cost curve that goes through  $(x_{0,1}^*, x_{0,2}^*)$ , there exists a path connecting the two points for which at every point the slope is less steep than the slope of the iso-cost curve, which in turn is less steep than the slope of the two iso-benefit curves. That is,

$$0 < -\frac{dz(x_1)}{dx_1} < \frac{c'_1(x_1)}{c'_2(x_2)} < \frac{b'_1(x_1)}{b'_2(x_2)} = \frac{h'_1(x_1)}{h'_2(x_2)}. \quad (24)$$

Hence, at every point along this path the individual would experience a gain in the cost dimension and a loss in both benefit dimensions. Exploiting path-independence, we can fix a feasible path by imposing conditions (24), and write the integrand for a given  $x_1$  as follows:

$$\begin{aligned} \phi'(x_1) &= \kappa \left[ b'_1(x_1) + b'_2(z(x_1)) \frac{dz(x_1)}{dx_1} \right] + \rho \left[ h'_1(x_1) + h'_2(z(x_1)) \frac{dz(x_1)}{dx_1} \right] \\ &\quad - \gamma \left[ c'_1(x_1) + c'_2(z(x_1)) \frac{dz(x_1)}{dx_1} \right] \\ &= \kappa b'_1(x_1) \left[ 1 - \frac{b'_2(z(x_1))}{b'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) \right] + \rho h'_1(x_1) \left[ 1 - \frac{h'_2(z(x_1))}{h'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) \right] \\ &\quad - \gamma c'_1(x_1) \left[ 1 - \frac{c'_2(z(x_1))}{c'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) \right], \end{aligned} \quad (25)$$

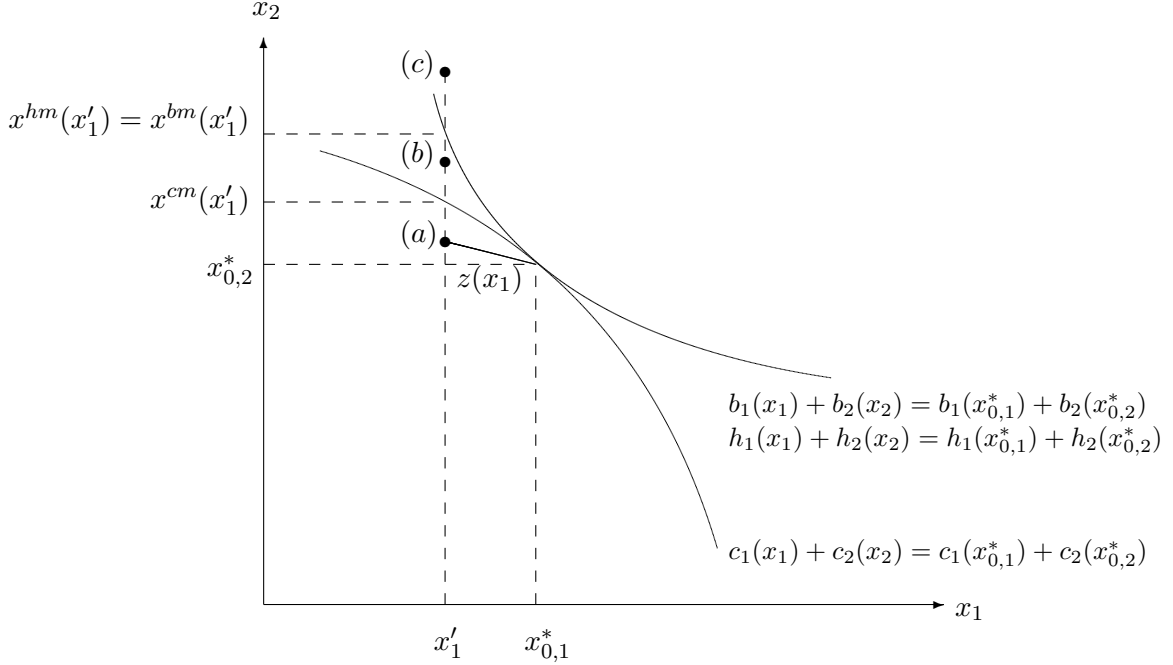


Figure 4: Illustration of possible deviations

where  $\kappa = 1 + \eta \beta \lambda$ ,  $\rho = \beta(1 + \eta \lambda)$ , and  $\gamma = 1 + \eta \beta$ . Conditions (24) give

$$0 < \frac{c'_2(z(x_1))}{c'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) < 1 \quad \text{and} \quad \frac{c'_2(z(x_1))}{c'_1(x_1)} > \frac{b'_2(z(x_1))}{b'_1(x_1)} = \frac{h'_2(z(x_1))}{h'_1(x_1)},$$

which implies that

$$1 - \frac{b'_2(z(x_1))}{b'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) > 1 - \frac{c'_2(z(x_1))}{c'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) > 0. \quad (26)$$

Hence,

$$\phi'(x_1) > \underbrace{[\kappa b'_1(x_1) + \rho h'_1(x_1) - \gamma c'_1(x_1)]}_{>0 \text{ for } x_1 < x_{0,1}^* \leq x_{max,1}^N; \text{ see eq. (4)}} \underbrace{\left[ 1 - \frac{c'_2(z(x_1))}{c'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) \right]}_{>0; \text{ see eq. (26)}} > 0.$$

(b) Here  $(x'_1, x'_2)$  lies between the iso-cost and iso-benefit curves, so there exists a path with

$$0 < \frac{c'_1(x_1)}{c'_2(z(x_1))} \leq -\frac{dz(x_1)}{dx_1} < \frac{b'_1(x_1)}{b'_2(z(x_1))} = \frac{h'_1(x_1)}{h'_2(z(x_1))}. \quad (27)$$

At every point along the path the individual would experience a loss in the cost and benefit dimensions. So we replace in equation (25)  $\kappa = 1 + \eta \beta \lambda$ ,  $\rho = \beta(1 + \eta \lambda)$ , and  $\gamma = 1 + \eta \beta \lambda$ . The first two terms remain positive, but conditions (27) now imply  $1 - \frac{c'_2(z(x_1))}{c'_1(x_1)} \left( -\frac{dz(x_1)}{dx_1} \right) \leq 0$ . Hence,  $\phi'(x_1) > 0$ .

(c) Here  $(x'_1, x'_2)$  lies above the iso-cost and iso-benefit curves, so there exists a path with

$$0 < \frac{c'_1(x_1)}{c'_2(z(x_1))} < \frac{b'_1(x_1)}{b'_2(z(x_1))} = \frac{h'_1(x_1)}{h'_2(z(x_1))} \leq -\frac{dz(x_1)}{dx_1}. \quad (28)$$

At every point along the path the individual would experience a gain in the benefit dimensions and a loss in the cost dimension. So we replace in equation (25)  $\kappa = 1 + \eta\beta$ ,  $\rho = \beta(1 + \eta)$ , and  $\gamma = 1 + \eta\beta\lambda$ . Rearranging terms gives

$$\begin{aligned}\phi'(x_1) &= \gamma c'_2(z(x_1)) \left[ \left( -\frac{dz(x_1)}{dx_1} \right) - \frac{c'_1(x_1)}{c'_2(z(x_1))} \right] - \kappa b'_2(z(x_1)) \left[ \left( -\frac{dz(x_1)}{dx_1} \right) - \frac{b'_1(x_1)}{b'_2(z(x_1))} \right] \\ &\quad - \rho h'_2(z(x_1)) \left[ \left( -\frac{dz(x_1)}{dx_1} \right) - \frac{h'_1(x_1)}{h'_2(z(x_1))} \right] \\ &> \underbrace{\left[ \gamma c'_2(z(x_1)) - \kappa b'_2(z(x_1)) - \rho h'_2(z(x_1)) \right]}_{>0 \text{ for } x_2 > x_{0,2}^* \geq x_{min,2}^N; \text{ see eq. (5)}} \underbrace{\left[ \left( -\frac{dz(x_1)}{dx_1} \right) - \frac{b'_1(x_1)}{b'_2(z(x_1))} \right]}_{\geq 0; \text{ see eq. (28)}} \geq 0.\end{aligned}$$

The second-to-last inequality exploits conditions (28).

## Part 2

We prove sufficiency by showing that a profitable deviation along the iso-cost curve or the iso-benefit curve exists. Without loss of generality, suppose that decisions involve the following relative bias (the other case is symmetric):

$$\frac{B'_{1,1}(x_{0,1}^*)}{B'_{1,2}(x_{0,2}^*)} > \frac{C'_{1,1}(x_{0,1}^*)}{C'_{1,2}(x_{0,2}^*)}. \quad (29)$$

We show that if decisions involve a relative bias then  $1 + \eta\beta - \beta(1 + \eta\lambda) > 0 \Leftrightarrow (1 - \beta)/\beta > \eta(\lambda - 1)$  is sufficient for a profitable joint deviation to exist.

**Case 1: Decisions involve immediate and delayed benefits.** Related to the observations in part 1 when condition (7) can hold, one can show that condition (29) admits cases where for  $(x_1, x_2) \gg 0$  we have  $b_1(x_1) > 0$ ,  $h_1(x_1) \geq 0$ ,  $b_2(x_2) \geq 0$ ,  $h_2(x_2) > 0$ ,  $c_1(x_1) > 0$ ,  $c_2(x_2) > 0$ .

We argue that self 1 has a profitable joint deviation along the iso-cost curve. To show this, we have to determine whether such a deviation results in losses or gains in the two benefit dimensions, which depends on the shape of the iso-benefit and iso-cost curves at  $x_{0,i}^*$ . We claim that

$$\frac{b'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*)} > \frac{c'_1(x_{0,1}^*)}{c'_2(x_{0,2}^*)} > \frac{h'_1(x_{0,1}^*)}{h'_2(x_{0,2}^*)}. \quad (30)$$

Condition (29) and equation (1) imply that  $\frac{b'_1(x_{0,1}^*) + \beta h'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*) + \beta h'_2(x_{0,2}^*)} > \frac{c'_1(x_{0,1}^*)}{c'_2(x_{0,2}^*)} = \frac{b'_1(x_{0,1}^*) + h'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*) + h'_2(x_{0,2}^*)}$ . From this it follows that  $\frac{b'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*)} > \frac{h'_1(x_{0,1}^*)}{h'_2(x_{0,2}^*)}$ , because

$$\frac{b'_1(x_{0,1}^*) + \beta h'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*) + \beta h'_2(x_{0,2}^*)} - \frac{b'_1(x_{0,1}^*) + h'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*) + h'_2(x_{0,2}^*)} \propto (1 - \beta) [b'_1(x_{0,1}^*) h'_2(x_{0,2}^*) - h'_1(x_{0,1}^*) b'_2(x_{0,2}^*)].$$

Now  $\frac{b'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*)} > \frac{h'_1(x_{0,1}^*)}{h'_2(x_{0,2}^*)}$  together with  $\frac{b'_1(x_{0,1}^*) + h'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*) + h'_2(x_{0,2}^*)} = \frac{c'_1(x_{0,1}^*)}{c'_2(x_{0,2}^*)}$  implies that  $\frac{b'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*)} > \frac{c'_1(x_{0,1}^*)}{c'_2(x_{0,2}^*)}$ , because

$$\frac{b'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*)} - \frac{c'_1(x_{0,1}^*)}{c'_2(x_{0,2}^*)} = \frac{b'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*)} - \frac{b'_1(x_{0,1}^*) + h'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*) + h'_2(x_{0,2}^*)} \propto b'_1(x_{0,1}^*) h'_2(x_{0,2}^*) - h'_1(x_{0,1}^*) b'_2(x_{0,2}^*).$$

Condition (30) implies that a move along the iso-cost curve to  $x'_1 > x_{0,1}^*$  and  $x^{cm}(x_1) < x_{0,2}^*$  causes a gain in the immediate benefit dimension and a loss in the delayed benefit dimension relative to sticking with the goals. The utility of self 1 from such a deviation is:

$$\begin{aligned} & b_1(x'_1) + \beta h_1(x'_1) + b_2(x^{cm}(x_1)) + \beta h_2(x^{cm}(x_1)) - [c_1(x'_1) + c_2(x^{cm}(x_1))] \\ & + \beta \underbrace{\mu(b_1(x'_1) + b_2(x^{cm}(x_1)) - b_1(x_{0,1}^*) - b_2(x_{0,2}^*))}_{>0} \\ & + \beta \underbrace{\mu(h_1(x'_1) + h_2(x^{cm}(x_1)) - h_1(x_{0,1}^*) - h_2(x_{0,2}^*))}_{<0}. \end{aligned}$$

Evaluating the impact on the utility of self 1 of a marginal deviation along the iso-cost curve at point  $(x_{0,1}^*, x_{0,2}^*)$ , we obtain (omitting arguments):

$$\begin{aligned} & (1 + \eta \beta) \left[ b'_1 + b'_2 \frac{dx_2}{dx_1} \Big|_{\sum c_i = \sum \hat{c}_i} \right] + \beta (1 + \eta \lambda) \left[ h'_1 + h'_2 \frac{dx_2}{dx_1} \Big|_{\sum c_i = \sum \hat{c}_i} \right] \\ = & (1 + \eta \beta) \left[ b'_1 - b'_2 \frac{c'_1}{c'_2} \right] + \beta (1 + \eta \lambda) \left[ h'_1 - h'_2 \frac{c'_1}{c'_2} \right] \end{aligned} \quad (31)$$

$$\begin{aligned} = & (1 + \eta \beta) c'_1 \left[ \frac{b'_1}{c'_1} - \frac{b'_2}{c'_2} \right] + \beta (1 + \eta \lambda) c'_1 \left[ \frac{h'_1}{c'_1} - \frac{h'_2}{c'_2} \right] \\ = & [1 + \eta \beta - \beta (1 + \eta \lambda)] c'_1 \left[ \frac{b'_1}{c'_1} - \frac{b'_2}{c'_2} \right], \end{aligned} \quad (32)$$

where the last line uses  $c'_i(x_{0,i}^*) = b'_i(x_{0,i}^*) + h'_i(x_{0,i}^*)$  to substitute for  $\frac{h'_i}{c'_i} = 1 - \frac{b'_i}{c'_i}$ . Condition (30) implies that  $\frac{b'_1}{c'_1} - \frac{b'_2}{c'_2} > 0$ . Hence, a joint deviation along the iso-cost curve increases the utility of self 1 if  $1 + \eta \beta > \beta (1 + \eta \lambda)$ .

**Case 2: One decision involves delayed benefits, the other delayed costs (in addition to immediate costs and benefits).** Consider a deviation to  $x'_1 > x_{0,1}^*$  and  $x^{cm}(x'_1) < x_{0,2}^*$ , i.e., a deviation that holds immediate costs constant. By similar arguments as for case 1, there is a loss in the delayed outcome dimension,  $\mu(h_1(x'_1) + h_2(x^{cm}(x'_1)) - h_1(x_{0,1}^*) - h_2(x_{0,1}^*)) < 0$ , and a gain in the immediate benefit dimension,  $\mu(b_1(x'_1) + b_2(x^{cm}(x'_1)) - b_1(x_{0,1}^*) - b_2(x_{0,1}^*)) > 0$ . The reason is that for the immediate outcomes on their own, the iso-benefit curve has a steeper slope than the iso-cost curve:  $\frac{b'_1(x_{0,1}^*)}{b'_2(x_{0,2}^*)} > \frac{c'_1(x_{0,1}^*)}{c'_2(x_{0,2}^*)}$ . To see this, note that the definition of  $x_{0,i}^*$ , namely  $b'_i(x_{0,i}^*) + h'_i(x_{0,i}^*) = c'_i(x_{0,i}^*)$ , implies  $b'_1(x_{0,1}^*) > c'_1(x_{0,1}^*)$  (because  $h'_1(\cdot) < 0$ ) and  $b'_2(x_{0,2}^*) < c'_2(x_{0,2}^*)$  (because  $h'_2(\cdot) > 0$ ). Evaluating the impact on the utility of self 1 of a marginal deviation along the iso-cost curve at point  $(x_{0,1}^*, x_{0,2}^*)$ , and rearranging using  $c'_i(x_{0,i}^*) = b'_i(x_{0,i}^*) + h'_i(x_{0,i}^*)$ , we again obtain equation (32). As just argued,  $\frac{b'_1(x_{0,1}^*)}{c'_1(x_{0,1}^*)} - \frac{b'_2(x_{0,2}^*)}{c'_2(x_{0,2}^*)} > 0$ . Hence, a joint deviation along the iso-cost curve increases the utility of self 1 if  $1 + \eta \beta > \beta (1 + \eta \lambda)$ .

## A.5 A sufficient condition for broad accounts to do as well as narrow accounts in the case where $x_{0,i}^* > x_{max,i}^N$

Matching goal levels  $x_{max,i}^N$  locally maximizes the utility of self 1 under broad accounts. Checking that this is a global maximum however is complicated by a large number of case distinctions. But one can rule out profitable joint deviations for particular orderings of the iso-cost and the two iso-benefit curves, which gives a sufficient condition for broad and narrow accounts to do equally well.

**Proposition 7** *Suppose either one or both tasks involve delayed benefits. Suppose  $x_{0,i}^* > x_{max,i}^N$ , so that self 0 sets goals  $(x_{max,1}^N, x_{max,2}^N)$  under narrow accounts. If for  $x_i \leq x_{max,i}$  and  $x_j \geq x_{max,j}$ ,  $\frac{c'_i(x_i)}{c'_j(x_j)} \leq \min \left\{ \frac{b'_i(x_i)}{b'_j(x_j)}, \frac{h'_i(x_i)}{h'_j(x_j)} \right\}$ , then  $(x_{max,1}^N, x_{max,2}^N)$  is broad-accounting implementable.*

**Proof.** We show that no profitable global joint deviation exists if  $\frac{c'_1(x_1)}{c'_2(x_2)} < \frac{b'_1(x_1)}{b'_2(x_2)} \leq \frac{h'_1(x_1)}{h'_2(x_2)}$  (the case where  $\frac{c'_1(x_1)}{c'_2(x_2)} < \frac{h'_1(x_1)}{h'_2(x_2)} < \frac{b'_1(x_1)}{b'_2(x_2)}$  is analogous). For deviations  $(x_1, x_2)$  with  $z(x_1) < x^{cm}(x_1)$ , or  $x^{cm}(x_1) \leq z(x_1) < x^{bm}(x_1)$ , or  $z(x_1) > x^{hm}(x_1)$  the analysis is analogous to part 1 (a)-(c) of the proof of proposition 3. To consider the remaining case  $x^{bm}(x_1) \leq z(x_1) \leq x^{hm}(x_1)$ , rearrange equation (25):

$$\begin{aligned} \phi'(x_1) &= \kappa b'(z(x_1)) \left[ \frac{b'_1(x_1)}{b'_2(x_2)} - \left( -\frac{dz(x_1)}{x_1} \right) \right] + \rho h'(z(x_1)) \left[ \frac{h'_1(x_1)}{h'_2(x_2)} - \left( -\frac{dz(x_1)}{x_1} \right) \right] \\ &\quad - \gamma c'(z(x_1)) \left[ \frac{c'_1(x_1)}{c'_2(x_2)} - \left( -\frac{dz(x_1)}{x_1} \right) \right]. \end{aligned}$$

Similar to the proof of proposition 3,  $\frac{h'_1(x_1)}{h'_2(x_2)} - \left( -\frac{dz(x_1)}{x_1} \right) \geq 0$ . Furthermore,  $\frac{c'_1(x_1)}{c'_2(x_2)} < -\frac{dz(x_1)}{x_1} \leq \frac{b'_1(x_1)}{b'_2(x_2)}$  implies

$$\begin{aligned} &\kappa b'(z(x_1)) \left[ \frac{b'_1(x_1)}{b'_2(x_2)} - \left( -\frac{dz(x_1)}{x_1} \right) \right] - \gamma c'(z(x_1)) \left[ \frac{c'_1(x_1)}{c'_2(x_2)} - \left( -\frac{dz(x_1)}{x_1} \right) \right] \\ &> [\kappa b'(z(x_1)) - \gamma c'(z(x_1))] \left[ \frac{c'_1(x_1)}{c'_2(x_2)} - \left( -\frac{dz(x_1)}{x_1} \right) \right] > 0, \end{aligned}$$

where  $\kappa b'(z(x_1)) - \gamma c'(z(x_1)) < 0$  follows from  $\kappa = 1 + \beta \eta$  and  $\gamma = 1 + \beta \lambda \eta$  and  $z(x_1) > x_{max,2}$ .

## B Goal setting with stochastic task outcomes

### B.1 Proof of Proposition 4

**Narrow accounts:** The utility of self 0 under narrow accounts is given by

$$\beta \left\{ h_L(x_i) + \pi [1 - \eta(\lambda - 1)(1 - \pi)] [h_H(x_i) - h_L(x_i)] - c(x_i) \right\} \quad (33)$$

The preferred decision of self 0 under narrow accounts is characterized by the first-order condition

$$h'_L(x_0^N) + \pi [1 - \eta(\lambda - 1)(1 - \pi)] [h'_H(x_0^N) - h'_L(x_0^N)] = c'(x_0^N). \quad (34)$$

Assumption 2 and  $h'_H(x_i) > h'_L(x_i)$  imply that  $x_0^N > 0$  and that  $x_0^N$  is strictly increasing in  $\pi$ . In addition,  $h''_H(x_i) - h''_L(x_i) < 0$  implies that the second-order condition is satisfied.

**Broad accounts:** The utility of self 0 from  $(x_1, x_2)$  is

$$\beta \left\{ \sum_i \left[ h_L(x_i) + \pi [1 - \eta(\lambda - 1)(1 - \pi)\{1 - \pi(1 - \pi)\}] [h_H(x_i) - h_L(x_i)] - c(x_i) \right] \right\} \quad (35)$$

$$+ \beta \eta \pi^2 (1 - \pi)^2 \left[ \mu(h_H(x_2) - h_H(x_1) + h_L(x_1) - h_L(x_2)) + \mu(h_H(x_1) - h_H(x_2) + h_L(x_2) - h_L(x_1)) \right].$$

We first argue that the preferred decisions of self 0 under broad accounts are symmetric. Note that for  $x_1 \neq x_2$ ,  $\mu(h_H(x_2) - h_H(x_1) + h_L(x_1) - h_L(x_2))$  is a gain if and only if  $\mu(h_H(x_1) - h_H(x_2) + h_L(x_2) - h_L(x_1))$  is a loss. Asymmetric decisions therefore reduce utility, because the last line of equation (35) is negative for  $x_1 \neq x_2$  and equal to zero for  $x_1 = x_2$ . Hence, the preferred decision of self 0 under broad accounts is characterized by the following first-order condition

$$h'_L(x_0^B) + \pi \left[ 1 - \eta(\lambda - 1)(1 - \pi)\{1 - \pi(1 - \pi)\} \right] \left[ h'_H(x_0^B) - h'_L(x_0^B) \right] = c'(x_0^B). \quad (36)$$

Assumption 2 and  $h'_H(x_i) > h'_L(x_i)$  imply that  $x_0^B > 0$  and  $x_0^B$  is strictly increasing in  $\pi$ . In addition,  $h''_H(x_i) - h''_L(x_i) < 0$  implies that the second-order conditions are satisfied.

**Comparison:** Note that  $[1 - \eta(\lambda - 1)(1 - \pi)\{1 - \pi(1 - \pi)\}] > [1 - \eta(\lambda - 1)(1 - \pi)]$ . From equations (33) - (36) it hence follows that  $x_0^B > x_0^N$  and that the utility of self 0 under broad accounts is always strictly larger than the utility of self 0 under narrow accounts.

## B.2 Derivation of implementable goals

We show the results from the text for the more general case  $h_L(x_i) > 0$ .

### B.2.1 Narrow accounts

**Lemma 4** *The maximal narrow-accounting implementable goal is defined by*

$$\beta \left\{ \pi [1 + \eta(1 + \pi(\lambda - 1))] h'_H(x_{max}^N) + (1 - \pi)(1 + \eta\lambda) h'_L(x_{max}^N) \right\} = (1 + \beta\eta) c'(x_{max}^N). \quad (37)$$

*The minimal narrow-accounting implementable is defined by*

$$\beta \left\{ \pi(1 + \eta) h'_H(x_{min}^N) + (1 - \pi)[1 + \eta(1 + \pi(\lambda - 1))] h'_L(x_{min}^N) \right\} = (1 + \beta\eta\lambda) c'(x_{min}^N). \quad (38)$$

Furthermore,  $x_{max}^N > x_0^N$  for  $\beta = 1$ .

**Proof.**

**Utility of self 1:** The consumption utility of self 1 in task  $i$  given decision  $x_i$  is

$$\beta \left\{ \pi h_H(x_i) + (1 - \pi) h_L(x_i) \right\} - c(x_i).$$

The gain-loss utility for a deviation to  $x_i > \hat{x}_i$



- in the cost dimension is  $\beta \eta \lambda (c(\hat{x}_i) - c(x_i))$ ;

- in the benefit dimension for  $h_L(x_i) < h_H(\hat{x}_i)$  is

$$\beta \eta \{ \pi h_H(x_i) + (1 - \pi) [1 + \pi (\lambda - 1)] h_L(x_i) - \pi [\pi + \lambda (1 - \pi)] h_H(\hat{x}_i) - (1 - \pi) h_L(\hat{x}_i) \};$$

- in the benefit dimension for  $h_L(x_i) \geq h_H(\hat{x}_i)$  is

$$\beta \eta \{ \pi h_H(x_i) + (1 - \pi) h_L(x_i) - \pi h_H(\hat{x}_i) - (1 - \pi) h_L(\hat{x}_i) \}.$$

The gain-loss utility for a deviation to  $x_i < \hat{x}_i$

- in the cost dimension is  $\beta \eta (c(\hat{x}_i) - c(x_i))$ ;

- in the benefit dimension for  $h_H(x_i) \geq h_L(\hat{x}_i)$  is

$$\beta \eta \{ \pi [1 + \pi (\lambda - 1)] h_H(x_i) + \lambda (1 - \pi) h_L(x_i) - \lambda \pi h_H(\hat{x}_i) - (1 - \pi) [\pi + \lambda (1 - \pi)] h_L(\hat{x}_i) \};$$

- in the benefit dimension for  $h_H(x_i) < h_L(\hat{x}_i)$  is

$$\beta \eta \{ \lambda \pi [1 + \pi (\lambda - 1)] h_H(x_i) + \lambda (1 - \pi) h_L(x_i) - \lambda \pi h_H(\hat{x}_i) - \lambda (1 - \pi) h_L(\hat{x}_i) \}.$$

Adding up consumption and gain-loss utility and differentiating, yields the minimal and maximal implementable goal, respectively. We go through the case distinctions.

**Maximal implementable goal:** Self 1 has no incentive to deviate and lower his decision level such that  $h_H(x_i) \geq h_L(\hat{x}_i)$ , as long as the goal does not exceed  $x_{max}^N$ , defined by

$$\beta \{ \pi [1 + \eta (1 - \pi + \pi \lambda)] h'_H(x_{max}^N) + (1 - \pi) (1 + \eta \lambda) h'_L(x_{max}^N) \} = (1 + \beta \eta) c'(x_{max}^N).$$

Self 1 has no incentive to deviate and lower his decision level such that  $h_H(x_i) < h_L(\hat{x}_i)$ , as long as the goal does not exceed  $x_{max,2}^N$ , defined by

$$\beta \{ \pi (1 + \eta \lambda) h'_H(x_{max,2}^N) + (1 - \pi) (1 + \eta \lambda) h'_L(x_{max,2}^N) \} = (1 + \beta \eta) c'(x_{max,2}^N).$$

Note that  $x_{max,2}^N > x_{max}^N$  for  $\pi < 1$  (with equality for  $\pi = 1$ ). Hence,  $x_{max}^N$  is the maximal implementable goal. For  $\beta = 1$ ,  $x_{max}^N > x_0^N$ . For small  $\beta$ ,  $x_{max}^N < x_0^N$  can arise.

**Minimal implementable goal:** Self 1 has no incentive to deviate and increase his decision such that  $h_L(x_i) < h_H(\hat{x}_i)$ , as long as the goal does not fall below  $x_{min}^N$ , defined by

$$\beta \{ \pi (1 + \eta) h'_H(x_{min}^N) + (1 - \pi) [1 + \eta (1 - \pi + \lambda \pi)] h'_L(x_{min}^N) \} = (1 + \beta \eta \lambda) c'(x_{min}^N).$$

Self 1 has no incentive to deviate and increase his decision level such that  $h_L(x_i) \geq h_H(\hat{x}_i)$ , as long as the goal does not fall below  $x_{min,2}^N$ , defined by

$$\beta (1 + \eta) \{ \pi h'_H(x_{min,2}^N) + (1 - \pi) h'_L(x_{min,2}^N) \} = (1 + \beta \eta \lambda) c'(x_{min,2}^N).$$

Note that  $x_{min,2}^N < x_{min}^N$  for  $\pi < 1$  (with equality for  $\pi = 1$ ). Hence,  $x_{min}^N$  is the minimal implementable goal. ■

### B.2.2 Unilateral deviations under broad accounts

We determine the upper (lower) bound for the goal, such that self 1 has no incentive to decrease (increase) his decision in one task. Without loss of generality, we consider deviations in task 2 and fix the decision in task 1. The following result shows that the range of potentially implementable goals is larger with symmetric goals ( $\hat{x}_1 = \hat{x}_2$ ) than with increasing goals ( $\hat{x}_1 < \hat{x}_2$ ) or decreasing goals ( $\hat{x}_1 > \hat{x}_2$ ). Together with the fact that self 0 has symmetric preferred decisions, this implies that asymmetric goals are never optimal.

**Lemma 5** *The upper bound on implementable goals for task 1 or 2, respectively, under broad accounts is greatest with symmetric goals,  $\hat{x}_1 = \hat{x}_2$ , and satisfies*

$$\beta \left\{ \begin{aligned} &\pi (1 + \eta [1 + (\lambda - 1) \pi (1 + (1 - \pi) (1 - 2\pi))]) h'_H(x_{max}^B) \\ &+ (1 - \pi) (1 + \eta [1 + (\lambda - 1) (1 - \pi (1 - \pi)^2)]) h'_L(x_{max}^B) \end{aligned} \right\} = (1 + \beta \eta) c'(x_{max}^B). \quad (39)$$

*The lower bound on implementable goals for task 1 or 2, respectively, under broad accounts is lowest with symmetric goals and satisfies*

$$\beta \left\{ \begin{aligned} &\pi (1 + \eta [1 + (\lambda - 1) \pi^2 (1 - \pi)]) h'_H(x_{min}^B) \\ &+ (1 - \pi) \left( 1 + \eta \left[ 1 + (\lambda - 1) \pi (1 + (1 - \pi) (1 - 2\pi)) \right] \right) h'_L(x_{min}^B) \end{aligned} \right\} = (1 + \beta \eta \lambda) c'(x_{min}^B). \quad (40)$$

Furthermore,  $x_{max}^B > x_0^B$  for  $\beta = 1$ .

**Proof.**

**Utility of self 1:** The consumption utility of self 1 from decisions  $x_2$  and  $x_1 = \hat{x}_1$  is

$$\beta \{ \pi h_H(x_2) + (1 - \pi) h_L(x_2) + \pi h_H(\hat{x}_1) + (1 - \pi) h_L(\hat{x}_1) \} - c(\hat{x}_1) - c(x_2).$$

The gain-loss utility in the cost dimension

- for  $x_2 > \hat{x}_2$  is  $\beta \eta \lambda (c(\hat{x}_2) - c(x_2))$ ;
- for  $x_2 < \hat{x}_2$  is  $\beta \eta (c(\hat{x}_2) - c(x_2))$ .

The only difference between the cases of symmetric, increasing, and decreasing goals is the gain-loss utility in the benefit dimension after a deviation of self 1 in task 2. To save space, we use  $\kappa$  to summarize the constant terms in the gain loss utility, such as factor  $\times h_H(\hat{x}_i)$  and factor  $\times h_L(\hat{x}_i)$ ; and we drop the proportionality factor  $\beta \eta$ . Further, we refer to the bound  $x_{max}^B$  ( $x_{min}^B$ ) simply as maximal (minimal) implementable goal.

### Symmetric goals ( $\hat{x}_1 = \hat{x}_2$ )

**Maximal implementable goal:** Consider first deviations to  $x_2 < \hat{x}_2$ , where  $\hat{x}_2 - x_2$  is small in the sense that  $h_H(x_2) \geq h_L(\hat{x}_2)$  (condition A),  $h_H(\hat{x}_1) + h_L(x_2) \geq h_L(\hat{x}_1) + h_L(\hat{x}_2)$  (condition B), and  $h_H(\hat{x}_1) + h_H(x_2) \geq h_L(\hat{x}_1) + h_L(\hat{x}_2)$  (condition C). The gain-loss utility in the benefit dimension is given by

$$\begin{aligned}
& (1 - \pi)^2 \left\{ \underbrace{\pi^2 h_H(x_2)}_c + \underbrace{\pi(1 - \pi) h_H(x_2)}_a + \underbrace{\pi(1 - \pi) h_L(x_2)}_b + \lambda(1 - \pi)^2 h_L(x_2) \right\} \\
& + 2(1 - \pi) \pi \left\{ \underbrace{\pi^2 h_H(x_2)}_a + \overbrace{\lambda \pi(1 - \pi) [h_H(x_2) + h_L(x_2)]}^I + \lambda(1 - \pi)^2 h_L(x_2) \right\} \\
& + \lambda \pi^2 \left\{ \pi^2 h_H(x_2) + \pi(1 - \pi) [h_H(x_2) + h_L(x_2)] + (1 - \pi)^2 h_L(x_2) \right\} + \kappa \\
= & \pi \left\{ 1 + (\lambda - 1) \pi [1 + (1 - \pi)(1 - 2\pi)] \right\} h_H(x_2) \\
& + (1 - \pi) \left\{ 1 + (\lambda - 1) [1 - \pi(1 - \pi)^2] \right\} h_L(x_2) + \kappa.
\end{aligned} \tag{41}$$

The terms underbraced by  $a - c$  are gains (i.e., not multiplied by  $\lambda$ ), because conditions A - C hold. Adding up the gain-loss utility in the benefit and cost dimensions as well as consumption utility, and differentiating yields the definition of  $x_{max}^B$  in equation (39).

For larger deviations, self 1 is more likely to suffer a loss compared to the case above. For example, if condition A is violated, the terms underbraced by  $a$  in equation (41) are multiplied by  $\lambda$ . Hence, this case gives a bound on implementable goals that is larger than  $x_{max}^B$ . A similar argument applies for deviations such that conditions B and C fail. Overall,  $x_{max}^B$  is the lowest of these bounds and, hence, the maximal implementable goal.

**Minimal implementable goal:** Consider deviations to  $x_2 > \hat{x}_2$ , where  $x_2 - \hat{x}_2$  is small in the sense that  $h_L(x_2) < h_H(\hat{x}_2)$  (condition J),  $h_L(\hat{x}_1) + h_H(x_2) < h_H(\hat{x}_1) + h_H(\hat{x}_2)$  (condition K), and  $h_L(\hat{x}_1) + h_L(x_2) < h_H(\hat{x}_1) + h_H(\hat{x}_2)$  (condition L). The gain-loss utility in the benefit dimension is given by:

$$\begin{aligned}
& (1 - \pi)^2 \left\{ \pi^2 h_H(x_2) + \pi(1 - \pi) h_H(x_2) + \pi(1 - \pi) h_L(x_2) + (1 - \pi)^2 h_L(x_2) \right\} \\
& + 2(1 - \pi) \pi \left\{ \pi^2 h_H(x_2) + \overbrace{\pi(1 - \pi) h_H(x_2) + \pi(1 - \pi) h_L(x_2)}^{IV} + \underbrace{\lambda(1 - \pi)^2 h_L(x_2)}_j \right\} \\
& + \pi^2 \left\{ \pi^2 h_H(x_2) + \underbrace{\lambda \pi(1 - \pi) h_H(x_2)}_k + \underbrace{\lambda \pi(1 - \pi) h_L(x_2)}_j + \underbrace{\lambda(1 - \pi)^2 h_L(x_2)}_l \right\} + \kappa \\
= & \pi \left\{ 1 + (\lambda - 1) \pi^2 (1 - \pi) \right\} h_H(x_2) \\
& + (1 - \pi) \left\{ 1 + (\lambda - 1) \pi [1 + (1 - \pi)(1 - 2\pi)] \right\} h_L(x_2) + \kappa.
\end{aligned} \tag{42}$$

The terms underbraced by  $j - l$  are losses (i.e., multiplied by  $\lambda$ ), because conditions K - L hold. Adding up the gain-loss utility in the benefit and cost dimensions as well as consumption utility, and differentiating yields the definition of  $x_{min}^B$  in equation (40).

For larger deviations, self 1 is more likely to experience a gain compared to the case above. For example, if condition J is violated, the terms underbraced by  $j$  in equation are multiplied by 1

instead of  $\lambda$ . Hence, this case gives a bound on implementable goals that is smaller than  $x_{min}^B$ . A similar argument applies for deviations such that conditions  $K$  and  $L$  fail. Overall,  $x_{min}^B$  is the highest of these bounds and, hence, the minimal implementable goal.

### Increasing goals ( $\hat{x}_1 < \hat{x}_2$ )

**Maximal implementable goal:** Consider deviations to  $x_2 < \hat{x}_2$ , where  $\hat{x}_2 - x_2$  is small in the sense that conditions  $A - C$  from above hold and, in addition,  $h_H(\hat{x}_1) + h_H(x_2) > h_L(\hat{x}_1) + h_H(\hat{x}_2)$  (condition  $D$ ), and  $h_L(\hat{x}_1) + h_H(x_2) > h_H(\hat{x}_1) + h_L(\hat{x}_2)$  (condition  $E$ ) or  $h_H(\hat{x}_1) + h_L(x_2) > h_L(\hat{x}_1) + h_H(\hat{x}_2)$  (condition  $F$ ). Note that at most one of the conditions ( $E$ ) and ( $F$ ) can hold for small deviations, because either  $h_L(\hat{x}_1) + h_H(\hat{x}_2)$  is greater, equal, or less than  $h_H(\hat{x}_1) + h_L(\hat{x}_2)$ . Start from the following expression:

$$\begin{aligned}
& (1 - \pi)^2 \left\{ \underbrace{\pi^2 h_H(x_2)}_c + \underbrace{\pi(1 - \pi) h_H(x_2)}_a + \underbrace{\pi(1 - \pi) h_L(x_2)}_b + \lambda(1 - \pi)^2 h_L(x_2) \right\} \\
& + (1 - \pi) \pi \left\{ \underbrace{\pi^2 h_H(x_2)}_a + \overbrace{\lambda \pi(1 - \pi) h_L(x_2) + \pi(1 - \pi) h_H(x_2)}^{II} + \lambda(1 - \pi)^2 h_L(x_2) \right\} \\
& + (1 - \pi) \pi \left\{ \underbrace{\pi^2 h_H(x_2)}_d + \overbrace{\pi(1 - \pi) h_L(x_2) + \lambda \pi(1 - \pi) h_H(x_2)}^{III} + \lambda(1 - \pi)^2 h_L(x_2) \right\} \\
& + \lambda \pi^2 \{ \pi^2 h_H(x_2) + \pi(1 - \pi) [h_H(x_2) + h_L(x_2)] + (1 - \pi)^2 h_L(x_2) \} + \kappa.
\end{aligned} \tag{43}$$

Similar to above, some of the gains in the terms underbraced by  $a - f$  turn into losses if some of the conditions  $A - F$  fail. Compared to the case with symmetric goals, losses are less likely. To see this, note first that condition  $D$  is weaker than condition  $A$ . Next, compare term  $I$  in equation (41) with terms  $II$  and  $III$  in equation (43). Because conditions  $E$  and  $F$  cannot both hold,  $2 \times I \geq II + III$  in all cases. All other terms in equations (41) and (43) coincide. Hence,  $x_{max}^B \geq x_{max,incr}^B$ .

**Minimal implementable goal:** Consider deviations to  $x_2 > \hat{x}_2$ , where  $x_2 - \hat{x}_2$  is small in the sense that conditions  $J - L$  from above hold and, in addition,  $h_L(\hat{x}_1) + h_H(x_2) < h_H(\hat{x}_1) + h_L(\hat{x}_2)$  (condition  $M$ ) or  $h_H(\hat{x}_1) + h_L(x_2) < h_L(\hat{x}_1) + h_H(\hat{x}_2)$  (condition  $N$ ). Note that conditions  $M$  and  $N$  cannot both hold. Start from the following expression:

$$\begin{aligned}
& (1 - \pi)^2 \{ \pi^2 h_H(x_2) + \pi(1 - \pi) h_H(x_2) + \pi(1 - \pi) h_L(x_2) + (1 - \pi)^2 h_L(x_2) \} \\
& + (1 - \pi) \pi \left\{ \pi^2 h_H(x_2) + \overbrace{\lambda \pi(1 - \pi) h_H(x_2) + \pi(1 - \pi) h_L(x_2)}^V + \underbrace{\lambda(1 - \pi)^2 h_L(x_2)}_j \right\} \\
& + (1 - \pi) \pi \left\{ \pi^2 h_H(x_2) + \overbrace{\pi(1 - \pi) h_H(x_2) + \lambda \pi(1 - \pi) h_L(x_2)}^{VI} + \underbrace{\lambda(1 - \pi)^2 h_L(x_2)}_j \right\} \\
& + \pi^2 \left\{ \pi^2 h_H(x_2) + \underbrace{\lambda \pi(1 - \pi) h_H(x_2)}_k + \underbrace{\lambda \pi(1 - \pi) h_L(x_2)}_j + \underbrace{\lambda(1 - \pi)^2 h_L(x_2)}_l \right\} + \kappa.
\end{aligned} \tag{44}$$

Similar to above, some of the losses in the terms underbraced by  $j - n$  turn into gains if some of the conditions  $J - N$  fail. Compared to the case with symmetric goals, losses are more likely. To see this, compare term  $IV$  in equation (42) with terms  $V$  and  $VI$  in equation (44). Because conditions  $M$  and  $N$  cannot both hold, and all other terms in equations (42) and (44) coincide,  $x_{min}^B \leq x_{min,incr}^B$ .

### Decreasing goals $\hat{x}_1 > \hat{x}_2$

By analogous arguments as for increasing goals,  $x_{min}^B < x_{min,decr}^B$  and  $x_{max}^B > x_{max,decr}^B$ . To see this note that nowhere in the derivations for increasing goals did we use  $\hat{x}_2 > \hat{x}_1$ . ■

## B.3 Comparison of narrow and broad accounting

### B.3.1 Proof of Lemma 3

Subtracting the term attached to  $h'_H(x_i)$  in equation (39) from the corresponding term in equation (37) yields  $A \equiv -\beta\eta(\lambda - 1)\pi^2(1 - \pi)(1 - 2\pi)$ .  $A < 0$  if and only if  $\pi < 1/2$ . This gives  $\underline{\pi} = 1/2$  for the special case where  $h_L(x_i) = 0$  for all  $x_i$ . The difference between the terms attached to  $h'_L(x_i)$  is  $B \equiv \beta\eta(1 - \pi)[1 + (\lambda - 1)\pi(1 - \pi)^2]$ .  $B > 0$  for  $\pi \in (0, 1)$  and  $B = 0$  for  $\pi \in \{0, 1\}$ . Recall that  $h'_H(x_i) > h'_L(x_i)$  for  $x_i > 0$  and that  $A > 0$  for  $\pi > 1/2$ . Hence, for  $\pi > 1/2$  we have  $A h'_H(x_i) + B h'_L(x_i) > (A + B) h'_L(x_i) > 0$ , which implies  $x_{max}^N > x_{max}^B$ . Thus, by the intermediate value theorem, the cutoff  $\underline{\pi} \in (0, 1/2]$ .

### B.3.2 Proof of Proposition 5

**Overview:** If broad and narrow accounts allow to implement the same decisions, then self 0 strictly prefers broad accounts (proposition 4). For  $\pi \leq \underline{\pi}$  we have  $x_{max}^N \leq x_{max}^B$  (lemma 3). So for  $\pi \leq \underline{\pi}$  a broad account does strictly better than narrow accounts (if no profitable joint deviation exists under broad accounts; see section 5.4.1). For  $\pi = 1$ , broad and narrow accounts do equally well, because  $x_{max}^B = x_{max}^N$  and risk pooling plays no role. The proof consists of showing that, starting from  $\pi = 1$ , a marginal decrease in  $\pi$  leads to a larger drop in utility under broad than under narrow accounts. That is, for  $\pi$  close to 1, narrow accounts yield strictly larger utility than broad accounts. The intermediate value theorem then implies the existence of a cutoff  $\hat{\pi} \in (\underline{\pi}, 1)$ , such that for  $\hat{\pi} < \pi < 1$  narrow accounting is strictly optimal.

**Broad accounts:** Note that for  $x_{max}^B < x_0^B$ , self 0 implements  $x_{max}^B$  (or possibly an even lower decision). Hence, the maximized utility of self 0 under broad accounts cannot exceed  $U_0^B(x_{max}^B, x_{max}^B | \cdot)$ .<sup>33</sup>

---

<sup>33</sup>The utility of self 0 is bounded above by  $U_0^B(x_{max}^B, x_{max}^B | x_{max}^B, x_{max}^B)$  because utility is strictly increasing for decision levels  $x_i < x_0^B$ . For our purpose of deriving conditions when narrow accounts dominate broad accounts, it does not matter whether or not this bound is tight. While  $x_{max}^B$  only rules out unilateral deviations by self 1, any possibilities for profitable joint deviations would only reduce the maximized utility by further restricting the set of decisions that self 0 can implement (see also section 5.4.1).

Our goal is to derive a formula for how this maximized utility of self 0 varies with  $\pi$ , at  $\pi = 1$ . For this we use the following interim results.

$$\begin{aligned} U_0^B(x_{max}^B, x_{max}^B | x_{max}^B, x_{max}^B) &= 2\beta \{h_L(x_{max}^B) + [h_H(x_{max}^B) - h_L(x_{max}^B)] B(\pi) - c(x_{max}^B)\}, \\ \frac{dU_0^B(\cdot)}{d\pi} &= 2\beta [h_H(x_{max}^B) - h_L(x_{max}^B)] B'(\pi) \\ &\quad + 2\beta \{[h'_H(x_{max}^B) - h'_L(x_{max}^B)] B(\pi) + h'_L(x_{max}^B) - c'(x_{max}^B)\} \frac{dx_{max}^B}{d\pi}, \end{aligned}$$

where  $B(\pi) = \pi [1 - \eta(1 - \pi)(1 - \pi(1 - \pi))(\lambda - 1)]$  and  $B'(\pi) = 1 - \eta(1 - 2\pi) [1 - 2\pi(1 - \pi)](\lambda - 1)$ . Next we derive  $\frac{dx_{max}^B}{d\pi}$ . Based on equation (39),  $\Phi^B(x, \pi) = 0$  implicitly defines  $x_{max}^B(\pi)$ , where<sup>34</sup>

$$\begin{aligned} \Phi^B(x, \pi) &\equiv \beta \left\{ \pi h'_H(x) \left( 1 + \eta [1 + (\lambda - 1)\pi(2 - 3\pi + 2\pi^2)] \right) \right. \\ &\quad \left. + (1 - \pi) h'_L(x) \left( 1 + \eta [1 + (\lambda - 1)(1 - \pi(1 - \pi)^2)] \right) \right\} - (1 + \beta\eta) c'(x). \end{aligned}$$

Implicit differentiation gives  $\frac{dx_{max}^B}{d\pi} = -\Phi_\pi^B(x, \pi) / \Phi_x^B(x, \pi)$ , where

$$\begin{aligned} \Phi_\pi^B(x, \pi) &= \beta \left\{ h'_H(x) \left[ 1 + \eta(1 + (\lambda - 1)\pi(4 - 9\pi + 8\pi^2)) \right] \right. \\ &\quad \left. - h'_L(x) \left[ 1 + \eta(1 + (\lambda - 1)\{1 + (1 - \pi)^3 - 3\pi(1 - \pi)^2\}) \right] \right\}, \\ \Phi_x^B(x, \pi) &= \beta \left\{ \pi h''_H(x) \left[ 1 + \eta [1 + (\lambda - 1)\pi(2 - 3\pi + 2\pi^2)] \right] \right. \\ &\quad \left. + (1 - \pi) h''_L(x) \left[ 1 + \eta [1 + (\lambda - 1)(1 - \pi(1 - \pi)^2)] \right] \right\} - (1 + \beta\eta) c''(x). \end{aligned}$$

To show our result, we need to derive  $\left. \frac{dU_0^B(x_{max}^B, x_{max}^B | x_{max}^B, x_{max}^B)}{d\pi} \right|_{\pi=1}$ . Using our above derivations,  $B(1) = 1$ ,  $B'(1) = 1 + \eta(\lambda - 1)$ , and

$$\left. \frac{dx_{max}^B}{d\pi} \right|_{\pi=1} = \beta \frac{h'_H(x_{max}^B) [1 + \eta(1 + 3(\lambda - 1))] - h'_L(x_{max}^B) (1 + \eta\lambda)}{(1 + \beta\eta) c''(x_{max}^B) - \beta(1 + \eta\lambda) h''_H(x_{max}^B)}.$$

Hence, the impact of a marginal change in  $\pi$  on the utility of self 0 at  $x_{max}^B$ , at  $\pi = 1$ , is

$$\begin{aligned} \left. \frac{dU_0^B(\cdot)}{d\pi} \right|_{\pi=1} &= \beta \left\{ 2 \overbrace{[h_H(x_{max}^B) - h_L(x_{max}^B)] [1 + \eta(\lambda - 1)]}^{\equiv \kappa^B(x_{max}^B)} \right. \\ &\quad \left. + 2\beta [h'_H(x_{max}^B) - c'(x_{max}^B)] \overbrace{\frac{h'_H(x_{max}^B) [1 + \eta(1 + 3(\lambda - 1))] - h'_L(x_{max}^B) (1 + \eta\lambda)}{(1 + \beta\eta) c''(x_{max}^B) - \beta(1 + \eta\lambda) h''_H(x_{max}^B)}}^{\psi^B(x_{max}^B)} \right\}. \end{aligned} \quad (45)$$

<sup>34</sup>Note that  $2 - 3\pi + 2\pi^2 = 1 + (1 - \pi)(1 - 2\pi)$ . The former representation is easier to work with here.

**Narrow accounts:** Note that  $U_0^N(x_0^N|x_0^N) \geq U_0^N(x_{max}^N|x_{max}^N)$ . So if  $U_0^N(x_{max}^N|x_{max}^N) > U_0^B(x_{max}^B, x_{max}^B|x_{max}^B, x_{max}^B)$ , then also  $U_0^N(x_0^N|x_0^N) > U_0^B(x_{max}^B, x_{max}^B|x_{max}^B, x_{max}^B)$ . Hence, it is sufficient to show the former inequality. The following interim results allow us to consider below how  $U_0^N(x_{max}^N|x_{max}^N)$  varies with  $\pi$ , at  $\pi = 1$ .

$$\begin{aligned} U_0^N(x_{max}^N|x_{max}^N) &= \beta \{ \pi [h_H(x) - h_L(x)] [1 - \eta(1 - \pi)(\lambda - 1)] + h_L(x) - c(x) \}, \\ \frac{dU_0^N(x_{max}^N|x_{max}^N)}{d\pi} &= \beta [h_H(x) - h_L(x)] [1 - \eta(\lambda - 1)(1 - 2\pi)] \\ &\quad + \beta \left\{ [h'_H(x) - h'_L(x)] \pi [1 - \eta(1 - \pi)(\lambda - 1)] + h'_L(x) - c'(x) \right\} \frac{dx_{max}^N}{d\pi}. \end{aligned}$$

Based on equation (37),  $\Phi^N(x, \pi) = 0$  implicitly defines  $x_{max}^N(\pi)$ , where

$$\Phi^N(x, \pi) \equiv \beta \{ \pi h'_H(x) [1 + \eta(1 + (\lambda - 1)\pi)] + (1 - \pi) h'_L(x) (1 + \eta\lambda) \} - (1 + \beta\eta) c'(x).$$

Implicit differentiation gives  $\frac{dx_{max}^N}{d\pi} = -\Phi_\pi^N(x, \pi)/\Phi_x^N(x, \pi)$ , where

$$\begin{aligned} \Phi_\pi^N(x, \pi) &= \beta \{ h'_H(x) [1 + \eta(1 + 2(\lambda - 1)\pi)] - h'_L(x) (1 + \eta\lambda) \}, \\ \Phi_x^N(x, \pi) &= \beta \{ \pi h''_H(x) [1 + \eta(1 + (\lambda - 1)\pi)] + (1 - \pi) h''_L(x) (1 + \eta\lambda) \} - (1 + \beta\eta) c''(x). \end{aligned}$$

To derive  $\frac{dU_0^N(x_{max}^N|x_{max}^N)}{d\pi} \Big|_{\pi=1}$ , use our above results to obtain

$$\frac{dx_{max}^N}{d\pi} \Big|_{\pi=1} = \beta \frac{h'_H(x_{max}^N) [1 + \eta(1 + 2(\lambda - 1))] - h'_L(x_{max}^N) (1 + \eta\lambda)}{(1 + \beta\eta) c''(x_{max}^N) - \beta(1 + \eta\lambda) h''_H(x_{max}^N)}.$$

Hence, the impact of a marginal change in  $\pi$  on the utility of self 0 at  $x_{max}^N$ , at  $\pi = 1$ , is

$$\begin{aligned} \frac{dU_0^N(\cdot)}{d\pi} \Big|_{\pi=1} &= \beta \left\{ \overbrace{[h_H(x_{max}^N) - h_L(x_{max}^N)] [1 + \eta(\lambda - 1)]}^{\equiv \kappa^N(x_{max}^N)} \right. \\ &\quad \left. + \beta \overbrace{[h'_H(x_{max}^N) - c'(x_{max}^N)] \frac{h'_H(x_{max}^N) [1 + \eta(1 + 2(\lambda - 1))] - h'_L(x_{max}^N) (1 + \eta\lambda)}{(1 + \beta\eta) c''(x_{max}^N) - \beta(1 + \eta\lambda) h''_H(x_{max}^N)}}^{\psi^N(x_{max}^N)} \right\}. \end{aligned} \quad (46)$$

**Comparison:** For  $\pi = 1$ ,  $x_{max}^N = x_{max}^B \equiv x_{max}$  and  $2U_0^N(x_{max}^N|\cdot) = U_0^B(x_{max}^B, x_{max}^B|\cdot)$ . Introducing a little bit of uncertainty, narrow accounts thus do strictly better than broad accounts if utility drops by less under narrow accounts, i.e., if

$$2 \frac{dU_0^N(x_{max}^N|\cdot)}{d\pi} \Big|_{\pi=1} < \frac{dU_0^B(x_{max}^B, x_{max}^B|\cdot)}{d\pi} \Big|_{\pi=1}. \quad (47)$$

Comparing equations (45) and (46), note that  $\kappa^N(x_{max}) = \kappa^B(x_{max})$  and that all terms in  $\psi^B(x_{max})$  and  $\psi^N(x_{max})$  coincide, except for  $\eta(1 + 3(\lambda - 1))$  in  $\psi^B(x_{max})$  and  $\eta(1 + 2(\lambda - 1))$  in  $\psi^N(x_{max})$ . Note further that at  $\pi = 1$ ,  $x_{max}^N = x_{max}^B \equiv x_{max} < x_0^B = x_0^N$ , and therefore  $h'_H(x_{max}) - c'(x_{max}) > 0$ . Finally,  $\eta(1 + 3(\lambda - 1)) > \eta(1 + 2(\lambda - 1))$  implies that inequality (47) holds.

### B.3.3 Proof of Corollary 2

Part 2 follows from part 1 and proposition 5. For part 1, we need to show that there exists no profitable joint deviation if the individual sets goals  $\hat{x} = x_0^B$  and  $x_{min}^B \leq x_0^B \leq x_{max}^B$ . By the same arguments as in the proof of proposition 3, we can restrict attention to deviations  $(x'_1, x'_2)$  of the kind  $x'_2 < x_0^B < x'_1$  and write the utility change from the deviation as a line integral  $\int_{x_0^B}^{x'_1} \phi'(x_1) dx_1$ . Hence, the aim is to show that  $\phi'(x_1) < 0$  for every point  $(x_1, z(x_1))$  on the path along which we are integrating. We abbreviate  $h_1 \equiv h(x_1)$ ,  $h_2 \equiv h(z(x_1))$ , etc., and use short-hands for the slopes of the path as well as the iso-benefit and iso-cost curves, respectively:  $-\frac{dz(x_1)}{dx_1} \equiv r_z$ ,  $\frac{h'_1}{h'_2} \equiv r_h$ , and  $\frac{c'_1}{c'_2} \equiv r_c$ . Taking the derivative of the utility of self 1 yields

$$\phi'(x_1) = \beta \pi [(1 + \eta \kappa_1) h'_1 - (1 + \eta \kappa_2) h'_2 r_z] - (1 + \beta \eta \kappa_3) [c'_1 - c'_2 r_z], \quad (48)$$

where the parameters  $\kappa_j \in \{1, 2, 3\}$  depend on the kind of deviation (see below).

a)  $z(x'_1) < x^{cm}(x'_1)$ .

We exploit path-independence of the line integral. There exists a path from  $(x_0^B, x_0^B)$  to  $(x'_1, x'_2)$  for which at every point  $0 < r_h < r_c < r_z$ . The parameters for the benefit dimension are  $\kappa_1 = 1 + (\lambda - 1) \pi^2$  and  $\kappa_2 = 1 + (\lambda - 1) \pi [1 + (1 - \pi)(1 - 2\pi)]$ . And  $\kappa_3 = 1$  because self 1 feels a gain in the cost dimension. Hence,

$$\begin{aligned} \phi'(x_1) &= \beta \pi h'_2 [(1 + \eta \kappa_1) r_h - (1 + \eta \kappa_2) r_z] + (1 + \beta \eta) c'_2 [r_z - r_c] \\ &< \beta \pi h'_2 [(1 + \eta \kappa_1) r_h - (1 + \eta \kappa_2) r_z] + \beta \pi h'_2 (1 + \eta \kappa_2) [r_z - r_c] \\ &\propto (1 + \eta \kappa_1) r_h - (1 + \eta \kappa_2) r_c < (1 + \eta \kappa_1) r_h - (1 + \eta \kappa_2) r_h \\ &\propto (\kappa_1 - \kappa_2) = -2 \pi (1 - \pi)^2 (\lambda - 1) < 0. \end{aligned}$$

The second line exploits that  $z(x_1) < x_0^B \leq x_{max}^B$  implies  $\beta \pi [1 + \eta \kappa_2] h'_2 > (1 + \beta \eta) c'_2$  (cf. lemma 4, setting  $h'_L(x) = 0$ ) and that  $r_z > r_c$ . The third line uses  $r_c > r_h$ .

b)  $x^{cm}(x'_1) \leq z(x'_1) < x^{hm}(x'_1)$ .

Along the path  $0 < r_h < r_z \leq r_c$ , self 1 feels a loss in the cost-dimension, i.e.  $\kappa_3 = \lambda$ . Hence,

$$\begin{aligned} \phi'(x_1) &= \beta \pi h'_2 [(1 + \eta \kappa_1) r_h - (1 + \eta \kappa_2) r_z] + (1 + \beta \eta \lambda) c'_2 [r_z - r_c] \\ &\leq \beta \pi h'_2 [(1 + \eta \kappa_1) r_h - (1 + \eta \kappa_2) r_z] < 0, \end{aligned}$$

because  $\kappa_2 > \kappa_1$ , where  $\kappa_1$  and  $\kappa_2$  are the same as in case a).

c)  $x^{hm}(x'_1) \leq z(x'_1) < x_0^B$ .

Along path  $0 < r_z \leq r_h < r_c$ , we have  $\kappa_1 = 1 + (\lambda - 1) \pi^2 (1 - \pi)$ ,  $\kappa_2 = 1 + (\lambda - 1) \pi (1 - \pi) (2 - \pi)$ , and  $\kappa_3 = \lambda$ . Hence,

$$\begin{aligned} \phi'(x_1) &= \beta \pi h'_1 \left[ (1 + \eta \kappa_1) - (1 + \eta \kappa_2) \frac{r_z}{r_h} \right] - (1 + \beta \eta \lambda) c'_1 \left[ 1 - \frac{r_z}{r_c} \right] \\ &< \beta \pi (1 + \eta \kappa_1) h'_1 \left[ 1 - \frac{r_z}{r_h} \right] - (1 + \beta \eta \lambda) c'_1 \left[ 1 - \frac{r_z}{r_c} \right] \\ &< [\beta \pi (1 + \eta \kappa_1) h'_1 - (1 + \beta \eta \lambda) c'_1] \left[ 1 - \frac{r_z}{r_c} \right] < 0. \end{aligned}$$



The second line exploits  $\kappa_2 > \kappa_1$ ; the third uses  $1 - r_z/r_c > 1 - r_z/r_h \geq 0$  and that  $x_1 > x_0^B \geq x_{min}^B$  implies  $\beta \pi [1 + \eta \kappa_1] h'_1 < (1 + \beta \eta \lambda) c'_1$  (cf. lemma 4, setting  $h'_L(x) = 0$ ).

### B.3.4 Proof of Corollary 3

**Part 1.** Subtract the term attached to  $h'_H(x_i)$  in equation (40) from the corresponding term in equation (38), yielding  $A \equiv -\beta \eta (\lambda - 1) \pi^3 (1 - \pi)$ . The difference between the terms attached to  $h'_L(x_i)$  is  $B \equiv -\beta \eta (\lambda - 1) \pi (1 - \pi)^2 (1 - 2\pi)$ . Recall that  $h'_H(x_i) > h'_L(x_i)$  for  $x_i > 0$ . Hence,  $A h'_H(x_i) + B h'_L(x_i) < (A + B) h'_L(x_i) < 0$ , which implies  $x_{min}^N < x_{min}^B$ .

**Part 2.** Rewrite equations (34) and (36) in the form  $c'(x_0^A) = \rho_L^A h'_L(x_0^A) + \rho_H^A h'_H(x_0^A)$  and equations (38) and (40) in the form  $c'(x_{min}^A) = \kappa_L^A h'_L(x_{min}^A) + \kappa_H^A h'_H(x_{min}^A)$ , where  $A \in \{B, N\}$  and

$$\begin{aligned} \rho_L^N &= (1 - \pi) [1 + \eta (\lambda - 1) \pi], & \rho_H^N &= \pi [1 - \eta (\lambda - 1) (1 - \pi)], \\ \kappa_L^N &= \frac{\beta(1-\pi)[1+\eta(1+(\lambda-1)\pi)]}{1+\beta\eta\lambda}, & \kappa_H^N &= \frac{\beta\pi(1+\eta)}{1+\beta\eta\lambda}, \\ \rho_L^B &= (1 - \pi) [1 + \eta (\lambda - 1) \{1 - \pi (1 - \pi)\}], & \rho_H^B &= \pi [1 - \eta (\lambda - 1) (1 - \pi) \{1 - \pi (1 - \pi)\}], \\ \kappa_L^B &= \frac{\beta(1-\pi)(1+\eta[1+(\lambda-1)\pi(1+(1-\pi)(1-2\pi))])}{1+\beta\eta\lambda}, & \kappa_H^B &= \frac{\beta\pi(1+\eta[1+(\lambda-1)\pi^2(1-\pi)])}{1+\beta\eta\lambda}. \end{aligned}$$

All  $\kappa_j^A$ -terms,  $j \in \{L, H\}$ ,  $A \in \{B, N\}$ , are increasing in  $\beta$ . So it is enough to show that  $\rho_j^A > \kappa_j^A$  at  $\beta = 1$ . For narrow accounts,  $\rho_L^N - \kappa_L^N = \eta(1 - \pi)(\lambda - 1)(1 + \eta \lambda \pi)/(1 + \eta \lambda) > 0$ . Assumption 2' directly gives  $\rho_H^N > \kappa_H^N$ . For broad accounts,

$$\begin{aligned} \rho_L^B - \kappa_L^B &= \frac{\eta(1 - \pi)(\lambda - 1)}{1 + \eta \lambda} [1 + (1 - \pi)(1 - 2\pi) + 2\pi^2(1 - \pi) + \eta \lambda (1 - \pi(1 - \pi))] > 0, \\ \rho_H^B - \kappa_H^B &= \frac{\eta(1 - \pi)(\lambda - 1)\pi}{1 + \eta \lambda} [\pi(1 + (1 - \pi)(1 - 2\pi)) - \eta \lambda (1 - \pi)(1 - \pi(1 - \pi))]. \end{aligned}$$

Note that

$$\rho_H^B - \kappa_H^B > 0 \quad \Leftrightarrow \quad \frac{\pi(1 + (1 - \pi)(1 - 2\pi))}{(1 - \pi)(1 - \pi(1 - \pi))} > \eta \lambda.$$

Assumption 2' implies that  $\frac{\pi}{1 - \pi} \geq \eta \lambda$ . To complete the proof, observe that for  $\pi \in (0, 1)$

$$\frac{\pi(1 + (1 - \pi)(1 - 2\pi))}{(1 - \pi)(1 - \pi(1 - \pi))} > \frac{\pi}{1 - \pi}.$$

**Part 3.** Follows from parts 1 and 2 together with proposition 4, because  $x_0^B < x_{max}^B$  at  $\beta = 1$ .

## B.4 The risk-incentive effect with $N > 2$ tasks

To extend the model to  $N > 2$  tasks, assume for simplicity that tasks are symmetric and that  $h_L(x) = 0$  for all  $x$ . Consider broad accounts with symmetric goals  $\hat{x}_1 = \dots = \hat{x}_n = \hat{x}$  for tasks  $n = 1, \dots, N$ . Suppose the individual sticks to his goal in all tasks, except for task 1. What impact would a deviation to  $x_1 < \hat{x}$  in task 1 have? Again, broad accounts have both a positive and negative effect on incentives, compared to narrow accounts.

**The positive incentive effect.** The individual expects that he will sometimes fail in task 1 but still succeeds. The outcome in task 1 then can help reduce a loss stemming from an expectation

to succeed in other tasks. Such “cross-subsidization” occurs in all events where (i) the individual expects to fail in task 1, (ii) actually succeeds in task 1, and (iii) the combined realized benefit from the other  $n - 1$  tasks,  $k \cdot h(\hat{x})$  (where  $k$  is the number of successes in the  $n - 1$  remaining tasks) falls short of the reference point for the other  $n - 1$  tasks,  $m \cdot h(\hat{x})$  (where  $m$  is the number of expected successes in the  $n - 1$  remaining tasks). That is, the probability that cross-subsidization occurs is  $\pi^2$  times the probability that  $m > k$ , given by

$$y \equiv \Pr(m > k) = \sum_{m=1}^{N-1} \underbrace{\left[ \binom{N-1}{m} \pi^m (1-\pi)^{N-1-m} \right]}_{\text{probability of event } m} \times \sum_{k=0}^{m-1} \underbrace{\left[ \binom{N-1}{k} \pi^k (1-\pi)^{N-1-k} \right]}_{\text{probability of } k < m \text{ for fixed } m}.$$

**The negative incentive effect.** Successes in tasks  $n \neq 1$  may already allow to avoid a loss stemming for the expectation to succeed in task 1. This occurs in all events where (i) the individual expects to succeed in task 1, (ii) actually fails in task 1, and (iii) the  $k$  successes in the  $n-1$  remaining tasks allow to meet the reference point  $(m + 1) h(\hat{x})$ . That is, the probability that such an event occurs is  $\pi(1 - \pi)$  times the probability that  $m < k$ , or  $\pi(1 - \pi)(1 - z - y)$ , where

$$z \equiv \Pr(m = k) = \sum_{m=0}^{N-1} \left[ \binom{N-1}{m} \pi^m (1-\pi)^{N-1-m} \right]^2.$$

Putting the two effects together, we obtain an expression that corresponds to  $\Pr(\text{loss}^B)$  in equation (14). Given an expectation to succeed in task 1 (which occurs with probability  $\pi$ ), a success in task 1 is only needed with probability  $1 - (1 - y - z)$  to prevent a loss. Given an expectation to fail in task 1 (which occurs with probability  $1 - \pi$ ), a success in task 1 can prevent a loss in other tasks with probability  $y$ . That is, the probability of a loss caused by unilateral deviation  $x_1 < \hat{x}$  is

$$\Pr(\text{loss}^B) = \pi^2 [1 - (1 - y - z)] + \pi(1 - \pi)y = \pi(y + \pi z).$$

Note that this formula nests the special case  $N = 2$ , where  $\Pr(\text{loss}^B) = \pi^2 [1 + (1 - \pi)(1 - 2\pi)]$  (cf. equation 18). For narrow accounts nothing changes relative to our analysis in section 5.3,  $\Pr(\text{loss}^N) = \pi^2$ . One can check numerically that broad accounts are less motivating than narrow accounts for  $\pi > \frac{1}{2}$ , because then  $\Pr(\text{loss}^N) > \Pr(\text{loss}^B)$ . Further, such a numerical exercise shows that  $|x_{max}^B - x_{min}^B|$  grows with the number of tasks  $N$  that are jointly evaluated. In this sense, increasing the breadth of an account translates into stronger (weaker) incentives for  $\pi < \frac{1}{2}$  ( $\pi > \frac{1}{2}$ ).

## C Proof of Proposition 6

We provide the proof only for the pre-task information scenario. The main driving forces are essentially the same in the interim-information scenario (derivations are available from the authors). Part 1: see the main text. To show part 2, we first derive the bounds to determine the maximal implementable goals under narrow accounting.<sup>35</sup> Next, we ask whether narrow-accounting implementable goals are also broad-accounting implementable, or if decision substitution prevents this.

<sup>35</sup>Our focus is on the maximal implementable goals because we are interested whether goals can help to overcome the individual’s self-control problem. The arguments for the minimal implementable goals are analogous.

## Narrow-accounting implementable goals

**Self t.I**, who observes the productive state and revises the inherited goal  $\hat{x}_t$  to  $x_t < \hat{x}_t$ , has utility

$$\begin{aligned} & \beta h(\hat{x}_t) - c(x_t) + \eta \gamma_{\tau,t} [\pi \lambda (h(x_t) - h(\hat{x}_t)) + (1 - \pi) (h(x_t) - 0)] \\ & - \eta \gamma_{\tau,t} [\pi (c(x_t) - c(\hat{x}_t)) + (1 - \pi) \lambda (c(x_t) - 0)] \end{aligned}$$

The first order condition implicitly defines  $\tilde{x}_{max,t}^N$ :

$$(\beta + \eta \gamma_{\tau,t} [1 + (\lambda - 1) \pi]) h'(\tilde{x}_{max,t}^N) = (1 + \eta \gamma_{\tau,t} [\lambda - (\lambda - 1) \pi]) c'(\tilde{x}_{max,t}^N). \quad (49)$$

If self 1.I maintains the goal for task 1 but revises the goal for the productive state of task 2 to  $x_2 < \hat{x}_2$ , he experiences anticipatory utility from comparing the new reference distributions with the inherited reference distribution. That is, for the benefits he compares  $h(\hat{x}_1)$  with  $(\pi \circ h(\hat{x}_1); (1 - \pi) \circ 0)$  and  $(\pi \circ h(x_2); (1 - \pi) \circ 0)$  with  $(\pi \circ h(\hat{x}_2); (1 - \pi) \circ 0)$ . Similarly for the costs. Moreover, he anticipates that the goal revision will affect the anticipatory utility that self 2.I feels after receiving pre-task information. The expected anticipatory gain-loss utility is  $-\eta \gamma_{\tau,2} \pi (1 - \pi) (\lambda - 1) (h(x_2) + c(x_2))$ . It results from comparisons in each state with the then inherited reference distributions  $(\pi \circ h(x_2); (1 - \pi) \circ 0)$  and  $(\pi \circ c(x_2); (1 - \pi) \circ 0)$ . The utility of self 1.I hence is given by

$$\begin{aligned} & \beta h(\hat{x}_1) - c(\hat{x}_1) + \beta \pi h(x_2) - \beta \pi c(x_2) + \eta \gamma_{\tau,1} (1 - \pi) [h(\hat{x}_1) - \lambda c(\hat{x}_1)] \\ & + \eta \gamma_{\tau,1} [\pi \lambda (h(x_2) - h(\hat{x}_2)) - \pi (c(x_2) - c(\hat{x}_2))] - \eta \beta \gamma_{\tau,2} \pi (1 - \pi) (\lambda - 1) (h(x_2) + c(x_2)) \end{aligned}$$

The first order condition implicitly defines  $\check{x}_{max,2}$ :

$$\begin{aligned} & (\beta + \eta \gamma_{\tau,1} \lambda - \eta \beta \gamma_{\tau,2} (1 - \pi) (\lambda - 1)) h'(\check{x}_{max,2}^N) \\ & = (\beta + \eta \gamma_{\tau,1} + \eta \beta \gamma_{\tau,2} (1 - \pi) (\lambda - 1)) c'(\check{x}_{max,2}^N). \end{aligned} \quad (50)$$

The same bound applies if self 1.I is in the unproductive state. Comparing equations (49) and (50), note that  $\gamma_{\tau,1} \geq \beta \gamma_{\tau,2}$  implies  $\check{x}_{max,2}^N \geq \tilde{x}_{max,2}^N$ . That is, the additional bound is not binding if anticipatory utility weighs at least as strongly as discounted future anticipatory utility.

### Are narrow-accounting implementable goals broad-accounting implementable?

Taking the derivative of the utility of self 1 in equation (21) with respect to  $x_1$ , dividing by  $c'_1$ , and using the fact that cost matching implies  $\frac{dx_2}{dx_1} = -\frac{c'_1}{c'_2}$ , we obtain

$$\begin{aligned}
& \beta \frac{h'_1}{c'_1} - 1 + \beta \pi \left( \frac{h'_2}{c'_2} - 1 \right) \frac{c'_2}{c'_1} \frac{dx_2}{dx_1} - \beta \eta \gamma_{\tau,2} \pi (1 - \pi) (\lambda - 1) \left( \frac{h'_2}{c'_2} + 1 \right) \frac{c'_2}{c'_1} \frac{dx_2}{dx_1} \\
& + \eta \gamma_{\tau,1} \left[ [1 + (\lambda - 1) \pi] \left( \frac{h'_1}{c'_1} + \pi \frac{h'_2}{c'_2} \frac{c'_2}{c'_1} \frac{dx_2}{dx_1} \right) - (1 - \pi) [1 + (\lambda - 1) (1 - \pi)] \right] \\
\propto & \beta \frac{h'_1}{c'_1} - 1 - \beta \pi \left( \frac{h'_2}{c'_2} - 1 \right) + \beta \eta \gamma_{\tau,2} \pi (1 - \pi) (\lambda - 1) \left( \frac{h'_2}{c'_2} + 1 \right) \\
& + \eta \gamma_{\tau,1} \left[ [1 + (\lambda - 1) \pi] \left( \frac{h'_1}{c'_1} - \pi \frac{h'_2}{c'_2} \right) - (1 - \pi) [1 + (\lambda - 1) (1 - \pi)] \right] \tag{51} \\
\leq & \beta (1 - \pi) \frac{h'_2}{c'_2} - (1 - \beta \pi) + \eta \gamma_{\tau,1} (1 - \pi) \left\{ [1 + (\lambda - 1) \pi] \frac{h'_2}{c'_2} - [1 + (\lambda - 1) (1 - \pi)] \right. \\
& \left. + \beta \pi (\lambda - 1) \left( \frac{h'_2}{c'_2} + 1 \right) \right\} \equiv \Psi(\pi).
\end{aligned}$$

The last step uses  $\gamma_{\tau,2} \geq \gamma_{\tau,1}$ , and that  $\frac{h'_2}{c'_2} \geq \frac{h'_1}{c'_1} > 1$  at  $(x_1 = x_{0,1}^*, x_2 = x_{0,2}^*)$  because  $x_{0,1}^* \geq x_{0,2}^*$ . Note that  $\Psi(1) = -(1 - \beta) < 0$ . Further, for  $\pi = 0$ , the expression in equation (51) becomes  $\frac{h'_1}{c'_1} (\beta + \eta \gamma_{\tau,1}) - (1 + \eta \lambda \gamma_{\tau,1}) > 0$ , where the inequality follows from applying the definition of  $\tilde{x}_{max,t}^N$  in equation (49) and the fact that  $x_{0,t}^* < \tilde{x}_{max,t}^N$ . Hence, by the intermediate value theorem, there exists a  $\hat{\pi} \in (0, 1)$ , such that for all  $\pi \geq \hat{\pi}$  decision substitution pays off. In this case, the preferred decisions of self 0 are not broad-accounting implementable and narrow accounting is optimal.

## References

- ABELER, J., A. FALK, L. GÖTTE, AND D. HUFFMAN (2011): “Reference Points and Effort Provision,” *American Economic Review*, 101(2), 470–492.
- ABELER, J., AND F. MARKLEIN (2010): “Fungibility, Labels, and Consumption,” CeDEx Discussion Paper 2010-13.
- AINSLIE, G. (2001): *Breakdown of Will*. Cambridge University Press, Cambridge.
- BÉNABOU, R., AND J. TIROLE (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.
- BERTRAND, M., S. MULLAINATHAN, AND E. SHAFIR (2006): “Behavioral Economics and Marketing in Aid of Decision Making Among the Poor,” *Journal of Public Policy & Marketing*, 25(1), 8–23.
- BISIN, A., AND K. HYNDMAN (2009): “Procrastination, Self-imposed Deadlines and Other Commitment Devices,” MPRA Working Paper Nr. 16235.
- 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.
- BRYAN, G., D. S. KARLAN, AND S. NELSON (2010): “Commitment Devices,” *Annual Review of Economics*, 2, 671–698.
- CAMERER, C. (2003): “Prospect Theory in the Wild: Evidence from the Field,” in *Advances in Behavioral Economics*, ed. by C. Camerer, G. Loewenstein, and M. Rabin, pp. 148–161. Princeton University Press.
- CAMERER, C., L. BABCOCK, G. LOEWENSTEIN, AND R. H. THALER (1997): “Labor Supply of New York City Cabdrivers: One Day at a Time,” *Quarterly Journal of Economics*, 112(2), 407–441.
- CARRILLO, J. D., AND M. DEWATRIPONT (2008): “Promises, Promises, ...,” *Economic Journal*, 118, 1453–1473.
- CRAWFORD, V., AND J. MENG (2011): “New York City Cabdrivers’ Labor Supply Revisited: Reference-Dependence Preferences with Rational-Expectations Targets for Hours and Income,” *American Economic Review*, forthcoming.

- DELLAVIGNA, S. (2009): “Psychology and Economics: Evidence from the Field,” *Journal of Economic Literature*, 47(2), 315–372.
- DORAN, G. T. (1981): “There’s a S.M.A.R.T. Way to Write Management’s Goals and Objectives,” *Management Review*, 70(11), 35–36.
- ELSTER, J. (2000): *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge University Press, Cambridge.
- FARBER, H. (2005): “Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers,” *Journal of Political Economy*, 113(1), 46–82.
- (2008): “Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers,” *American Economic Review*, 98(3), 1069–1082.
- FEHR, E., AND L. GÖTTE (2007): “Do Workers Work More if Wages are High? Evidence from a Randomized Field Experiment,” *American Economic Review*, 97(1), 298–317.
- FISHBACH, A., R. DHAR, AND Y. ZHANG (2006): “Subgoals as Substitutes or Complements: The Role of Goal Accessibility,” *Journal of Personality and Social Psychology*, 91(2), 232–242.
- FREDERICK, S., AND G. LOEWENSTEIN (2003): “Hedonic Adaptation,” in *Well-Being: The Foundations of Hedonic Psychology*, ed. by D. Kahneman, E. Diener, and N. Schwarz, chap. 16, pp. 302–329. Russell Sage Foundation, New York.
- GILL, D., AND V. PROWSE (2011): “A Structural Analysis of Disappointment Aversion in a Real Effort Competition,” *American Economic Review*, forthcoming.
- GNEEZY, U., A. KAPTEYN, AND J. POTTERS (2003): “Evaluation Periods and Asset Prices in a Market Experiment,” *Journal of Finance*, 58(2), 821–837.
- GNEEZY, U., AND J. POTTERS (1997): “An Experiment on Risk Taking and Evaluation Periods,” *Quarterly Journal of Economics*, 112(2), 631–845.
- GOLLWITZER, P. M., AND P. SHEERAN (2006): “Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes,” in *Advances in Experimental Social Psychology*, ed. by M. Zanna, vol. 38, chap. 2, pp. 69–119. Academic Press, New York.
- GÖTTE, L., D. HUFFMAN, AND E. FEHR (2004): “Loss Aversion and Labor Supply,” *Journal of the European Economic Association*, 2(2-3), 216–228.
- HEATH, C., R. P. LARRICK, AND G. WU (1999): “Goals as Reference Points,” *Cognitive Psychology*, 38, 79–109.
- HEATH, C., AND J. B. SOLL (1996): “Mental Budgeting and Consumer Decisions,” *Journal of Consumer Research: An Interdisciplinary Quarterly*, 23(1), 40–52.

- HERWEG, F., AND D. MÜLLER (2011): “Performance of Procrastinators: On the Value of Deadlines,” *Theory and Decision*, 70, 329–366.
- HERWEG, F., D. MÜLLER, AND P. WEINSCHENK (2010): “Binary Payment Schemes: Moral Hazard and Loss Aversion,” *American Economic Review*, 100(5), 2451–2477.
- HOLLENBECK, J. R., C. R. WILLIAMS, AND H. J. KLEIN (1989): “An Empirical Examination of the Antecedents of Commitment to Difficult Goals,” *Journal of Applied Psychology*, 74(1), 18–23.
- HSIAW, A. (2010a): “Goal Bracketing and Self-control,” mimeo, Harvard University.
- (2010b): “Goal-Setting, Social Comparison, and Self-control,” mimeo, Harvard University.
- KAHNEMAN, D., AND A. TVERSKY (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47(2), 263–91.
- (1984): “Choices, Values, and Frames,” *American Psychologist*, 39(4), 341–350.
- KHAN, U., AND R. DHAR (2007): “Where There is a Way, is There a Will? The Effect of Future Choices on Self-Control,” *Journal of Experimental Psychology*, 136(2), 277–288.
- KÓSZEGI, B. (2010): “Utility from Anticipation and Personal Equilibrium,” *Economic Theory*, 44(3), 415–444.
- KÓSZEGI, B., AND M. RABIN (2006): “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics*, 121(4), 1133–1165.
- (2007): “Reference-Dependent Risk Attitudes,” *American Economic Review*, 97(4), 1047–1073.
- (2009): “Reference-Dependent Consumption Plans,” *American Economic Review*, 99(3), 909–936.
- KOCH, A. K., AND J. NAFZIGER (2009a): “Motivational Goal Bracketing,” IZA Discussion Paper No 4471.
- (2009b): “Self-rewards and Personal Motivation,” Working Paper, Aarhus University.
- (2011): “Self-regulation through Goal Setting,” *Scandinavian Journal of Economics*, 113(1), 212–227.
- LAIBSON, D. (1997): “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics*, 112(2), 443–477.
- LATHAM, G. P., AND G. A. YUKL (1975): “A Review of Research on the Application of Goal Setting in Organizations,” *The Academy of Management Journal*, 18(4), pp. 824–845.

- LOCKE, E. A., AND G. P. LATHAM (1990): *A Theory of Goal Setting & Task Performance*. Prentice Hall, Englewood Cliffs, NJ.
- LOEWENSTEIN, G. (1999): "Because It Is There: The Challenge of Mountaineering for Utility Theory," *Kyklos*, 52(3), 315–343.
- LOEWENSTEIN, G., AND D. PRELEC (1992): "Anomalies in Intertemporal Choice: Evidence and an Interpretation," *Quarterly Journal of Economics*, 107(2), 573–597.
- MATTHEY, A., AND N. DWENGER (2008): "Don't Aim too High: The Potential Costs of High Aspirations," mimeo, Max-Planck-Institute Jena.
- ODEAN, T. (1998): "Are Investors Reluctant to Realize Their Losses?," *Journal of Finance*, 53(5), 1775–1798.
- O'DONOGHUE, T., AND M. RABIN (1999): "Doing It Now or Later," *American Economic Review*, 89(1), 103–124.
- 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.
- PRENDERGAST, C. (2007): "The Motivation and Bias of Bureaucrats," *American Economic Review*, 97(1), 180–196.
- RABIN, M., AND G. WEIZSÄCKER (2009): "Narrow Bracketing and Dominated Choices," *American Economic Review*, 99(4), 1508–1543.
- RACHLIN, H. (1995): "Self-control: Beyond Commitment," *Behavioral and Brain Sciences*, 18(1), 109–159.
- READ, D., G. LOEWENSTEIN, AND M. RABIN (1999): "Choice Bracketing," *Journal of Risk and Uncertainty*, 19(1-3), 171–197.
- SHEFRIN, H. M., AND R. H. THALER (1988): "The Behavioral Life-Cycle Hypothesis," *Economic Inquiry*, 26(4), 609–643.
- SHIELDS, M. A., AND M. WARD (2001): "Improving Nurse Retention in the National Health Service in England: The Impact of Job Satisfaction on Intentions to Quit," *Journal of Health Economics*, 20(5), 677–701.
- STROTZ, R. H. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23(3), 165–180.
- SUVOROV, A., AND J. VAN DE VEN (2008): "Goal Setting as a Self-Regulation Mechanism," Working Papers w0122, Center for Economic and Financial Research (CEFIR).
- THALER, R. (1999): "Mental Accounting Matters," *Journal of Behavioral Decision Making*, 12, 183–206.



- THALER, R. H. (1980): "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior & Organization*, 1(1), 39–60.
- (1985): "Mental Accounting and Consumer Choice," *Marketing Science*, 4(3), 199–214.
- THALER, R. H. (1990): "Saving, Fungibility, and Mental Accounts," *Journal of Economic Perspectives*, 4(1), 193–205.
- THALER, R. H., A. TVERSKY, D. KAHNEMAN, AND A. SCHWARTZ (1997): "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test," *Quarterly Journal of Economics*, 112(2), 647–661.
- TVERSKY, A., AND D. KAHNEMAN (1981): "The Framing of Decisions and the Psychology of Choice," *Science, New Series*, 211(4481), 453–458.
- (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106(4), 1039–1061.
- WERTENBROCH, K. (1998): "Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice," *Marketing Science*, 17(4), 317–337.