Planning to Self-Control

Claudio Kretz*

E-mail: claudio.kretz@kit.edu

August 29, 2024

In economics, plans normally reflect how decision makers desire or predict to behave in the future. We suggest that plans play a more active role as (intra-psychic) devices to exert self-control over own future choices when tastes are changing. We provide axiomatic foundations for decision makers who act as if planning to self-control in a simple two-period setting. We show that our model is behaviorally equivalent to a general self-control cost representation that allows for magnitude effects: self-control increases when upping the stakes of a decision. The simplest possible specification – a fixed cost of self-control – produces such behavior. When applied to a simple consumption-savings problem, present-biased decision makers use self-control (to overcome present bias) only if they are sufficiently wealthy, producing over-consumption by the poor – and hence a poverty trap.

Keywords: *plans – self-control – dynamic inconsistency – magnitude effect – poverty trap*

1. Introduction

We make plans on a daily basis. Indeed, planning is predominant in thoughts about the future.[1] Recent research in psychology and human decision processes suggests that planning acts as a form of 'pragmatic prospection'. Individuals engage in it "so as to guide actions to bring about desirable outcomes" (Baumeister et al., 2016, p. 3). That is, plans

*I am indebted to Clemens Puppe for his continued support and his many suggestions. I thank Hans Gersbach and Klaus Nehring for fruitful discussions on the subject of this paper. I thank Nick Netzer, Yusufcan Masatlioglu and participants at the 2022 BRIC conference in Prague and the Brown Bag seminar at KIT for helpful comments. Mert Ergin provided excellent research assistance for this project.

[1] Baumeister et al. (2020) report that close to three quarters of thoughts about the future involve plans and that thinking about the future is common in individuals.

play an active part in steering future own choices. For example, implementation intentions in the form of simple plans have been found to help overcome self-regulatory problems and lead to better self-control (Gollwitzer, 1999; Gollwitzer and Sheeran, 2006; Gollwitzer et al., 2010; Gollwitzer and Oettingen, 2011).

To illustrate, consider an individual contemplating potential after-work activities at lunch-time. Suppose that she may either watch a movie or go for a run. At lunch-time, when she is still energized, she prefers to exercise. At the same time, she anticipates her preference to change come the evening. This creates an obvious incentive to commit to the after-work run (for example, by arranging to meet with a friend). If external commitment is out of reach (her friend might be busy), however, she is left to her own devices. By making a *plan*, she may induce herself to go for the run. That is, planning may serve as an *internal* commitment device allowing to exert *self-control*.

Yet planning has received little explicit treatment in economics. Generally speaking, in intertemporal decision problems, plans capture what decision makers desire or predict about own future choices (e.g., how much to save and consume in the future) but do not actually influence these choices. For example, a present-biased saver may plan to start saving more tomorrow. However, this plan does not actually affect her savings decision tomorrow. Thus, if her future self is present-bias again, she will keep under-saving.[2] In contrast, in this paper, we present a model in which plans serve as internal commitment devices that enable the decision maker to exert self-control in subsequent choices.[3]

We characterize *Planning To Self-Control* in terms of three simple and intuitive Axioms on choice between and from decision problems. Our model generalizes temptation-based models of self-control pioneered by Gul and Pesendorfer (2001) to allow for magnitude effects: self-control increases when upping the stakes of a decision problem. For example, in intertemporal choice tasks, individuals are consistently found to switch from sooner, smaller rewards to larger, later ones when rewards are scaled up.[4] We show that – given a particularly simple specification, our model produces such behavior for agents that are present-biased unless using self-control to overcome it.[5] When applied to a simple infinite-horizon consumption-savings model, it produces a poverty trap at the

[2]In such cases, when subsequent choices do not conform to the plan made beforehand, the decision maker is said to be *dynamically inconsistent*.

[3]The fact that anticipating such a change of preference creates an incentive to commit has been well understood since the seminal work by Strotz (1955). However, as already pointed out by Thaler and Shefrin (1981), even in the absence of external commitment devices, decision makers typically have at their disposal strategies short of external commitments: self-control.

[4]We discuss the empirical evidence in Section 4 below.

[5]Recent research in neuroscience links intertemporal decisions in general, and the magnitude effect in particular, to self-control (Figner et al., 2010; Ballard et al., 2017, 2018).

individual level: decision makers forego self-control and over-consume (running down their wealth) unless they are endowed with sufficient initial wealth.

Preview of Main Results

Consider an individual faced with a decision problem A. At the planning stage she evaluates alternatives in A according to some utility function u but anticipates to choose from A according to v at the choice stage. By making a plan, the decision maker is able to restrict subsequent choices. That is, plan $P \subseteq A$ induces choice $c(A) = \max(P, v)$.[6] For instance, a restaurant goer may plan to 'choose a vegetarian dish'. If, say, the restaurant offers steak, salad and a veggie lasagna, this restricts her to choose between the salad and lasagna. At the same time, plan $P \subseteq A$ comes at a cost $\kappa(P, A)$. When evaluating decision problem A, the decision maker makes an optimal plan by trading off its cost against the benefit of (increased) self-control:

$$\begin{aligned} U(A) = \ & \max_{P \subseteq A} u(x_P) - \kappa(P, A) \\ & \text{s.t. } x_P = \max(P, v). \end{aligned} \tag{$\star$}$$

We show that *Planning to Self-Control* is equivalent to three simple and intuitive Axioms on preferences over decision problems and subsequent choices from them: decision makers (1) always weakly prefer to commit, (2) strictly prefer to commit only if this rids them of a self-control problem, and (3) are made strictly better off when adding an alternative to a decision problem A that is preferred to their choice from A both at the planning and the choice stage.

Decision makers who evaluate decision problems according to ($\star$) act as if being unconstrained in their ability to induce self-control: every alternative $x \in A$ is choosable given appropriate planning (e.g., consider $P = \{x\}$). The minimal planning cost to induce $x \in A$ may be interpreted as its self-control cost $C(x, A)$. We show that Planning To Self-Control is behaviorally equivalent to a decision maker who, at every decision problem A, acts as if choosing the optimal level of self-control:

$$U(A) = \max_{x \in A} u(x) - C(x, A). \tag{$\star\star$}$$

The restrictions our model puts on $C(x, A)$ allow for an interpretation in terms of self-control costs but are weak enough to nest other axiomatic models of self-control.

[6]For every $f : X \to \mathbb{R}$ and all $A \in \mathcal{A}$, we define $\max(A, f) = \{x \in A | \nexists y \in A : f(y) > f(x)\}$.

For example, in Gul and Pesendorfer (2001), $C(x, A) = \max_A v - v(x)$. Here, v is interpreted as a measure of how much other alternatives tempt the decision maker at the choice stage. Subsequent work (Takeoka, 2008; Noor and Takeoka, 2010, 2015) has considered more general functional forms for the cost function but has retained convexity (in foregone temptation utility) implying that decision makers use less self-control when more tempting alternatives are added to A.[7] This seems at odds with the intuition that self-control may become more attractive as the *stakes* of a decision increase because the costs of self-control decrease relative to its benefit. The simplest possible specification in our model, a fixed cost $C(x, A) = k > 0$ if $v(x) < \max_A v$ and $C(x, A) = 0$ else, implies such a *magnitude effect* for self-control.[8] To see this, note that if, say $x = \max(A, u)$ and $y = \max(A, v)$, then $(\star\star)$ reduces to $U(A) = \max\{u(x) - k, u(y)\}$. Thus, self-control is worthwhile if (and only if) the utility stakes $u(x) - u(y)$ exceed k.[9]

For example, reconsider our individual making plans for after-work activities at lunchtime. To put a twist on the story, suppose now that, initially, she is not very concerned about missing a run (r) when only faced with the alternative to watch a movie (m) and ends up in front of the screen. As a third option, to go out for drinks (d), becomes available to her, however, the stakes of the problem increase. While she expects to prefer the social occasion come the evening, she is profoundly concerned about its alcohol-related health risks in advance, so that $v(d) > v(m) > v(r)$ but $u(d) \ll u(m) < u(r)$. Consequently, she may use self-control in this situation seeing that its benefits have increased substantially.[10] Indeed, a fixed cost of self-control would imply such behavior given that $u(r) - u(d) > k > u(r) - u(m)$.

The possibly most well-known and empirically best documented case of increasing

[7]For details, see our discussion of temptation-driven self-control models in Appendix A.

[8]While simple, we believe that a fixed cost of self-control is also compelling on intuitive grounds as it may reflect the cost of engaging a 'self-control system' in the human brain. Findings that exertion of self-control is linked to heightened activity in certain areas of the human pre-frontal cortex (Figner et al., 2010; Ballard et al., 2017, 2018) provide tentative evidence in support of the existence of such a system.

[9]In the finite choice setting we use for our axiomatization below, utility is identified only up to positive monotone transformations hence may carry no *cardinal* information. Still $u(x) - u(y)$, as an (ordinal) measure of the stakes at menu A, is guaranteed to increase as more options z are added to A that present self-control problems (in the sense that if $v(z) > v(y)$, then $u(z) < u(y)$).

[10]Note that this constitutes a violation of the Weak Axiom of Revealed Preference (WARP) caused by increasing self-control: $c(\{m, r\}) = m$ but $c(\{d, m, r\}) = r$ where $u(r) > u(m)$. How is this possible? Note that, from a self-control perspective, watching the movie is a fundamentally different alternative in the two choice problems. This is due to the fact that it requires no self-control when the only alternative is to go for the run, while it *does* require self-control when being presented with the option to go out for drinks. At the same time, going for the run requires self-control in both instances. Thus, the *marginal* cost of self-controlling from m to r may decrease as another option d becomes available. Given a fixed cost of self-control, for example, the marginal cost is zero when d is present.

self-control is that of magnitude effects in intertemporal choice tasks: individuals tend to choose a smaller immediate reward (s) when the stakes are low but self-control to a larger later one (l) when the stakes are high. For example, if, say $s = \$30$ and $l = \$50$, you might prefer receiving s immediately over receiving l in a year. At the same time, if $s = \$150$ and $l = \$250$, you prefer to wait a year to receive the additional \$100. While this behavior is hard to square with discounted utility maximization, it may be perfectly reasonable if you are struggling with self-control problems related to immediate rewards.[11] When little is at stake, choosing the immediate reward is of little consequence. Yet, as the stakes grow, self-control becomes a worthwhile exercise. Indeed, we show in Section 4 below that this is natural for decision makers who exhibit present-bias unless engaging in self-control at a fixed cost to overcome it. More generally, in a simple infinite-horizon consumption-savings problem, such decision makers over-consume (due to present-bias) unless they are sufficiently wealthy to make self-control attractive. This produces a poverty trap (at the individual level).

Relation to the Literature

Recent work in psychology and human decision processes has identified several strategies and processes engaged by individuals to facilitate self-control (Duckworth et al., 2014, 2016; Hennecke and Bürgler, 2020).[12] This 'process model' of self-control distinguishes between preventive strategies (e.g., commitment) seeking to avoid situations involving conflicts, and interventive, intra-psychic ones employed to deal with self-control problems if they occur (cf. Inzlicht et al., 2021).[13] While, traditionally, goal setting has been considered central among the latter, goal achievement has been found to improve substantially when forming implementation intentions through making plans (Gollwitzer,

[11]While not inconsistent with discounted utility maximization per se, such choices would imply unreasonable curvature of the utility function (see Noor, 2011).

[12]Earlier findings that individuals subjected to cognitively laborious tasks (requiring self-control to stay focused) exhibited lower levels of self-control in subsequent experiments initially led the literature to theorize about the existence of a limited stock of cognitive (energy) resource that gets depleted ('ego-depletion') when exercising self-control (Baumeister et al., 1998, 2007). For a recent axiomatic treatment of choice behavior given a limited stock of 'willpower', see Masatlioglu et al. (2020). However, it should be noted that the term is not always used consistently. For example, Bermúdez et al. (2023) seem to refer to any kind of internal psychological mechanisms for resolving self-control problems as 'willpower' (as opposed to more externally-rooted devices such as commitment or extrinsic incentives).

[13]In this sense, our model is primarily concerned with an intra-psychic/interventive strategy for self-control: planning. Findings by Bermúdez et al. (2023) suggest that this internal psychological aspect is also more representative of every-day notions of self-control. However, the situational/preventive aspect features indirectly in our model to the effect that menu preferences reflect the decision maker's desire to pre-commit (in order to rid herself of self-control problems).

1999; Gollwitzer and Sheeran, 2006; Gollwitzer et al., 2010; Gollwitzer and Oettingen, 2011).[14] Masicampo and Baumeister (2011) argue that this might be due to the fact that plans can effectively turn control of goal pursuit over to automatic unconscious processes that can be called upon when the need arises. In support of the effectiveness of plans, Sjåstad and Baumeister (2018) show that willingness to plan is associated with good self-control. Other measures of propensity to plan have been shown to be predictive of goal achievement (Ludwig et al., 2018) and good credit scores (Lynch et al., 2010). In an early study of self-control in children, Mischel and Patterson (1976) find that four-year old children are better at sticking with a boring task (and resisting a distraction) when given a simple if-then plan.

In economics, the study of conflicting self-interests as exemplified by dynamically inconsistent tastes and the resulting desire to pre-commit to a course of action goes back at least to Strotz (1955). Pointing out that focusing on pre-commitment only provides part of the story, Thaler and Shefrin (1981) develop an early theory of self-control. Gul and Pesendorfer (2001) pioneer an axiomatic treatment based on the idea that decision makers take self-control into account when ranking decision problems (menus). In the strand of literature their work has inspired (see, e.g., Gul and Pesendorfer, 2004, 2006; Noor and Takeoka, 2010, 2015) self-control is costly since decision makers need to resist more tempting alternatives. While this literature considers costs that are convex (in foregone temptation utility) and lead to a loss of self-control when adding more tempting options to a menu, our model allows for more general specifications consistent with increasing self-control. Moreover, this literature takes as behavioral primitive preferences over menus of *lotteries* and invokes an Independence assumption. While we build on the same general framework, our axiomatic treatment is set in a (risk-free) finite-choice environment. *Prima facie*, the notions of self-control and risk seem unrelated to us. Thus, we believe it is of interest to develop an axiomatic foundation in a deterministic setting. While, in our model, self-control is enacted by making plans, Nehring (2006) considers a more general approach in which decision makers optimize over preferences according to which they choose subsequently. He characterizes this model in terms of the 'positive' component of the the 'Set Betweenness' Axiom of Gul and Pesendorfer (2001): that the union of two menus can never be strictly preferred to both of them. While our model fails the fully-fledged 'Set Betweenness' Axiom, it always satisfies the weakening considered by Nehring (2006). Thus, in terms of generality, our model ranks in between the two.

[14] In a related vein, Taylor et al. (1998) report that (mental) process simulations improved exam performance by psychology students substantially better than outcome simulations.

A natural interpretation of representation ($\star$) is that of a planning-stage self gaming a choice-stage self. The planning self is faced with selecting a self-control action $P \subseteq A$ (the plan) to which the choice-stage self reacts optimally (choosing $x_P = max(P, v)$). This connects our work to games-of-multiple-selves models in the literature. For example, Fudenberg and Levine (2006, 2012) capture self-control in terms of equilibria of a game played by a long-run self (the "planner") and multiple sequential short-run selves (the "doers"). The "planner" can steer future selves through some costly self-control action entering their utility functions. Thus, in their model, self-control actions affect choice-stage *preferences*. Hsiaw (2013) and Koch and Nafziger (2011) study the role of goals. They show that strategically setting goals can help attenuate present bias assuming that they provide reference points for future utility. In contrast, we capture self-control actions in terms of *partial* (internal) *commitment* devices: plans restrict the choice-stage self to choose from a sub-menu according to *stable* preferences. In general, while game-of-multiple-selves models derive their results from assumptions on utility functions and associated costs of self-control, our axiomatic treatment takes observable choice behavior as primitives of the model and derives the representation from testable axioms.

Like we do in this paper, Benhabib and Bisin (2005) argue that self-control can be enacted through internal commitment mechanisms, although they do not model them explicitly. In analogy to our analysis of a fixed cost of self-control, they assume that decisions otherwise made through automatic responses can be overridden through activation of controlled processing (that is, internal commitment acts through cognitive control) at some given cost. Interestingly, when applied to a consumption-savings problem where automatic decisions are driven by (stochastic) temptation of immediate consumption, the optimal self-control behavior is determined by a cut-off rule: the decision maker makes use of self-control when enough is at stake; i.e. when the temptation is sufficiently large. In contrast, in the simple consumption-savings model we present below, stakes are determined by the decision maker's wealth. That is, self-control is exerted only by the sufficiently wealthy.[15]

The rest of the paper is structured as follows. Section 2 presents the framework and discusses our Axioms. Section 3 contains our two Representation Theorems (Theorem 1 for ($\star$) and Theorem 2 for ($\star\star$)). Section 4 applies the assumption of a fixed cost of self-control to intertemporal consumption(-savings) choices. Here, we show that the fact that self-control increases in the stakes of a decision translates into a magnitude effect and leads to poverty traps in infinite-horizon problems. Lastly, Section 5 concludes. All

[15]Unlike us, Benhabib and Bisin assume cognitive control costs to be proportional to the wealth stakes of the problem, thereby focusing on varying temptation.

proofs are contained in the Appendix.

2. The Model

Preliminaries

We consider a simple dynamic setting with two stages: a planning stage and a choice stage. Let X be the set of alternatives at the choice stage; we assume that $|X| < \infty$. Define $\mathcal{A} = 2^X \backslash \{\emptyset\}$ to be the collection of all non-empty decision problems (henceforth also: *menus*) over X. We generically denote alternatives from X by x, y, z and menus from $\mathcal{A}$ by A, B. At the planning stage, the decision maker ranks menus according to some weak order $\succsim \subseteq \mathcal{A} \times \mathcal{A}$. When restricted to singletons, $\succsim$ reveals the decision maker's planning-stage (commitment) preference over alternatives. For simplicity, we assume that it does not display any indifferences. For ease of notation, we also denote singleton sets $\{x\}$ simply by x and write $x \succsim y$ instead of $\{x\} \succsim \{y\}$.

Axiom 0.1. Weak Order. *$\succsim$ is complete and transitive. Its restriction to singletons is anti-symmetric.*

At the choice stage, the decision maker chooses from menus. We model this by means of a (non-empty and singleton-valued) choice function $c : \mathcal{A} \to X$, $A \mapsto c(A) \in A$. Dynamic choice behavior is captured by the tuple $(\succsim, c(\cdot))$.

We denote by $\geq$ the anticipated default choice-stage preference over alternatives in the absence of self-control. Note, however, that since the decision maker might optimally bring self-control to bear on observed choice behavior, $\geq$ is not revealed by observing $c(\cdot)$ alone. That is, in general, $\geq$ does not rationalize $c(\cdot)$.[16] Indeed, $c(\cdot)$ may not be rationalized by *any* preference relation as optimal planning can lead to plausible violations of the Weak Axiom of Revealed Preference (WARP). We elaborate further below. At the same time, $\geq$ is revealed by observing $(\succsim, c(\cdot))$ jointly. Consider any $x \succ y$. First, note that choosing x from the menu $\{x, y\}$ may require costly planning; thus $x \succsim \{x, y\}$. Second, the default choice (in the absence of costly planning) from $\{x, y\}$ is guaranteed to be no worse than y; thus $\{x, y\} \succsim y$. Consequently, $x \succsim \{x, y\} \succsim y$.

If $x \sim \{x, y\}$, the decision maker anticipates to choose x from $\{x, y\}$ without the need for self-control (i.e. by default: $x > y$). In this case, planning-stage and choice-stage

[16] We say that a preference relation R rationalizes choice function $c(\cdot)$ if, for all menus A, $c(A) = \{x \in A | \forall y \in A : xRy\}$.

preferences agree; we say that x *dominates* y:

$$x \gg y :\iff x \sim \{x, y\} \succ y.$$

On the other hand, if $x \succ \{x, y\}$, the decision maker anticipates that x is choosable only under costly planning because y is preferred at the choice-stage: $y > x$. We denote such anticipated preference reversals by:

$$x \gtreqless y :\iff x \succ \{x, y\} \succsim y.$$

If these relations are to reveal a consistent choice-stage (default) *ranking* of alternatives, the decision maker must have a transitive perception of both dominance and reversals.

Axiom 0.2. Transitive Reversals and Dominance. $[x \gtreqless y$ *and* $y \gtreqless z] \implies x \gtreqless z$. $[x \gg y$ *and* $y \gg z] \implies x \gg z$.

In turn, this ensures that $\geq = (> \cup -)$, is a linear order; where, for all $x \neq y$,

$$x > y :\iff [x \gg y \text{ or } y \gtreqless x]$$

and, for all $x \in X$: $x - x$.[17] Indeed, vindicating the intuitions presented above, $x \gtreqless y \iff [x \succ y$ and $x < y]$ and $x \gg y \iff [x \succ y$ and $x > y]$.

Main Axioms

Anticipating a preference reversal $x \gtreqless y$ presents the decision maker with a self-control problem. (Henceforth, we will also simply refer to $x \gtreqless y$ as a self-control problem.) Planning allows to exert self-control but is costly. Thus, in making an optimal plan, the decision makers weighs said costs against the benefits of (increased) self-control. Observed choices as captured by the choice function $c(\cdot)$ reflect optimal resolution of this trade-off. The intuition driving the main Axioms of our model is that the costliness of a plan is linked to the self-control problems it helps overcome. That is, planning is the costlier the more self-control it allows to exert. For example, when considering binary menus, this is implicit in our definition of a self-control problem: $x \gtreqless y \iff x \succ \{x, y\} \succsim y$. Indeed, plan $P = \{x, y\}$ (which we may also think of as making no plan) does not induce self-control ($y = \max(\{x, y\}, >)$) but is free of cost.[18] Consequently, $\{x, y\} \succsim$

[17] See Lemma 1 in the Appendix.

[18] For every $P \subseteq X \times X$ and all $A \in \mathcal{A}$, we define $\max(A, P) = \{x \in A | \nexists y \in A : yPx\}$.

y. On the other hand, plan $P = \{x\}$ allows to exert self-control ($x = \max(\{x\}, >)$) but is costly; thus $\{x, y\} \prec x$.

Generally speaking, if planning costs are tied to the self-control problems it help overcome, the decision maker cannot be left worse off by committing to a sub-menu given it contains what is chosen from the original one. To see this, suppose we observe the decision maker choosing alternative $x \in A$ from menu $A \cup B$. Then, in a revealed sense, the decision maker prefers to plan so as to induce x in menu $A \cup B$.[19] Potentially, this involves self-controlling away from alternatives $y \in B$. The need to do so is removed, however, when committing to the sub-menu A. Thus, planning to induce x in A comes at a lesser cost. Therefore, A should be no worse to the decision maker than $A \cup B$. This is Axiom 1.

Axiom 1. Weak Preference for Commitment. $c(A \cup B) \in A \implies A \succsim A \cup B$.

Moreover, if preference for commitment is strict, this must be due to the fact that it rids the decision maker of some self-control problem that need not be overcome by planning when committing. In other words, it is self-control problems that make planning costly. This is Axiom 2.

Axiom 2. Costly Planning/Self-Control. $A \succ A \cup B \implies \exists y \in B : c(A) \gtrless y$.

Equivalently, the counter-positive of Axiom 2 states that if, for all $y \in B$, $c(A) > y$ or $y \succ c(A)$, then $A \cup B \succsim A$. This reflects the intuition that additional options can never hurt as long as the decision maker either does not need to plan to self-control away from them or has no incentive to do so.

In contrast, when $x \gg y$, there is no self-control problem. The decision maker strictly prefers x both at the planning stage and at the choice stage. Axiom 3 states that adding x to some menu from which y is chosen must make the decision makers better off. Roughly speaking, this reflects the intuition that there exists some plan for menu $x \cup A$ inducing x that is no more costly than the optimal plan inducing y in menu A.[20] As x is strictly preferred at the planning stage, the decision maker values the addition of x to A.

Axiom 3. Unequivocally Better Choice. $x \gg c(A) \implies x \cup A \succ A$.

[19] We say that $P \subseteq A$ induces x (in menu A) if $x = \max(P, >)$.

[20] To be precise, we cannot make this inference from the fact that the decision maker strictly prefers adding x to A. The latter only implies that, if planning to induce x in $A \cup \{x\}$ is more costly than the optimal plan for A (inducing y), then this is dominated by the corresponding increase in planning-stage utility that x offers over y. This fact is reflected by Property 3.b) of a planning/self-control cost function in our model (cf. Definition 1/2).

A fortiori, when $\{x, y\} \subseteq A$ and $x \gg y$, then $y \neq c(A)$. That is, alternatives that are dominated by some other available alternative are never chosen.

As we show below our axioms are not only necessary but also sufficient for a decision maker to be Planning to Self-Control. Before we turn to our representation, we discuss our Axioms further.

Discussion of Main Axioms

Axiom 1 implies a weakened version of the 'Set-Betweenness' Axiom that is central in the characterization of temptation-self-control preferences in Gul and Pesendorfer (2001). 'Set-Betweenness' requires that the union of any two menus $A \succsim B$ rank between them: $A \succsim A \cup B \succsim B$. Indeed, Axiom 1 implies its 'positive' part: $A \succsim A \cup B$. Nehring (2006) shows that it is necessary and sufficient for $\succsim$ to be rationalized in terms of a general model of 'second-order preference'.[21] However, the corresponding 'negative' part fails. To see why $A \succsim B \succ A \cup B$ is reasonable within our model, suppose that the decision maker self-controls to x in menus A and B (that is $x = c(A) = c(B)$, which requires that $x \in A \cap B$, of course). Planning to induce x in the menu $A \cup B$ is more costly, however, seeing that the decision maker now needs to self-control away from alternatives in A *and* B at the same time. In contrast, this is not possible if $A \cup B$ contains only two alternatives. Indeed for all $x \succsim y$, Axiom 2 implies that $\{x, y\} \succsim y$. Axiom 1 implies that $x \succsim \{x, y\}$. In other words, 'Set-Betweenness' holds for binary menus.

Axiom 1 implies that if $A \succ y$ for some $y \in A$, then $y \neq c(A)$. For example, if $x \succsim \{x, y\} \succ y$, then we must have $x = c(\{x, y\})$. On the other hand, if $\{x, y\} \sim y$, our Axioms do not put any restrictions on choice from $\{x, y\}$. Note that if $\{x, y\} \sim y$ and $x = c(\{x, y\})$, the decision maker plans to self-control to x at a cost that exactly equals the benefit of self-control to her. Hence she is indifferent between self-control and no self-control ($y = c(\{x, y\})$).

Other than the ones discussed above, our Axioms imply only weak (consistency) restrictions on choice from menus. In particular, $c(\cdot)$ may fail to satisfy WARP known to characterize choice behavior that is rationalized by some preference relation. In the presence of self-control problems it is natural to allow for violations of WARP. To see this, reconsider our introductory example from above. While our decision maker self-controls to go for the run (r) in the presence of an alternative she considers seriously

[21] Nehring considers a second condition ('singleton monotonicity') on menu preferences. When adding to some menu an alternative that is preferred to all alternatives in the menu, this never leaves the decision makers worse off. Formally, if for all $y \in A$, $x \succsim y$, then $x \cup A \succsim A$. Nehring calls this subclass of preferences 'self-command' preferences. Our Axiom 2 implies 'singleton monotonicity'. Thus, our model is a special case of 'self-command' preference.

detrimental to her health (to go for drinks: d), she optimally decides to forego self-control when only ending up watching a movie (m). This choice pattern, $r = c(\{d, m, r\})$ and $m = c(\{m, r\})$, constitutes a violation of WARP. Seeing that $r \gtreqless m$, it is an example of *increasing* self-control: the decision maker has more self-control at the menu $\{d, m, r\}$ than at the sub-menu $\{m, r\}$. To give another example of where such behavior can be plausible, consider an individual struggling with being tempted by desserts. It may be easier to resist having a dessert altogether when there are 10 of them available (including one that is particularly unhealthy but tempting) as compared to when there is only one (moderately unhealthy but tempting one).

At the same time, our Axioms are equally consistent with violations of WARP resulting from decreasing self-control. Such choice patters are generally consistent with generalizations of Gul and Pesendorfer (2001) (see also Appendix A for more details). For instance, suppose some product is available at different prices (e.g., because it is available from different brands). There is a high-price (h), medium-price (m) and low-price (l) variant. At the planning stage, a consumer regards all variants as perfect substitutes (i.e., there are no perceived differences in terms of quality) so that she ranks them according to price: $l \succ m \succ h$. However, she expects her ranking to reverse at the time of choice: $h \succ m \succ l$ (e.g., driven by product packaging and presentation or impulsive inferences about quality). Suppose further that a corner store only offers l and m, a bigger supermarket sells all three variants. While our consumer may optimally (plan to) self-control to the low-cost option l in the corner store ($l = c(\{l, m\})$), doing so in the supermarket ($\{l, m, h\}$) might require a higher planning effort so as to exclude both m and h at the choice stage. Thus, she may only choose to self-control to the medium-price option ($m = c(\{l, m, h\})$) but optimally forego a higher level of self-control (to l). As $l \gtreqless m$, this choice pattern constitutes a loss of self-control when adding the high-price option h to the menu $\{l, m\}$.

3. Representation

Planning to Self-Control

Our representation builds on the idea that plans allow the decision maker to restrict subsequent choices. We model this by *identifying* plans with the choice restrictions (commitment) they entail. That is, for every decision problem A and every non-empty $P \subseteq A$, we call P a *plan* for A. If plan P is made, the decision maker chooses the best available alternative that is consistent with P. Given our Axioms above, (default) preferences at the choice stage are represented by some utility function v. Thus, plan $P \subseteq A$ *induces* $x_P := \max(P, v)$ at the choice stage. At the same time, it comes at a cost $\kappa(P, A)$. Our Axioms imply restrictions on the planning-cost function $\kappa(\cdot, \cdot)$ (cf. Definition 1); we discuss them in detail below. Intuitively speaking, planning incurs a cost (if and) only if it allows to exert self-control and is the costlier the more it does so. As preferences over decision problems $\succsim$ are a weak order, they are representable by some utility function U. We let u be the restriction of U to singleton menus; that is, for all $x \in X$, $u(x) := U(\{x\})$. Thus, u represents the decision maker's planning-stage (commitment) preference over alternatives. Our first main result shows that the tuple $(\succsim, c(\cdot))$ satisfies our Axioms if and only if it is *Planning to Self-Control* (PTSC). That is, $U(A)$ is the indirect utility from maximizing planning-stage utility $u(x_P)$ net of planning costs $\kappa(P, A)$ and optimal planning rationalizes subsequent choice behavior: $c(A) = x_{P^\star}$ for some optimal $P^\star$.

Theorem 1. *If and only if $(\succsim, c(\cdot))$ satisfies Axioms 0.1,0.2,1-3, there exist strictly increasing*[22] *utility functions $u, v : X \to \mathbb{R}$ and planning-cost function κ such that:*

1. *$\succsim$ is represented by*

$$\begin{aligned} U(A) = \ & \max_{P \subseteq A} u(x_P) - \kappa(P, A) \\ & \text{s.t. } x_P = \max(P, v). \end{aligned} \tag{$\star$}$$

2. *$c(A) = x_{P^\star}$ for some $P^\star$ that solves $(\star)$.*

Definition 1. *We say that $\kappa : \mathcal{A} \times \mathcal{A} \ni (P \subseteq A, A) \mapsto \kappa(P, A) \geq 0$ is a planning-cost function if*

1. $\kappa(P, A) > 0 \iff \exists y \in A \backslash P : v(y) > v(x_P)$
2. $x_P \in A \implies \kappa(P \cap A, A) \leq \kappa(P, A \cup B)$
3. a) $u(x_{P \cup B}) \leq u(x_P) \implies \kappa(P \cup B, A \cup B) \leq \kappa(P, A)$

   b) $u(x_{P \cup B}) > u(x_P) \implies \kappa(P \cup B, A \cup B) - \kappa(P, A) < u(x_{P \cup B}) - u(x_P)$

The first property requires that P incurs a cost at menu A if and only if the alternative x_P it induces is not the default choice; that is, if there exists some $y \in A \backslash P$ that is chosen over x_P unless it is excluded from P (as $v(y) > v(x_P)$). Second, plan $P \subseteq A \cup B$ is

[22] That is, $u, v : X \to \mathbb{R}$ allow only for trivial indifferences: $u(x) = u(y) \implies x = y$ and $v(x) = v(y) \implies x = y$.

no more costly when projected onto the sub-decision problem A as long as it induces the same alternative: note that if $x_P \in A$, then $x_{P\cap A} = x_P$. In particular, this implies that some fixed P is (weakly) less costly at smaller menus as compared to larger ones. That is, if $P \subseteq A$, then $\kappa(P, A) \leq \kappa(P, A \cup B)$ for every menu B. Equivalently put, the planning cost associated with restricting choices to some given P (weakly) increases as more alternatives are available at the choice stage. This reflects the intuition that P may need to exclude additional alternatives when choosing from $A \cup B$ as compared to A (namely those in B; unless $P \subseteq B$). On the other hand, when modifying plan $P \subseteq A$ so as to be consistent with all alternatives that are added to menu A, that is when considering $P \cup B \subseteq A \cup B$, the associated planning cost should be no higher. The third property establishes this for the case that $x_P = x_{P\cup B}$ or $x_P \gtrless x_{P\cup B}$ (part a)). However, if $x_{P\cup B} \gg x_P$ (i.e. if B contains an alternative that dominates x_P), then our Axioms only allow us to ascertain that the cost differential $\kappa(P \cup B, A \cup B) - \kappa(P, A)$ is bounded above by the planning-stage utility differential $u(x_{P\cup B}) - u(x_P)$ (part b)).[23] Note that an immediate consequence of Property 3a) is that if $P' \subseteq P \subseteq A$ induce the same alternative, then $\kappa(P', A) \geq \kappa(P, A)$ (let $B \subseteq A \backslash P'$).

Planning is costly. For this reason, rational decision makers use it only to the extent necessary to induce self-control. We say that plan $P \subseteq A$ is *efficient* if for all alternatives $y \in A\backslash P$ excluded by P it holds that self-control is both necessary ($v(y) > v(x_P)$) and beneficial ($u(y) < u(x_P)$), that is, $x_P \gtrless y$.[24] Without loss of generality, we can restrict a PTSC decision maker to choose among all efficient plans in representation ($\star$). To see this, note that for any $P \subseteq A$, we can construct an efficient plan $P' \supseteq P$ such that $A\backslash P' = \{y \in A : x_P \gtrless y\} = \{y \in A : u(x_P) > u(y) \textit{ and } v(x_P) < v(y)\}$. If $x_{P'} = x_P$, the cost of plan P' is (weakly) less than that of P (let $B = P'\backslash P \subseteq A$ and use property 3a) for planning-cost functions) but induces the same alternative. Hence P can be optimal only if P' is. Else if $x_{P'} \neq x_P$, then $x_{P'}$ must dominate x_P ($u(x_{P'}) > u(x_P)$ and $v(x_{P'}) > v(x_P)$). By property 3b) for planning-cost functions, the cost of plan P' exceeds that of P (if at all) by less than the planning-stage utility benefit of $x_{P'}$ with respect to x_P: $u(x_{P'}) - \kappa(P', A) < u(x_P) - \kappa(P, A)$. Here, P cannot be optimal.

In our model, plan $P \subseteq A$ induces $x_P = \max(P, v)$ at the choice stage. Prima facie, this seems to assume that decision makers (rationally expect to) stick to the plan they made beforehand. This seems implausible. However, it is not the only possible interpretation. Alternatively, we can identify with $P \subseteq A$ the plan for which P is the

[23]Note that, by definition, $v(x_{P\cup B}) \geq v(x_P)$, with equality only if $v(x_{P\cup B}) = v(x_P)$. If $u(x_{P\cup B}) > u(x_P)$, then $x_{P\cup B} \neq x_P$; hence $x_{P\cup B} \gg x_P$.

[24]In other words, all $y \in A\backslash P$ present the decision maker with a self-control problem: $x_P \gtrless y$.

collection of alternatives that are acceptable (deviations) given said plan. After all, what is important about a plan in our model is to what extent it restricts subsequent choices but not how it is represented in the decision maker's mind.[25] Given this interpretation, however, it is important to note that our concept of optimality (for plans) implicitly assumes that every non-empty choice restriction $P \subseteq A$ is induced by *some* plan the decision makers can make. In effect, this ensures that self-control is *potentially unlimited* (but possibly too costly). Fudenberg and Levine make an analogous assumption for their 'planner-doer' model: every action by a future self can be elicited by the long-run self through some appropriate self-control action (2006, Assumption 2).

As can be expected in our finite choice context, the additive form in ($\star$) is not identified. That is, while Theorem 1 shows that a representation of this from always exists, it does not ensure that U is of the prescribed form whenever U represents $\succsim$. In general, there are multiple solutions to ($\star$). However, the optimal *efficient* plan $P^\star$ that solves ($\star$) and rationalizes observed choice behavior (i.e. $c(A) = x_{P^\star}$) is unique.[26]

Equivalent Self-Control Cost Model

Plans are instrumental in inducing self-control. That is, a PTSC decision maker plans in order to induce alternatives that would otherwise not be chosen. Consider decision problem A. For every alternative $x \in A$, let $P_x := \{y \in A : v(y) \leq v(x)\}$. Note that if $P \subseteq A$ is some plan inducing x, we need to have $P \subseteq P_x$; hence $\kappa(P, A) \geq \kappa(P_x, A)$.[27] Thus, P_x is the cost-minimal plan inducing x. From an abstract point of view, we may also think of the decision maker as optimizing directly over eventual choices $x \in A$ incurring a self-control cost $C(x, A) = \kappa(P_x, A)$. Then, she evaluates menus according to

$$U(A) = \max_{x \in A} u(x) - C(x, A). \qquad (\star\star)$$

Indeed, Theorem 2 shows that this is an equivalent representation of planning-stage menu preferences. Moreover, if $x^\star$ solves ($\star\star$), then there exists some (efficient) plan $P^\star$ solving ($\star$) such that $x^\star = x_{P^\star}$ (hence $x^\star = c(A)$). Vice versa, if $P^\star$ solves ($\star$), then $x_{P^\star}$ solves ($\star\star$).

Theorem 2. *Let κ be a planning-cost function, then there exists a self-control-cost function C such that:*

[25] That is, we adopt an extensional definition of plans.

[26] If we do not restrict ourselves to efficient plans in ($\star$), there may be several equally costly plans that induce the choice $c(A)$. However, as noted above, restriction to efficient plans is without loss of generality.

[27] To see this, let $B = P_x \backslash P \subseteq A \backslash P$ and use Property 3a) for $\kappa(\cdot, \cdot)$.

1. $\max_{P\subseteq A} u(x_P) - \kappa(P,A) = \max_{x\in A} u(x) - C(x,A)$

2. a) $P^\star$ *solves* $\max_{P\subseteq A} u(x_P) - \kappa(P,A) \implies x_{P^\star}$ *solves* $\max_{x\in A} u(x) - C(x,A)$

   b) $x^\star$ *solves* $\max_{x\in A} u(x) - C(x,A) \implies \exists P^\star \subseteq A : x_{P^*} = x^\star$ *and* $P^\star$ *solves* $\max_{P\subseteq A} u(x_P) - \kappa(P,A)$

Vice versa, if C is a self-control-cost function, then there exists a planning-cost function κ such that 1. and 2. hold.

Definition 2. *We say that $C : X \times \mathcal{A} \ni (x \in A, A) \mapsto C(x,A) \geq 0$ is a self-control-cost function if*

1. $C(x,A) > 0 \iff \exists y \in A : v(y) > v(x)$

2. $C(x,A) \leq C(x, A\cup B)$ *and* $C(x,A) < C(x,A\cup B) \implies \exists y \in B : v(y) > v(x)$

3. *if $v(x) > v(y)$, then:*

   a) $u(x) < u(y) \implies C(x, x\cup A) \leq C(y,A)$

   b) $u(x) > u(y) \implies C(x, x\cup A) - C(y,A) < u(x) - u(y)$

In analogy to planning-cost functions, self-control costs increase in the number of self-control problems that need to be overcome. First, the cost of self-controlling to $x \in A$ is strictly positive if and only if at least one alternative $y \in A$ is preferred at the choice stage. Second, self-control costs are greater at larger menus; and strictly so only if the larger menu contains an alternative that is preferred at the choice stage. Lastly, suppose that x is choice-stage preferred to $y \in A$. Then x incurs a smaller self-control cost when added to menu A than y does in menu A given that x presents the decision makers with a self-control problem vis-à-vis y: $y \gtrless x$. If x dominates y, the self-control cost for x in $x \cup A$ exceeds that for y in A by strictly less than the additional planning-stage utility offered by x.

Theorem 2 is helpful as it facilitates comparison of the PTSC model with other models of self-control by capturing those in terms of additional assumptions on the self-control-cost function in ($\star\star$). Theorem 2 characterizes the constrained optimization problem ($\star$) in terms of an unconstrained one. This is of particular interest in applications. To simplify both the verbal and technical exposition, we mostly suppress the planning aspect for the rest of the paper and formulate our assumptions in terms of (generic) self-control costs. While our definition of a self-control-cost function involves minimal criteria to be consistent with our Axioms, it allows for a large variety of functional forms. For example, one interesting class of functions arises when considering self-control costs driven by the need to resist temptation. In Appendix A we show how some prominent temptation-driven models of self-control correspond to specific assumptions on self-control costs. Another particularly simple case arises when costs are fixed; that is, do not vary by 'how much' self-control need be exerted. We turn to this case next and show that it generates well-known stylized facts about intertemporal decision making.

4. A Fixed Cost Implies Increasing Self-Control: Magnitude Effect and Poverty Traps in Intertemporal Choice

In general, Representation ($\star\star$) allows for self-control costs to increase in the 'amount' of self-control needed to follow through with a certain choice. That is, self-control may become more costly when including alternatives that pose additional self-control problems (cf. Definition 2, in particular, Property 2).[28] In this section we consider the particular case of a fixed cost of self-control; that is, when all self-control is equally costly. Intuitively speaking, this could reflect the existence of a self-control system in the human brain that can be activated at some given cognitive cost.[29] Once engaged, this system takes over all decision making (at no additional variable cost).

Formally speaking, we say that the cost of self-control is fixed if there exists some $k > 0$ such that for all $A \in \mathcal{A}$ and $x \in A$:

$$C(x, A) = \begin{cases} 0 & \text{if } x = \max(A, v) \\ k & \text{else} \end{cases}.$$

A fixed cost greatly simplifies the self-control decision faced at some decision problem A. As all self-control is equally costly, the decision maker never exhibits intermediate levels of it. The self-control decision reduces to a binary choice between full self-control at cost k (choosing the planning-stage optimum $\max(A, u)$), and no self-control (choosing $\max(A, v)$). Let $W(A) := \max_{x \in A} u(x)$ be the indirect utility given (full) self-control but net of cost k and $V(A) := u(x_A) = u(\max(A, v))$ denote the indirect utility given no self-control. Self-control is optimal at menu A if and only if

$$W(A) - V(A) \geq k. \tag{1}$$

[28] Equivalently, planning is more costly when more self-control problems are present (cf. Definition 1, in particular, Property 2).

[29] Presumably, the latter might best be understood as an opportunity cost for not engaging those parts of the brain involved in self-control in other cognitive activities (Boureau et al., 2015).

Importantly, note that, while v determines choices given no self-control, these are still evaluated according u when considering whether to self-control (or not). The overall indirect utility at menu A – reflecting self-control behavior – can be conveniently expressed as

$$U(A) = \max\{W(A) - k, V(A)\}.$$

Consequently, the decision maker gains self-control as the (utility) stakes of decision problem A, $W(A) - V(A)$ increase.

Although a simplification, the fixed-cost model is general enough to account for a variety of stylized facts about intertemporal choice behavior. In what follows we consider applications to typical intertemporal choice tasks in experimental settings and to consumption-savings decisions.

The Magnitude Effect

Consider an individual faced with choosing between pairs of dated money rewards. Let the tuple $(m, t) \in \mathbb{R} \times \mathbb{N}$ denote an alternative that pays $\$m$ at time t. In a typical experimental setting, individuals are asked to choose between a smaller sooner (s, t) and a larger later $(l, t+1)$ reward. Individuals exhibit *present bias* when preferring the sooner reward when it is immediate but choose the larger reward when all payoffs are delayed by the same amount of time $\tau \geq 1$. If the decision is made at time t, that is:

$$(s, t) \succ_t (l, t+1) \textit{ and } (s, t+\tau) \prec_t (l, t+\tau+1) \qquad (\textit{present bias})$$

where $0 < s < l$.

For example, an individual may prefer receiving \$30 immediately to receiving \$50 in a week but prefer \$50 in a year and a week to \$30 in a year. For an overview of the early literature finding present bias, see, for example, Frederick et al. (2002). In more recent studies, the status of present bias for *money* rewards has been somewhat contested (see, e.g., Andreoni and Sprenger, 2012; Sutter et al., 2013; Meyer, 2015; Sun and Potters, 2022; however, cf. Meier and Sprenger, 2010; Benhabib et al., 2010; Andersen et al., 2013; Augenblick et al., 2015). Yet present bias is consistently found in studies involving *real* rewards (see, e.g., McClure et al., 2007; Augenblick and Rabin, 2019).

The possibly most well documented finding about intertemporal choice behavior, however, is that preferences between smaller, sooner and larger, later rewards change systematically when increasing the stakes of a decision. For example, an individual preferring \$30 immediately to \$50 in a week may switch to the larger, later one when the rewards are \$150 and \$250 respectively (i.e. scaled up by factor 5). Such *magnitude effects* have been found both for money and real rewards across a variety of settings (Thaler, 1981; Green et al., 1997; Kirby, 1997; Benhabib et al., 2010; Andersen et al., 2013; Sutter et al., 2013; Meyer, 2015; Sun and Potters, 2022, see, e.g.,).[30]

Formally, individuals exhibit a magnitude effect if

$$(s,t) \succ_t (l,t+1) \textit{ and } \exists\lambda > 1 : (\lambda s,t) \prec_t (\lambda l,t+1). \qquad (\textit{magnitude effect})$$

While, in principle, magnitude effects can be accounted for in the standard (exponentially discounted utility) model by curvature of the utility function, the extreme curvature this would require creates implausible predictions (Noor, 2011). At the same time, magnitude effects are natural in a model with self-control problems that can be overcome at a fixed cost. For simplicity, consider a discounted value maximizer who, given self-control, is an exponential discounter and, given no self-control, discounts quasi-hyperbolically. That is, she evaluates dated rewards according to $u_t(m,t+\tau) = D_\delta(\tau)\cdot m$ and $v_t(m,t+\tau) = D_{\beta,\delta}(\tau)\cdot m$ where

$$D_\delta(\tau) = \delta^\tau \text{ and } D_{\beta,\delta}(\tau) = \begin{cases} \beta\delta^\tau & \tau \geq 1 \\ 1 & \tau = 0 \end{cases}$$

for some $0 \leq \beta < 1$ and $0 < \delta \leq 1$.

Suppose that

$$0 \leq \beta\delta l < s < \delta l \qquad (2)$$

Then, the decision maker would prefer the later reward $(l,t+1)$ given self-control but choose the immediate reward (s,t) under no self-control. If both rewards are delayed, however, the later reward is preferred no matter what the self-control decision. Thus, the decision maker exhibits present bias if choosing to forego self-control in the former case. This is case if the benefit of self-control, $\delta l - s$, falls short of the self-control cost k.[31]

Now suppose that rewards are scaled up by some factor $\lambda > 1$. Note that this leaves preferences regarding $(\lambda s,t)$ and $(\lambda l,t+1)$ given self-control and no self-control unaffected

[30] Studies that elicit money discount rates through indifference statements which ask subjects to specify an amount x such that $(s,t) \sim_t (x,t+1)$, refer to the magnitude effect as an increasing money discount factor $\frac{s}{x}$ (decreasing money discount rate $\frac{x}{s}-1$).

[31] Here, linear utility implies that the benefit of self-control is the discounted value it produces in excess of the no-self-control option. Vice versa, we may interpret the cost of self-control k as the premium the decision makers demands to make self-control worthwhile.

(and still governed by Equation (2)) while making self-control worthwhile if

$$\lambda(\delta l - s) \geq k. \tag{3}$$

Thus, as the stakes rise, the decision maker will eventually find it optimal to exert self-control (no matter what the cost of self-control k). For instance, suppose $\beta = 0.5$, $\delta = 1$, $k = 80$ and reconsider our example from above. As $0.5 \cdot 50 < 30 < 50$, Equation (2) is satisfied. Moreover, since $50 - 30 < k = 80 < 5 \cdot (50 - 30) = 250 - 150$, the decision maker foregoes self-control in the low-stakes ($\lambda = 1$: $(\$30, t) \succ_t (50\$, t+1)$) but gains self-control in the high-stakes ($\lambda = 5$: $(\$150, t) \prec_t (250\$, t+1)$) condition (cf. Equation (3)).

More generally, when choosing at time t from a menu of dated rewards $\mathcal{M}_t = \{(m^{(i)}, t + \tau^{(i)}), i \in \mathcal{I}\}$ (where $\mathcal{I}$ is some index set) such that an immediate reward is preferred under no self-control: $(m^{(i^*)}, t) = \max(\mathcal{M}_t, D_{\beta,\delta}(\tau) \cdot m)$ for some $i^* \in \mathcal{I}$. Then, the choice from $\mathcal{M}$ (given optimal self-control) maximizes the discounted value $D(\tau, m) \cdot m$ according to the (magnitude-dependent) discount function

$$D(\tau, m) = \begin{cases} \delta^\tau - \frac{k}{m} & \tau \geq 1 \\ 1 & \tau = 0 \end{cases}. \tag{4}$$

Benhabib et al. (2010) show that such a fixed component is a better fit to their experimental data than both hyperbolic and quasi-hyperbolic discounting.[32] Discounting according to (4) produces a magnitude effect as the discount factor is increasing in the size of the reward. That is, larger rewards are discounted at lower rates.

Optimal Self-Control in Consumption-Savings Decisions

In this subsection, we consider a decision maker with Epstein-Zin preferences and a fixed cost of self-control $k > 0$ facing a simple consumption-savings problem. In each period t, the decision maker decides how much of her wealth w (carried over from the previous period) to consume and how much to save (at gross interest rate $R \geq 1$) for the future. As for the previous subsection, we assume that preferences given self-control and no self-control differ only in terms of how they discount the future. While the decision maker discounts exponentially at rate $0 < \delta \leq 1$ given self-control, no self-control behavior is based on quasi-hyperbolic β-δ discounting where $0 \leq \beta < 1$.

[32] The hypothesis of exponential discounting can be rejected as it does not allow for present bias which they find in their data.

To build intuition, we start with the simple case of two periods. In the second period, the decision maker simply consumes her savings carried over from the first period. Every initial wealth level $w > 0$ presents the decision maker with a menu of first-period consumption levels $c \in [0, w]$. In the second period, she consumes her savings carried over from the first period: $R(w-c)$. Thus, the decision maker solves a single self-control problem in the first period. By slight abuse of notation, we write $U(w) = U([0, w])$ to denote the value function of this problem. Given a fixed cost of self-control, we have

$$U(w) = \max\{W(w) - k, V(w)\} \tag{5}$$

where

$$W(w) = \max_{c\in[0,w]} [c^\sigma + \delta(R(w-c))^\sigma]^{\frac{1}{\sigma}} \tag{6}$$

and

$$V(w) = [c_{NSC}^\sigma + \delta(R(w-c_{NSC}))^\sigma]^{\frac{1}{\sigma}} \tag{7}$$

such that c_{NSC} solves

$$\max_{c\in[0,w]} [c^\sigma + \beta\delta(R(w-c))^\sigma]^{\frac{1}{\sigma}}. \tag{8}$$

where $\delta R \leq 1$.

The parameter $\sigma < 1$ captures how readily optimal growth c_{t+1}/c_t responds to changes in the interest rate R: $\sigma = 1 - 1/\gamma$ where $\gamma = d\ln\left(\frac{c_{t+1}}{c_t}\right)/d\ln R > 0$ is the Elasticity of Intertemporal Substitution (EIS). When σ decreases, the decision maker becomes less inclined to readjust to a changing interest rate. In the limit, as σ tends to $-\infty$ (i.e. γ tends to 0) there is extreme consumption smoothing to the effect that $c_{t+1} \approx c_t$ in the optimum (irrespective of R). In the opposite case, as σ tends to 1 (i.e. γ tends to $+\infty$) consumption is perfectly substitutable across time.[33] When $\beta = 1$, the model reduces to the standard consumption-savings (pie-eating) problem (hence we assume $\beta < 1$).

As preferences are homothetic, the optimal solutions c_{SC} to (6) and c_{NSC} to (8) are to consume a constant fraction (depending on R, σ, β, δ) out of wealth. That is, there exist $0 < \mu_{SC}, \mu_{NSC} < 1$ such that $c_{SC} = \mu_{SC} \cdot w$ and $c_{NSC} = \mu_{NSC} \cdot w$. As the decision maker is less patient in the absence of self-control ($\beta < 1$), she over-consumes: $\mu_{NSC} > \mu_{SC}$.[34] Since utility is homogeneous of degree one, indirect utilities $W(w), V(w)$ are affine in

[33] In this limit case, decisions are made based maximizing discounted values (of consumption streams). Thus, our discussion of the magnitude effect in the last subsection is a limiting case of the model presented here.

[34] As may be easily verified by computing the First Order Conditions (FOCs): $\mu_{SC} = \frac{\delta^{-\gamma}R^{1-\gamma}}{1+\delta^{-\gamma}R^{1-\gamma}} < \frac{(\beta\delta)^{-\gamma}R^{1-\gamma}}{1+(\beta\delta)^{-\gamma}R^{1-\gamma}} = \mu_{NSC}$.

wealth. That is, there exist $b_{SC}, b_{NSC} > 0$ such that $W(w) = b_{SC} \cdot w$ and $V(w) = b_{NSC} \cdot w$. As b_{NSC} reflects sub-optimal resolution of the intertemporal problem (based on $\beta\delta < \delta$), we have $b_{NSC} < b_{SC}$. Consequently, the (net) benefit of self-control

$$W(w) - V(w) = (b_{SC} - b_{NSC}) \cdot w$$

is an increasing affine function in wealth.

As self-control is worthwhile only if its benefits exceed its cost $k > 0$, optimal self-control behavior is given by a simple cut-off rule: self-control is exerted if and only if the wealth stakes exceed the critical value $\bar{w} = \frac{k}{b_{SC} - b_{NSC}}$:

$$c^*(w) = \begin{cases} \mu_{NSC} \cdot w & w \leq \bar{w} \\ \mu_{SC} \cdot w & w > \bar{w} \end{cases}.$$

Thus, poor decision makers over-consume (relative to the self-control benchmark enacted by richer decision makers). As the marginal utility of an additional unit of wealth is higher when using self-control than when not ($b_{,SC} > b_{NSC}$), the value function is kinked upwards at $\bar{w}$:

$$V(w) = \begin{cases} b_{NSC} \cdot w & w \leq \bar{w} \\ b_{SC} \cdot w - k & w > \bar{w} \end{cases}.$$

Infinite Horizon: Poverty Traps

Suppose now that the problem is infinitely lived. Depending on her wealth stock w, the decision maker decides how much to consume in the current period and how much to save. Her savings $R(w - c)$ become the wealth endowment of the continuation problem faced in the next period. As the decision maker faces the same problem in every period, we have

$$U(w) = \max\{W(w) - k, V(w)\} \tag{9}$$

where W, V obey the Bellman Equations

$$W(w) = \max_{c \in [0,w]} [(1 - \delta)c^\sigma + \delta U(R(w - c))^\sigma]^{\frac{1}{\sigma}} \tag{10}$$

and

$$V(w) = [(1 - \delta)c_{NSC}^\sigma + \delta U(R(w - c_{NSC}))^\sigma]^{\frac{1}{\sigma}} \tag{11}$$

such that c_{NSC} solves

$$\max_{c\in[0,w]} \left[(1-\delta)c^{\sigma} + \beta\delta U(R(w-c))^{\sigma}\right]^{\frac{1}{\sigma}}.^{35} \tag{12}$$

Note that multiplying utility from current consumption by $(1-\delta)$ in Equations (10) and (11) is a convenient normalization in the infinite-horizon context as it ensures that self-control/commitment utility is measured in units of stationary consumption: $U(\bar{c}, \bar{c}, \dots) = \bar{c}$.[36] The term $(1-\delta)$ appears in Equation (12) as well to ensure that $\beta\delta$ is the Marginal Rate of Intertemporal Substitution (hence may be interpreted as the rate of time preference) in the no-self-control problem.

Under conditions we identify in the Appendix, the solution to the infinite-horizon problem is simple and mirrors the two-period case discussed above: when the decision maker is relatively poor, little is at stake financially and she optimally decides to forego self-control. This leads to over-consumption and under-saving, thus running down her wealth stock more. In turn, this makes self-control even less attractive in the future and the pattern repeats. As a consequence, wealth decreases and diminishes asymptotically. The decision maker is stuck in a poverty trap: Insufficient wealth makes self-control unattractive and lack of self-control leads to lower wealth (due to over-consumption). On the other hand, when decision makers are sufficiently wealthy, the stakes are high and self-control is optimal. In turn, this induces them to save enough to make self-control worthwhile in the future. As a marginal dollar of wealth is used sub-optimally when not self-controlling, the value function is kinked upwards at a critical wealth level. That is, there exist some $a > 0$ and $b_{SC} > b_{NSC} > 0$ such that

$$U(w) = \begin{cases} b_{NSC} \cdot w & \textit{if } w < \bar{w} \\ -a + b_{SC} \cdot w & \textit{if } w \geq \bar{w} \end{cases} \tag{13}$$

for $\bar{w} = \frac{a}{b_{SC} - b_{NSC}}$.

Here, a captures the cumulative cost of exerting self-control in the current and all future periods ($a = \frac{k}{1-R^{-1}}$, cf. Equation (20) in the Appendix). For $w > \bar{w}$, the marginal

[35] That is, in every period t, preferences over consumption c_t and continuation problems A_{t+1} given self-control (u_t) and given no self-control (v_t) are represented by: $u_t(c_t, A_{t+1}) = [(1-\delta)c_t^{\sigma} + \delta U_{t+1}(A_{t+1})^{\sigma}]^{\frac{1}{\sigma}}$ and $v_t(c_t, A_{t+1}) = [(1-\delta)c_t^{\sigma} + \beta\delta U_{t+1}(A_{t+1})^{\sigma}]^{\frac{1}{\sigma}}$.

[36] By slight abuse of notation, we write $U(\bar{c}, \bar{c}, \dots)$ to denote the utility derived from being committed to the stationary consumption stream $(\bar{c}, \bar{c}, \dots)$. More formally, let $A_{\bar{c}}$ denote the (degenerate) menu that commits to current consumption $\bar{c}$ and continuation problem $A_{\bar{c}}$. Then $U(A_{\bar{c}}) = [(1-\delta)c^{\sigma} + \delta U(A_{\bar{c}})^{\sigma}]^{\frac{1}{\sigma}}$; thus, $U(A_{\bar{c}}) = \bar{c} = U(\bar{c}, \bar{c}, \dots)$. Since there are no self-control decisions to be made, we have $U(\bar{c}, \bar{c}, \dots) = V(\bar{c}, \bar{c}, \dots) = W(\bar{c}, \bar{c}, \dots)$.

utility of wealth, b_{SC}, is equal to that of a standard decision maker without self-control problems.[37] The fact that the marginal utility of wealth is lower for poor decision makers ($b_{NSC} < b_{SC}$) reflects the fact that they over-consume (under-save) due to a lack of self-control.

While the kink introduces a non-differentiability at $w = \bar{w}$, U is differentiable everywhere else. Moreover, at $w = \bar{w}$ the right and left derivatives exist and are given by b_{SC} and b_{NSC} respectively. As $b_{SC} > b_{NSC}$, it is clear that $R(w - c) = \bar{w}$ for no solution c to Problems (10) and (12). Intuitively, for every c such that $R(w - c) = \bar{w}$, the fact that $b_{SC} > b_{NSC}$ creates a wedge between the marginal utility of an additional dollar saved, $\left[-(1-\delta)c^{\sigma-1} + \delta U(\bar{w})^{\sigma-1} b_{SC} R\right]((1-\delta)c^{\sigma} + \delta U(\bar{w})^{\sigma})$, and that of an additional dollar consumed, $\left[(1-\delta)c^{\sigma-1} - \delta U(\bar{w})^{\sigma-1} b_{NSC} R\right]((1-\delta)c^{\sigma} + \delta U(\bar{w})^{\sigma})$. As at least one of them must be positive, the decision maker is better off by consuming slightly more or less. In particular, this implies that any optimal solution c^* to Problems (10) and (12) satisfies a First-Order Condition (FOC). However, potentially, we may need to consider two candidate FOCs; one for $R(w-c) > \bar{w}$ and one for $R(w-c) < \bar{w}$. Thus, for current wealth levels close to $\bar{w}$, there are two candidate solutions; one for which the DM saves enough to reach wealth levels that allow for self-control in the future and one where she saves less and ends up with future wealth that entails optimally foregoing self-control. On the other hand, when w is small ($w \gtrapprox 0$) or very large ($w \gg \bar{w}$), one of the FOCs does not have a solution. Intuitively speaking, very poor decision makers never save enough to pass $\bar{w}$ tomorrow while very rich decision makers never consume so much so as to fall below $\bar{w}$.

To analyze this in more detail, note that the general FOC to Problems (10) and (12) is given by

$$(1-\beta)c^{\sigma-1} = \xi U(R(w-c))^{\sigma-1} U'(R(w-c))R$$

where $\xi = \delta$ (for Equation (10)) or $\xi = \beta\delta$ (for Equation (12)). Given Equation (13) for U, this results in the following two candidate FOCs

$$c^{\sigma-1} = \frac{\xi}{1-\delta}(b_{NSC} R(w-c))^{\sigma-1} b_{NSC} R \text{ where } R(w-c) < \bar{w} \tag{14}$$

and

$$c^{\sigma-1} = \frac{\xi}{1-\delta}(-a + b_{SC} R(w-c))^{\sigma-1} b_{SC} R \text{ where } R(w-c) > \bar{w}. \tag{15}$$

As the solution to (14) entails a future level of wealth in the No-Self-Control region of U, we denote its solution by using the subscript *NSC* (and, likewise, the subscript

[37] That is, the solution to the benchmark problem where $k = 0$ is given by $V(w) = b_{SC} \cdot w$.

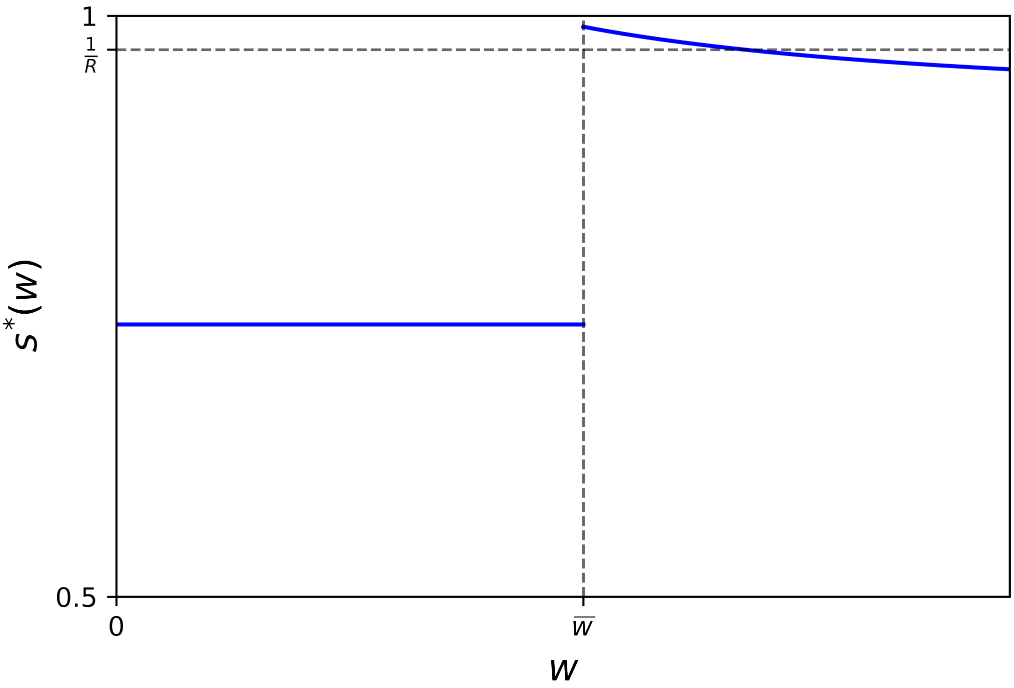

Figure 1: Optimal savings rate for $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$ and $R = 1.03$.

SC for the solution to (15)). Note, however that this refers to the *future* exertion of self-control. The solutions depend on the current exertion of self-control only through their dependence on the discount parameter ($\xi = \delta$ given self-control and $\xi = \beta\delta$ given no self-control). To make this explicit, we denote the candidate solutions to (10) and (12) by $c_{\xi,NSC}$ and $c_{\xi,SC}$.

Solving Equations (14) and (15) for c and noting that $1/(\sigma - 1) = -\gamma$, we obtain

$$c_{\xi,NSC} = \mu(\xi, b_{NSC}) \cdot w \text{ and } c_{\xi,SC} = \mu(\xi, b_{SC}) \cdot \left[-\frac{a}{b_{SC} R} + w. \right] \tag{16}$$

where, for all $0 < \xi \leq 1$ and all $b > 0$,

$$\mu(\xi, b) := \frac{(\frac{\xi}{1-\delta})^{-\gamma} (bR)^{1-\gamma}}{1 + (\frac{\xi}{1-\delta})^{-\gamma} (bR)^{1-\gamma}} \tag{17}$$

denotes the Marginal Propensity to Consume (MPC) depending on discount factor ξ and marginal utility of (future) wealth b.

As we show in the Appendix, the optimal consumption policy $c^*(w)$ is such that (i) the decision maker exerts self-control if and only if $w \geq \bar{w}$ (ii) under self-control, the solution is given by $c_{\delta,SC}$ (i.e. such that $R(w-c) > \bar{w}$) (iii) under no self-control, the solution is given by $c_{\beta\delta,NSC}$ (i.e. such that $R(w-c) < \bar{w}$). That is,

$$c^*(w) = \begin{cases} \mu(\beta\delta, b_{NSC}) \cdot w & \textit{if } w < \bar{w} \\ \mu(\delta, b_{SC}) \cdot \left(-\frac{a}{b_{SC}R} + w\right) & \textit{if } w \geq \bar{w} \end{cases}.$$

The optimal savings rate $s^*(w) = 1 - \frac{c^*(w)}{w}$ is given by

$$s^*(w) = \begin{cases} 1 - \mu(\beta\delta, b_{NSC}) & \text{if } w < \bar{w} \\ \underbrace{1 - \mu(\delta, b_{SC})}_{\substack{\text{savings rate} \\ \text{w/o SC problem}}} + \underbrace{\mu(\delta, b_{SC}) \frac{a}{b_{SC}R} \frac{1}{w}}_{\text{excess savings}} & \text{if } w \geq \bar{w}\ . \end{cases}$$

Note that the savings rate in the benchmark problem ($k = 0$) is $\frac{1}{R}(\delta R)^\gamma = 1 - \mu(\delta, b_{SC})$.[38] So the need to self-control creates excess savings. As the amount of additional savings is fixed, the savings rate is decreasing for $w \geq \bar{w}$ and asymptotes to the benchmark level as $w \to +\infty$. Figure 4 depicts the savings rate for our exemplary parameter combination.

As $1 - \mu(\delta, b_{NSC}) < 1 - \mu(\beta, b_{SC}) \leq \frac{1}{R}$ for $\delta R \leq 1$,[39] the model implies poverty traps for the poor. Decision makers whose initial wealth is low ($w < \bar{w}$) run down their wealth exponentially at rate $R(1 - \mu(\delta, b_{NSC})) < 1$. On the other hand, affluent decision makers ($w \geq \bar{w}$) are not be prone to asymptotically diminishing wealth. If $\delta R < 1$, their wealth converges to the steady state level

$$w_{SS} = \frac{R - (\delta R)^\gamma}{1 - (\delta R)^\gamma} \frac{a}{b_{SC}R}.$$

For our exemplary parameter combination Figure 4 illustrates the optimal wealth dynamics $w^*_{t+1}(w_t)$; Figure 4 depicts the optimal wealth paths for initial wealth levels below $\bar{w}$ (blue), above $\bar{w}$ but below w_{SS} (orange) and above w_{SS} (green).

In steady state, we have

$$c_{SS} = c^*(w_{SS}) = \frac{k}{b_{SC}R} \frac{R - (\delta R)^\gamma}{1 - (\delta R)^\gamma}.$$[40]

[38] As we show in the Appendix, we have $b_{SC} = (1-\delta)^{-\frac{\gamma}{1-\gamma}} \left[1 - R^{-1}(R\delta)^\gamma\right]^{\frac{1}{1-\gamma}}$. Consequently, $1 - \mu(\delta, b_{SC}) = \frac{1}{R}(\delta R)^\gamma$.

[39] See the derivation following Equation (25) in the Appendix.

[40] Thus, steady state consumption is increasing in the self-control cost k. However, note that, as $\bar{w} = \frac{k}{1-\frac{1}{R}} \frac{1}{b_{SC} - b_{NSC}}$ (and b_{SC}, b_{NSC} are independent of k), the initial wealth level required to converge to

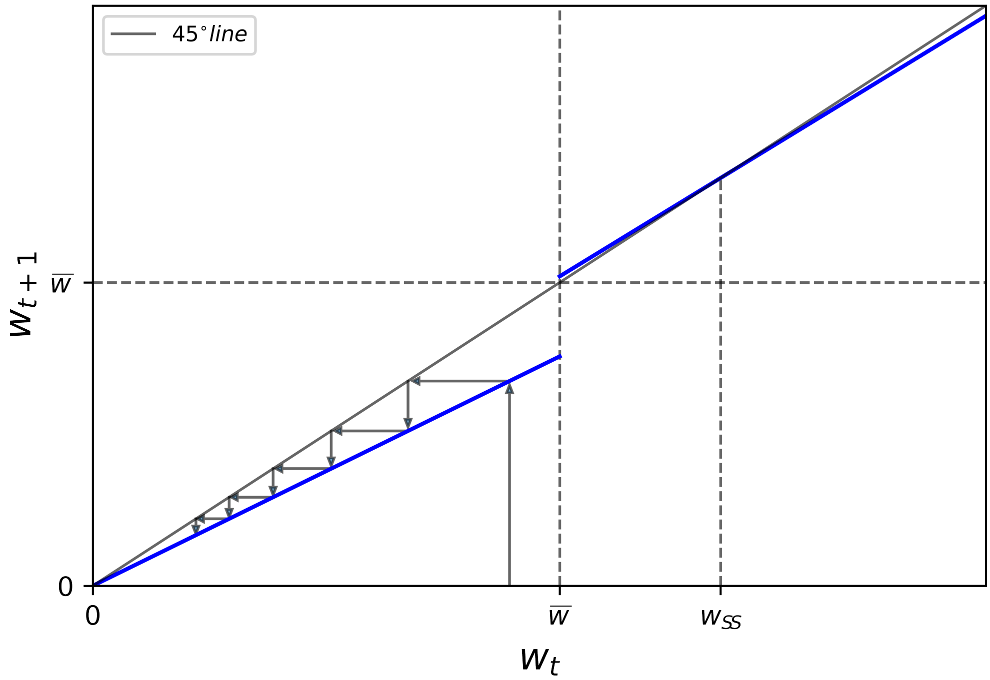

Figure 2: Wealth dynamics for $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$ and $R = 1.03$.

If $\delta R = 1$, affluent decision makers keep on accumulating wealth without converging to a finite steady state level. However, asymptotically, their wealth grows at a diminishing rate (i.e. the gross growth rate approaches 1 as $w \to +\infty$).

5. Conclusion

In this paper, we have shown three simple and intuitive Axioms on dynamic choice behavior in a simple two-period setting to be equivalent to a decision maker who is strategically Planning to Self-Control. Planning is optimal to the effect that, for every decision problem (menu), it maximizes planning-stage utility of the choice it induces, net of a planning cost. Planning to Self-Control is behaviorally equivalent to model with self-control costs that allows for more general specifications of the cost function that have not been consider in other axiomatic models of self-control. Under a particularly simple and intuitive such cost specification, a fixed cost, self-control is increasing in the stakes of the problem. When applied to intertemporal problems faced by a present-biased decision maker, this produces a magnitude effect: the empirical finding that reverse

this steady state increases as well.

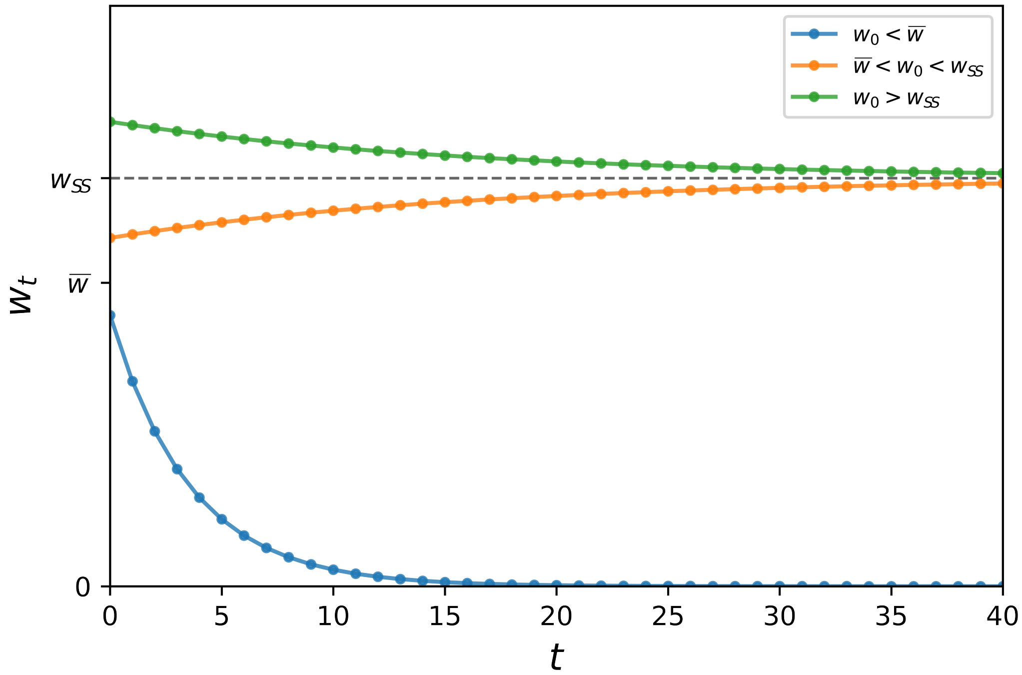

Figure 3: Wealth paths for $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$, $R = 1.03$ and three initial wealth levels w_0

their preferences for smaller sooner rewards vs. larger later ones when both rewards are scaled up. In a simple consumption-savings problem, increasing self-control means that self-control is exerted only by decision makers with sufficiently high wealth. The poor forego self-control and over-consume, thus running down their wealth even more. This results in a poverty trap.

References

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2013). Discounting behaviour and the magnitude effect: Evidence from a field experiment in denmark. *Economica*, 80:670–697.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102:3333–3356.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, 130:1067–1115.

Augenblick, N. and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, 86:941–975.

Ballard, I. C., Aydogan, G., Kim, B., and McClure, S. M. (2018). Causal evidence for the dependence of the magnitude effect on dorsolateral prefrontal cortex. *Nature: Scientific Reports*, 8.

Ballard, I. C., Kim, B., Liatsis, A., Aydogan, G., Cohen, J. D., and McClure, S. M. (2017). More is meaningful: The magnitude effect in intertemporal choice depends on self-control. *Psychological Science*, 28:1443–1454.

Baumeister, R. F., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74:1252–1265.

Baumeister, R. F., Hofmann, W., Summerville, A., Reiss, P. T., and Vohs, K. D. (2020). Everyday thoughts in time: Experience sampling studies of mental time travel. *Personality and Social Psychology Bulletin*, 46:1631–1648.

Baumeister, R. F., Vohs, K. D., and Oettingen, G. (2016). Pragmatic prospection: How and why people think about the future. *Review of General Psychology*, 20:3–16.

Baumeister, R. F., Vohs, K. D., and Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16:351–355.

Benhabib, J. and Bisin, A. (2005). Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption-saving decisions. *Games and Economic Behavior*, 52:460–492.

Benhabib, J., Bisin, A., and Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, 69:205–223.

Bermúdez, J. P., Murray, S., Chartrand, L., and Barbosa, S. (2023). What's inside is all that counts? the contours of everyday thinking about self-control. *Review of Philosophy and Psychology*, 14:33–55.

Boureau, Y.-L., Sokol-Hessner, P., and Daw, N. D. (2015). Deciding how to decide: Self-control and meta-decision making. *Trends in Cognitive Sciences*, 19:700–710.

Dekel, E., Lipman, B. L., and Rustichini, A. (2009). Temptation-driven preferences. *The Review of Economic Studies*, 76:937–971.

Duckworth, A. L., Gendler, T. S., and Gross, J. J. (2014). Self-control in school-age children. *Educational Psychologist*, 49:199–217.

Duckworth, A. L., Gendler, T. S., and Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science*, 11:35–55.

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., and Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, 13:538–539.

Frederick, S., Loewenstein, G., and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40:351–401.

Fudenberg, D. and Levine, D. K. (2006). A dual-self model of impulse control. *The American Economic Review*, 96:1449–1476.

Fudenberg, D. and Levine, D. K. (2012). Timing and self-control. *Econometrica*, 80:1–42.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *The American Psychologist*, 54:493–503.

Gollwitzer, P. M., Gawrilow, C., and Oettingen, G. (2010). The power of planning: Self-control by effective goal striving. In Hassin, R. R., Ochsner, K., and Trope, Y., editors, *Self Control in Society, Mind, and Brain*, pages 279–296. Oxford University Press.

Gollwitzer, P. M. and Oettingen, G. (2011). Planning promotes goal striving. In Vohs, K. D. and Baumeister, R. F., editors, *Handbook of Self-Regulation: Research, Theory, and Applications*, pages 162–185. Guilford Press.

Gollwitzer, P. M. and Sheeran, P. (2006). Implementation intentions and goal achievement. *Advances in Experimental Social Psychology*, 38:69–119.

Green, L., Myerson, J., and McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, 25:715–723.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69:1403–1435.

Gul, F. and Pesendorfer, W. (2004). Self-control and the theory of consumption. *Econometrica*, 72:119–158.

Gul, F. and Pesendorfer, W. (2005). The revealed preference theory of changing tastes. *Review of Economic Studies*, 72:429–448.

Gul, F. and Pesendorfer, W. (2006). The simple theory of temptation and self-control. *Working Paper*.

Hennecke, M. and Bürgler, S. (2020). Many roads lead to rome: Self-regulatory strategies and their effects on self-control. *Social and Personality Psychology Compass*, 14.

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

Inzlicht, M., Werner, K. M., Briskin, J. L., and Roberts, B. W. (2021). Integrating models of self-regulation. *Annual Review of Psychology*, 72:319–345.

Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, 126:54–70.

Koch, A. K. and Nafziger, J. (2011). Self-regulation through goal setting. *Scandinavian Journal of Economics*, 113:212–227.

Ludwig, R. M., Srivastava, S., and Berkman, E. T. (2018). Planfulness: A process-focused construct of individual differences in goal achievement. *Collabra: Psychology*, 4.

Lynch, J. G., Netemeyer, R. G., Spiller, S. A., and Zammit, A. (2010). A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *Journal of Consumer Research*, 37:108–128.

Masatlioglu, Y., Nakajima, D., and Ozdenoren, E. (2020). Willpower and compromise effect. *Theoretical Economics*, 15:279–317.

Masicampo, E. J. and Baumeister, R. F. (2011). Consider it done! plan making can eliminate the cognitive effects of unfulfilled goals. *Journal of Personality and Social Psychology*, 101:667–683.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience*, 27:5796–5804.

Meier, S. and Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2:193–210.

Meyer, A. G. (2015). The impacts of elicitation mechanism and reward size on estimated rates of time preference. *Journal of Behavioral and Experimental Economics*, 58:132–148.

Mischel, W. and Patterson, C. J. (1976). Substantive and structural elements of effective plans for self-control. *Journal of Personality and Social Psychology*, 34(5):942.

Nehring, K. (2006). Self-control through second-order preferences. *Working Paper*.

Noor, J. (2011). Intertemporal choice and the magnitude effect. *Games and Economic Behavior*, 72:255–270.

Noor, J. and Takeoka, N. (2010). Uphill self-control. *Theoretical Economics*, 5:127–158.

Noor, J. and Takeoka, N. (2015). Menu-dependent self-control. *Journal of Mathematical Economics*, 61:1–20.

Sjåstad, H. and Baumeister, R. F. (2018). The future and the will: Planning requires self-control, and ego depletion leads to planning aversion. *Journal of Experimental Social Psychology*, 76:127–141.

Stovall, J. E. (2010). Multiple temptations. *Econometrica*, 78:349–376.

Stovall, J. E. (2018). Temptation with uncertain normative preference. *Theoretical Economics*, 13:145–174.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23:165–180.

Sun, C. and Potters, J. (2022). Magnitude effect in intertemporal allocation tasks. *Experimental Economics*, 25:593–623.

Sutter, M., Kocher, M. G., and Glätzle-Rützler, D. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, 103:510–531.

Takeoka, N. (2008). Temptation, certainty effect, and diminishing self-control. *Working Paper*.

Taylor, S. E., Pham, L. B., Rivkin, I. D., and Armor, D. A. (1998). Harnessing the imagination: Mental simulation, self-regulation, and coping. *American Psychologist*, pages 429–439.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, 8:201–207.

Thaler, R. H. and Shefrin, H. M. (1981). An economic theory of self-control. *Source: Journal of Political Economy*, 89:392–406.

Appendix

A. Relation to Temptation Models of Self-Control

In the axiomatic treatment of self-control problems pioneered by Gul and Pesendorfer (2001) self-control problems result from temptations which must be resisted at the choice stage. When choosing some alternative $x \in A$, the decision maker incurs a self-control cost from foregoing the most tempting alternative in A as measured in terms of a temptation utility. Their model and two prominent generalizations thereof are special cases of our self-control-cost representation

$$U(A) = \max_{x \in A} u(x) - C(x, A) \tag{$\star\star$}$$

when (1) the choice-stage utility v is interpreted as temptation and (2) self-control costs depend only on maximal temptations in A (and the associated shortfall in temptation utility).

1. **Gul and Pesendorfer (2001)**
   develop the base case in which costs are given by the temptation utility foregone by not choosing the most tempting alternative:

   $$C(x, A) = \max_{y \in A} v(y) - v(x).$$

2. **Takeoka (2008); Noor and Takeoka (2010)**
   allow for marginal costs to be increasing. That is, they consider

   $$C(x, A) = \phi\left(\max_{y \in A} v(y) - v(x)\right)$$

   for some strictly increasing and *convex* function $\phi(\cdot) \geq 0$ with $\phi(0) = 0$.

3. **Noor and Takeoka (2015)**
   consider menu-dependent costs

   $$C(x, A) = \psi\left(\max_{y \in A} v(y)\right)\left(\max_{y \in A} v(y) - v(x)\right)$$

   for some (weakly) increasing $\psi(\cdot) \geq 0$ such that $\psi(l) > 0$ for all $l > \min_{x \in X} v(x)$.

We verify the properties of a self-control-cost function (cf. Definition 2) for $C(x, A)$ as defined above.

1. **Gul and Pesendorfer (2001)**
   As
   $$C(x, A) = \max_{y \in A} v(y) - v(x)$$
   is a special case of both Takeoka (2008); Noor and Takeoka (2010) and Noor and Takeoka (2015), the proof is included there.

2. **Takeoka (2008); Noor and Takeoka (2010)**
   Let
   $$C(x, A) = \phi\left(\max_{y \in A} v(y) - v(x)\right)$$
   for some strictly increasing and *convex* function $\phi(\cdot) \geq 0$ with $\phi(0) = 0$.

   a) Clearly, $\phi(\max_{y \in A} v(y) - v(x)) > 0$ only if $\max_{y \in A} v(y) - v(x) > 0$; thus, $v(y) > v(x)$ for some $y \in A$. Vice versa, if there exists some $y \in A$ such that $v(y) - v(x)$, then $\max_{y \in A} v(y) - v(x) > 0$. As ϕ is strictly increasing and $\phi(0) = 0$, we must have $C(x, A) = \phi(\max_{y \in A} v(y) - v(x)) > 0$.

   b) As ϕ is non-decreasing, we have $\phi(\max_{y \in A} v(y) - v(x)) \leq \phi(\max_{y \in A \cup B} v(y) - v(x))$.

   c) We show that $C(x, x \cup A) \leq C(y, A)$ whenever $v(x) > v(y)$ which is sufficient. Consider two cases. If $v(x) \geq \max_{z \in A} v(z)$, then $C(x, x \cup A) = \phi(0) = 0$ and the claim is obvious. Else, $\max_{z \in x \cup A} v(z) = \max_{z \in A} v(z)$. As ϕ is non-decreasing, $C(x, x \cup A) = \phi(\max_{z \in A} v(z) - v(x)) \leq \phi(\max_{z \in A} v(z) - v(y)) = C(y, A)$.

3. **Noor and Takeoka (2015)**
   Let
   $$C(x, A) = \psi\left(\max_{y \in A} v(y)\right)\left(\max_{y \in A} v(y) - v(x)\right)$$
   for some non-decreasing $\psi(\cdot) \geq 0$.

   a) Clearly, $\psi(\max_{y \in A} v(y))(\max_{y \in A} v(y) - v(x)) > 0$ only if $\max_{y \in A} v(y) - v(x) > 0$; thus, $v(y) > v(x)$ for some $y \in A$. Vice versa, if there exists some $y \in A$ such that $v(y) - v(x)$, then $\max_{y \in A} v(y) > v(x) \geq \min_{z \in X} v(z)$. Thus, $\psi(\max_{y \in A} v(y)) > 0$. Consequently, $C(x, A) = \psi(\max_{y \in A} v(y))(\max_{y \in A} v(y) - v(x)) > 0$.

b) As ψ is non-decreasing, it holds that $\psi(\max_{y\in A} v(y))(\max_{y\in A} v(y) - v(x)) \leq \psi(\max_{y\in A\cup B} v(y))(\max_{y\in A} v(y) - v(x)) \leq \psi(\max_{y\in A\cup B} v(y))(\max_{y\in A\cup B} v(y) - v(x))$.

c) We show that $C(x, x \cup A) \leq C(y, A)$ whenever $v(x) > v(y)$ which is sufficient. Consider two cases. If $v(x) \geq \max_{z\in A} v(z)$, then $C(x, x \cup A) = \psi(\max_{z\in A} v(z)) \cdot 0 = 0$ and the claim is obvious. Else, $\max_{z\in x\cup A} v(z) = \max_{z\in A} v(z)$. Thus, $C(x, x \cup A) = \psi(\max_{z\in A} v(z))(\max_{z\in A} v(z) - v(x)) \leq \psi(\max_{z\in A} v(z))(\max_{z\in A} v(z) - v(y)) = C(y, A)$.

Noor and Takeoka (2010) consider our more general representation ($\star\star$) in their introduction but give no axiomatic foundation for it. However, they do axiomatize what they call a 'general self-control' model where $C(x, A) = \tau(x, \max_{y\in A} v(y))$ for some $\tau(\cdot,\cdot)$ that is weakly increasing in its second argument and satisfies: (i) $\tau(x, v(y)) > 0 \implies v(x) < v(y)$; (ii) $[u(x) > u(y)$ and $v(x) < v(y)] \implies \tau(x, v(y)) > 0$. This does not define a self-control-cost function in the sense of Definition 2. While, like the models presented above, it is more specific than our model to the effect that it allows costs only to depend on the temptation maximum (instead of all temptations in menu A), it puts less restrictions on the structure of costs at any given menu. For example, it allows for the possibility that $\tau(x, \max_{z\in A} v(z)) < \tau(z, \max_{z\in A} v(z))$ for $\{x, y\} \in A$ even if $x \gtrless y$ (that is, if $v(x) < v(y)$ and $u(x) > u(y)$). Thus, the less tempting alternative x may incur a smaller self-control cost (even though $\max_{z\in A} v(z) - v(x) > \max_{z\in A} v(z) - v(y)$). We would argue that this is inconsistent with the intuition of self-control costs being caused by resisting temptation.

While sharing the axiomatic approach, the temptation models above differ from our work as they are developed in a lottery setting and impose the Independence Axiom.[41] We do not need to invoke the expected utility assumption. A notable exception in this regard is Gul and Pesendorfer (2006). They axiomatize a generic temptation model in a finite choice setting. Under a regularity condition, they derive a 'strict and generic' representation $U(A) = \zeta(\max_{x\in A} \omega(x), \max_{y\in A} v(y))$ where $\zeta(\cdot,\cdot)$ is strictly increasing in the first and strictly decreasing in the second argument. Thus, we can equivalently write $U(A) = \max_{x\in A} \zeta(\omega(x), \max_{y\in A} v(y))$. Again, this may be put in the form of ($\star\star$) by letting $C(x, A) = \zeta(\omega(x), v(x)) - \zeta(\omega(x), \max_{y\in A} v(y))$ and noting that $u(x) = \tau(\omega(x), v(x))$. As for the general model in Noor and Takeoka (2010), however, this does

[41] Further examples include Dekel et al. (2009) and Stovall (2010) who generalize G&P so as to include subjective uncertainty about temptations; Stovall (2018) additionally introduces subjective uncertainty about commitment utility. Since we model choice-stage behavior in terms of choice *functions*, our model does not incorporate uncertainty about choices.

not define a self-control-cost function (cf. Definition 2). Again, it is more specific than our model to the effect that costs may depend only on maximal temptations but less demanding on the cost structure at a given menu A such that – somewhat counter-intuitively – costs may be smaller for a less tempting alternative.[42]

Our self-control-cost representation ($\star\star$) lends itself to an indirect utility interpretation where menu A is evaluated by $U(A) = \max_{x\in A} u(x, A)$, the utility received from the best alternative according to $u(x, A) = u(x) - C(x, A)$. Gul and Pesendorfer (2001) point out that this allows to model a desire for commitment (and the underlying problems of self-control) without the need to invoke time-inconsistent preferences. Rather, self-control problems are explainable in terms of preferences being menu-dependent. The experienced utility of choosing x from some menu A (which presents the decision maker with self-control problems) is less than that of receiving x choice-free (i.e. from the commitment menu $\{x\}$). This creates a desire for commitment even if the decision maker's (menu-dependent) preferences remain unchanged as time passes.

In contrast, our PTSC representation ($\star$) has a natural interpretation in terms of dynamic inconsistency. We may think of the decision maker as if anticipating a change of preference from u to v. Yet, unlike for Strotz models (Strotz, 1955; Gul and Pesendorfer, 2005), PTSC agents may not just sit idly by but can constrain their own choices (made according to v) to conform to a previously made plan. As Theorem 2 above shows, this interpretation is indistinguishable from the time-consistent (indirect utility) version in terms of the behavioral observables in our model (menu preferences and choice from menus). Thus, temptation models of self-control are equally consistent with a dynamically inconsistent decision maker optimally planning to self-control. It is interesting to note, however, that a prediction that is implicit in our PTSC representation is that observed choices from menus should purely reflect choice-stage preferences (as represented by v) when the planning-stage is absent. In other words, choice behavior may systematically differ when decision makers have previously had the chance to plan (optimal self-control) as compared to a situation when they are not (no self-control). This would be irreconcilable with the alternative view of menu-dependent yet time-consistent preferences.

While most temptation models do not model choice from menus explicitly, the indirect utility interpretation for preference over menus $U(A) = \max_{x\in A} u(x, A)$ suggests choice according to $u(x, A)$. For example, in Gul and Pesendorfer (2001), where $u(x, A) = u(x)+v(x)-\max_{y\in A} v(y)$, the decision maker can be thought of as choosing according to

[42]Specifically, when $\{x, y\} \subseteq A$ such that $x \gtrless y$, we may have $C(x, A) = \zeta(\omega(x), v(x)) - \zeta(\omega(x), \max_{z\in A} v(z)) < \zeta(\omega(y), v(y)) - \zeta(\omega(y), \max_{z\in A} v(z)) = C(y, A)$ if $\omega(x) < \omega(y)$.

$u(x)+v(x)$ (a compromise between commitment and temptation utility). Thus, as choice is rationalizable by a standard (menu-independent) utility function, it satisfies WARP. The convex (Takeoka, 2008; Noor and Takeoka, 2010) and menu-dependent (Noor and Takeoka, 2015) models allow for violations of WARP. However, both generalizations only allow for cost specifications that feature cost differentials which are increasing in menu size. As a consequence, these models can produce failures of WARP related to decreasing self-control but cannot incorporate those connected to increasing self-control.

To make these statements precise, let $x \gtrless y$. We say that choice function $c(\cdot)$ exhibits *decreasing self-control* if there exist $A, B \in \mathcal{A}$ with $\{x, y\} \subseteq A$ such that $x = c(A)$ and $y = c(A \cup B)$. While x is chosen from menu A (using costly self-control), the decision maker loses self-control at the larger menu $A \cup B$. Note that this constitutes a violation of WARP (Sen's Condition α). Analogously, we say that $c(\cdot)$ exhibits *increasing self-control* if there exist $A, B \in \mathcal{A}$ with $\{x, y\} \subseteq A$ such that $y = c(A)$ and $x = c(A \cup B)$. Here, the decision maker gains self-control as she moves from A to the larger menu $A \cup B$. Again, this violates WARP. As we argue in the main text the latter violation is of particular behavioral interest as it reflects an exercise of self-control that is positively responsive to the *magnitude* or *stakes* of the decision problem. Intuitively, when there is little at stake, the returns to self-control might be too small to put up the effort. As the stakes increase, however, so do the benefits associated with self-control eventually making it worthwhile. As we write in the main text above, such a *magnitude* effect is empirically well established in intertemporal settings with choice between a smaller, sooner reward and a larger, later reward, for example.

To see that the aforementioned temptation models are unable to capture this, consider the cost differential for two alternatives such that $x \gtrless y$; that is, $u(x) > u(y)$ and $v(x) < v(y)$. For the base model in Gul and Pesendorfer (2001), it is given by $\max_{z \in A} v(z) - v(x) - (\max_{z \in A} v(z) - v(y)) = v(y) - v(x)$, a constant that is independent of the menu. Thus, either x is the better choice globally (if $u(x) - u(y) > v(y) - v(x)$) or else y is (hence no violation of WARP is possible). For the convex and menu-dependent model, the cost differential is $\phi(\max_{z \in A} v(z) - v(x)) - \phi(\max_{z \in A} v(z) - v(y))$ and $\psi(\max_{z \in A} v(z))(v(y) - v(x))$ respectively. Due to the convexity of ϕ and the non-decreasingness of ψ both increase when alternatives are added to menu A. In contrast, our definition of a self-control-cost function in representation ($\star\star$) allows for choices to exhibit increasing self-control. For a particularly simple example, consider the case of a fixed cost $k > 0$. We discuss this case in detail and consider several applications to intertemporal choice in Section 4 above. Importantly, a fixed of self-control are inconsistent with the temptation models above. Note that –considered as a function over foregone temptation utility– it is discontinuous at 0. This also introduces a non-convexity.

B. Proof of Theorem 1

B.1. Sufficiency of Axioms 0.1-3

We define for all $x, y \in X$: $x \geq y :\iff [x \gg y \text{ or } y \gtrless x \text{ or } x = y]$. Remember that $x \gg y :\iff x \sim \{x,y\} \succ y$ and $y \gtrless x :\iff y \succ \{x,y\} \succsim x$.

Lemma 1. *$\geq$ is a linear order on X.*

Proof. By definition, $\gg$ and $\gtrless$ are asymmetric and irreflexive. The identity relation $=$ is symmetric and reflexive. Thus, $x \succ y \iff [y \gtrless x \text{ or } x \gg y]$ and $x = y \iff x = y$ (where $\succ$ and $=$ denote the asymmetric and symmetric part of $\geq$). By consequence, $\geq$ is anti-symmetric.

As $\gg$, $\gtrless$ and $=$ are transitive, so is their disjunction $\geq$.

Lastly, to see completeness, let $x, y \in X$. If $x = y$, then $x \geq y$. Else, $x \neq y$. As the restriction of $\succ$ to singletons is a linear order, we have $x \succ y$ or $y \succ x$. W.l.o.g. consider the first case. By Axiom 2, $\{x,y\} \succsim y$. If $x = c(\{x,y\})$, then $x \succsim \{x,y\}$ by Axiom 1. Else if $y = c(\{x,y\})$, then $x \succ y \succsim \{x,y\}$ by Axiom 1. Thus, $x \succsim \{x,y\}$ in both cases. Consequently, we have $x \succsim \{x,y\} \succsim y$. As $x \succ y$, we have either $x \sim \{x,y\} \succ y$ or $x \succ \{x,y\} \succsim y$. Thus, $x \succ y$ or $y \succ x$. $\square$

As $\geq$ is a linear order on X, there exists a utility function $v : X \to \mathbb{R}$ representing it.

Lemma 2. *There exists a utility function $U : \mathcal{A} \to \mathbb{R}$ representing $\succsim$ such that for all $x, y \in X$ and all $B, C \in \mathcal{A}$ for which $y \gtrless x$, $y = c(y \cup C)$ and $B \subseteq C$ hold, we have $U(x) - U(x \cup B) \leq U(y) - U(y \cup C)$.*

Proof. As $\succsim$ is a weak order and $\mathcal{A}$ is finite, there exists some $\widetilde{U} : \mathcal{A} \to \mathbb{R}$ representing it.[43] Defining $U(A) = \exp(\gamma \cdot \widetilde{U}(A))$, we show that the desired property holds if $\gamma > 0$ is appropriately chosen. Note that U – it being a positive monotone transformation of $\widetilde{U}$ – represents $\succsim$.

Let $x, y \in X$ and $B, C \in \mathcal{A}$ be as stated above. By Axiom 1, $U(y) \geq U(y \cup C)$.

Consider two cases: (1) If $\widetilde{U}(x) - \widetilde{U}(x \cup B) \leq 0$, we have $U(x) - U(x \cup B) \leq 0 \leq U(y) - U(y \cup C)$ for every $\gamma > 0$. In this case, let $\gamma_{x,y,B,C} = 1 > 0$. (2) Else $\widetilde{U}(x) - \widetilde{U}(x \cup B) > 0$. By Axiom 2, $x \gtrless z$ for some $z \in B \subseteq C$; thus, $y \gtrless x \gtrless z$. By transitivity (Axiom 0.2), $y \gtrless z$. By Axiom 1, $\widetilde{U}(y \cup C) \leq \widetilde{U}(\{y, z\}) < \widetilde{U}(y)$. Consequently, there exists

[43]For example, let $U(A) = |WT(A)|$; where $WT(A) = \{B \in \mathcal{A} : A \succsim B\} \subseteq \mathcal{A}$.

some $\bar{U} \in \mathbb{R}$ such that $\widetilde{U}(y \cup C) < \bar{U} < \widetilde{U}(y)$ and $\widetilde{U}(x) < \bar{U}$. For every $\gamma > 0$, we have $U(y) - U(y \cup C) > \exp(\gamma\widetilde{U(y)}) - \exp(\gamma\bar{U})$. Moreover, strict monotonicity, convexity and continuous differentiability of $\exp(\gamma\cdot)$ imply $\exp(\gamma\widetilde{U(y)}) - \exp(\gamma\bar{U}) \geq \gamma\exp(\gamma\bar{U})(\widetilde{U}(y) - \bar{U})$ and $\gamma\exp(\gamma\widetilde{U}(x))(\widetilde{U}(x) - \widetilde{U}(x \cup B)) \geq U(x) - U(x \cup B)$. Thus, $U(y) - U(y \cup C) > U(x) - U(x \cup B)$ if

$$\gamma\exp(\gamma\bar{U})(\widetilde{U}(y) - \bar{U}) \geq \gamma\exp(\gamma\widetilde{U}(x))(\widetilde{U}(x) - \widetilde{U}(x \cup B))$$
$$\iff \gamma \geq \frac{1}{\bar{U} - \widetilde{U}(x)} \ln\left(\frac{\widetilde{U}(x) - \widetilde{U}(x \cup B)}{\widetilde{U}(y) - \bar{U}}\right).$$

Let $\gamma_{x,y,B,C} = \max\left\{\frac{1}{\bar{U}-\widetilde{U}(x)} \ln\left(\frac{\widetilde{U}(x)-\widetilde{U}(x\cup B)}{\widetilde{U}(y)-\bar{U}}\right), 1\right\} > 0$.

Thus, the desired property holds for every $\gamma \geq \max\limits_{x,y,B,C} \gamma_{x,y,B}, C > 0$ where x, y, B, C are as stated in the lemma. □

For every $x \in X$, define $u(x) = U(x)$ where U is as stated in Lemma 2. For all $P \in \mathcal{A}$, define $x_P = \max(P, v) = \operatorname{argmax}_{y \in P} v(y)$. We use Lemma 2 to define a planning-cost function.

Lemma 3. *Let $\succsim$ be represented by U as given through Lemma 2. For all $(P \subseteq A, A) \in \mathcal{A} \times \mathcal{A}$, let $\mathcal{Y}(P, A) = \{(y, C \supseteq \{z \in A \backslash P : v(z) > v(x_P)\}) \in X \times \mathcal{A} : [y \gtreqless x_P$ or $y = x_P]$ and $y = c(y \cup C)\}$ and define*

$$\kappa(P, A) = \begin{cases} K & \mathcal{Y}(P, A) = \emptyset \\ \min\limits_{(y,C) \in \mathcal{Y}(P,A)} (U(y) - U(y \cup C)) & \mathcal{Y}(P, A) \neq \emptyset \end{cases}$$

where $K > \max\limits_{A,B \in \mathcal{A}} |U(A) - U(B)| \geq 0$.

Then, κ is a planning-cost function.

Proof. Let $\emptyset \neq P \subseteq A \in \mathcal{A}$. Note that for all $(y, C) \in \mathcal{Y}(P, A)$: $U(y) - U(y \cup C) \geq 0$ by Axiom 1. Thus, $\kappa(P, A) \geq 0$. We verify the properties of a planning-cost function:

1. First, suppose that $\{z \in A \backslash P : v(z) > v(x_P)\} = \emptyset$. As $c(\{x_P\}) = x_P$, we have $(x_P, \emptyset) \in \mathcal{Y}(P, A)$. Hence $\kappa(P, A) \leq U(x_P) - U(x_P) = 0$. Second, suppose there exists some $z \in A \backslash P$ such that $v(z) > v(x_P)$. If $\mathcal{Y}(P, A) = \emptyset$, the claim is immediate. Else, consider any $(y, C) \in \mathcal{Y}(P, A)$. Note that $v(y) \leq v(x_P) < v(z)$. By Axiom 3, we cannot have $u(y) < u(z)$. Thus, $u(y) > u(z)$. Seeing that $y \gtreqless z \in B$, we have by Axiom 1: $U(y) > U(\{y, z\}) \geq U(y \cup C)$ hence

$U(y) - U(y \cup C) > 0$. As this holds for all $(y, C) \in \mathcal{Y}(P, A)$, we have $\kappa(A, P) = \min_{(y,C)\in\mathcal{Y}(P,A)}(U(y) - U(y \cup C)) > 0$.

2. If $x_P \in A$, then $x_{P\cap A} = x_P$. If $\mathcal{Y}(P, A \cup B) = \emptyset$, then $\kappa(P \cap A, A) \leq K = \kappa(P, A \cup B)$. Else, consider any $(y, C) \in \mathcal{Y}(P, A \cup B)$. Then, $y = x_P = x_{P\cap A}$ or $y \gtrless x_P = x_{P\cap A}$, $y = c(y \cup C)$ and $C \supseteq \{z \in (A \cup B)\backslash P : v(z) > v(x_P)\} \supseteq \{z \in A\backslash(P \cap A) : v(z) > v(x_{P\cap A})\}$ seeing that $A\backslash(P \cap A) = A\backslash P \subseteq (A \cup B)\backslash P$. Thus, $(y, C) \in \mathcal{Y}(P\cap A, A)$. Consequently, $\kappa(P\cap A, A) = \min_{(y,C)\in\mathcal{Y}(P\cap A,A)}(U(y)-U(y\cup C)) \leq \min_{(y,C)\in\mathcal{Y}(P,A\cup B)}(U(y) - U(y \cup C)) = \kappa(P, A \cup B)$.

3. Note that $(A \cup B)\backslash(P \cup B) = A\backslash B$.
   a) If $u(x_P) = u(x_{P\cup B})$, then $x_{P\cup B} = x_P$. We have $\mathcal{Y}(P \cup B, A \cup B) = \mathcal{Y}(P, A)$; thus, $\kappa(P \cup B, A \cup B) = \kappa(P, A)$. Else, $u(x_{P\cup B}) < u(x_P)$ implies $x_{P\cup B} \neq x_P$. Thus, we must have $x_P \gtrless x_{P\cup B}$. As $\{z \in (A \cup B)\backslash(P \cup B) : v(z) > v(x_{P\cup B})\} \subseteq \{z \in A\backslash P : v(z) > v(x_P)\}$, we have $\mathcal{Y}(P \cup B, A \cup B) \supseteq \mathcal{Y}(P, A)$. Consequently, $\kappa(P \cup B, A \cup B) = \min_{(y,C)\in\mathcal{Y}(P\cup B,A\cup B)} U(y) - U(y \cup B) \leq \min_{(y,C)\in\mathcal{Y}(P,A)} U(y) - U(y \cup B) = \kappa(P, A)$.
   b) As $u(x_{P\cup B}) > u(x_P)$, we have $x_{P\cup B} \neq x_P$. Consequently, $v(x_{P\cup B}) > v(x_P)$. If $\mathcal{Y}(A, P) = \emptyset$, then $\kappa(P\cup B, A\cup B) \leq K = \kappa(P, A)$; thus, $\kappa(P\cup B, A\cup B) - \kappa(P, A) \leq 0 < U(x_{P\cup B})-U(x_P)$. Else, let $(y^*, B^*) = \text{argmax}_{(y,B)\in\mathcal{Y}(P,A)} U(y) - U(y \cup B)$. We show that there exists some $(y', B') \in \mathcal{Y}(P \cup B, A \cup B)$ s.t. $U(y') - U(y' \cup B') - (U(y^*) - U(y^* \cup B^*)) \leq U(x_{P\cup B}) - U(x_P)$.

      Clearly, $v(x_{P\cup B}) > v(x_P) \geq v(y^*)$. If $y^* \gtrless x_{P\cup B}$, then $(y^*, B^*) \in \mathcal{Y}(P \cup B, A \cup B)$ and the claim holds trivially (noting that $U(x_{P\cup B}) - U(x_P) > 0$). Else, we have $x_{P\cup B} \gg y^*$. By Axiom 3, $U(x_{P\cup B}\cup y^*\cup B^*) > U(y^*\cup B^*)$; thus, $x_{P\cup B} = c(x_{P\cup B}\cup y^*\cup B^*)$ (by Axiom 1). Note that $U(x_{P\cup B}\cup y^*\cup B^*)-U(y^*\cup B^*) > 0$ implies that $U(x_{P\cup B}) - U(y^*) - (U(x_{P\cup B} \cup y^* \cup B^*) - U(y^* \cup B^*)) < U(x_{P\cup B}) - U(y^*) \leq U(x_{P\cup B}) - U(x_P)$. Hence the claim follows from letting $y' = x_{P\cup B}$ and $B' = y^* \cup B^*$.

      Lastly, the claim implies that $\min_{(y,B)\in\mathcal{Y}(P\cup B,A\cup B)}(U(y)-U(y\cup B))-(U(y^*)-U(y^*\cup B^*)) \leq U(y') - U(y' \cup B') - (U(y^*) - U(y^* \cup B^*)) < U(x_{P\cup B}) - U(x_P)$. That is, $\kappa(P \cup B, A \cup B) - \kappa(P, A) < U(x_{P\cup B}) - U(x_P)$.

□

Let κ be as defined in Lemma 3 and consider any $A \in \mathcal{A}$. Let $x^\star = c(A)$ and define $P^\star = \{y \in A : v(y) \leq v(x^\star)\}$. By definition, $x_{P^\star} = x^\star = c(A)$. As $(x^\star, A\backslash\{x^\star\}) \in$

$\mathcal{Y}(P^\star, A)$, we have $\kappa(P^\star, A) \leq U(x^\star) - U(A)$. Thus, $u(x_{P^\star}) - \kappa(P^\star, A) \geq U(A)$. Now consider any $P \subseteq A$ and any $(y, C) \in \mathcal{Y}(P, A)$. As $y = c(y \cup C)$ and $C \supseteq \{z \in A \backslash P : v(z) > v(x_P)\}$, Axiom 3 implies that $\{z \in A \backslash P : v(z) > v(x_P)\} = \{z \in A \backslash P : v(z) > v(x_P) \text{ and } u(z) < u(x_P)\}$. Using the fact that $y = x_P$ or $y \gtreqless x_P$, $y = c(y \cup C)$ and $C \supseteq \{z \in A \backslash P : x_P \gtreqless z\}$ together with Lemma 2: $U(x_P) - (U(y) - U(y \cup C)) \leq U(x_P) - (U(x_P) - U(x_P \cup \{z \in A \backslash P : x_P \gtreqless z\})) = U(x_P \cup \{z \in A \backslash P : x_P \gtreqless z\})$. Now Axiom 2 implies $U(x_P \cup \{z \in A \backslash P : x_P \gtreqless z\}) \leq U(A)$. We obtain $u(x_P) - \kappa(P, A) = U(x_P) - \min_{(y,C) \in \mathcal{Y}(P,A)} U(y) - U(y \cup C) \leq U(A)$. Hence $P^\star$ solves $(\star)$ and $U(A) = u(x_{P^\star}) - \kappa(P^\star, A) = \max_{P \subseteq A} u(x_P) - \kappa(P, A)$.

B.2. Necessity of Axioms 0.1-3

Suppose that $\succsim$ is represented by $U(A) = \max_{P \subseteq A} u(x_P) - \kappa(P, A)$ where κ is a planning-cost function and that $c(A) = x_{P^\star}$ for some solution $P^\star \subseteq A$. Note that $U(\{x\}) = u(x)$ for all $x \in X$. Thus, u represents the restriction of $\succsim$ to singletons; that is, $x \succsim y \iff u(x) \geq u(y)$. Note that the first property in the definition of a planning-cost function implies that $U(x) > U(\{x, y\}) \geq U(y) \iff [u(x) > u(y) \text{ and } v(x) < v(y)]$ as well as that $U(x) = U(\{x, y\}) > U(y) \iff [u(x) > u(y) \text{ and } v(x) > v(y)]$. Thus, $x \geq y \iff: [x \gg y \text{ or } y \gtreqless x \text{ or } x = y] \iff v(x) \geq v(y)$.

Axiom 0.1: $\succsim$ is represented by some utility function, therefore a weak order. As the restriction of $\succsim$ to singletons is represented by some strictly increasing $u : X \to \mathbb{R}$, it is a linear order.

Axiom 0.2: Let $x \gtreqless y$ and $y \gtreqless z$. Then $u(x) > u(y) > u(z)$ and $v(z) > v(y) > v(x)$. Hence $u(x) > u(z)$ and $v(z) > v(x)$; therefore, $x \gtreqless z$. Let $x \gg y$ and $y \gg z$. Then $u(x) > u(y) > u(z)$ and $v(x) > v(y) > v(z)$. Hence $u(x) > u(z)$ and $v(x) > v(z)$; therefore $x \gg z$.

Axiom 1: Let $c(A \cup B) = x \in A$. Suppose $P^\star \subseteq A \cup B$ solves $(\star)$ such that $x = x_{P^\star}$. Then, $x = x_{P^\star \cap A}$ and $P^\star \cap A \subseteq A$. By the second property of a planning-cost function: $\kappa(P^\star \cap A, A) = \kappa(P^\star \cap A, (A \cup B) \cap A) \leq \kappa(P^\star, A \cup B)$. Consequently, $U(A) \geq U(x) - \kappa(P^\star \cap A, A) \geq U(x) - \kappa(P^\star, A \cup B) = U(A \cup B)$.

Axiom 2: Let $P^\star$ solve $(\star)$ such that $x_{P^\star} = c(A)$ and assume that for all $y \in B$: $u(x) < u(y)$ or $v(x) > v(y)$ (note that we may assume w.l.o.g. that $x \notin B$). Define $P' = P^\star \cup B \subseteq A \cup B$. Then $x_{P'} = x_{P^\star}$ or $u(x_{P'}) > u(x_{P^\star})$. Using the third property of a planning-cost function, we have $U(A \cup B) \geq u(x_{P'}) - \kappa(P', A \cup B) = u(x_{P'}) - \kappa(P^\star \cup B, A \cup B) \geq$ [44] $u(x_{P^\star}) - \kappa(P^\star, A) = U(A)$.

[44] If $x_{P'} = x_{P^\star}$, we use part a) of the third property of a planning-cost function; if $u(x_{P'}) > u(x_{P^\star})$, we

Axiom 3: Let $P^\star \subseteq A$ solve ($\star$) such that $x_{P^\star} = c(A)$ and suppose $x \gg x_{P^\star}$, i.e. $u(x) > u(x_{P^\star})$ and $v(x) > v(x_{P^\star})$. Seeing that $x = \text{argmax}_{y \in x \cup P^\star} v(y)$, we have $U(x \cup A) \geq u(x) - \kappa(x \cup P^\star, x \cup A) > u(x_{P^\star}) - \kappa(P^\star, A) = U(A)$ (where the last strict inequality uses property 3.b) of planning-cost function κ).

C. Proof of Theorem 2

Let κ be a planning-cost function. For all $A \in \mathcal{A}$ and $x \in A$, let $P_x = \{y \in A : v(y) \leq v(x)\}$ and note that $x_{P_x} = x$. We define $C(x, A) = \kappa(P_x, A)$. We show that C is indeed a self-control-cost function:

1. We have $C(x, A) > 0 \iff \kappa(P_x, A) > 0 \iff \exists y \in A \backslash P_x : v(y) > v(x) \iff \exists y \in A : v(y) > v(x)$.

2. First, using $\{y \in A \cup B : v(y) > v(x)\} \subseteq P_x \cup B$: $C(x, A) = \kappa(P_x, A) \leq \kappa(\{y \in A \cup B : v(y) > v(x)\}, A \cup B) = C(x, A \cup B)$ (seeing that $P_x = \{y \in A \cup B : v(y) > v(x)\} \cap A$). Second, suppose that for all $y \in B : v(y) > v(x)$ (again, assume w.l.o.g. that $y \neq x$). Then, $C(x, A) = \kappa(P_x, A) = \kappa(P_x \cup B, A \cup B) = \kappa(\{y \in A \cup B : v(y) > v(x)\}, A \cup B) = C(x, A \cup B)$.

3. Let $y \in A$ and suppose $v(x) > v(y)$. If $u(x) < u(y)$, we have $C(y, A) = \kappa(P_y, A) \geq \kappa(x \cup P_y, x \cup A) \geq \kappa(\{z \in x \cup A : v(z) \leq v(x)\}, x \cup A) = C(x, x \cup A)$. Else if $u(x) > u(y)$, we have $C(x, x \cup A) - C(y, A) = \kappa(\{z \in x \cup A : v(z) \leq v(x)\}, x \cup A) - \kappa(P_y, A) \leq \kappa(x \cup P_y, x \cup A) - \kappa(P_y, A) < u(x) - u(y)$.

Suppose $P^\star$ solves $\max_{P \subseteq A} u(x_P) - \kappa(P, A)$ and let $x^\star = x_{P^\star}$. Assume that there exists some $y \in A$ such that $u(y) - C(y, A) > u(x^\star) - C(x^\star, A)$. Then $u(y) - \kappa(P_y, A) > u(x^\star) - \kappa(P^\star, A)$ contradicts optimality of $P^\star$. Vice versa, if $x^\star$ solves $\max_{x \in A} u(x) - C(x, A)$, let $P^\star = P_{x^\star}$. Assume that there exists some $P \subseteq A$ such that $u(x_P) - \kappa(P, A) > u(x^\star) - \kappa(P^\star, A)$. We must have $P \subseteq \{y \in A : v(y) \leq v(x_P)\} = P_{x_P}$. Thus, $u(x_P) - C(x_P, A) = u(x_P) - \kappa(P_{x_P}, A) \geq u(x_P) - \kappa(P, A) > u(x^\star) - \kappa(P_{x^\star}, A) = u(x^\star) - C(x^\star, A)$ contradicting optimality of $x^\star$.

Lastly, if $P^\star$ solves $\max_{P \subseteq A} u(x_P) - \kappa(P, A)$ and $x^\star = x_{P^\star}$, we have $\max_{P \subseteq A} u(x_P) - \kappa(P, A) = u(x^\star) - \kappa(P^\star, A) = u(x^\star) - C(x^\star, A) = \max_{x \in A} u(x) - C(x, A)$.

employ part b).

D. Characterizing When The Single-Kinked Value Function (13) is a Solution to Equations (9)-(12)

Below, we identify necessary and sufficient conditions for the value function to exhibit a single kink separating a self-control and no-self-control region in the infinite-horizon case (cf. Equation (13)). The kink makes it necessary to consider two candidate solutions when optimal future wealth is close to $\bar{w}$; in this case, $c_{\xi,NSC}$ and $c_{\xi,SC}$ simultaneously satisfy the FOCs (14) and (15). Yet when w is small (close to zero), optimal consumption is given by $c_{\xi,NSC}$; when w is large ($w \gg \bar{w}$) it is given by $c_{\xi,SC}$. This follows from the fact that in these cases only one of the FOCs has a solution. Moreover, a (corner) solution for which future wealth sits exactly at $\bar{w}$ is never optimal (see main text).

Moreover, using (16), note that indirect utilities

$$\begin{aligned}&\left[(1-\delta)c^{\sigma}_{\xi,NSC} + \delta(b_{NSC}R(w - c_{\xi,NSC}))^{\sigma}\right]^{\frac{1}{\sigma}}\\ =&\left[(1-\delta)\mu(\xi, b_{NSC})^{\sigma} + \delta(b_{NSC}R(1-\mu(\xi, b_{NSC}))^{\sigma}\right]^{\frac{1}{\sigma}} \cdot w\end{aligned}$$

and

$$\begin{aligned}&\left[(1-\delta)c^{\sigma}_{\xi,SC} + \delta(-a + b_{SC}R(w - c_{\xi,SC}))^{\sigma}\right]^{\frac{1}{\sigma}}\\ =&\left[(1-\delta)\mu(\xi, b_{SC})^{\sigma} + \delta(Rb_{SC}(1-\mu(\xi, b_{SC}))^{\sigma}\right]^{\frac{1}{\sigma}} \cdot \left[-\frac{a}{b_{SC}R} + w\right]\end{aligned}$$

are linearly increasing in w. At the same time, the cost of self-control is fixed at $k > 0$. Thus, optimal self-control behavior is such that when w is large (tends to $+\infty$), the (utility) benefits of self-control are large and eventually surpass k; when w is small (tends to 0), (utility) benefits of self-control are small and below k. Consequently, for small enough wealth levels w, the decision maker chooses not to self-control hence $U(w) = V(w)$. As the solution to the no-self-control problem (11) for small w is given by $c_{\delta,NSC}$, we must have

$$\begin{aligned}b_{NSC} \cdot w &= \left[(1-\delta)\mu(\beta\delta, b_{NSC})^{\sigma} + \delta(b_{NSC}R(1-\mu(\beta\delta, b_{NSC})))^{\sigma}\right]^{\frac{1}{\sigma}} \cdot w\\ \iff b_{NSC} &= \left[(1-\delta)\mu(\beta\delta, b_{NSC})^{\sigma} + \delta(b_{NSC}R)^{\sigma}(1-\mu(\beta\delta, b_{NSC}))^{\sigma}\right]^{\frac{1}{\sigma}}. \qquad (18)\end{aligned}$$

Analogously, if w is sufficiently large, self-control is worthwhile to the decision maker (hence $U(w) + k = W(w)$) and the solution to the self-control problem (10) is given by

$c_{\beta,SC}$; hence

$$k - a + b_{SC} \cdot w = \left[(1-\delta)\mu(\delta, b_{SC})^{\sigma} + \delta\left(b_{SC}R(1-\mu(\delta, b_{SC}))\right)^{\sigma}\right]^{\frac{1}{\sigma}} \cdot \left[-\frac{a}{b_{SC}R} + w\right].$$

Matching coefficients, we must have

$$b_{SC} = \left[(1-\delta)\mu(\delta, b_{SC})^{\sigma} + \delta(b_{SC}R)^{\sigma}(1-\mu(\delta, b_{SC}))^{\sigma}\right]^{\frac{1}{\sigma}} \tag{19}$$

and

$$k - a = -\frac{a}{R} \implies a = \frac{R}{R-1}k. \tag{20}$$

Using Equation (17) above, we observe that for all $0 < \xi \leq 1$ and $b > 0$:

$$\begin{aligned}
&(1-\beta)\mu(\xi, b)^{\sigma} + \xi(bR(1-\mu(\xi, b)))^{\sigma} \\
=&(1-\beta)\mu(\xi, b)^{\sigma}\left[1 + \frac{\xi}{1-\beta}(bR)^{\sigma}\left(\frac{1-\mu(\xi, b)}{\mu(\xi, b)}\right)^{\sigma}\right] \\
=&(1-\beta)\mu(\xi, b)^{\sigma}\left[1 + \left(\frac{\xi}{1-\beta}\right)^{\gamma}(bR)^{\gamma-1}\right] \\
=&(1-\beta)\mu(\xi, b)^{\sigma-1}.
\end{aligned} \tag{21}$$

We use this to simplify (19) further. As $\frac{\sigma-1}{\sigma} = \frac{1}{1-\gamma}$ and $\frac{1}{\sigma} = -\frac{\gamma}{1-\gamma}$, we obtain

$$b_{SC} = (1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\delta, b_{SC})^{\frac{1}{1-\gamma}}.$$

Thus,

$$b_{SC}^{1-\gamma} = (1-\delta)^{-\gamma}\frac{\left(\frac{\delta}{1-\delta}\right)^{-\gamma}(b_{SC}R)^{1-\gamma}}{1 + \left(\frac{\delta}{1-\delta}\right)^{-\gamma}(b_{SC}R)^{1-\gamma}}$$

hence

$$\beta^{\gamma}R^{\gamma-1} = 1 - \mu(\delta, b_{SC}). \tag{22}$$

Solving for b_{SC}, we obtain

$$\begin{aligned}
b_{SC} =&\frac{1}{R}\left(\frac{\delta}{1-\delta}\right)^{\frac{\gamma}{1-\gamma}}\left[R(R\delta)^{-\gamma} - 1\right]^{\frac{1}{1-\gamma}} \\
=&(1-\delta)^{-\frac{\gamma}{1-\gamma}}\left[1 - R^{-1}(R\delta)^{\gamma}\right]^{\frac{1}{1-\gamma}}.
\end{aligned} \tag{23}$$

Note that as long as $R > 1$ and $\delta R \leq 1$, we have $b_{SC} > 0$ for all $\gamma > 0$.

In similar fashion, we can rewrite (18) as

$$
\begin{aligned}
b_{NSC} &= (1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\beta\delta, b_{NSC}) \left[1 + \frac{\delta}{1-\delta}(b_{NSC}R)^{1-\frac{1}{\gamma}} \left(\frac{1-\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{NSC})}\right)^{1-\frac{1}{\gamma}}\right]^{-\frac{\gamma}{1-\gamma}} \\
&= (1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\beta\delta, b_{NSC}) \left[1 + \frac{\delta}{1-\delta}\left(\frac{\beta\delta}{1-\delta}\right)^{\gamma-1} (b_{NSC}R)^{\gamma-1}\right]^{-\frac{\gamma}{1-\gamma}} \\
&= (1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\beta\delta, b_{NSC}) \mu(\alpha, b_{NSC})^{\frac{\gamma}{1-\gamma}}
\end{aligned}
$$

where α is such that $\alpha := \beta^{1-\frac{1}{\gamma}}\delta$. Simplifying, we obtain

$$
1 = \delta^{-\gamma} R^{1-\gamma} \frac{1}{\left(1 + \left(\frac{\beta\delta}{1-\delta}\right)^{-\gamma} (b_{NSC}R)^{1-\gamma}\right)^{1-\gamma} \left(1 + \left(\frac{\alpha}{1-\delta}\right)^{-\gamma} (b_{NSC}R)^{1-\gamma}\right)^{\gamma}},
$$

thus

$$
(1-\mu(\beta\delta, b_{NSC}))^{1-\gamma}(1-\mu(\alpha, b_{NSC}))^{\gamma} = \delta^{\gamma} R^{\gamma-1}. \tag{24}
$$

Comparing (22) and (24), we note that

$$
(1-\mu(\delta, b_{SC})) = (1-\mu(\beta\delta, b_{NSC}))^{1-\gamma}(1-\mu(\alpha, b_{NSC}))^{\gamma}. \tag{25}
$$

As $\alpha = \beta^{-\frac{1}{\gamma}}\beta\delta > \delta$, we have $1-\mu(\beta\delta, b_{NSC}) < 1-\mu(\alpha, b_{NSC})$ (cf. Equation (17)). Thus, $1-\mu(\delta, b_{SC}) = (1-\mu(\beta\delta, b_{NSC}))\left(\frac{1-\mu(\alpha, b_{NSC})}{1-\mu(\beta\delta, b_{NSC})}\right)^{\gamma} > 1-\mu(\beta\delta, b_{NSC})$.

Optimal decision under no-self-control ($\beta\delta$-discounting)

Above, we identified two candidate solutions for problem (12): $c_{\beta\delta,NSC}$ and $c_{\beta\delta,SC}$. Note that $c_{\beta\delta,SC}$ is optimal if and only if

$$
\begin{aligned}
&[(1-\delta)(\mu(\beta\delta, b_{NSC}))^{\sigma} + \beta\delta(b_{NSC}R(1-\mu(\beta\delta, b_{NSC})))^{\sigma}]^{\frac{1}{\sigma}} \cdot w \\
\le\, &[(1-\delta)\mu(\beta\delta, b_{SC})^{\sigma} + \beta\delta(b_{SC}R(1-\mu(\beta\delta, b_{SC})))^{\sigma}]^{\frac{1}{\sigma}} \cdot \left[-\frac{a}{b_{SC}R} + w\right].
\end{aligned}
$$

Using (21), we obtain

$$\begin{aligned}
\frac{a}{b_{SC}R}\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}} &\leq \left[\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}} - \mu(\beta\delta, b_{NSC})^{\frac{1}{1-\gamma}}\right] \cdot w \\
\iff w &\geq \frac{a}{b_{SC}R}\frac{\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}}}{\mu(\beta\delta, b_{SC})^{\frac{1}{1-\gamma}} - \mu(\beta\delta, b_{NSC})^{\frac{1}{1-\gamma}}} \\
&= \frac{a}{b_{SC}R}\frac{1}{1-\left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})}\right)^{\frac{1}{1-\gamma}}} := \bar{w}_{\beta\delta}.
\end{aligned} \tag{26}$$

Seeing that $\mu(\beta\delta, b_{SC}) > \mu(\beta\delta, b_{NSC}) \iff \gamma < 1$, we note that $\bar{w}_{\beta\delta} > \frac{a}{b_{SC}R} > 0$.

Thus, we have

$$V(w) = \begin{cases} b_{NSC} \cdot w & \text{if } w < \bar{w}_{\beta\delta} \\ (1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta, b_{SC})\mu(\alpha, b_{SC})^{\frac{1}{1-\gamma}}\left[-\frac{a}{b_{SC}R} + w\right] & \text{if } w \geq \bar{w}_{\beta\delta} \end{cases}, \tag{27}$$

where $(1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta, b_{SC})\mu(\alpha, b_{SC})^{\frac{1}{1-\gamma}} > b_{NSC}$. Note that $\bar{w}_{\beta\delta}$ is the cut-off when discounting with $\beta\delta$. However, V evaluates the optimal $\beta\delta$-based decision by discounting with δ. Thus, V is discontinuous at $\bar{w}_{\beta\delta}$ with $V(\bar{w}_{\beta\delta}) > b_{NSC}\bar{w}_{\beta\delta}$.

Optimal decision under self-control (δ-discounting)

Analogous to the no-self-control case above, $c_{\delta,SC}$ is optimal in (10) if and only if

$$\begin{aligned}
&[(1-\delta)(\mu(\delta, b_{NSC}))^{\sigma} + \delta(b_{NSC}R(1-\mu(\delta, b_{NSC})))^{\sigma}]^{\frac{1}{\sigma}} \cdot w \\
\leq &[(1-\delta)\mu(\delta, b_{SC})^{\sigma} + \delta(b_{SC}R(1-\mu(\delta, b_{SC})))^{\sigma}]^{\frac{1}{\sigma}} \cdot \left[-\frac{a}{b_{SC}R} + w\right].
\end{aligned}$$

Using (21) and (19), we obtain

$$\begin{aligned}
\frac{a}{R} &\leq \left[b_{SC} - (1-\delta)^{\frac{\gamma}{\gamma-1}}\mu(\delta, b_{NSC})^{\frac{1}{1-\gamma}}\right] \cdot w \\
\iff w &\geq \frac{a}{b_{SC}R}\frac{\mu(\delta, b_{SC})^{\frac{1}{1-\gamma}}}{\mu(\delta, b_{SC})^{\frac{1}{1-\gamma}} - \mu(\delta, b_{NSC})^{\frac{1}{1-\gamma}}} \\
&= \frac{a}{b_{SC}R}\frac{1}{1-\left(\frac{\mu(\delta, b_{NSC})}{\mu(\delta, b_{SC})}\right)^{\frac{1}{1-\gamma}}} := \bar{w}_{\delta}.
\end{aligned} \tag{28}$$

Again, as $\mu(\delta, b_{SC}) > \mu(\delta, b_{NSC}) \iff \gamma < 1$, we note that $\bar{w}_{\delta} > \frac{a}{b_{SC}R} > 0$.

Thus, we have

$$W(w)-k=\begin{cases}-k+(1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\delta,b_{NSC})^{\frac{1}{1-\gamma}}\cdot w & \text{if } w\le \bar{w}_\delta\\ -a+b_{SC}\cdot w & \text{if } w>\bar{w}_\delta\end{cases}, \tag{29}$$

where $a>k$ and $(1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\delta,b_{NSC})^{\frac{1}{1-\gamma}}<b_{SC}$. Thus, W is kinked upwards at $\bar{w}_\delta$.

Note that

$$\begin{aligned}&\bar{w}_\delta\le\bar{w}_{\beta\delta}\\ \iff&\left(\frac{\mu(\delta,b_{NSC})}{\mu(\delta,b_{SC})}\right)^{\frac{1}{1-\gamma}}\le\left(\frac{\mu(\beta\delta,b_{NSC})}{\mu(\beta\delta,b_{SC})}\right)^{\frac{1}{1-\gamma}}.\end{aligned}$$

Moreover, for all $\xi\in(0,1)$:

$$\frac{\mu(\xi,b_{NSC})}{\mu(\xi,b_{SC})}=\left(\frac{b_{NSC}}{b_{SC}}\right)^{1-\gamma}\frac{1+\left(\frac{\xi}{1-\delta}\right)^{-\gamma}(b_{SC}R)^{1-\gamma}}{1+\left(\frac{\xi}{1-\delta}\right)^{-\gamma}(b_{NSC}R)^{1-\gamma}}. \tag{30}$$

As $b_{SC}>b_{NSC}$, this term is strictly decreasing in ξ when $\gamma<1$ and strictly increasing in ξ when $\gamma>1$. Consequently, $\left(\frac{\mu(\xi,b_{NSC})}{\mu(\xi,b_{SC})}\right)^{\frac{1}{1-\gamma}}$ is strictly decreasing in ξ and we have $\bar{w}_\delta<\bar{w}_{\beta\delta}$. Intuitively, as the decision maker places higher weight on the future when self-controlling, she will accrue enough savings to induce self-control tomorrow at a lower critical wealth level ($\bar{w}_\delta$) than when having no-self-control ($\bar{w}_{\beta\delta}$).

Remember that $\alpha:=\beta^{1-\frac{1}{\gamma}}\delta$. Thus, by construction, $\frac{\mu(\delta,b)}{1-\mu(\delta,b)}=\left(\frac{\mu(\alpha,b)}{1-\mu(\alpha,b)}\right)^{\gamma}\left(\frac{\mu(\beta\delta,b)}{1-\mu(\beta\delta,b)}\right)^{1-\gamma}$ for all $b>0$.

Lemma 4. *For all $b>0$, $0<\beta<1$ and $\gamma>0$, $\gamma\neq1$, we have $\mu(\delta,b)^{\frac{1}{1-\gamma}}>\mu(\beta\delta,b)\mu(\alpha,b)^{\frac{\gamma}{1-\gamma}}$.*

Proof. To begin with, note that for all real numbers $r>s>0$ and every $\lambda\in(0,1)$, we have $s<s^{1-\lambda}r^{\lambda}<r$.

For $0<\gamma<1$, we have $\alpha>\delta$. By definition of α, it holds $\delta=(\beta\delta)^{1-\gamma}\alpha^{\gamma}$. As $\ln\left(1+x^{-\gamma}(1-\delta)^{\gamma}(bR)^{1-\gamma}\right)$ is a strictly convex function in $\ln(x)$[45], we have

$$\left(1+\left(\frac{\beta\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{1-\gamma}\left(1+\left(\frac{\alpha}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{\gamma}>1+\left(\frac{\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}.$$

[45]Note that $\frac{d\ln\left(1+\exp(-\gamma\ln(x))(1-\delta)^{\gamma}(bR)^{1-\gamma}\right)}{d\ln(x)}=-\gamma\frac{\exp(-\gamma\ln(x))(1-\delta)^{\gamma}(bR)^{1-\gamma}}{1+\exp(-\gamma\ln(x))(1-\delta)^{\gamma}(bR)^{1-\gamma}}$ is strictly increasing in $\ln(x)$.

Seeing that

$$\mu(\delta,b)=\frac{1-\mu(\delta,b)}{(1-\mu(\beta\delta,b))^{1-\gamma}(1-\mu(\alpha,b))^{\gamma}}\mu(\beta\delta,b)^{1-\gamma}\mu(\alpha,b)^{\gamma}$$

and

$$\begin{aligned}\frac{1-\mu(\delta,b)}{(1-\mu(\beta\delta,b))^{1-\gamma}(1-\mu(\alpha,b))^{\gamma}} &= \frac{\left(1+\left(\frac{\beta\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{1-\gamma}\left(1+\left(\frac{\alpha}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}\right)^{\gamma}}{1+\left(\frac{\delta}{1-\delta}\right)^{-\gamma}(bR)^{1-\gamma}}\\ &> 1,\end{aligned}$$

we have $\mu(\delta,b)^{\frac{1}{1-\gamma}}>\mu(\beta\delta,b)\mu(\alpha,b)^{\frac{\gamma}{1-\gamma}}$.

For $\gamma>1$, we have $\beta\delta<\alpha=\beta^{1-\frac{1}{\gamma}}\delta<\delta$ and

$$\frac{\mu(\alpha,b)}{1-\mu(\alpha,b)}=\left(\frac{\mu(\delta,b)}{1-\mu(\delta,b)}\right)^{\frac{1}{\gamma}}\left(\frac{\mu(\beta\delta,b)}{a-\mu(\beta\delta,b)}\right)^{1-\frac{1}{\gamma}}.$$

Analogous to the above case, we obtain $\mu(\alpha,b)>\mu(\delta,b)^{\frac{1}{\gamma}}\mu(\beta\delta,b)^{1-\frac{1}{\gamma}}$. Thus, $\mu(\delta,b)^{\frac{1}{1-\gamma}}>\mu(\beta\delta,b)\mu(\alpha,b)^{\frac{\gamma}{1-\gamma}}$. □

Self-control vs. no self-control

U as given by Equation (13) is indeed a solution to the Bellman equations (9)-(11) if the decision maker optimally chooses to forego self-control for all $w<\bar{w}<\bar{w}_{\beta\delta}$ while self-controlling for all $w\geq\bar{w}\geq\bar{w}_{\delta}$. Given Equations (27) and (29) we derived for V and W above, this is the case if and only if (i) $\bar{w}_{\delta}\leq\bar{w}$ and (ii) $\bar{w}<\bar{w}_{\beta\delta}$ and $V(w)\leq -a+b_{SC}\cdot w$ for all $w\geq\bar{w}_{\beta\delta}$. To see this, note that – by definition – $\bar{w}_{\delta}\leq\bar{w}<\bar{w}_{\beta\delta}$ means that $W(w)-k\geq V(w) \iff w\geq\bar{w}$ for all $w\in[\bar{w}_{\delta},\bar{w}_{\beta\delta}]$. As $W(w)-k$ is linear with a single kink at $\bar{w}_{\delta}$ and $W(0)-k=-k<0=V(0)$, $\bar{w}_{\delta}\leq\bar{w}$ implies that $W(w)-k\leq V(w)$ for all $0\leq w\leq\bar{w}_{\delta}$ as well. Lastly, as $(1-\delta)^{-\frac{\gamma}{1-\gamma}}\mu(\beta\delta,b_{SC})\mu(\alpha,b_{SC})^{\frac{1}{1-\gamma}}<b_{SC}$, we have $-a+b_{SC}\cdot w\geq V(w)$ for all $w\geq\bar{w}_{\beta\delta}$ if and only if $-a+b_{SC}\cdot\bar{w}_{\beta\delta}\geq V(\bar{w}_{\beta\delta})$. Moreover, the latter can hold only if $\bar{w}<\bar{w}_{\beta\delta}$. Indeed, if $\bar{w}\geq\bar{w}_{\beta\delta}$, then $V(\bar{w}_{\beta\delta})>b_{NSC}\cdot\bar{w}_{\beta\delta}\geq -a+b_{SC}\cdot\bar{w}_{\beta\delta}$ seeing that V exhibits an upward jump at $w=\bar{w}_{\beta\delta}$, cf. Equation (27).

Consequently, Condition (ii) can be equivalently stated as $-a+b_{SC}\cdot\bar{w}_{\beta\delta}\geq V(\bar{w}_{\beta\delta})$.

Using Equations (23) and (26), that is,

$$-a + b_{SC} \cdot \bar{w}_{\beta\delta} \geq \left[-\frac{a}{b_{SC}R} + \bar{w}_{\beta\delta} \right] \left[(1-\delta)\mu(\beta\delta, b_{SC})^{\sigma} + \delta(1-\mu(\beta\delta, b_{SC}))^{\sigma}(Rb_{SC})^{\sigma} \right]^{\frac{1}{\sigma}}$$

$$\iff \left(-\frac{a}{b_{SC}R} \right) (1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\delta, b_{SC})^{\frac{1}{1-\gamma}} \left[R - \frac{1}{1 - \left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}} \right]$$

$$\geq \left(-\frac{a}{b_{SC}R} \right) \left(- \frac{\left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}}{1 - \left(\frac{\mu(\beta\delta, b_{NSC})}{\mu(\beta\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}} \right) (1-\delta)^{-\frac{\gamma}{1-\gamma}} \mu(\beta\delta, b_{SC}) \mu(\alpha, b_{SC})^{\frac{\gamma}{1-\gamma}}$$

$$\iff R - (R-1) \left(\frac{\mu(\beta\delta, b_{SC})}{\mu(\beta\delta, b_{NSC})} \right)^{\frac{1}{1-\gamma}} \geq \left(\frac{\mu(\beta\delta, b_{SC})^{1-\gamma} \mu(\alpha, b_{SC})^{\gamma}}{\mu(\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}.$$

Using Equation (30), Condition (i) can be expressed as

$$\frac{a}{b_{SC}R} \frac{1}{1 - \left(\frac{\mu(\delta, b_{NSC})}{\mu(\delta, b_{SC})} \right)^{\frac{1}{1-\gamma}}} \leq \frac{a}{b_{SC} - b_{NSC}}$$

$$\iff \frac{1}{R} \frac{1}{b_{SC} - b_{NSC} \left(\frac{1 + \left(\frac{\delta}{1-\delta} \right)^{-\gamma} (b_{SC}R)^{1-\gamma}}{1 + \left(\frac{\delta}{1-\delta} \right)^{-\gamma} (b_{NSC}R)^{1-\gamma}} \right)^{\frac{1}{1-\gamma}}} \leq \frac{1}{b_{SC} - b_{NSC}}.$$

Finally, we note that both conditions are satisfied for the parameter combination $\beta = 0.1$, $\delta = 0.9$, $\gamma = 0.8$, $k = 0.1$ and $R = 1.03$ considered in the main text for which $b_{SC} \approx 0.04774$, and $b_{NSC} \approx 0.00395$. Moreover, we have $a \approx 3.43333$ and $\bar{w} \approx 78.41362$ in this case.