NAIVETÉ ABOUT TEMPTATION AND SELF-CONTROL: FOUNDATIONS FOR NAIVE QUASI-HYPERBOLIC DISCOUNTING

By

David S. Ahn, Ryota Iijima, and Todd Sarver

August 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2099

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Naiveté about Temptation and Self-Control: Foundations for Naive Quasi-Hyperbolic Discounting[*]

David S. Ahn[†] Ryota Iijima[‡] Todd Sarver[§]

August 9, 2017

Abstract

We introduce and characterize a recursive model of dynamic choice that accommodates naiveté about present bias. The model incorporates costly self-control in the sense of Gul and Pesendorfer (2001) to overcome the technical hurdles of the Strotz representation. The important novel condition is an axiom for naiveté. We first introduce appropriate definitions of absolute and comparative naiveté for a simple two-period model, and explore their implications for the costly self-control model. We then extend this definition for infinite-horizon environments, and discuss some of the subtleties involved with the extension. Incorporating the definition of absolute naiveté as an axiom, we characterize a recursive representation of naive quasi-hyperbolic discounting with self-control for an individual who is jointly overoptimistic about her present-bias factor and her ability to resist instant gratification. We study the implications of our proposed comparison of naiveté for the parameters of the recursive representation. Finally, we discuss the obstacles that preclude more general notions of naiveté, and illuminate the impossibility of a definition that simultaneously incorporates both random choice and costly self-control.

KEYWORDS: Naive, sophisticated, self-control, quasi-hyperbolic discounting

[*] Ahn and Sarver acknowledge the financial support of the National Science Foundation through Grants SES-1357719 and SES-1357955. We also thank Yves Le Yaouanq, Pietro Ortoleva, Tomasz Strzalecki, and numerous seminar participants.

[†] Department of Economics, University of California, Berkeley, 530 Evans Hall #3880, Berkeley, CA 94720-3880. Email: dahn@econ.berkeley.edu

[‡] Yale University, Department of Economics, 30 Hillhouse Ave, New Haven, CT 06510. Email: ryota.iijima@yale.edu

[§] Duke University, Department of Economics, 213 Social Sciences/Box 90097, Durham, NC 27708. Email: todd.sarver@duke.edu.

1 Introduction

Naiveté about dynamically inconsistent behavior seems intuitively realistic and has important consequences for economic analysis. Behavioral models of agents with overoptimistic beliefs about their future decisions are now prevalent tools that feature across a variety of applications. Naiveté is an inherently dynamic phenomenon that implicates today's projections regarding future behavior. When the domain of choice is itself temporal, as in consumption over time, yet another layer of dynamics is introduced since naiveté then involves current assessments of future trade-offs.

Of course, complicated long-run dynamic problems are central to many economic settings that have nothing to do with naiveté. The standard approach to manageably analyze such problems is through a recursive representation of dynamic choice. The development of modern finance or macroeconomics seems unimaginable without the endemic recursive techniques that are now a standard part of the graduate curriculum. Despite the general importance of behavior over time in economics and its particular importance for applications of naiveté, a recursive dynamic model of a naive agent making choices over time remains outstanding. This paper remedies that gap, providing the appropriate environment and conditions to characterize a system of recursive equations that parsimoniously represents naive behavior over an infinite time horizon.

An immediate obstacle to developing a dynamic model of naiveté is that the ubiquitous Strotz model of dynamic inconsistency is poorly suited for recursive representations. Even assuming full sophistication, the Strotz model is well-known to be discontinuous and consequently ill-defined for environments with more than two periods of choice (Peleg and Yaari (1973); Gul and Pesendorfer (2005)).[1] This is because a Strotzian agent lacks any self-control to curb future impulses and therefore is highly sensitive to small changes in the characteristics of tempting options. Our approach instead follows Gul and Pesendorfer (2004), Noor (2011), and Krusell, Kuruşçu, and Smith (2010) in considering self-control in a dynamic environment. The moderating effects of even a small amount of self-control allows escape from the technical issues of the Strotz model. In addition to its methodological benefits, incorporating self-control into models of temptation has compelling substantive motivations per se, as argued in the seminal paper by Gul and Pesendorfer (2001). For methodological and substantive reasons, we employ the self-control model to represent dynamic naive choice.

An important foundational step in route to developing a recursive representation for naive agents is formulating appropriate behavioral definitions of naiveté. Our first order of business is to introduce definitions of absolute and comparative naiveté for individuals

[1] One workaround to finesse this impossibility is to restrict the set of decision problems and preferences parameters, e.g., by imposing lower bounds on risk aversion, the present-bias parameter, and uncertainty about future income (Harris and Laibson (2001)). We take a different approach in this paper.

who can exert costly self-control in the face of temptation. While definitions of absolute sophistication for self-control preferences have been proposed by Noor (2011) and definitions of absolute and comparative naiveté for Strotz preferences have been proposed by Ahn, Iijima, Le Yaouanq, and Sarver (2016),[2] no suitable definitions of naiveté for self-control currently exist.

We first explore these concepts in a simple two-stage environment with ex-ante rankings of menus and ex-post choice from menus to sharpen intuitions. We then proceed to extend these intuitions to infinite-horizon environments. We propose a system of equations to recursively represent naive quasi-hyperbolic discounting over time, building on earlier related recursive representations for fully sophisticated choice by Gul and Pesendorfer (2004) and Noor (2011). These equations capture an agent who is naive about both her present-bias and her ability to resist the impulse for immediate gratification. Incorporating an infinite-horizon version of our definition of absolute naiveté as an axiom, we provide a behavioral characterization of our proposed model. To our knowledge, this provides the first recursive model of dynamic naive choice. The model is applied to a simple consumption-saving problem to illustrate how naiveté influences consumption choice in the recursive environment.

We conclude by discussing the scope of our proposed definition of naiveté with self-control and its relationship to other proposals. We relate our definition to the definition of naiveté for consequentialist behavior proposed by Ahn, Iijima, Le Yaouanq, and Sarver (2016) and show that the two approaches are equivalent for deterministic Strotz preferences. However, we also argue for the impossibility of a comprehensive definition of naiveté that is suitable for both random choice and self-control: No definition can correctly accommodate both the deterministic self-control model and the random Strotz model, and no definition can accommodate random self-control.

2 Prelude: A Two-Stage Model

2.1 Primitives

To establish intuition, we commence our analysis with a two-stage model in this section before we proceed to the infinite-horizon recursive model in the next section. Let C denote a compact and metrizable space of outcomes and $\Delta(C)$ denote the set of lotteries (countably-additive Borel probability measures) over C, with typical elements $p, q, \ldots \in \Delta(C)$. Slightly abusing notation, we identify c with the degenerate lottery $\delta_c \in \Delta(C)$. Let $\mathcal{K}(\Delta(C))$ denote the family of nonempty compact subsets of $\Delta(C)$ with typical elements

[2] See also the recent theoretical analysis by Freeman (2016) that uses procrastination to uncover naiveté within Strotzian models of dynamic inconsistency.

$x, y, \ldots \in \mathcal{K}(\Delta(C))$. An expected-utility function is a continuous affine function $u : \Delta(C) \to \mathbb{R}$, that is, a continuous function such that, for all lotteries p and q, $u(\alpha p + (1-\alpha)q) = \alpha u(p) + (1-\alpha)u(q)$. Write $u \approx v$ when u and v are ordinally equivalent expected-utility functions, that is, when u is a positive affine transformation of v.

We study a pair of behavioral primitives that capture choice at two different points in time. The first is a preference relation $\succsim$ on $\mathcal{K}(\Delta(C))$. This ranking of menus is assumed to occur in the first period ("ex ante") before the direct experience of temptation but while (possibly incorrectly) anticipating its future occurrence. As such, it allows inferences about the individual's projection of her future behavior. The second is a choice correspondence $\mathcal{C} : \mathcal{K}(\Delta(C)) \rightrightarrows \Delta(C)$ with $\mathcal{C}(x) \subset x$ for all $x \in \mathcal{K}(\Delta(C))$. The behavior encoded in $\mathcal{C}$ occurs the second period ("ex post") and is taken while experiencing temptation.

These primitives are a special case of the domain used in Ahn, Iijima, Le Yaouanq, and Sarver (2016) to study naiveté without self-control and in Ahn and Sarver (2013) to study unforeseen contingencies.[3] The identification of naiveté and sophistication in our model relies crucially on observing both periods of choice data. Clearly, multiple stages of choice are required to identify time-inconsistent behavior. In addition, the ex-ante ranking of non-singleton option sets is required to elicit beliefs about future choice and hence to identify whether an individual is naive or sophisticated. This combination of ex-ante choice of option sets (or equivalently, commitments) and ex-post choice is therefore also common in the empirical literature that studies time inconsistency and naiveté.[4] Perhaps most closely related is a recent experiment by Toussaert (2016) that elicited ex-ante menu preferences and ex-post choices of the subjects and found evidence for the self-control model of Gul and Pesendorfer (2001).

2.2 Naiveté about Temptation with Self-Control

We introduce the following behavioral definitions of sophistication and naiveté that account for the possibility of costly self-control.

Definition 1 *An individual is* sophisticated *if, for all lotteries p and q with $\{p\} \succ \{q\}$,*

$$\mathcal{C}(\{p, q\}) = \{p\} \iff \{p, q\} \succ \{q\}.$$

[3] In these papers the second-stage choice is allowed to be random. While we feel this is an important consideration when there is uncertainty about future behavior, in this paper we restrict attention to deterministic choice in each period. This restriction is not solely for the sake of exposition: We argue in Section 4 that no definition of naiveté can satisfactorily accommodate both self-control and random choice.

[4] Examples include DellaVigna and Malmendier (2006); Shui and Ausubel (2005); Giné, Karlan, and Zinman (2010); Kaur, Kremer, and Mullainathan (2015); Augenblick, Niederle, and Sprenger (2015).

An individual is naive *if, for all lotteries p and q with $\{p\} \succ \{q\}$,*

$$\mathcal{C}(\{p,q\}) = \{p\} \implies \{p,q\} \succ \{q\}.$$

An individual is strictly naive *if she is naive and not sophisticated.*[5]

This definition of sophistication was introduced by Noor (2011, Axiom 7) and a similar condition was used by Kopylov (2012). To our knowledge, the definition of naiveté is new. Both definitions admit simple interpretations: An individual is sophisticated if she correctly anticipates her future choices and exhibits no unanticipated preference reversals, whereas a naive individual my have preference reversals that she fails to anticipate. More concretely, consider both sides of the required equivalence in the definition of sophistication. On the right, a strict preference for $\{p, q\}$ over $\{q\}$ reveals that the individual believes that she will choose the alternative p over q if given the option ex post. On the left, the ex-ante preferred option p is actually chosen. That is, her anticipated and actual choices align. A sophisticated individual correctly forecasts her future choices and therefore strictly prefers to add an ex-ante superior option p to the singleton menu $\{q\}$ if and only if it will be actually chosen over q ex post.

In contrast, a naive individual might exhibit the ranking $\{p, q\} \succ \{q\}$, indicating that she anticipates choosing the ex-ante preferred option p, yet ultimately choose q over p in the second period. Thus a naive individual may exhibit unanticipated preference reversals. However, our definition of naiveté still imposes some structure between believed and actual choices. Any time the individual will actually choose in a time-consistent manner ($\{p\} \succ \{q\}$ and $\mathcal{C}(\{p, q\}) = \{p\}$) she correctly predicts her consistent behavior; she does not anticipate preference reversals when there are none. Rather than permitting arbitrary incorrect beliefs for a naive individual, our definition is intended to capture the most pervasive form of naiveté that has been documented empirically and used in applications: underestimation of the future influence of temptation.[6]

Ahn, Iijima, Le Yaouanq, and Sarver (2016) proposed definitions of sophistication and naiveté for individuals who are consequentialist in the sense that they are indifferent between any two menus that share the same anticipated choices, as for example in the case of the Strotz model of changing tastes. Specifically, Ahn, Iijima, Le Yaouanq, and Sarver (2016) classify an individual as naive if $x \succsim \{p\}$ for all x and $p \in \mathcal{C}(x)$, and as sophisticated

[5] Definition 1 can be stated in terms of non-singleton menus. That is, an individual is *sophisticated* if for all menus x, y such that $\{p\} \succ \{q\}$ for all $p \in y$ and $q \in x$, $\mathcal{C}(x \cup y) \subset y \iff x \cup y \succ x$. An individual is *naive* if for all menus x, y such that $\{p\} \succ \{q\}$ for all $p \in y$ and $q \in x$, $\mathcal{C}(x \cup y) \subset y \implies x \cup y \succ x$.

[6] Our definition classifies an individual as naive if she makes *any* unanticipated preference reversals, which is sometimes also referred to as "partial naiveté" in the literature on time inconsistency. Some papers in this literature reserve the term "naive" for the case of complete ignorance of future time inconsistency. This extreme of complete naiveté is the special case of our definition where $\{p, q\} \succ \{q\}$ any time $\{p\} \succ \{q\}$.

if $x \sim \{p\}$ for all x and $p \in \mathcal{C}(x)$. In the presence of self-control, these conditions are too demanding. An individual who chooses salad over cake may still strictly prefer to go to a restaurant that does not serve dessert to avoid having to exercise self-control and defeat the temptation to eat cake. That is, costly self-control may decrease the value of a menu that contains tempting options so that $\{p\} \succ x$ for $p \in \mathcal{C}(x)$ is possible for a sophisticated, or even a naive, individual. Definition 1 instead investigates the marginal impact of making a new option p available in the ex-ante and ex-post stages. Section 4.1 formally analyzes the relationship between these two sets of definitions and shows that Definition 1 is applicable more broadly to preferences both with and without self-control.[7]

With the definition of absolute naiveté in hand, we can now address the comparison of naiveté across different individuals. Our approach is to compare the number of violations of sophistication: A more naive individual exhibits more unexpected preference reversals than a less naive individual.

Definition 2 *Individual 1 is* more naive *than individual 2 if, for all lotteries p and q,*

$$\big[\{p,q\} \succ_2 \{q\} \text{ and } \mathcal{C}_2(\{p,q\}) = \{q\}\big] \implies \big[\{p,q\} \succ_1 \{q\} \text{ and } \mathcal{C}_1(\{p,q\}) = \{q\}\big].$$

A more naive individual has more instances where she desires the addition of an option ex ante that ultimately goes unchosen ex post. Our interpretation of this condition is that any time individual 2 anticipates choosing the ex-ante superior alternative p over q (as reflected by $\{p,q\} \succ_2 \{q\}$) but in fact chooses q ex post, individual 1 makes the same incorrect prediction. Note that any individual is trivially more naive than a sophisticate: If individual 2 is sophisticated, then it is never the case that $\{p,q\} \succ_2 \{q\}$ and $\mathcal{C}_2(\{p,q\}) = \{q\}$; hence Definition 2 is vacuously satisfied.

As an application of these concepts, consider a two-stage version of the self-control representation of Gul and Pesendorfer (2001).

Definition 3 *A* self-control representation *of $(\succsim, \mathcal{C})$ is a triple $(u, v, \hat{v})$ of expected-utility functions such that the function $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ defined by*

$$U(x) = \max_{p \in x} \big[u(p) + \hat{v}(p)\big] - \max_{q \in x} \hat{v}(q)$$

represents $\succsim$ and

$$\mathcal{C}(x) = \operatorname*{argmax}_{p \in x}[u(p) + v(p)].$$

[7] However, the definitions proposed in Ahn, Iijima, Le Yaouanq, and Sarver (2016) have the advantage that they are readily extended to random choice driven by uncertain temptations, so long as the individual is consequentialist (exhibits no self-control).

The first function u reflects virtuous or normative utilities, for example how healthy different foods are. The second function $\hat{v}$ reflects how tempting the individual expects each options to be, for example how delicious different foods are. The interpretation is that the individual expects to maximize $u(p)$ minus the cost $[\max_{q \in x} \hat{v}(q) - \hat{v}(p)]$ of having to exert self-control to refrain from eating the most tempting option. She therefore anticipates choosing the option that maximizes the compromise $u(p) + \hat{v}(p)$ of the virtuous and (anticipated) temptation utility among the available options in menu x. The divergence between u and $u + \hat{v}$ captures the individual's perception of how temptation will influence her future choices. For a potentially naive individual, her actual ex-post choices are not necessarily those anticipated ex ante. Instead, the actual self-control cost associated with choosing p from the menu x is $[\max_{q \in x} v(q) - v(p)]$, where the actual temptation v can differ from anticipated temptation $\hat{v}$. The decision maker's ex-post choices are therefore governed by the utility function $u + v$ rather than $u + \hat{v}$.

The following definition offers a structured comparison of two utility functions w and w' and formalizes the a notion of greater congruence with the commitment utility u. Recall that $w \approx w'$ denotes ordinal equivalence of expected-utility functions, i.e., one is a positive affine transformation of the other.

Definition 4 *Let u, w, w' be expected-utility functions. Then w is* more u-aligned *than w', written as $w \gg_u w'$, if $w \approx \alpha u + (1 - \alpha) w'$ for some $\alpha \in [0, 1]$.*

We now provide a functional characterization of our absolute and comparative definitions of naiveté for the self-control representation. Our result begins with the assumption that the individual has a two-stage self-control representation, which is a natural starting point since the primitive axioms on choice that characterize this representation are already well established.[8] We say a pair $(\succsim, \mathcal{C})$ is *regular* if there exist lotteries p and q such that $\{p\} \succ \{q\}$ and $\mathcal{C}(\{p, q\}) = \{p\}$. Regularity excludes preferences where the choices resulting from actual temptation in the second period are exactly opposed to the commitment preference.

Theorem 1 *Suppose $(\succsim, \mathcal{C})$ is regular and has a self-control representation $(u, v, \hat{v})$. Then the individual is naive if and only if $u + \hat{v} \gg_u u + v$ (and is sophisticated if and only if $u + \hat{v} \approx u + v$).*

If the decision maker is naive, then she believes that her future choices will be closer to the virtuous ones. This overoptimism about virtuous future behavior corresponds to a

[8] Specifically, $(\succsim, \mathcal{C})$ has a (two-stage) self-control representation $(u, v, \hat{v})$ if and only if $\succsim$ satisfies the axioms of Gul and Pesendorfer (2001, Theorem 1) and $\mathcal{C}$ satisfies the weak axiom of revealed preference, continuity, and independence.

particular alignment of these utility functions:

$$u + \hat{v} \approx \alpha u + (1 - \alpha)(u + v).$$

The individual optimistically believes that her future choices will include an unwarranted weight on the virtuous preference u. Although the behavioral definition of naiveté permits incorrect beliefs, it does place some structure on the relationship between anticipated and actual choices. For example, it excludes situations like a consumer who thinks she will find sweets tempting when in fact she will be tempted by salty snacks. Excluding such orthogonally incorrect beliefs is essential in relating $\hat{v}$ to v and deriving some structure in applications.

Note that our behavioral definition of naiveté places restrictions on the utility functions $u + \hat{v}$ and $u + v$ governing anticipated and actual choices, but it does not apply directly to the alignment of the temptation utilities $\hat{v}$ and v themselves. This seems natural since our focus is on naiveté about the choices that result from temptation, not about when individuals are tempted per se. Example 1 below illustrates the distinction: It is possible for an individual to be overly optimistic about choice, as captured by $u + \hat{v} \gg_u u + v$, while simultaneously being overly pessimistic about how often she will be tempted, as captured by $v \gg_u \hat{v}$.

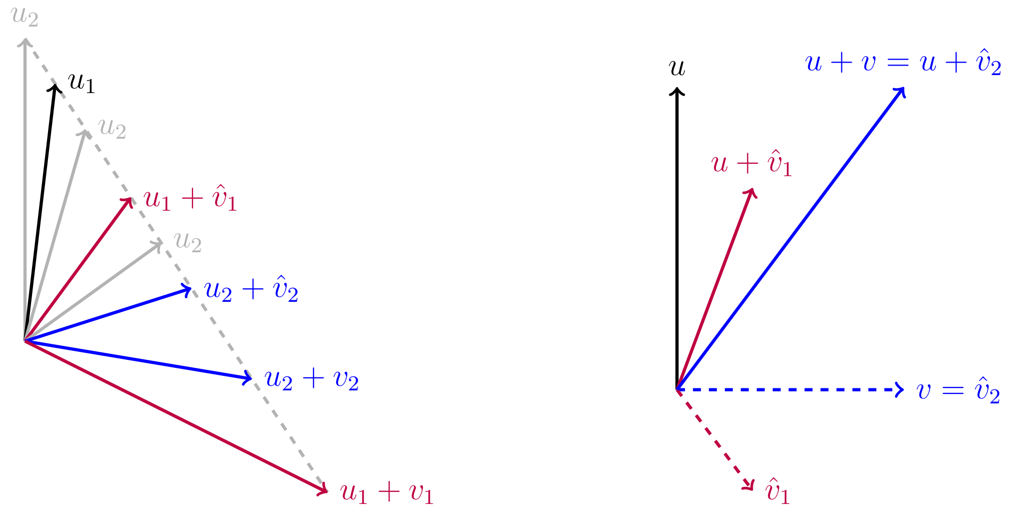
Our behavioral comparison of naiveté is necessary and sufficient for linear alignment of the actual and believed utilities across individuals. In particular, the more naive individual has a more optimistic view of her future behavior ($u_1 + \hat{v}_1 \gg_{u_1} u_2 + \hat{v}_2$), while actually making less virtuous choices ($u_2 + v_2 \gg_{u_1} u_1 + v_1$). We say $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ are *jointly regular* if there exist lotteries p and q such that $\{p\} \succ_i \{q\}$ and $\mathcal{C}_i(\{p, q\}) = \{p\}$ for $i = 1, 2$.

Theorem 2 *Suppose $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ are naive, jointly regular, and have self-control representations $(u_1, v_1, \hat{v}_1)$ and $(u_2, v_2, \hat{v}_2)$. Then individual 1 is more naive than individual 2 if and only if either*

$$u_1 + \hat{v}_1 \gg_{u_1} u_2 + \hat{v}_2 \gg_{u_1} u_2 + v_2 \gg_{u_1} u_1 + v_1,$$

or individual 2 is sophisticated ($u_2 + \hat{v}_2 \approx u_2 + v_2$).

Figure 1a illustrates the conditions in Theorems 1 and 2. Naiveté implies that, up to an affine transformation, the anticipated compromise between commitment and temptation utility $u_i + \hat{v}_i$ for each individual is a convex combination of the commitment utility u_i and the actual compromise utility $u_i + v_i$. Moreover, if individual 1 is more naive than individual 2, then the "wedge" between the believed and actual utilities governing choices, $u_i + \hat{v}_i$ and $u_i + v_i$, respectively, is smaller for individual 2. These relationships



(a) Theorem 2: Alignment of believed and actual utilities implied by comparative naiveté.

(b) Example 1: Individual 1 can be more naive than individual 2 even if $\hat{v}_2 \gg_u \hat{v}_1$ ($u_1 = u_2 = u$ and $v_1 = v_2 = v$).

Figure 1. Comparing naiveté

provide functional meaning to the statement that beliefs about the influence of temptation are more accurate for individual 2 than 1. Figure 1a also illustrates several different possible locations of u_2 relative to the other utility functions. There is some freedom in how the normative utilities of the two individuals are aligned, which permits meaningful comparisons of the degree of naiveté of individuals even when they do not have identical ex-ante commitment preferences.[9]

There is an obvious connection between the choices an individual anticipates making and her demand for commitment: If an individual anticipates choosing a less virtuous alternative from a menu, she will exhibit a preference for commitment. However, for self-control preferences, there will also be instances in which an individual desires commitment even though she anticipates choosing the most virtuous option in the menu. This occurs when she finds another option in the menu tempting, but expects to resist that temptation. Although our comparative measure concerns the relationship between the anticipated and actual choices by individuals, it does not impose restrictions on whether one individual or another is tempted more often. The following example illustrates the distinction.

Example 1 Fix any u and v that are not affine transformations of each other. Let $(u, v, \hat{v}_1)$ and $(u, v, \hat{v}_2)$ be self-control representations for the ex-ante preferences of indi-

[9] There are, of course, some restrictions on the relationship between u_1 and u_2 in Theorem 2. The assumption that $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ are jointly regular implies there exist lotteries p and q such that $u_i(p) > u_i(q)$ and $(u_i + v_i)(p) > (u_i + v_i)(q)$ for $i = 1, 2$. When individual 2 is strictly naive, this implies that u_2 lies in the arc between $-(u_1 + v_1)$ and $u_2 + \hat{v}_2$ in Figure 1a, which can be formalized as $u_2 + \hat{v}_2 \gg_{u_2} u_2 + v_2 \gg_{u_2} u_1 + v_1$.

viduals 1 and 2, respectively, where $\hat{v}_1 = (1/3)(v - u)$ and $\hat{v}_2 = v$. Then,

$$u + \hat{v}_1 = \frac{2}{3}u + \frac{1}{3}v \approx \frac{1}{2}u + \frac{1}{2}(u + v).$$

Since $\hat{v}_2 = v_2 = v_1 = v$, this implies that the condition in Theorem 2 is satisfied:

$$u + \hat{v}_1 \gg_u u + \hat{v}_2 = u + v_2 = u + v_1.$$

Thus the two individuals make the same ex-post choices, individual 2 is sophisticated, and individual 1 is naive. In particular, individual 1 is more naive than individual 2, even though her anticipated temptation utility diverges further from her commitment utility than that of individual 2, $\hat{v}_2 \gg_u \hat{v}_1$.[10] Figure 1b illustrates these commitment and temptation utilities. □

It is worthwhile to note that the self-control representation has been applied to a variety of settings, including habit formation, social preferences, and non-Bayesian belief updating.[11] Thus our results are also applicable to these specific settings to characterize the particular implications of absolute and comparative naiveté. While naiveté in self-control models has been relatively less explored in the literature, we are not the first study that formalizes it. The welfare effects of naiveté within a special case of the self-control representation were examined by Heidhues and Kőszegi (2009). In the next section, we illustrate the implications of our definitions for their proposed model.

2.3 Naiveté about the Cost of Exerting Self-Control

Heidhues and Kőszegi (2009) proposed the following special case of the self-control representation.

Definition 5 *A* Heidhues-Kőszegi representation *of* $(\succsim, \mathcal{C})$ *is tuple* $(u, \bar{v}, \gamma, \hat{\gamma})$ *of expected-utility functions* u *and* $\bar{v}$ *and scalars* $\gamma, \hat{\gamma} \geq 0$ *such that the function* $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ *defined by*

$$U(x) = \max_{p \in x} \big[u(p) + \hat{\gamma}\bar{v}(p)\big] - \max_{q \in x} \hat{\gamma}\bar{v}(q)$$

represents $\succsim$ *and*

$$\mathcal{C}(x) = \operatorname*{argmax}_{p \in x}[u(p) + \gamma\bar{v}(p)].$$

[10] Gul and Pesendorfer (2001, Theorem 8) characterized a comparative measure of preference for commitment. In the case where individuals 1 and 2 have the same commitment utility u, their results show that $\hat{v}_2 \gg_u \hat{v}_1$ if and only if individual 1 has greater preference for commitment than individual 2: That is, for any menu x, if there exists $y \subset x$ such that $y \succ_2 x$ then there exists $y' \subset x$ such that $y' \succ_1 x$. Their comparative measure could easily be applied in conjunction with ours to impose restrictions on both the relationship between $\hat{v}_1$ and $\hat{v}_2$ and the relationship between $u + \hat{v}_1$ and $u + \hat{v}_2$.

[11] Lipman and Pesendorfer (2013) provide a comprehensive survey.

The Heidhues-Kőszegi representation can be written as a self-control representation $(u, v, \hat{v})$ by taking $v = \gamma\bar{v}$ and $\hat{v} = \hat{\gamma}\bar{v}$. The interpretation of this representation is that the individual correctly anticipates which alternatives will be tempting but may incorrectly anticipate the magnitude of temptation and hence the cost of exerting self-control. Put differently, temptation may have a greater influence on future choice than the individual realizes, but she will not have any unexpected temptations.

The following proposition characterizes the Heidhues-Kőszegi representation within the class of two-stage self-control representations. We say that $\succsim$ *has no preference for commitment* if $\{p\} \succ \{q\}$ implies $\{p\} \sim \{p, q\}$.

Proposition 1 *Suppose $(\succsim, \mathcal{C})$ is has a self-control representation $(u, v, \hat{v})$, and suppose there exists some pair of lotteries p and q such that $\{p\} \sim \{p, q\} \succ \{q\}$. Then the following are equivalent:*

1. *Either $\succsim$ has no preference for commitment or, for any lotteries p and q,*

$$\{p\} \sim \{p, q\} \succ \{q\} \implies \mathcal{C}(\{p, q\}) = \{p\}.$$

2. *$(\succsim, \mathcal{C})$ has a Heidhues-Kőszegi representation $(u, \bar{v}, \gamma, \hat{\gamma})$.*

To interpret the behavioral condition in this proposition, recall that $\{p\} \sim \{p, q\} \succ \{q\}$ implies that q is *not* more tempting than p. In contrast, $\{p\} \succ \{p, q\} \succ \{q\}$ implies that q is more tempting than p but the individual anticipates exerting self-control and resisting this temptation. Condition 1 in Proposition 1 still permits preference reversals in the latter case, but rules out reversals in the former case. In other words, the individual may hold incorrect beliefs about how tempting an alternative is, but she will never end up choosing an alternative that she does not expect to find tempting at all.[12]

The implications of absolute and comparative naiveté for the Heidhues-Kőszegi representation follow as immediate corollaries of Theorems 1 and 2. To simplify the statement of the conditions in this result, we assume that the function $\bar{v}$ is *independent* of u, meaning it is not constant and it is not the case that $\bar{v} \approx u$. Note that this assumption is without loss of generality.[13]

Corollary 1 *Suppose $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ are jointly regular and have Heidhues-Kőszegi representations $(u, \bar{v}, \gamma_1, \hat{\gamma}_1)$ and $(u, \bar{v}, \gamma_2, \hat{\gamma}_2)$, where $\bar{v}$ is independent of u.*

[12] The exception is the case where $\succsim$ has no preference for commitment. In this case, the individual anticipates no temptation whatsoever ($\hat{\gamma} = 0$), yet may in fact be tempted ($\gamma > 0$).

[13] If $(u, \bar{v}, \gamma, \hat{\gamma})$ is a Heidhues-Kőszegi representation of $(\succsim, \mathcal{C})$ and $\bar{v}$ is not independent of u, there is an equivalent representation $(u, \bar{v}', 0, 0)$, where $\bar{v}'$ is an arbitrary non-constant function with $\bar{v}' \not\approx u$.

1. *Individual i is naive if and only if $\hat{\gamma}_i \leq \gamma_i$ (and is sophisticated if and only if $\hat{\gamma}_i = \gamma_i$).*
2. *When both individuals are naive, individual 1 is more naive than individual 2 if and only if either $\hat{\gamma}_1 \leq \hat{\gamma}_2 \leq \gamma_2 \leq \gamma_1$ or individual 2 is sophisticated ($\hat{\gamma}_2 = \gamma_2$).*

3 Infinite Horizon

3.1 Primitives

Now having some intuition gained from the two-period model, we consider a fully dynamic model with infinitely many discrete time periods. We represent the environment recursively. Let C be a compact metric space for consumption in each period. Gul and Pesendorfer (2004) prove there exists a space Z homeomorphic to $\mathcal{K}(\Delta(C \times Z))$, the family of compact subsets of $\Delta(C \times Z)$. Each menu $x \in Z$ represents a continuation problem. We study choices over $\Delta(C \times Z)$. For notational ease, we identify each degenerate lottery with its sure outcome, that is, we write (c, x) for the degenerate lottery $\delta_{(c,x)}$ returning (c, x) with probability one. To understand the domain, consider a deterministic $(c, x) \in C \times Z$. The first component c represents current consumption, while the second component $x \in Z$ represents a future continuation problem. Therefore preferences over (c, x) capture how the decision maker trades off immediate consumption against future flexibility.

At each period $t = 1, 2, \ldots$, the individual's behavior is summarized by a preference relation $\succsim_t$ on $\Delta(C \times Z)$.[14] The dependence of behavior on the date t allows for the possibility that sophistication can vary over time. In Sections 3.2, 3.3, and 3.4, we will study preferences that are time-invariant, so $p \succsim_t q \iff p \succsim_{t+1} q$. This implicitly assumes that sophistication and self-control are stationary. Stationarity is an understandably common assumption, as it allows for a fully recursive representation of behavior, which we believe will help the application of the model to financial and macroeconomic environments. In Section 3.5, we will relax stationarity to allow for increasing sophistication over time.

Note that imposing time-invariance of the preference relation does *not* assume dynamic consistency or sophistication. The structure of the recursive domain elicits both actual choices today *and* preferences over tomorrow's menus (through the second component Z of continuation problems), but imposes no relationship between them. There can be tension between today's choices and what the decision maker believes will be chosen tomorrow. For example, suppose $(c, \{p\}) \succ_t (c, \{q\})$. This means that p is a more virtuous than q because the consumer strictly prefers to commit to it for tomorrow, keeping today's

[14] Alternatively, we could take a choice correspondence as primitive and impose rationalizability as an axiom as in Noor (2011).

consumption constant. Moreover, if $(c, \{p, q\}) \succ_t (c, \{q\})$, then she believes she will select p over q tomorrow. Now suppose $q \succ_t p$, so the consumer succumbs to temptation and chooses q over p today. Then her beliefs about her future behavior do not align with her immediate choices. For stationary preferences, this also implies $q \succ_{t+1} p$ and hence the consumer exhibits an unanticipated preference reversal. This is exactly why the domain $\Delta(C \times Z)$ is the appropriate environment to study sophistication.

3.2 Stationary Quasi-Hyperbolic Discounting

Recall the self-control representation consists of normative utility U and a (perceived) temptation utility $\hat{V}$. With the dynamic structure, we can sharpen U and $\hat{V}$ into specific functional forms. In particular, we exclude static temptations over immediate consumption, like eating chocolate instead of salad, and make self-control purely dynamic. Temptation is only about the tradeoff between a better option today versus future opportunities.

As a foil for our suggested naive representation, we describe a self-control version of the (β, δ) quasi-hyperbolic discounting model of Gul and Pesendorfer (2005) and Krusell, Kuruşçu, and Smith (2010), which is a special case of a model characterized by Noor (2011).[15] As mentioned, the ability to construct well-defined recursive representations for this environment is an important advantage for the continuous self-control model over the Strotz model.

Definition 6 *A sophisticated quasi-hyperbolic discounting representation of $\{\succsim_t\}_{t\in\mathbb{N}}$ consists of continuous functions $u : C \to \mathbb{R}$ and $U, V : \Delta(C \times Z) \to \mathbb{R}$ satisfying the following system of equations:*

$$U(p) = \int_{C\times Z} (u(c) + \delta W(x))\, dp(c, x)$$
$$V(p) = \gamma \int_{C\times Z} (u(c) + \beta\delta W(x))\, dp(c, x)$$
$$W(x) = \max_{q\in x}(U(q) + V(q)) - \max_{q\in x} V(q)$$

and such that, for all $t \in \mathbb{N}$,

$$p \succsim_t q \iff U(p) + V(p) \geq U(q) + V(q),$$

where $0 \leq \beta \leq 1$, $0 < \delta < 1$, and $\gamma \geq 0$.

[15] This is a special case of what Noor (2011) refers to as "quasi-hyperbolic self-control" (see his Definition 2.2 and Theorems 4.5 and 4.6). He permits the static felicity function in the expression for V to be another function v and allows $\beta > 1$.

The tension between time periods in the quasi-hyperbolic self-control model is more transparent when we explicitly compute the choice that maximizes the utility $U + V$ for a family of deterministic consumption streams, where the only nontrivial flexibility is in the first period. Recall that

$$\begin{aligned} U(p) + V(p) &= \int_{C\times Z} \Big((1+\gamma)u(c) + (1+\gamma\beta)\delta W(x)\Big)\, dp(c,x) \\ &= (1+\gamma)\int_{C\times Z} \left(u(c) + \frac{1+\gamma\beta}{1+\gamma}\delta W(x)\right) dp(c,x). \end{aligned}$$

For a deterministic consumption stream $(c_t, c_{t+1}, \dots)$, the indirect utility is simple:

$$W(c_{t+1}, c_{t+2}, \dots) = U(c_{t+1}, c_{t+2}, \dots) = \sum_{i=1}^{\infty} \delta^{i-1} u(c_{t+i}).$$

Thus choice at period t for a deterministic consumption stream within a menu of such streams is made to maximize

$$u(c_t) + \frac{1+\gamma\beta}{1+\gamma}\sum_{i=1}^{\infty}\delta^i u(c_{t+i}). \tag{1}$$

The relationship between the self-control and Strozian models in the dynamic case is essentially similar to the two-period model, but with additional structure. The parameter γ measures the magnitude of the temptation for immediate consumption. As $\gamma \to \infty$, this model converges to the Strotzian version of the (β, δ) quasi-hyperbolic discounting with the same parameters.[16] However, there are technical difficulties in developing even sophisticated versions of Strotzian models with infinite horizons and nontrivial future choice problems, as observed by Peleg and Yaari (1973) and Gul and Pesendorfer (2005). While admitting the Strotz model as a limit case, the small perturbation to allow just a touch of self-control through a positive γ allows for recursive formulations and makes the self-control model amenable to application, e.g., Gul and Pesendorfer (2004) and Krusell, Kuruşçu, and Smith (2010). Alternate perturbations can also recover continuity, for example, Harris and Laibson (2013) introduce random duration of the "present" time period towards which the agent is tempted to transfer consumption.

Of course, the preceding model is fully sophisticated, so it cannot capture the effects of naiveté. We now introduce a recursive formulation of the $(\beta, \hat{\beta}, \delta)$ model of O'Donoghue and Rabin (2001). A leading application of the $(\beta, \hat{\beta}, \delta)$ model is procrastination on a single project like the decision to enroll in a 401(k). Such stopping problems are statistically

[16] In fact, when preferences are restricted to full commitment streams, Equation (1) shows that the observed choices of the quasi-hyperbolic self-control model over budget sets of consumption streams can be rationalized by a normalized quasi-hyperbolic Strotzian representation with present bias factor $\frac{1+\gamma\beta}{1+\gamma}$.

convenient because continuation values are trivial once the task is completed. On the other hand, many natural decisions are not stopping problems but perpetual ones, such as how much to contribute each period to the 401(k) after enrollment. To our knowledge, the $(\beta, \hat{\beta}, \delta)$ model has not yet been applied in recursive infinite-horizon settings, and we hope this model takes steps to bridge that gap.

Definition 7 *A* naive quasi-hyperbolic discounting representation *of* $\{\succsim_t\}_{t\in\mathbb{N}}$ *consists of continuous functions* $u : C \to \mathbb{R}$ *and* $U, \hat{V}, V : \Delta(C \times Z) \to \mathbb{R}$ *satisfying the following system of equations:*

$$\begin{aligned}
U(p) &= \int_{C\times Z} (u(c) + \delta \hat{W}(x))\, dp(c, x) \\
V(p) &= \gamma \int_{C\times Z} (u(c) + \beta\delta \hat{W}(x))\, dp(c, x) \\
\hat{V}(p) &= \hat{\gamma} \int_{C\times Z} (u(c) + \hat{\beta}\delta \hat{W}(x))\, dp(c, x) \\
\hat{W}(x) &= \max_{q\in x}(U(q) + \hat{V}(q)) - \max_{q\in x} \hat{V}(q)
\end{aligned}$$

and such that, for all $t \in \mathbb{N}$*,*

$$p \succsim_t q \iff U(p) + V(p) \geq U(q) + V(q),$$

where $\beta, \hat{\beta} \in [0, 1]$*,* $0 < \delta < 1$*, and* $\gamma, \hat{\gamma} \geq 0$ *satisfy*

$$\frac{1 + \hat{\gamma}\hat{\beta}}{1 + \hat{\gamma}} \geq \frac{1 + \gamma\beta}{1 + \gamma}.$$

In the basic two-stage model, naiveté is captured by the divergence between the anticipated temptation V realized in the second period and the temptation $\hat{V}$ anticipated in the first period. In the dynamic environment, $\hat{V}$ appears as a component of the continuation utility $\hat{W}$ while the actual temptation V is used to make today's choice. That is, the consumer believes tomorrow she will maximize $U + \hat{V}$ even while she chooses to maximize $U + V$ today. Moreover, in the dynamic setting the wedge between $\hat{V}$ and V is given a specialized parametric form as the difference between $\hat{\beta}$ and β. So all of the temptation and naiveté is purely temporal, rather than a result of static tastes.

We note that the values of γ and β are not individually identified, because they influence the individual's choice at the current period only through weighting instantaneous utility u and continuation payoff $\delta\hat{W}$ by $1+\gamma$ and $1+\gamma\beta$, respectively. Due to this lack of uniqueness, if a naive quasi-hyperbolic discounting representation exists, we can always find another equivalent representation with $\beta \leq \hat{\beta}$ and $\gamma \geq \hat{\gamma}$. In addition, the presence of the additional parameters γ and $\hat{\gamma}$ makes the parametric characterization of naiveté more subtle. That is, a simple comparison of β and $\hat{\beta}$ is insufficient to identify naiveté in this model because it does not control for naiveté regarding the intensity parameter γ.

3.3 Characterization

The naive version of the quasi-hyperbolic model is new, so its foundations are obviously outstanding. Related axiomatizations of sophisticated dynamic self-control do exist, e.g., Gul and Pesendorfer (2004) and Noor (2011), and we borrow some of their conditions. Recall that (c, x) refers to the degenerate lottery $\delta_{(c,x)}$. Mixtures of menus are defined pointwise: $\lambda x + (1 - \lambda)y = \{\lambda p + (1 - \lambda)q : p \in x, q \in y\}$. The first six axioms are standard in models of dynamic self-control and appear in Gul and Pesendorfer (2004) and Noor (2011).

Axiom 1 (Weak Order) $\succsim_t$ *is a complete and transitive binary relation.*

Axiom 2 (Continuity) *The sets* $\{p : p \succsim_t q\}$ *and* $\{p : q \succsim_t p\}$ *are closed.*

Axiom 3 (Independence) $p \succ_t q$ *implies* $\lambda p + (1 - \lambda)r \succ_t \lambda q + (1 - \lambda)r$*.*

Axiom 4 (Set Betweenness) $(c, x) \succsim_t (c, y)$ *implies* $(c, x) \succsim_t (c, x \cup y) \succsim_t (c, y)$*.*

Axiom 5 (Indifference to Timing) $\lambda(c, x) + (1 - \lambda)(c, y) \sim_t (c, \lambda x + (1 - \lambda)y)$*.*

Axiom 6 (Separability) $\frac{1}{2}(c, x)+\frac{1}{2}(c', y) \sim_t \frac{1}{2}(c, y)+\frac{1}{2}(c', x)$ *and* $(c'', \{\frac{1}{2}(c, x)+\frac{1}{2}(c', y)\}) \sim_t (c'', \{\frac{1}{2}(c, y) + \frac{1}{2}(c', x)\})$*.*

These first six axioms guarantee that preferences over continuation problems, defined by $(c, x) \succsim_t (c, y)$, can be represented by a self-control representation $(U_t, \hat{V}_t)$. For this section, we restrict attention to stationary preferences. The following stationarity axiom links behavior across time periods and implies the same $(U, \hat{V})$ can be used to represent preferences over continuation problems in every period.

Axiom 7 (Stationarity) $p \succsim_t q \iff p \succsim_{t+1} q$*.*

The next two axioms are novel and provide more structure on the temptation utility V. Before introducing them, some notation is required. For any $p \in \Delta(C \times Z)$, let p^1 denote the marginal distribution over C and p^2 denote the marginal distribution over Z. For any marginal distributions p^1 and q^2, let $p^1 \times q^2$ denote their product distribution.

In particular, $p^1 \times p^2$ is the measure that has the same marginals on C and Z as p, but removes any correlation between the two dimensions. The prior axioms make any correlation irrelevant, so $p \sim_t p^1 \times p^2$. Considering marginals is useful because it permits the replacement of a stream's marginal distribution over continuation problems, holding fixed the marginal distribution over current consumption.

Axiom 8 (Present Bias) *If $q \succ_t p$ and $(c, \{p\}) \succsim_t (c, \{q\})$, then $p \succ_t p^1 \times q^2$.*

In many dynamic models without present bias, an individual prefers p to q in the present if and only if she holds the same ranking when committing for some future period:

$$p \succsim_t q \iff (c, \{p\}) \succsim_t (c, \{q\}). \tag{2}$$

Clearly, this condition is would not be satisfied by an individual who is present biased, as the prototypical experiment on present bias finds preferences reversals occur with temporal distancing. Axiom 8 relaxes this condition: Equation (2) can be violated by preferring q to p today while preferring p to q when committing for the future, but only if q offers better immediate consumption and p offers better future consumption—this is the essence of present bias. Thus replacing the marginal distribution p^2 over continuation values with the marginal q^2 makes the lottery strictly worse, as formalized in our axiom.

The next axiom rules out temptations when there is no intertemporal tradeoff. As a consequence, all temptations involve rates of substitution across time, and do not involve static temptations at a single period.

Axiom 9 (No Temptation by Atemporal Choices) *If $p^1 = q^1$ or $p^2 = q^2$, then $(c, \{p, q\}) \succsim_t (c, \{p\})$.*

Correctly anticipating all future choices corresponds to the sophistication condition defined previously in Section 2.2. The following conditions directly apply the definitions for sophistication and naiveté introduced in the two-period model on the projection of preferences on future menus. Some subtleties do arise in extending the two-stage definitions of naiveté to general environments. In particular, the analog of a "commitment" consumption in an infinite horizon is not obvious, especially when considering a recursive representation. For example, the notion of a commitment as a singleton choice set in the subsequent period is arguably too weak in a recursive representation because such a choice set may still include nontrivial choices at later future dates. It fixes a single lottery over continuation problems in its second component Z, but leaves open what the choice from that period onward will be, since Z is itself just a parameterization of $\mathcal{K}(\Delta(C \times Z))$. Instead, the appropriate analog of a commitment should fully specify static consumption levels at all dates, that is, a commitment is an element of $\Delta(C^{\mathbb{N}})$. It is important to observe that $\Delta(C^{\mathbb{N}})$ is a strict subset of $\Delta(C \times Z)$.

The following definitions extend the concepts from the two-period model, substituting $\Delta(C^{\mathbb{N}})$ as a fully committed stream of consumption levels.

Axiom 10 (Sophistication) *For all $p, q \in \Delta(C^{\mathbb{N}})$ with $(c, \{p\}) \succ_t (c, \{q\})$,*

$$p \succ_{t+1} q \iff (c, \{p, q\}) \succ_t (c, \{q\}).$$

Axiom 11 (Naiveté) *For all $p, q \in \Delta(C^{\mathbb{N}})$ with $(c, \{p\}) \succ_t (c, \{q\})$,*

$$p \succ_{t+1} q \implies (c, \{p, q\}) \succ_t (c, \{q\}).$$

In words, if a virtuous alternative is chosen in the subsequent period, that choice was correctly anticipated, but the converse may not hold. The individual may incorrectly anticipate making a virtuous choice in the future.

In the two-period model, there is only one immediate future choice period. In the dynamic model, there are many periods beyond $t + 1$. Therefore, Axiom 11 may appear too weak because it only implicates conjectures at period t regarding choices in period $t+1$, but leaves open the possibility of naive conjectures regarding choices in some period $t + \tau$ with $\tau > 1$. However, the other axioms that are invoked in our representation will render these additional implications redundant. For example, consider the following, stronger definition of niaveté: For every $\tau \geq 1$ and $p, q \in \Delta(C^{\mathbb{N}})$,

$$(\underbrace{c, \ldots, c}_{\tau \text{ periods}}, \{p, q\}) \succ_t (\underbrace{c, \ldots, c}_{\tau \text{ periods}}, \{q\})$$

whenever

$$(\underbrace{c, \ldots, c}_{\tau \text{ periods}}, \{p\}) \succ_t (\underbrace{c, \ldots, c}_{\tau \text{ periods}}, \{q\}) \quad \text{and} \quad p \succ_{t+\tau} q.$$

Together with our other axioms, this stronger condition is implied by Axiom 11.

The following representation result characterizes sophisticated and naive stationary quasi-hyperbolic discounting. We say a profile of preference relations $\{\succsim_t\}_{t \in \mathbb{N}}$ is *nontrivial* if, for every $t \in \mathbb{N}$, there exist $c, c' \in C$ and $x \in Z$ such that $(c, x) \succ_t (c', x)$.

Theorem 3

1. *A profile of nontrivial relations $\{\succsim_t\}_{t \in \mathbb{N}}$ satisfies Axioms 1–10 if and only if it has a sophisticated quasi-hyperbolic discounting representation $(u, \gamma, \beta, \delta)$.*

2. *A profile of nontrivial relations $\{\succsim_t\}_{t\in\mathbb{N}}$ satisfies Axioms 1–9 and 11 if and only if it has a naive quasi-hyperbolic discounting representation $(u, \gamma, \hat{\gamma}, \beta, \hat{\beta}, \delta)$.*

3.4 Comparatives

We now study the comparison of naiveté in infinite-horizon settings. The following definition is an adaptation of our comparative from the two-period setting to the dynamic environment. Recalling the earlier intuition, a more naive individual today at period t has more instances where she incorrectly anticipates making a more virtuous choice tomorrow at period $t+1$ (captured by the relation $(c, \{p, q\}) \succ_t^1 (c, \{q\})$), while in reality she will make the less virtuous choice at $t+1$ (captured by the relation $q \succ_{t+1}^1 p$).

Definition 8 *Individual 1 is* more naive *than individual 2 if, for all $p, q \in \Delta(C^{\mathbb{N}})$,*

$$\big[(c, \{p, q\}) \succ_t^2 (c, \{q\}) \text{ and } q \succ_{t+1}^2 p\big] \implies \big[(c, \{p, q\}) \succ_t^1 (c, \{q\}) \text{ and } q \succ_{t+1}^1 p\big].$$

The following theorem characterizes comparative naiveté for individuals who have quasi-hyperbolic discounting representations. Recall that if individual 2 is sophisticated, i.e., $\frac{1+\hat{\gamma}^2\hat{\beta}^2}{1+\hat{\gamma}^2} = \frac{1+\gamma^2\beta^2}{1+\gamma^2}$, then individual 1 is trivially more naive. Otherwise, if individual 2 is strictly naive, then our comparative measure corresponds to a natural ordering of the present bias factors.

We say $\{\succsim_t^1\}_{t\in\mathbb{N}}$ and $\{\succsim_t^2\}_{t\in\mathbb{N}}$ are *jointly nontrivial* if, for every $t \in \mathbb{N}$, there exist $c, c' \in C$ and $x \in Z$ such that $(c, x) \succ_t^i (c', x)$ for $i = 1, 2$. Joint nontriviality ensures that both u^1 and u^2 are non-constant and that they agree on the ranking $u^i(c) > u^i(c')$ for some pair of consumption alternatives.

Theorem 4 *Suppose $\{\succsim_t^1\}_{t\in\mathbb{N}}$ and $\{\succsim_t^2\}_{t\in\mathbb{N}}$ are jointly nontrivial and admit naive quasi-hyperbolic discounting representations. Then individual 1 is more naive than individual 2 if and only if either individual 2 is sophisticated or $u^1 \approx u^2$, $\delta^1 = \delta^2$, and*

$$\frac{1+\hat{\gamma}^1\hat{\beta}^1}{1+\hat{\gamma}^1} \geq \frac{1+\hat{\gamma}^2\hat{\beta}^2}{1+\hat{\gamma}^2} \geq \frac{1+\gamma^2\beta^2}{1+\gamma^2} \geq \frac{1+\gamma^1\beta^1}{1+\gamma^1}.$$

3.5 Extension: Diminishing Naiveté

In this section we relax the stationarity assumption (Axiom 7) used in Theorem 3. There are many ways to formulate a non-stationary model, but motivated by recent research emphasizing individuals' learning about their self-control over time we consider the following representation.[17]

Definition 9 *A* quasi-hyperbolic discounting representation with diminishing naiveté *of $\{\succsim_t\}_{t\in\mathbb{N}}$ consists of continuous functions $u : C \to \mathbb{R}$ and $U_t, \hat{V}_t, V_t : \Delta(C \times Z) \to \mathbb{R}$ for each t satisfying the following system of equations:*

$$\begin{aligned}
U_t(p) &= \int_{C\times Z} (u(c) + \delta \hat{W}_t(x))\, dp(c,x) \\
V_t(p) &= \gamma \int_{C\times Z} (u(c) + \beta\delta \hat{W}_t(x))\, dp(c,x) \\
\hat{V}_t(p) &= \hat{\gamma}_t \int_{C\times Z} (u(c) + \hat{\beta}_t\delta \hat{W}_t(x))\, dp(c,x) \\
\hat{W}_t(x) &= \max_{q\in x}(U_t(q) + \hat{V}_t(q)) - \max_{q\in x} \hat{V}_t(q)
\end{aligned}$$

and such that, for all $t \in \mathbb{N}$,

$$p \succsim_t q \iff U_t(p) + V_t(p) \geq U_t(q) + V_t(q),$$

where $\beta, \hat{\beta}_t \in [0,1]$, $0 < \delta < 1$, and $\gamma, \hat{\gamma}_t \geq 0$ satisfy

$$\frac{1+\hat{\gamma}_t\hat{\beta}_t}{1+\hat{\gamma}_t} \geq \frac{1+\hat{\gamma}_{t+1}\hat{\beta}_{t+1}}{1+\hat{\gamma}_{t+1}} \geq \frac{1+\gamma\beta}{1+\gamma}.$$

In this formulation, the individual's anticipation updates to become more accurate over time, as expressed by the condition $\frac{1+\hat{\gamma}_t\hat{\beta}_t}{1+\hat{\gamma}_t} \geq \frac{1+\hat{\gamma}_{t+1}\hat{\beta}_{t+1}}{1+\hat{\gamma}_{t+1}} \geq \frac{1+\gamma\beta}{1+\gamma}$. One subtle epistemic consideration is the individual's view of her future updating, in addition to the attendant higher-order beliefs about how her future selves will anticipate future updating. This model suppresses these complications and takes the simplification that the individual is myopic about her future updating. She does not expect to actually revise her anticipation in future, since the continuation value function is used to evaluate the future problems. In other words, she is unaware of the possibility that her understanding can be misspecified.

The following axiom states that the individual's period-t self is more naive than her period-$(t+1)$ self, that is, she becomes progressively less naive about her future behavior over time.

Axiom 12 (Diminishing Naiveté) *For all $p, q \in \Delta(C^{\mathbb{N}})$,*

$$\big[(c,\{p,q\}) \succ_{t+1} (c,\{q\}) \text{ and } q \succ_{t+2} p\big] \implies \big[(c,\{p,q\}) \succ_t (c,\{q\}) \text{ and } q \succ_{t+1} p\big]$$

[17] Kaur, Kremer, and Mullainathan (2015) find evidence that sophistication about self-control improves over time. Ali (2011) analyzes a Bayesian individual who updates her belief about temptation strength over time.

We will focus in this section on preference profiles that maintain the same actual present bias over time. The only variation over time is in the increasing accuracy of beliefs about present bias in future periods.[18] We therefore impose the following stationarity axiom for preferences over commitment streams of consumption.

Axiom 13 (Commitment Stationarity) *For $p, q \in \Delta(C^{\mathbb{N}})$,*

$$p \succsim_t q \iff p \succsim_{t+1} q.$$

Relaxing Axiom 7 (Stationarity) and instead using Axioms 12 and 13, we obtain the following characterization result for the quasi-hyperbolic discounting model with diminishing naiveté.

Theorem 5 *A profile of nontrivial relations $\{\succsim_t\}_{t\in\mathbb{N}}$ satisfies Axioms 1–6, 8–9, and 11–13 if and only if it has a quasi-hyperbolic discounting representation with diminishing naiveté $(u, \gamma, \hat{\gamma}_t, \beta, \hat{\beta}_t, \delta)_{t\in\mathbb{N}}$.*

3.6 Application: Consumption-Saving Problem

As a simple exercise in the recursive environment, we apply our stationary naive quasi-hyperbolic discounting representation to a consumption-saving problem. The per-period consumption utility obeys constant relative risk aversion, that is,

$$u(c) = \begin{cases} \frac{c^{1-\sigma}}{1-\sigma} & \text{for } \sigma \neq 1 \\ \log c & \text{for } \sigma = 1, \end{cases}$$

where $\sigma > 0$ is the coefficient of relative risk aversion. Let $R > 0$ denote the gross interest rate.

Slightly abusing notation, let $\hat{W}(m)$ denote the anticipated continuation value as a function of wealth $m \geq 0$. It obeys

$$\hat{W}(m) = \max_{\hat{c}\in[0,m]} \left[(1+\hat{\gamma})u(\hat{c}) + \delta(1+\hat{\gamma}\hat{\beta})\hat{W}(R(m-\hat{c}))\right] \\ - \hat{\gamma} \max_{\tilde{c}\in[0,m]} \left[u(\tilde{c}) + \delta\hat{\beta}\hat{W}(R(m-\tilde{c}))\right]. \quad (3)$$

[18] More general representations are also possible. In the proof of Theorems 4 and 5 in Appendix A.5, we first characterize a more general representation in Proposition 5 in which both actual and anticipated present bias can vary over time.

The consumption choice at m is given by

$$c(m) \in \operatorname*{argmax}_{c\in[0,m]} \left[u(c) + \delta \frac{1+\gamma\beta}{1+\gamma} \hat{W}(R(m-c)) \right].$$

In the proposition below we focus on a solution in which the value function takes the same isoelastic form as u. We do not know whether there exist solutions that do not have this form. However, the restriction seems natural in this exercise, since the solution of this form is uniquely optimal under the benchmark case of exponential discounting (i.e., $\frac{1+\gamma\beta}{1+\gamma} = \frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}} = 1$).

Proposition 2 *Assume that $(1+\hat{\gamma}\hat{\beta})\delta R^{1-\sigma} < 1$.*[19] *Then there exist unique $A > 0$ and $B \in \mathbb{R}$ such that*

$$\hat{W}(m) = Au(m) + B$$

is a solution to Equation (3). Moreover, the optimal policy c for this value function satisfies $c(m) = \lambda m$ for some $\lambda \in (0,1)$, and:

1. *If $\sigma < 1$, then A is increasing and λ is decreasing in $\hat{\beta}$.*
2. *If $\sigma = 1$, then A and λ are constant in $\hat{\beta}$.*
3. *If $\sigma > 1$, then A is decreasing and λ is increasing in $\hat{\beta}$.*

In all cases, λ is decreasing in β.

While increasing β always leads to a lower current consumption level $c(m)$, the effect of increasing $\hat{\beta}$ depends on the value of σ. As an analogy, it is worthwhile to point out that increasing $\hat{\beta}$ leads to the same implication as increasing the interest rate R. Recall that, under standard exponential discounting, as R becomes higher, the current consumption increases if $\sigma > 1$, is constant if $\sigma = 1$, and decreases if $\sigma < 1$. This is because a higher interest rate implies two conflicting forces: The first is the intertemporal substitution effect that makes the current consumption lower, and the second is the income effect that raises the current consumption. The first effect dominates when the intertemporal elasticity of substitution $1/\sigma$ is higher than 1, and the second effect dominates if $1/\sigma$ is less than 1.

[19] This assumption is used to guarantee the unique existence of a solution.

4 Connections and Impossibilities

4.1 Relating the Strotz and Self-Control Naiveté Conditions

Ahn, Iijima, Le Yaouanq, and Sarver (2016) consider naiveté in a class of Strotz preferences where the individual always maximizes the temptation utility v in the ex-post stage, rather than maximizing $u+v$ as in the self-control model. The following is a version of the two-stage Strotz model that is adapted to our deterministic choice correspondence domain. For any expected-utility function w, let $B_w(x)$ denote the set of w-maximizers in x, that is, $B_w(x) = \operatorname{argmax}_{p \in x} w(p)$.

Definition 10 *A* Strotz representation *of* $(\succsim, \mathcal{C})$ *is a triple* $(u, v, \hat{v})$ *of expected-utility functions such that the function* $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ *defined by*

$$U(x) = \max_{p \in B_{\hat{v}}(x)} u(p)$$

represents $\succsim$ *and*

$$\mathcal{C}(x) = B_u(B_v(x)).$$

The following are the definitions of naiveté and sophistication for Strotz preferences from Ahn, Iijima, Le Yaouanq, and Sarver (2016), adapted to the current domain.

Definition 11 *An individual is* Strotz sophisticated *if, for all menus* x,

$$x \sim \{p\}, \quad \forall p \in \mathcal{C}(x).$$

An individual is Strotz naive *if, for all menus* x,

$$x \succsim \{p\}, \quad \forall p \in \mathcal{C}(x).$$

The definition of Strotz naiveté is too restrictive in the case of self-control preferences. The following result shows the exact implications of this definition for the self-control representation.

Proposition 3 *Suppose* $(\succsim, \mathcal{C})$ *is regular and has a self-control representation* $(u, v, \hat{v})$*. Then the individual is Strotz naive (Definition 11) if and only if* $\hat{v} \gg_u u + v$.

One interesting implication of Proposition 3 is that the Heidhues–Koszegi representation of Definition 3 can never be Strotz naive, and hence it requires alternate definitions like those provided in this paper for nonparametric foundations.

It is important to note that the case of $\hat{v} \approx u + v$ does not correspond to Strotz-sophisticated. In fact, Strotz-sophistication automatically fails whenever there are lotteries p, q such that $\{p\} \succ \{p, q\} \succ \{q\}$ because there is no selection in $x = \{p, q\}$ that is indifferent to x.

Although the implications of Strotz-naivete are too strong when applied to the self-control representation, the implications of naiveté proposed in this paper are suitable for Strotz representations. This is because Strotz representations are a limit case of self-control representations. To see this, parameterize a family of representations $(u, \gamma v, \gamma \hat{v})$ and take γ to infinity. Then the vectors v and $\hat{v}$ dominate the smaller u vector in determining actual and anticipated choice. Moreover, since choices are almost driven entirely by temptation, the penalty for self-control diminishes since no self-control is actually exerted. Given appropriate continuity in the limit, our definitions of naiveté for self-control representations should therefore also have the correct implications for Strotz representations. Indeed they do.

Proposition 4 *Suppose $(\succsim, \mathcal{C})$ is regular and has a Strotz representation $(u, \hat{v}, v)$ such that v is non-constant. Then, the following are equivalent:*

1. *the individual is naive (resp. sophisticated)*
2. *the individual is Strotz naive (resp. Strotz sophisticated)*
3. *$\hat{v} \gg_u v$ (resp. $\hat{v} \approx v$)*

4.2 Impossibility of a Unified Definition of Naiveté for Self-Control and Random Strotz Preferences

Ahn, Iijima, Le Yaouanq, and Sarver (2016) propose a single definition of naiveté suitable for both Strotz representations and the more general class of random Strotz representations. Proposition 4 showed that our definitions of naiveté under self-control for the general class of deterministic self-control preferences, when applied to deterministic Strotz preferences, viewed as a special limit case with large intensity of temptation, yield the same parametric restrictions as the definition of naiveté proposed by Ahn, Iijima, Le Yaouanq, and Sarver (2016) for the general class of random Strotz preference, with deterministic Strotz being a special deterministic case. This begs the question of whether a single definition exists that can be applied across both general classes of random Strotz and of self-control representations. This is impossible. The following example shows that no suitable definition of naiveté or sophistication can be applied to both consequentialist and nonconsequentialist models once random choice is permitted.

Example 2 Suppose $\succsim$ has a self-control representation $(u, \hat{v})$:

$$U(x) = \max_{p \in x}[u(p) + \hat{v}(p)] - \max_{q \in x} \hat{v}(q).$$

By Theorem 1 in Dekel and Lipman (2012), $\succsim$ also has the following random Strotz representation:[20]

$$U(x) = \int_0^1 \max_{p \in B_{\hat{v}+\alpha u}(x)} u(p)\, d\alpha.$$

Let $x^*_{SC} = B_{u+\hat{v}}(x)$ and $x^*_{RS} = \int_0^1 B_u(B_{\hat{v}+\alpha u}(x))\, d\alpha$. These would be the (average) choice sets of a sophisticated individual for these two different representations for the same ex-ante preference $\succsim$. Note that the second representation results in stochastic anticipated ex-post choices. A natural primitive for ex-post stochastic decisions is a random choice correspondence $\mathcal{C} : \mathcal{K}(\Delta(C)) \rightrightarrows \Delta(\Delta(C))$ that specifies a set of possible random selections for the agent, satisfying the feasibility constraint $\mathcal{C}(x) \subset \Delta(x)$. For any $\lambda^x \in \mathcal{C}(x)$, let $m(\lambda^x) = \int_x p\, d\lambda^x(p)$ denote the mean of λ^x and let $m(\mathcal{C}(x)) = \{m(\lambda^x) : \lambda^x \in \mathcal{C}(x)\}$ denote the set of means induced by $\mathcal{C}(x)$.[21]

Using the desired functional characterizations of sophistication and naiveté, if the individual does in fact exert self-control with a fixed anticipated temptation utility $\hat{v}$, then she is sophisticated if $m(\mathcal{C}(x)) = x^*_{SC}$, and she is naive if the lotteries in $m(\mathcal{C}(x))$ are worse than those in x^*_{SC}. If instead she does not anticipate exerting self-control and anticipates choosing according to the utility function $\hat{v} + \alpha u$ where α is distributed uniformly on $[0, 1]$, then she is sophisticated if $m(\mathcal{C}(x)) = x^*_{RS}$ and she is naive if the lotteries in $m(\mathcal{C}(x))$ are worse than those in x^*_{RS}.

The difficulty arises because the lotteries in x^*_{SC} are generally better than those in x^*_{RS}.[22] For example, suppose $x = \{p, q\}$ where $u(p) > u(q)$, $\hat{v}(q) > \hat{v}(p)$, and $(u+\hat{v})(p) > (u + \hat{v})(q)$. Then $x^*_{SC} = \{p\}$, whereas $x^*_{RS} \subset \{\beta p + (1 - \beta)q : \beta \in (0, 1)\}$. Hence $u(x^*_{SC}) > u(x^*_{RS})$. Suppose the choice correspondence satisfies

$$u(x^*_{SC}) > u(m(\mathcal{C}(x))) > u(x^*_{RS}).$$

[20] The intuition for this equivalence is straightforward. Let $f_x(\alpha) \equiv \max_{p \in x}(\hat{v} + \alpha u)(p)$. Note that the self-control representation is defined by precisely $U(x) = f_x(1) - f_x(0)$. By the Envelope Theorem, we also have

$$f_x(1) - f_x(0) = \int_0^1 f'_x(\alpha)\, d\alpha = \int_0^1 u(p(\alpha))\, d\alpha,$$

where $p(\alpha) \in \operatorname{argmax}_{q \in x}(\hat{v} + \alpha u)(q)$ for all $\alpha \in [0, 1]$.

[21] We assume that the set all selections $\lambda^x \in \mathcal{C}(x)$ is observable to make the proposed tension even stronger: Even with information about the full choice correspondence (as opposed to only observing a selection function from that correspondence), we cannot determine whether the individual is naive or sophisticated.

[22] Dekel and Lipman (2012, Theorem 5) made a similar observation.

If the individual actually has a self-control representation, then she should be classified as naive. However, if she actually has a random Strotz representation, then she is overly pessimistic and should not be classified as naive. □

There are obviously instances in which the individual would be classified as naive regardless of her actual representation, that is, when $u(x^*_{SC}) > u(x^*_{RS}) \geq u(m(\mathcal{C}(x)))$. Thus there are sufficient conditions for naiveté (see, e.g., Proposition 3 or Ahn, Iijima, Le Yaouanq, and Sarver (2016, Theorem 9)), but a tight characterization is not possible.

4.3 Impossibility of any Definition of Naiveté for Random Self-Control Preferences

Another approach to incorporate stochastic choice is to consider random temptations within the self-control representation. However, as observed by Stovall (2010) and Dekel and Lipman (2012), this type of representation is generally not uniquely identified from ex-ante preferences. The following example shows that this lack of identification precludes a sensible definition of naiveté for random self-control preferences. This impossibility is true even if Strotz and random Strotz preferences are excluded a priori from the analysis.

Example 3 Suppose $\succsim$ has a self-control representation $(u, \hat{v})$:

$$U(x) = \max_{p \in x}[u(p) + \hat{v}(p)] - \max_{q \in x} \hat{v}(q).$$

Fix any $\alpha \in (0, 1)$ and let $\hat{v}_1 = \frac{1}{1-\alpha}(\alpha u + \hat{v})$ and $\hat{v}_2 = \frac{1}{\alpha}\hat{v}$. Note that

$$u + \hat{v}_1 = \frac{1}{1-\alpha}(u + \hat{v}) \quad \text{and} \quad u + \hat{v}_2 = \frac{1}{\alpha}(\alpha u + \hat{v}),$$

and therefore U can also be expressed as a (nontrivially) random self-control representation:

$$U(x) = (1-\alpha)\left(\max_{p \in x}[u(p) + \hat{v}_1(p)] - \max_{q \in x} \hat{v}_1(q)\right) + \alpha\left(\max_{p \in x}[u(p) + \hat{v}_2(p)] - \max_{q \in x} \hat{v}_2(q)\right).$$

Let $x^* = B_{u+\hat{v}}(x)$ and $x^{**} = (1-\alpha)B_{u+\hat{v}}(x) + \alpha B_{\alpha u + \hat{v}}(x)$. These would be the (average) choice sets of a sophisticated individual for these respective representations.

Similar to the issues discussed in the previous section, the difficulty arises because the lotteries in x^* are generally better than those in x^{**}. For example, suppose $x = \{p, q\}$ where $(u + \hat{v})(p) > (u + \hat{v})(q)$ and $(\alpha u + \hat{v})(q) > (\alpha u + \hat{v})(p)$. Then $x^* = \{p\}$ and

$x^{**} = \{(1-\alpha)p + \alpha q\}$. Hence $u(x^*) > u(x^{**})$. If the choice correspondence satisfies

$$u(x^*) > u(m(\mathcal{C}(x))) > u(x^{**}),$$

then we again have the problem of not knowing how to properly classify this individual. Under the first self-control representation, we should classify her as naive. However, under the second random self-control representation, she is overly pessimistic and we should not classify her as naive. □

While a tight characterization of naiveté accommodating both the random Strotz and random self-control models is impossible, some interpretable sufficient conditions that imply naivete for both models are possible, and indeed some were proposed by Ahn, Iijima, Le Yaouanq, and Sarver (2016). However, as the examples in this section show, the problem is in finding tight conditions that are also necessary for naiveté for both models.

As a final note, one could also take an alternative perspective on this issue. Instead of asking when behavior should *definitively* be classified as naive versus sophisticated, as we have done in this section, one could instead ask when behavior could be *rationalized* as naive (or sophisticated) for *some* random self-control or random Strotz representation of the preference $\succsim$. The examples in this section show that there is some overlap of these regions: Some distributions of actual choices can be rationalized as both naive and sophisticated (and also pessimistic), depending on whether ex-ante preferences are represented by a self-control, random self-control, or random Strotz representation. Le Yaouanq (2015, Section 4) contains a more detailed discussion of this approach.

A Proofs

A.1 Preliminaries

The following lemma will be used repeatedly in the proofs of our main results.

Lemma 1 *Let u, w, w' be expected-utility functions defined on $\Delta(C)$ such that u and w' are not ordinally opposed.*[23] *If for all lotteries p and q we have*

$$\big[u(p) > u(q) \text{ and } w'(p) > w'(q)\big] \implies w(p) > w(q),$$

then $w \gg_u w'$.

In the case of finite C, it is easy to show that Lemma 1 follows from Lemma 3 in Dekel and Lipman (2012), who also noted the connection to the Harsanyi Aggregation Theorem. Our analysis of dynamic representations defined on infinite-horizon decision problems requires the more general domain of compact outcome spaces. We include a short proof of Lemma 1 for the case of compact C to show that no technical problems arise in extending their result to our more general domain. Our proof is based on the following slight variation of Farkas' Lemma.[24]

Lemma 2 *Suppose $f_1, f_2, g : \Delta(C) \to \mathbb{R}$ are continuous and affine, and suppose f_1 and f_2 are not ordinally opposed. Then the following are equivalent:*

1. *For all $p, q \in \Delta(C)$: $[f_1(p) > f_1(q)$ and $f_2(p) > f_2(q)] \implies g(p) \geq g(q)$.*
2. *There exist scalars $a, b \geq 0$ and $c \in \mathbb{R}$ such that $g = af_1 + bf_2 + c$.*

Proof of Lemma 2: It is immediate that 2 implies 1. To show 1 implies 2, we first argue that 1 implies the same implication holds when the strict inequalities are replaced with weak inequalities:

$$[f_1(p) \geq f_1(q) \text{ and } f_2(p) \geq f_2(q)] \implies g(p) \geq g(q). \tag{4}$$

The argument relies on the assumption that f_1 and f_2 are not ordinally opposed and is similar to the use of constraint qualification in establishing the Kuhn-Tucker Theorem. Suppose $p, q \in \Delta(C)$ satisfy $f_1(p) \geq f_1(q)$ and $f_2(p) \geq f_2(q)$. Since f_1 and f_2 are not ordinally opposed, there exist $p^*, q^* \in \Delta(C)$ such that $f_1(p^*) > f_1(q^*)$ and $f_2(p^*) > f_2(q^*)$. Let $p^\alpha \equiv \alpha p^* + (1-\alpha)p$ and $q^\alpha \equiv \alpha q^* + (1-\alpha)q$. Since these functions are affine, $f_1(p^\alpha) > f_1(q^\alpha)$ and $f_2(p^\alpha) > f_2(q^\alpha)$

[23] That is, there exist lotteries p and q such that both $u(p) > u(q)$ and $w'(p) > w'(q)$.

[24] There are two small distinctions between this result and the classic version of Farkas' Lemma. First, Farkas' Lemma deals with linear functions defined on a vector space whereas we restrict to linear functions defined on the convex subset $\Delta(C)$ of the vector space $ca(C)$ of all finite signed measures on C. Second, in condition 1 we only assume the conclusion that $g(p) \geq g(q)$ when the corresponding inequalities for f_1 and f_2 are strict. Together with our assumption that f_1 and f_2 are not ordinally opposed, we show in the proof that this condition implies the same conclusion for the case where the inequalities are weak.

for all $\alpha \in (0, 1]$. Condition 1 therefore implies $g(p^\alpha) \geq g(q^\alpha)$ for all $\alpha \in (0, 1]$. By continuity $g(p) \geq g(q)$. This establishes the condition in Equation (4).

Fix any $\bar{c} \in C$ and define $\bar{f}_1(p) \equiv f_1(p) - f_1(\delta_{\bar{c}})$, $\bar{f}_2(p) \equiv f_2(p) - f_2(\delta_{\bar{c}})$, and $\bar{g}(p) \equiv g(p) - g(\delta_{\bar{c}})$. Note that Equation (4) holds for f_1, f_2, g if and only if it holds for $\bar{f}_1, \bar{f}_2, \bar{g}$. Each of these functions can be extended to a continuous linear function on the space $ca(C)$ of all finite signed measures on C: Since the mapping $c \mapsto \bar{f}_1(\delta_c)$ is continuous in the topology on C, the function $F_1(p) \equiv \int \bar{f}_1(\delta_c) dp$ for $p \in ca(C)$ is a well-defined continuous linear functional that extends $\bar{f}_1$. Define F_2 and G analogously. We next show that for any $p, q \in ca(C)$:

$$[F_1(p) \geq F_1(q) \text{ and } F_2(p) \geq F_2(q)] \implies G(p) \geq G(q). \tag{5}$$

To establish this condition, fix any $p, q \in ca(C)$ and suppose $F_i(p) \geq F_i(q)$ for $i = 1, 2$. Let $p' = p - p(C)\delta_{\bar{c}}$ and $q' = q - q(C)\delta_{\bar{c}}$. Then $p'(C) = q'(C) = 0$, and we also have $F_i(p') \geq F_i(q')$ since $\bar{f}_i(\delta_{\bar{c}}) = 0$. Equivalently, $F_i(p' - q') \geq 0$. There exist $p'', q'' \in \Delta(C)$ and $\alpha \geq 0$ such that $p' - q' = \alpha(p'' - q'')$. By linearity, $F_i(p'') \geq F_i(q'')$, which implies $f_i(p'') \geq f_i(q'')$ for $i = 1, 2$. Equation (4) therefore implies $g(p'') \geq g(q'')$, which implies $G(p'') \geq G(q'')$ and consequently $G(p') \geq G(q')$ and $G(p) \geq G(q)$. This establishes Equation (5).

By the Convex Cone Alternative Theorem (an infinite-dimensional version of Farkas' Lemma) (Aliprantis and Border (2006, Corollary 5.84)), Equation (5) implies there exist $a, b \geq 0$ such that $G = aF_1 + bF_2$. Thus $\bar{g} = a\bar{f}_1 + b\bar{f}_2$, and hence $g = af_1 + bf_2 + c$, where $c = g(\delta_{\bar{c}}) - af_1(\delta_{\bar{c}}) - bf_2(\delta_{\bar{c}})$. ■

Proof of Lemma 1: By Lemma 2, the conditions in this lemma imply that there exist scalars $a, b \geq 0$ and $c \in \mathbb{R}$ such that $w = au + bw' + c$. Since there must exist some p and q such that $u(p) > u(q)$ and $w'(p) > w'(q)$, the function w cannot be constant. This implies $a + b > 0$. Thus $w \approx \alpha u + (1 - \alpha)w'$ for $\alpha = a/(a + b) \in [0, 1]$. ■

A.2 Proof of Theorem 1

Sufficiency: To establish sufficiency, suppose the individual is naive. Then, for any lotteries p and q,

$$\begin{aligned}
\big[u(p) > u(q) \text{ and } (u+v)(p) > (u+v)(q)\big] &\implies \mathcal{C}(\{p,q\}) = \{p\} \succ \{q\} \\
&\implies \{p,q\} \succ \{q\} \qquad \text{(by naiveté)} \\
&\implies (u+\hat{v})(p) > (u+\hat{v})(q).
\end{aligned}$$

Regularity requires that u and $u + v$ not be ordinally opposed. Therefore, Lemma 1 implies $u + \hat{v} \gg_u u + v$.

If in addition the individual is sophisticated, then an analogous argument leads to

$$\big[u(p) > u(q) \text{ and } (u+\hat{v})(p) > (u+\hat{v})(q)\big] \implies (u+v)(p) > (u+v)(q)$$

for any lotteries p and q, which ensures $u+v \gg_u u+\hat{v}$ by Lemma 1. Thus $u+v \approx u+\hat{v}$.

Necessity: To establish necessity, suppose $u+\hat{v} \approx \alpha u + (1-\alpha)(u+v)$ for $\alpha \in [0,1]$ and take any lotteries p and q. Then

$$\begin{aligned}
\big[\{p\} \succ \{q\} \text{ and } \mathcal{C}(\{p,q\}) = \{p\}\big] &\implies \big[u(p) > u(q) \text{ and } (u+v)(p) > (u+v)(q)\big] \\
&\implies \big[u(p) > u(q) \text{ and } (u+\hat{v})(p) > (u+\hat{v})(q)\big] \\
&\implies \{p,q\} \succ \{q\},
\end{aligned}$$

and thus the individual is naive. If in addition $u+v \approx u+\hat{v}$, then one can analogously show

$$\big[\{p\} \succ \{q\} \text{ and } \{p,q\} \succ \{q\}\big] \implies \mathcal{C}(\{p,q\}) = \{p\},$$

and thus the individual is sophisticated.

A.3 Proof of Theorem 2

We first make an observation that will be useful later in the proof. Since each individual is assumed to be naive, Theorem 1 implies $u_i + \hat{v}_i \approx \alpha_i u_i + (1-\alpha_i)(u_i+v_i)$ for some $\alpha_i \in [0,1]$, and consequently, for any lotteries p and q,

$$\big[(u_i+\hat{v}_i)(p) > (u_i+\hat{v}_i)(q) \text{ and } (u_i+v_i)(q) > (u_i+v_i)(p)\big] \implies u_i(p) > u_i(q).$$

Therefore, for any lotteries p and q,

$$\begin{aligned}
&\big[(u_i+\hat{v}_i)(p) > (u_i+\hat{v}_i)(q) \text{ and } (u_i+v_i)(q) > (u_i+v_i)(p)\big] \\
&\iff \big[u_i(p) > u_i(q) \text{ and } (u_i+\hat{v}_i)(p) > (u_i+\hat{v}_i)(q) \text{ and } (u_i+v_i)(q) > (u_i+v_i)(p)\big] \qquad (6) \\
&\iff \big[\{p,q\} \succ_i \{q\} \text{ and } \mathcal{C}_i(\{p,q\}) = \{q\}\big].
\end{aligned}$$

Sufficiency: Suppose individual 1 is more naive than individual 2. By Equation (6), this can equivalently be stated as

$$\begin{aligned}
&\big[(u_2+\hat{v}_2)(p) > (u_2+\hat{v}_2)(q) \text{ and } (u_2+v_2)(q) > (u_2+v_2)(p)\big] \\
&\qquad \implies \big[(u_1+\hat{v}_1)(p) > (u_1+\hat{v}_1)(q) \text{ and } (u_1+v_1)(q) > (u_1+v_1)(p)\big].
\end{aligned}$$

If individual 2 is sophisticated then the conclusion of the theorem is trivially satisfied, so suppose not. Then individual 2 must be strictly naive, and hence there must exist lotteries p and q such that $(u_2+\hat{v}_2)(p) > (u_2+\hat{v}_2)(q)$ and $(u_2+v_2)(q) > (u_2+v_2)(p)$. Thus the functions $(u_2+\hat{v}_2)$ and

$-(u_2+v_2)$ are not ordinally opposed. Therefore, by Lemma 2, there exist scalars $a, \hat{a}, b, \hat{b} \geq 0$ and $c, \hat{c} \in \mathbb{R}$ such that

$$\begin{aligned} u_1+\hat{v}_1 &= \hat{a}(u_2+\hat{v}_2) - \hat{b}(u_2+v_2) + \hat{c}, \\ -(u_1+v_1) &= a(u_2+\hat{v}_2) - b(u_2+v_2) + c. \end{aligned}$$

Taking b times the first expression minus $\hat{b}$ times the second, and taking a times the first expression minus $\hat{a}$ times the second yields the following:

$$\begin{aligned} b(u_1+\hat{v}_1) + \hat{b}(u_1+v_1) &= (\hat{a}b - a\hat{b})(u_2+\hat{v}_2) + (b\hat{c} - \hat{b}c), \\ a(u_1+\hat{v}_1) + \hat{a}(u_1+v_1) &= (\hat{a}b - a\hat{b})(u_2+v_2) + (a\hat{c} - \hat{a}c). \end{aligned} \tag{7}$$

Claim 1 *Since $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ are jointly regular, $\hat{a}b > a\hat{b}$. In particular, $\hat{a} > 0$, $b > 0$, and $\frac{b}{\hat{b}+b} > \frac{a}{\hat{a}+a}$.*

Proof: Joint regularity requires there exist lotteries p and q such that $u_i(p) > u_i(q)$ and $(u_i+v_i)(p) > (u_i+v_i)(q)$ for $i = 1, 2$. Since both individuals are naive, by Theorem 1 this also implies $(u_i+\hat{v}_i)(p) > (u_i+\hat{v}_i)(q)$. Thus

$$\begin{aligned} \hat{a}(u_2+\hat{v}_2)(p) - \hat{b}(u_2+v_2)(p) &= (u_1+\hat{v}_1)(p) - \hat{c} \\ &> (u_1+\hat{v}_1)(q) - \hat{c} = \hat{a}(u_2+\hat{v}_2)(q) - \hat{b}(u_2+v_2)(q), \\ a(u_2+\hat{v}_2)(q) - b(u_2+v_2)(q) &= -(u_1+v_1)(q) - c \\ &> -(u_1+v_1)(p) - c = a(u_2+\hat{v}_2)(p) - b(u_2+v_2)(p). \end{aligned}$$

Rearranging terms, these equations imply

$$\begin{aligned} \hat{a}(u_2+\hat{v}_2)(p-q) &> \hat{b}(u_2+v_2)(p-q) \\ b(u_2+v_2)(p-q) &> a(u_2+\hat{v}_2)(p-q). \end{aligned}$$

Multiplying these inequalities, and using the fact that $(u_2+\hat{v}_2)(p-q) > 0$ and $(u_2+v_2)(p-q) > 0$ by the regularity inequalities for individual 2, we have $\hat{a}b > a\hat{b}$. This implies $(\hat{a}+a)b > a(\hat{b}+b)$, and hence $\frac{b}{\hat{b}+b} > \frac{a}{\hat{a}+a}$. ■

By Claim 1, Equation (7) implies

$$\begin{aligned} u_2+\hat{v}_2 &\approx \hat{\alpha}(u_1+\hat{v}_1) + (1-\hat{\alpha})(u_1+v_1), \\ u_2+v_2 &\approx \alpha(u_1+\hat{v}_1) + (1-\alpha)(u_1+v_1), \end{aligned}$$

where

$$\hat{\alpha} = \frac{b}{\hat{b}+b} > \frac{a}{\hat{a}+a} = \alpha.$$

Since $u_1+\hat{v}_1$ is itself an affine transformation of a convex combination of u_1 and u_1+v_1, we have

$$u_1 + \hat{v}_1 \gg_{u_1} u_2 + \hat{v}_2 \gg_{u_1} u_2 + v_2 \gg_{u_1} u_1 + v_1,$$

as claimed.

Necessity: If individual 2 is sophisticated, then trivially individual 1 is more naive than individual 2. Consider now the case where individual 2 is strictly naive and

$$u_1 + \hat{v}_1 \gg_{u_1} u_2 + \hat{v}_2 \gg_{u_1} u_2 + v_2 \gg_{u_1} u_1 + v_1,$$

which can equivalently be stated as

$$\begin{aligned} u_2 + \hat{v}_2 &\approx \hat{\alpha}(u_1 + \hat{v}_1) + (1 - \hat{\alpha})(u_1 + v_1), \\ u_2 + v_2 &\approx \alpha(u_1 + \hat{v}_1) + (1 - \alpha)(u_1 + v_1), \end{aligned}$$

for $\hat{\alpha} > \alpha$. Then, for any lotteries p and q,

$$\begin{aligned} &\big[(u_2 + \hat{v}_2)(p) > (u_2 + \hat{v}_2)(q) \text{ and } (u_2 + v_2)(q) > (u_2 + v_2)(p)\big] \\ &\implies \hat{\alpha}(u_1 + \hat{v}_1)(p - q) + (1 - \hat{\alpha})(u_1 + v_1)(p - q) \\ &\qquad\qquad > 0 > \alpha(u_1 + \hat{v}_1)(p - q) + (1 - \alpha)(u_1 + v_1)(p - q) \\ &\implies \big[(u_1 + \hat{v}_1)(p) > (u_1 + \hat{v}_1)(q) \text{ and } (u_1 + v_1)(q) > (u_1 + v_1)(p)\big]. \end{aligned}$$

By Equation (6), this condition is equivalent to individual 1 being more naive than 2.

A.4 Proof of Proposition 1

Proof of 2 ⇒ 1: The relation $\succsim$ has no preference for commitment when $\hat{\gamma} = 0$. Otherwise, when $\hat{\gamma} > 0$, $\{p\} \sim \{p, q\} \succ \{q\}$ is equivalent to $u(p) > u(q)$ and $\bar{v}(p) \geq \bar{v}(q)$. Thus $(u + \gamma\bar{v})(p) > (u + \gamma\bar{v})(q)$ for any $\gamma \geq 0$, and hence $\mathcal{C}(\{p, q\}) = \{p\}$.

Proof of 1 ⇒ 2: If $\succsim$ has no preference for commitment, let $\bar{v} = v$, $\gamma = 1$, and $\hat{\gamma} = 0$. In the alternative case where $\succsim$ has a preference for commitment (so $\hat{v}$ is non-constant and $\hat{v} \not\approx u$), condition 1 requires that for any p and q,

$$\begin{aligned} [u(p) > u(q) \text{ and } \hat{v}(p) \geq \hat{v}(q)] &\iff \{p\} \sim \{p, q\} \succ \{q\} \\ &\implies \mathcal{C}(\{p, q\}) = \{p\} \\ &\iff (u + v)(p) > (u + v)(q). \end{aligned} \tag{8}$$

We assumed there exist some pair of lotteries p and q such that $\{p\} \sim \{p, q\} \succ \{q\}$. Therefore, u and $\hat{v}$ are not ordinally opposed. Thus, by Lemma 1, $u + v \gg_u \hat{v}$. That is, $u + v \approx \alpha u + (1 - \alpha)\hat{v}$ for some $0 \leq \alpha \leq 1$.

Note that $u \not\approx \hat{v}$ since $\succsim$ has a preference for commitment, and $u \not\approx -\hat{v}$ since the two functions are not ordinally opposed. Therefore, there must exist lotteries p and q such that $u(p) > u(q)$ and $\hat{v}(p) = \hat{v}(q)$. By Equation (8), this implies $(u+v)(p) > (u+v)(q)$. Hence $u + v \not\approx \hat{v}$, that is, $\alpha > 0$. We therefore have $u + v \approx u + \frac{1-\alpha}{\alpha}\hat{v}$. Let $\bar{v} = \hat{v}$, $\gamma = \frac{1-\alpha}{\alpha}$, and $\hat{\gamma} = 1$.

A.5 Proof of Theorems 3 and 5

We begin by proving a general representation result using the following weaker form of stationarity.

Axiom 14 (Weak Commitment Stationarity) *For $p, q \in \Delta(C^{\mathbb{N}})$,*

$$(c, \{p\}) \succsim_t (c, \{q\}) \iff (c, \{p\}) \succsim_{t+1} (c, \{q\}).$$

Axiom 14 permits the actual present bias to vary over time. After proving the following general result, we add Axiom 7 (Stationarity) to prove Theorem 3, and we add Axioms 12 (Diminishing Naiveté) and 13 (Commitment Stationarity) to prove Theorem 5.

Proposition 5 *A profile of nontrivial relations $\{\succsim_t\}_{t\in\mathbb{N}}$ satisfies Axioms 1–6, 8–9, 11, and 14 if and only if there exist continuous functions $u : C \to \mathbb{R}$ and $U_t, \hat{V}_t, V_t : \Delta(C \times Z) \to \mathbb{R}$ satisfying the following system of equations:*

$$\begin{aligned}
U_t(p) &= \int_{C\times Z} (u(c) + \delta \hat{W}_t(x))\, dp(c,x) \\
V_t(p) &= \gamma_t \int_{C\times Z} (u(c) + \beta_t \delta \hat{W}_t(x))\, dp(c,x) \\
\hat{V}_t(p) &= \hat{\gamma}_t \int_{C\times Z} (u(c) + \hat{\beta}_t \delta \hat{W}_t(x))\, dp(c,x) \\
\hat{W}_t(x) &= \max_{q\in x}(U_t(q) + \hat{V}_t(q)) - \max_{q\in x} \hat{V}_t(q)
\end{aligned}$$

and such that, for all $t \in \mathbb{N}$,

$$p \succsim_t q \iff U_t(p) + V_t(p) \geq U_t(q) + V_t(q),$$

where $\beta_t, \hat{\beta}_t \in [0,1]$, $0 < \delta < 1$, and $\gamma_t, \hat{\gamma}_t \geq 0$ satisfy

$$\frac{1 + \hat{\gamma}_t \hat{\beta}_t}{1 + \hat{\gamma}_t} \geq \frac{1 + \gamma_{t+1}\beta_{t+1}}{1 + \gamma_{t+1}}. \tag{9}$$

Moreover, $\{\succsim_t\}_{t\in\mathbb{N}}$ also satisfies Axiom 10 if and only if Equation (9) holds with equality.

A.5.1 Proof of Proposition 5

We only show the sufficiency of the axioms. Axioms 1–3 imply there exist continuous functions $f_t : C \times Z \to \mathbb{R}$ for $t \in \mathbb{N}$ such that

$$p \succsim_t q \iff \int f_t(c,x)\, dp(c,x) \geq \int f_t(c,x)\, dq(c,x).$$

The first part of Axiom 6 (Separability) implies that f is separable, so

$$f_t(c,x) = f_t^1(c) + f_t^2(x)$$

for some continuous functions f_t^1 and f_t^2. In addition, Axiom 5 (Indifference to Timing) implies f_t is linear in the second argument: $\lambda f_t(c,x) + (1-\lambda) f_t(c,y) = f_t(c, \lambda x + (1-\lambda) y)$. Equivalently,

$$\lambda f_t^2(x) + (1-\lambda) f_t^2(y) = f_t^2(\lambda x + (1-\lambda) y).$$

Next, Axiom 14 (Weak Commitment Stationarity) implies that, for any $p, q \in \Delta(C^{\mathbb{N}})$,

$$f_t^2(\{p\}) \geq f_t^2(\{q\}) \iff f_{t+1}^2(\{p\}) \geq f_{t+1}^2(\{q\}).$$

By the linearity of f_t^2, this implies that, for any $t, t' \in \mathbb{N}$, the restrictions of f_t^2 and $f_{t'}^2$ to deterministic consumption streams in $C^{\mathbb{N}}$ are identical up to a positive affine transformation. Therefore, by taking an affine transformation of each f_t, we can without loss of generality assume that $f_t^2(\{p\}) = f_{t'}^2(\{p\})$ for all $t, t' \in \mathbb{N}$ and for all $p \in \Delta(C^{\mathbb{N}})$.

Define a preference $\succsim_t^*$ over Z by $x \succsim_t^* y$ if and only if $f_t^2(x) \geq f_t^2(y)$ or, equivalently, $(c,x) \succsim_t (c,y)$. Note that this induced preference does not depend on the choice of c by separability. Axioms 1-5 imply that the induced preference over menus Z satisfies Axioms 1-4 in Gul and Pesendorfer (2001). Specifically, the linearity of f_t^2 in the menu (which we obtained using the combination of Axioms 3 and 5) implies that $\succsim_t^*$ satisfies the independence axiom for mixtures of menus (Gul and Pesendorfer, 2001, Axiom 3). Their other axioms are direct translations of ours. Thus, for each $t \in \mathbb{N}$, there exist continuous and linear functions U_t, $\hat{V}_t : \Delta(C \times Z) \to \mathbb{R}$ such that

$$x \succsim_t^* y \iff \max_{p \in x}(U_t(p) + \hat{V}_t(p)) - \max_{q \in x} \hat{V}_t(q) \geq \max_{p \in y}(U_t(p) + \hat{V}_t(p)) - \max_{q \in y} \hat{V}_t(q).$$

Since both f_t^2 and this self-control representation are linear in menus, they must be the same up to an affine transformation. Taking a common affine transformation of U_t and $\hat{V}_t$ if necessary, we therefore have

$$f_t^2(x) = \max_{p \in x}(U_t(p) + \hat{V}_t(p)) - \max_{q \in x} \hat{V}_t(q). \tag{10}$$

By Equation (10), $f_t^2(\{p\}) = U_t(p)$ for all $p \in \Delta(C \times Z)$. Thus the second part of Axiom 6

(Separability) implies that U_t is separable, so

$$U_t(c,x) = u_t^1(c) + u_t^2(x) \tag{11}$$

for some continuous functions u_t^1 and u_t^2.

Claim 2 *There exist scalars $\theta^u_{t,i}, \alpha^u_{t,i}$ for $i = 1,2$ with $\theta^u_{t,2} \geq \theta^u_{t,1} > 0$ such that $u_t^i = \theta^u_{t,i} f_t^i + \alpha^u_{t,i}$.*

Proof: Axiom 8 (Present Bias) ensures that (i) $u_t^1 \approx f_t^1$ and (ii) $u_t^2 \approx f_t^2$. To show (i), take any p, q such that $p^2 = q^2$ and $f_t^1(q^1) > f_t^1(p^1)$.[25] Then $q \succ_t p$ and $p \sim_t p^1 \times q^2$, which implies $(c, \{q\}) \succ_t (c, \{p\})$ by Axiom 8. Thus $u_t^1(q^1) > u_t^1(p^1)$, and the claim follows since f_t^1 is non-constant (by nontriviality). To show (ii), take any p, q such that $p^1 = q^1$ and $f_t^2(q^2) > f_t^2(p^2)$. Then $q \succ_t p$ and $p \prec_t p^1 \times q^2$, which implies $(c, \{q\}) \succ_t (c, \{p\})$ by Axiom 8. Thus $u_t^2(q^2) > u_t^2(p^2)$, and the claim follows since f_t^2 is non-constant (by Equation (10) and u_t^1 non-constant).

Thus we can write $u_t^i = \theta^u_{t,i} f_t^i + \alpha^u_{t,i}$ for some constants $\theta^u_{t,i}, \alpha^u_{t,i}$ with $\theta^u_{t,i} > 0$ for $i = 1,2$. Finally, toward a contradiction, suppose that $\theta^u_{t,2} < \theta^u_{t,1}$. Then, since f_t^1 and f_t^2 are non-constant, we can take p, q such that $f_t^1(p^1) > f_t^1(q^1)$, $f_t^2(p^2) < f_t^2(q^2)$, and

$$\frac{\theta^u_{t,2}}{\theta^u_{t,1}} < \frac{f_t^1(p^1) - f_t^1(q^1)}{f_t^2(q^2) - f_t^2(p^2)} < 1.$$

The first inequality implies $\theta^u_{t,1} f_t^1(q^1) + \theta^u_{t,2} f_t^2(q^2) < \theta^u_{t,1} f_t^1(p^1) + \theta^u_{t,2} f_t^2(p^2)$, and hence $U_t(p) > U_t(q)$ or, equivalently, $(c, \{p\}) \succ_t (c, \{q\})$. The second inequality implies $f_t^1(p^1) + f_t^2(p^2) < f_t^1(q^1) + f_t^2(q^2)$, and hence $q \succ_t p$. Axiom 8 therefore requires that $p \succ_t p^1 \times q^2$. However, since $f_t^2(p^2) < f_t^2(q^2)$, we have $p^1 \times q^2 \succ_t p$, a contradiction. Thus we must have $\theta^u_{t,2} \geq \theta^u_{t,1}$. ■

Claim 3 *For all $t, t' \in \mathbb{N}$, $\theta^u_{t,2} = \theta^u_{t',2} \in (0,1)$ and $u_t^1(c) + \alpha^u_{t,2} = u_{t'}^1(c) + \alpha^u_{t',2}$ for all $c \in C$.*

Proof: Note that by Equations (10) and (11) and Claim 2, for any $(c_0, c_1, c_2, \dots) \in C^{\mathbb{N}}$,[26]

$$\begin{aligned} f_t^2(c_0, c_1, c_2, \dots) &= U_t(c_0, c_1, c_2, \dots) \\ &= u_t^1(c_0) + u_t^2(c_1, c_2, \dots) \\ &= u_t^1(c_0) + \alpha^u_{t,2} + \theta^u_{t,2} f_t^2(c_1, c_2, \dots). \end{aligned} \tag{12}$$

Following the same approach as Gul and Pesendorfer (2004, page 151), we show $\theta^u_{t,2} < 1$ using continuity. Fix any $c \in C$ and let $x^c = \{(c, c, c, \dots)\} = \{(c, x^c)\}$. Fix any other consumption

[25] We write $f_t^1(p^1)$ to denote $\int f_t^1(c)\, dp^1(c)$, and write $f_t^2(p^2)$ to denote $\int f_t^2(x)\, dp^2(x)$. We adopt similar notational conventions for u_t^1 and u_t^2.

[26] Our notation here is slightly informal. More precisely, for any $(c_0, c_1, c_2, \dots) \in C^{\mathbb{N}}$, there exists $x^i \in Z$ for $i = 0, 1, 2, \dots$ such that $x^i = \{(c_i, x^{i+1})\}$. To simplify notation, we write $(c_0, c_1, c_2, \dots)$ to indicate the menu $x^0 = \{(c_0, x^1)\} = \{(c_0, \{(c_1, x^2)\})\} = \cdots$.

stream $y = \{(c_0, c_1, c_2, \dots)\} \in Z$ such that $f_t^2(y) \neq f_t^2(x^c)$. Let $y^1 = \{(c, y)\}$ and define y^n inductively by $y^n = \{(c, y^{n-1})\}$. Then $y^n \to z^c$, and therefore by continuity,

$$f_t^2(y^n) - f_t^2(x^c) = \big(\theta_{t,2}^u\big)^n \big(f_t^2(y) - f_t^2(x^c)\big) \to 0,$$

which requires that $\theta_{t,2}^u < 1$.

Recall that $f_t^2(\{p\}) = f_{t'}^2(\{p\})$ for all $t, t' \in \mathbb{N}$ and for all $p \in \Delta(C^{\mathbb{N}})$ or, equivalently, $U_t|_{\Delta(C^{\mathbb{N}})} = U_{t'}|_{\Delta(C^{\mathbb{N}})}$. Therefore, by Equation (12), we must have $\theta_{t,2}^u = \theta_{t',2}^u$ and $u_t^1(c) + \alpha_{t,2}^u = u_{t'}^1(c) + \alpha_{t',2}^u$ for all $c \in C$, as claimed. ■

To begin constructing the representation, set $\delta \equiv \theta_{t,2}^u \in (0,1)$ and

$$u(c) \equiv u_t^1(c) + \alpha_{t,2}^u = \theta_{t,1}^u f_t^1(c) + \alpha_{t,1}^u + \alpha_{t,2}^u.$$

Claim 3 ensures that δ and u are well-defined, as they do not depend on the choice of t. Set $\hat{W}_t(x) \equiv f_t^2(x)$ and hence, by Equation (11) and Claim 2,

$$U_t(c, x) = \theta_{t,1}^u f_t^1(c) + \alpha_{t,1}^u + \theta_{t,2}^u f_t^2(x) + \alpha_{t,2}^u = u(c) + \delta \hat{W}_t(x),$$

so the first displayed equation in Proposition 5 is satisfied.

By Claim 2, we have $0 < \theta_{t,1}^u / \theta_{t,2}^u \leq 1$. Therefore, there exist $\gamma_t \geq 0$ and $\beta_t \in [0,1]$ such that

$$\frac{1 + \gamma_t \beta_t}{1 + \gamma_t} = \frac{\theta_{t,1}^u}{\theta_{t,2}^u}.$$

Note that there are multiple values of γ_t and β_t that satisfy this equality, so these parameters are not individually identified from preferences. Next, defining V_t as in the second displayed equation in Proposition 5, we have

$$\begin{aligned}(U_t + V_t)(c, x) &= (1 + \gamma_t) u(c) + (1 + \gamma_t \beta_t) \delta \hat{W}_t(x) \\ &= (1 + \gamma_t) \Big(u(c) + \frac{1 + \gamma_t \beta_t}{1 + \gamma_t} \delta \hat{W}_t(x) \Big) \\ &= (1 + \gamma_t) \Big(\theta_{t,1}^u f_t^1(c) + \theta_{t,1}^u f_t^2(x) + \alpha_{t,1}^u + \alpha_{t,2}^u \Big),\end{aligned}$$

which is a positive affine transformation of $f_t(c, x)$. Thus

$$p \succsim_t q \iff U_t(p) + V_t(p) \geq U_t(q) + V_t(q),$$

The next claims are used to establish the desired form for $\hat{V}_t$.

Claim 4 *The function $\hat{V}_t$ is separable for all t, so $\hat{V}_t(c, x) = \hat{v}_t^1(c) + \hat{v}_t^2(x)$.*

Proof: It suffices to show that correlation does affect the value assigned to a lottery p by the function $\hat{V}_t$. That is, we only need to show $\hat{V}_t(p) = \hat{V}_t(p^1 \times p^2)$ for all lotteries p.[27] We will show that non-equality leads to a contradiction of Axiom 9 (No Temptation by Atemporal Choices) by considering two cases. For now, restrict attention to lotteries in the set

$$A = \Big\{ p \in \Delta(C \times Z) : \min_{c \in C} u_t^1(c) < u_t^1(p^1) < \max_{c \in C} u_t^1(c) \Big\}.$$

Case (i): $\hat{V}_t(p) > \hat{V}_t(p^1 \times p^2)$. By the continuity of $\hat{V}_t$, there exists $q^1 \in \Delta(C)$ such that $u_t^1(q^1) > u_t^1(p^1)$ and $\hat{V}_t(p) > \hat{V}_t(q^1 \times p^2)$. The first inequality implies $U_t(q^1 \times p^2) > U_t(p)$. By the self-control representation in Equation (10), this implies $(c, \{q^1 \times p^2\}) \succ_t (c, \{q^1 \times p^2, p\})$, in violation of Axiom 9.

Case (ii): $\hat{V}_t(p) < \hat{V}_t(p^1 \times p^2)$. By the continuity of $\hat{V}_t$, there exists $q^1 \in \Delta(C)$ such that $u_t^1(q^1) < u_t^1(p^1)$ and $\hat{V}_t(p) < \hat{V}_t(q^1 \times p^2)$. The first inequality implies $U_t(p) > U_t(q^1 \times p^2)$. By the self-control representation in Equation (10), this implies $(c, \{p\}) \succ_t (c, \{p, q^1 \times p^2\})$, in violation of Axiom 9.

We have now shown that $\hat{V}_t(p) = \hat{V}_t(p^1 \times p^2)$ for all $p \in A$. Since u_t^1 is non-constant, A is dense in $\Delta(C \times Z)$. By the continuity of $\hat{V}_t$, we therefore have $\hat{V}_t(p) = \hat{V}_t(p^1 \times p^2)$ for all $p \in \Delta(C \times Z)$. ■

Claim 5 *There exist scalars $\theta_{t,i}^v \geq 0$ and $\alpha_{t,i}^v \in \mathbb{R}$ for $i = 1, 2$ such that $\hat{v}_t^1 = \theta_{t,1}^v u + \alpha_{t,1}^v$ and $\hat{v}_t^2 = \theta_{t,2}^v \delta \hat{W}_t + \alpha_{t,2}^v$.*

Proof: Axiom 9 (No Temptation by Atemporal Choices) ensures that (i) $\hat{v}_t^1 \approx u$ or $\hat{v}_t^1$ is constant, and (ii) $\hat{v}_t^2 \approx \delta \hat{W}_t$ or $\hat{v}_t^2$ is constant. To show (i), take any p, q with $p^2 = q^2$ and $u(p^1) > u(q^1)$. Then $u_t^1(p^1) > u_t^2(q^1)$ and hence $(c, \{p\}) \sim_t (c, \{p, q\}) \succ_t (c, \{q\})$ by Axiom 9, which requires that $\hat{v}_t^1(p^1) \geq \hat{v}_t^1(q^1)$. Since u is non-constant, the desired claim follows. Part (ii) is analogously shown by taking any p, q with $p^1 = q^1$ and $\delta \hat{W}_t(p^2) > \delta \hat{W}_t(q^2)$. Then $u_t^2(p^2) > u_t^2(q^2)$ and hence $(c, \{p\}) \sim_t (c, \{p, q\}) \succ_t (c, \{q\})$ by Axiom 9, which implies $\hat{v}_t^2(p^2) \geq \hat{v}_t^2(q^2)$. ■

By Equation (10), changing $\hat{V}_t$ by the addition of a scalar does not alter the function f_t^2. Therefore, we can without loss of generality assume that $\alpha_{t,1}^v = \alpha_{t,2}^v = 0$. We next characterize the implications of naiveté and sophistication.

Claim 6

1. *If $\{\succsim_t\}_{t \in \mathbb{N}}$ satisfies Axiom 11 (Naiveté), then* $\frac{1+\gamma_{t+1}\beta_{t+1}}{1+\gamma_{t+1}} \leq \frac{1+\theta_{t,2}^v}{1+\theta_{t,1}^v} \leq 1$.

[27] To see that this condition is sufficient for separability, fix any $\bar{c} \in C$ and $\bar{x} \in Z$, and define $\hat{v}_t^1(c) \equiv \hat{V}_t(c, \bar{x})$ and $\hat{v}_t^2(x) \equiv \hat{V}_t(\bar{c}, x) - \hat{V}_t(\bar{c}, \bar{x})$. For any (c, x), let $p = \frac{1}{2}\delta_{(c,x)} + \frac{1}{2}\delta_{(\bar{c},\bar{x})}$ and $q = \frac{1}{2}\delta_{(c,\bar{x})} + \frac{1}{2}\delta_{(\bar{c},x)}$. Then $p^1 \times p^2 = q^1 \times q^2$, so $\hat{V}_t(p) = \hat{V}_t(q)$. Thus $\hat{V}_t(c, x) + \hat{V}_t(\bar{c}, \bar{x}) = \hat{V}_t(c, \bar{x}) + \hat{V}_t(\bar{c}, x)$ or, equivalently, $\hat{V}_t(c, x) = \hat{v}_t^1(c) + \hat{v}_t^2(x)$.

2. If $\{\succsim_t\}_{t\in\mathbb{N}}$ satisfies Axiom 10 (Sophistication), then $\frac{1+\gamma_{t+1}\beta_{t+1}}{1+\gamma_{t+1}} = \frac{1+\theta^v_{t,2}}{1+\theta^v_{t,1}} \leq 1$.

Proof: To prove 1, note that for all $p, q \in \Delta(C^{\mathbb{N}})$,

$$\begin{aligned}
&\big[U_t(p) > U_t(q) \text{ and } (U_{t+1}+V_{t+1})(p) > (U_{t+1}+V_{t+1})(q)\big] \\
&\qquad\qquad \implies \big[(c,\{p\}) \succ_t (c,\{q\}) \text{ and } p \succ_{t+1} q\big] \\
&\qquad\qquad \implies (c,\{p,q\}) \succ_t (c,\{q\}) \qquad\qquad \text{(by Axiom 11)} \\
&\qquad\qquad \implies (U_t+\hat{V}_t)(p) > (U_t+\hat{V}_t)(q).
\end{aligned}$$

Since U_t and $U_{t+1}+V_{t+1}$ both rank constant consumption streams $(c,c,c,\dots) \in C^{\mathbb{N}}$ in accordance with u, they are not ordinally opposed on $\Delta(C^{\mathbb{N}})$. Therefore, by Lemma 1, there exists $\alpha \in [0,1]$ such that

$$(U_t+\hat{V}_t)\big|_{\Delta(C^{\mathbb{N}})} \approx (\alpha U_t + (1-\alpha)(U_{t+1}+V_{t+1}))\big|_{\Delta(C^{\mathbb{N}})}.$$

Thus, for any $(c_0,c_1,c_2,\dots) \in C^{\mathbb{N}}$,

$$\begin{aligned}
(U_t+\hat{V}_t)(c_0,c_1,c_2,\dots) &= (u+\hat{v}^1_t)(c_0) + (\delta\hat{W}_t + \hat{v}^2_t)(c_1,c_2,\dots) \\
&= (1+\theta^v_{t,1})u(c_0) + (1+\theta^v_{t,2})\delta\hat{W}_t(c_1,c_2,\dots)
\end{aligned}$$

must be a positive affine transformation of

$$\begin{aligned}
&(\alpha U_t + (1-\alpha)(U_{t+1}+V_{t+1}))(c_0,c_1,c_2,\dots) \\
&\quad = (\alpha u + (1-\alpha)(1+\gamma_{t+1})u)(c_0) + (\alpha\delta\hat{W}_t + (1-\alpha)(1+\gamma_{t+1}\beta_{t+1})\delta\hat{W}_{t+1})(c_1,c_2,\dots) \\
&\quad = (\alpha + (1-\alpha)(1+\gamma_{t+1}))u(c_0) + (\alpha + (1-\alpha)(1+\gamma_{t+1}\beta_{t+1}))\delta\hat{W}_t(c_1,c_2,\dots),
\end{aligned}$$

where the last equality follows because $\hat{W}_t$ and $\hat{W}_{t+1}$ agree on $C^{\mathbb{N}}$. This is only possible if

$$\frac{1+\theta^v_{t,2}}{1+\theta^v_{t,1}} = \frac{\alpha + (1-\alpha)(1+\gamma_{t+1}\beta_{t+1})}{\alpha + (1-\alpha)(1+\gamma_{t+1})},$$

which implies

$$\frac{1+\gamma_{t+1}\beta_{t+1}}{1+\gamma_{t+1}} \leq \frac{1+\theta^v_{t,2}}{1+\theta^v_{t,1}} \leq 1.$$

To prove 2, note that we now have stronger restrictions on the utility functions in the representation: For all $p, q \in \Delta(C^{\mathbb{N}})$,

$$\begin{aligned}
&\big[U_t(p) > U_t(q) \text{ and } (U_{t+1}+V_{t+1})(p) > (U_{t+1}+V_{t+1})(q)\big] \\
&\qquad \iff \big[(c,\{p\}) \succ_t (c,\{q\}) \text{ and } p \succ_{t+1} q\big] \\
&\qquad \iff \big[(c,\{p\}) \succ_t (c,\{q\}) \text{ and } (c,\{p,q\}) \succ_t (c,\{q\})\big] \qquad \text{(by Axiom 10)} \\
&\qquad \iff \big[U_t(p) > U_t(q) \text{ and } (U_t+\hat{V}_t)(p) > (U_t+\hat{V}_t)(q)\big].
\end{aligned}$$

Applying Lemma 1 twice, we obtain

$$(U_t + \hat{V}_t)\big|_{\Delta(C^{\mathbb{N}})} \approx (U_{t+1} + V_{t+1})\big|_{\Delta(C^{\mathbb{N}})},$$

which implies

$$\frac{1 + \gamma_{t+1}\beta_{t+1}}{1 + \gamma_{t+1}} = \frac{1 + \theta^v_{t,2}}{1 + \theta^v_{t,1}} \leq 1,$$

as claimed. ■

Set $\hat{\gamma}_t = \theta^v_{t,1} \geq 0$. If $\hat{\gamma}_t = 0$, then set $\hat{\beta}_t \equiv 0$. Otherwise, set $\hat{\beta}_t \equiv \theta^v_{t,2}/\theta^v_{t,1} = \theta^v_{t,2}/\hat{\gamma}_t$. By Claim 6, $\theta^v_{t,2} \leq \theta^v_{t,1}$ and therefore $\hat{\beta}_t \in [0,1]$. In addition, we have $\hat{\gamma}_t\hat{\beta}_t = \theta^v_{t,2}$ in both the case of $\hat{\gamma}_t = 0$ and $\hat{\gamma}_t > 0$. Thus

$$\hat{V}_t(c,x) = \theta^v_{t,1}u(c) + \theta^v_{t,2}\delta\hat{W}_t(x) = \hat{\gamma}_t u(c) + \hat{\gamma}_t\hat{\beta}_t\delta\hat{W}_t(x),$$

so the third displayed equation in Proposition 5 is satisfied. Note also that by Claim 6,

$$\frac{1 + \hat{\gamma}_t\hat{\beta}_t}{1 + \hat{\gamma}_t} = \frac{1 + \theta^v_{t,2}}{1 + \theta^v_{t,1}} \geq \frac{1 + \gamma_{t+1}\beta_{t+1}}{1 + \gamma_{t+1}},$$

with equality if $\{\succsim_t\}_{t\in\mathbb{N}}$ satisfies Axiom 10 (Sophistication). This completes the proof of Proposition 5.

A.5.2 Proof of Theorem 3

We only show the sufficiency of the axioms.

Part 2: The assumptions in this part of the theorem are the same as in Proposition 5, except that Axiom 14 (Weak Commitment Stationarity) is replaced with the stronger condition of Axiom 7 (Stationarity). The profile of relations $\{\succsim_t\}_{t\in\mathbb{N}}$ therefore has a representation $(u, \gamma_t, \hat{\gamma}_t, \beta_t, \hat{\beta}_t, \delta)_{t\in\mathbb{N}}$ as in Proposition 5, with the additional condition that for any $p, q \in \Delta(C \times Z)$ and any $t, t' \in \mathbb{N}$,

$$U_t(p) + V_t(p) \geq U_t(q) + V_t(q) \iff U_{t'}(p) + V_{t'}(p) \geq U_{t'}(q) + V_{t'}(q).$$

Therefore, for any fixed $t \in \mathbb{N}$, setting $(u, \gamma, \hat{\gamma}, \beta, \hat{\beta}, \delta) = (u, \gamma_t, \hat{\gamma}_t, \beta_t, \hat{\beta}_t, \delta)$ and $(U, \hat{V}, V, \hat{W}) = (U_t, \hat{V}_t, V_t, \hat{W}_t)$ gives a naive quasi-hyperbolic discounting representation for $\{\succsim_t\}_{t\in\mathbb{N}}$.

Part 1: By replacing Axiom 11 (Naiveté) with the more restrictive Axiom 10 (Sophistication), Proposition 5 implies that

$$\frac{1 + \hat{\gamma}\hat{\beta}}{1 + \hat{\gamma}} = \frac{1 + \hat{\gamma}_t\hat{\beta}_t}{1 + \hat{\gamma}_t} = \frac{1 + \gamma_{t+1}\beta_{t+1}}{1 + \gamma_{t+1}} = \frac{1 + \gamma\beta}{1 + \gamma}.$$

It is therefore without loss of generality to set $\gamma = \hat{\gamma}$ and $\beta = \hat{\beta}$, giving a sophisticated quasi-hyperbolic discounting representation $(u, \gamma, \beta, \delta)$.

A.5.3 Proof of Theorem 5

We only show the sufficiency of the axioms. Note first that if $\{\succsim_t\}_{t\in\mathbb{N}}$ satisfies Axioms 1–3 and 5, then Axiom 13 (Commitment Stationarity) implies Axiom 14 (Weak Commitment Stationarity). Therefore, the profile of preferences has a representation $(u, \gamma_t, \hat{\gamma}_t, \beta_t, \hat{\beta}_t, \delta)_{t\in\mathbb{N}}$ as in Proposition 5. By Axiom 13, for any $p, q \in \Delta(C^{\mathbb{N}})$ and $t, t' \in \mathbb{N}$,

$$U_t(p) + V_t(p) \geq U_t(q) + V_t(q) \iff U_{t'}(p) + V_{t'}(p) \geq U_{t'}(q) + V_{t'}(q).$$

Thus, for any $(c_0, c_1, c_2, \dots) \in C^{\mathbb{N}}$,

$$(U_t + V_t)(c_0, c_1, c_2, \dots) = (1 + \gamma_t)u(c_0) + (1 + \gamma_t\beta_t)\sum_{i=1}^{\infty}\delta^i u(c_i)$$

must be a positive affine transformation of

$$(U_{t'} + V_{t'})(c_0, c_1, c_2, \dots) = (1 + \gamma_{t'})u(c_0) + (1 + \gamma_{t'}\beta_{t'})\sum_{i=1}^{\infty}\delta^i u(c_i),$$

which is only possible if

$$\frac{1 + \gamma_t\beta_t}{1 + \gamma_t} = \frac{1 + \gamma_{t'}\beta_{t'}}{1 + \gamma_{t'}}.$$

Thus it is without loss of generality to assume that $\gamma \equiv \gamma_t = \gamma_{t'}$ and $\beta \equiv \beta_t = \beta_{t'}$ for all $t, t' \in \mathbb{N}$.

We prove that the individual's beliefs become more accurate over time by mapping into an appropriate version of the two-period environment from Section 2 and applying the comparative naivaté result from Theorem 2.

We construct preferences over menus $\mathcal{K}(\Delta(C^{\mathbb{N}})) \subset Z$ and choice correspondences from these menus as follows: For each time period $t \in \mathbb{N}$, define an induced preference $\succsim_t^*$ over $\mathcal{K}(\Delta(C^{\mathbb{N}}))$ by $x \succsim_t^* y$ if and only if $\hat{W}_t(x) \geq \hat{W}_t(y)$ or, equivalently, $(c, x) \succsim_t (c, y)$. Define an induced choice function from menus in $\mathcal{K}(\Delta(C^{\mathbb{N}}))$ by

$$\mathcal{C}_{t+1}(x) \equiv \operatorname*{argmax}_{p\in x}\big[U_{t+1}(p) + V_{t+1}(p)\big] = \operatorname*{argmax}_{p\in x}\big[U_t(p) + V_t(p)\big],$$

where the second inequality follows because $\gamma_t = \gamma_{t+1}$ and $\beta_t = \beta_{t+1}$ imply $U_t(p) = U_{t+1}(p)$ and $V_t(p) = V_{t+1}(p)$ for all $p \in \Delta(C^{\mathbb{N}})$. By construction, $(U_t, V_t, \hat{V}_t)$ (more precisely, the restrictions of these functions to $\Delta(C^{\mathbb{N}})$) is a self-control representation for $(\succsim_t^*, \mathcal{C}_{t+1})$.

Note that, for any $p, q \in \Delta(C^{\mathbb{N}})$,

$$\big[(c, \{p, q\}) \succ_t (c, \{q\}) \text{ and } q \succ_{t+1} p\big] \iff \big[\{p, q\} \succsim_t^* \{q\} \text{ and } \mathcal{C}_{t+1}(\{p, q\}) = \{q\}\big].$$

Therefore, Axiom 12 (Diminishing Naiveté) is equivalent to $(\succsim_t^*, \mathcal{C}_{t+1})$ being more naive than $(\succsim_{t+1}^*, \mathcal{C}_{t+2})$ according to Definition 2 for all $t \in \mathbb{N}$. In addition, nontriviality implies there exist $c, c' \in C$ such that $u(c) > u(c')$. Letting $p = \delta_{(c,c,c,\dots)}$ and $q = \delta_{(c',c',c',\dots)}$, this implies $U_t(p) > U_t(q)$ and $V_t(p) > V_t(q)$. Thus $\{p\} \succ_t^* \{q\}$ and $\mathcal{C}_{t+1}(\{p,q\}) = \{p\}$ for all $t \in \mathbb{N}$, so the joint regularity condition from Theorem 2 is satisfied. We can therefore apply the theorem to conclude that, for every $t \in \mathbb{N}$, either

$$U_t + \hat{V}_t \gg_{U_t} U_{t+1} + \hat{V}_{t+1} \gg_{U_t} U_{t+1} + V_{t+1} \gg_{U_t} U_t + V_t,$$

or the individual is sophisticated at $t+1$:

$$U_{t+1} + \hat{V}_{t+1} \approx U_{t+1} + V_{t+1}.$$

Note that it should be understood in these expressions that we are referring to the restrictions of these functions to $\Delta(C^{\mathbb{N}}) \subset \Delta(C \times Z)$. Following similar arguments to those used to prove Claim 6 in the proof of Proposition 5, these conditions translate immediately to the following: Either

$$\frac{1 + \hat{\gamma}_t \hat{\beta}_t}{1 + \hat{\gamma}_t} \geq \frac{1 + \hat{\gamma}_{t+1} \hat{\beta}_{t+1}}{1 + \hat{\gamma}_{t+1}} \geq \frac{1 + \gamma_{t+1} \beta_{t+1}}{1 + \gamma_{t+1}} \geq \frac{1 + \gamma_t \beta_t}{1 + \gamma_t}$$

or

$$\frac{1 + \hat{\gamma}_{t+1} \hat{\beta}_{t+1}}{1 + \hat{\gamma}_{t+1}} = \frac{1 + \gamma_{t+1} \beta_{t+1}}{1 + \gamma_{t+1}}.$$

Since $\gamma_t = \gamma$ and $\beta_t = \beta$ for all $t \in \mathbb{N}$, in either case we have

$$\frac{1 + \hat{\gamma}_t \hat{\beta}_t}{1 + \hat{\gamma}_t} \geq \frac{1 + \hat{\gamma}_{t+1} \hat{\beta}_{t+1}}{1 + \hat{\gamma}_{t+1}} \geq \frac{1 + \gamma \beta}{1 + \gamma}.$$

This completes the proof.

A.6 Proof of Theorem 4

Similar to the proof of Theorem 5, this proof consists of mapping the recursive environment into an appropriate version the two-period environment from Section 2 and applying the comparative naiveté result in Theorem 2.

We construct preferences over menus $\mathcal{K}(\Delta(C^{\mathbb{N}})) \subset Z$ and choice correspondences from these menus as follows: For individuals $i = 1, 2$, take $U^i, \hat{V}^i, V^i, \hat{W}^i$ as in the naive quasi-hyperbolic discounting representations. Define an induced preference $\succsim_i^*$ over $\mathcal{K}(\Delta(C^{\mathbb{N}}))$ by $x \succsim_i^* y$ if and only if $\hat{W}^i(x) \geq \hat{W}^i(y)$ or, equivalently, $(c, x) \succsim_t^i (c, y)$. Define an induced choice function from menus in $\mathcal{K}(\Delta(C^{\mathbb{N}}))$ by

$$\mathcal{C}_i(x) \equiv \operatorname*{argmax}_{p \in x} \left[U^i(p) + V^i(p)\right].$$

By construction, $(U^i, V^i, \hat{V}^i)$ (more precisely, the restrictions of these functions to $\Delta(C^{\mathbb{N}})$) is a self-control representation for $(\succsim_i^*, \mathcal{C}_i)$.

Note that, for any $p, q \in \Delta(C^{\mathbb{N}})$,

$$\big[(c, \{p, q\}) \succ_t^i (c, \{q\}) \text{ and } q \succ_{t+1}^i p\big] \iff \big[\{p, q\} \succsim_i^* \{q\} \text{ and } \mathcal{C}_i(\{p, q\}) = \{q\}\big].$$

Thus $\{\succsim_t^1\}_{t\in\mathbb{N}}$ is more naive than $\{\succsim_t^2\}_{t\in\mathbb{N}}$ according to Definition 8 if and only if $(\succsim_1^*, \mathcal{C}_1)$ is more naive than $(\succsim_2^*, \mathcal{C}_2)$ according to Definition 2. In addition, joint nontriviality implies there exist $c, c' \in C$ such that $u^i(c) > u^i(c')$ for $i = 1, 2$. Letting $p = \delta_{(c,c,c,\dots)}$ and $q = \delta_{(c',c',c',\dots)}$, this implies $U^i(p) > U^i(q)$ and $V^i(p) > V^i(q)$. Thus $\{p\} \succ_i^* \{q\}$ and $\mathcal{C}_i(\{p, q\}) = \{p\}$ for $i = 1, 2$, so the joint regularity condition from Theorem 2 is satisfied. We can therefore apply the theorem to conclude that individual 1 is more naive than individual 2 if and only if either individual 2 is sophisticated or

$$U^1 + \hat{V}^1 \gg_{U^1} U^2 + \hat{V}^2 \gg_{U^1} U^2 + V^2 \gg_{U^1} U^1 + V^1. \tag{13}$$

Note that it should be understood in this expression that we are referring to the restrictions of these functions to $\Delta(C^{\mathbb{N}}) \subset \Delta(C \times Z)$.

The proof is completed by showing that Equation (13) is equivalent to the conditions in the statement of the theorem. To see this, note first that $U^2 + V^2 \gg_{U^1} U^1 + V^1$ (restricted to $\Delta(C^{\mathbb{N}})$) if and only if there exists $\alpha \in [0, 1]$ such that

$$(U^2 + V^2)\big|_{\Delta(C^{\mathbb{N}})} \approx (\alpha U^1 + (1-\alpha)(U^1 + V^1))\big|_{\Delta(C^{\mathbb{N}})}.$$

Thus, for any $(c_0, c_1, c_2, \dots) \in C^{\mathbb{N}}$,

$$(U^2 + V^2)(c_0, c_1, c_2, \dots) = (1 + \gamma^2)u^2(c_0) + (1 + \gamma^2\beta^2)\sum_{i=1}^{\infty}(\delta^2)^i u^2(c_i)$$

must be a positive affine transformation of

$$\begin{aligned}&(\alpha U^1 + (1-\alpha)(U^1 + V^1))(c_0, c_1, c_2, \dots)\\ &\qquad = \big(\alpha + (1-\alpha)(1+\gamma^1)\big)u^1(c_0) + \big(\alpha + (1-\alpha)(1+\gamma^1\beta^1)\big)\sum_{i=1}^{\infty}(\delta^1)^i u^1(c_i).\end{aligned}$$

This is equivalent to $u^1 \approx u^2$, $\delta^1 = \delta^2$, and

$$\frac{1 + \gamma^2\beta^2}{1 + \gamma^2} = \frac{\alpha + (1-\alpha)(1+\gamma^1\beta^1)}{\alpha + (1-\alpha)(1+\gamma^1)} \geq \frac{1 + \gamma^1\beta^1}{1 + \gamma^1}.$$

By analogous arguments, since $u^1 \approx u^2$ and $\delta^1 = \delta^2$,

$$\begin{aligned}U^2 + \hat{V}^2 \gg_{U^1} U^2 + V^2 &\iff \frac{1 + \hat{\gamma}^2\hat{\beta}^2}{1 + \hat{\gamma}^2} \geq \frac{1 + \gamma^2\beta^2}{1 + \gamma^2},\\ U^1 + \hat{V}^1 \gg_{U^1} U^2 + \hat{V}^2 &\iff \frac{1 + \hat{\gamma}^1\hat{\beta}^1}{1 + \hat{\gamma}^1} \geq \frac{1 + \hat{\gamma}^2\hat{\beta}^2}{1 + \hat{\gamma}^2}.\end{aligned}$$

This completes the proof.

A.7 Proof of Proposition 2

Under any value function of the form $\hat{W}(m) = Au(m) + B$ such that $A > 0$ and $B \in \mathbb{R}$, one can explicitly solve

$$\operatorname*{argmax}_{c\in[0,m]} \Big[(1+\hat{\gamma})u(c) + \delta(1+\hat{\gamma}\hat{\beta})\hat{W}(R(m-c))\Big] = \frac{1}{1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{1/\sigma}}m$$

$$\operatorname*{argmax}_{c\in[0,m]} \Big[u(c) + \delta\hat{\beta}\hat{W}(R(m-c))\Big] = \frac{1}{1+(\delta'\hat{\beta}A)^{1/\sigma}}m$$

where we define $\delta' := \delta R^{1-\sigma} < 1$. Consider first the case $\sigma \neq 1$. Then any such a value function needs to satisfy

$$\begin{aligned}
\hat{W}(m) = \max_{\hat{c}\in[0,m]} &\Big[(1+\hat{\gamma})u(\hat{c}) + \delta(1+\hat{\gamma}\hat{\beta})\hat{W}(R(m-\hat{c}))\Big] - \hat{\gamma}\max_{\tilde{c}\in[0,m]}\Big[u(\tilde{c}) + \delta\hat{\beta}\hat{W}(R(m-\tilde{c}))\Big] \\
&= (1+\hat{\gamma})\frac{1}{\left(1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{1/\sigma}\right)^{1-\sigma}}\frac{m^{1-\sigma}}{1-\sigma} \\
&\quad + \delta(1+\hat{\gamma}\hat{\beta})\left(A\frac{(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{1-\sigma/\sigma}}{\left(1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{1/\sigma}\right)^{1-\sigma}}\frac{(Rm)^{1-\sigma}}{1-\sigma} + B\right) \\
&\quad - \hat{\gamma}\frac{1}{\left(1+(\delta'\hat{\beta}A)^{1/\sigma}\right)^{1-\sigma}}\frac{m^{1-\sigma}}{1-\sigma} - \hat{\gamma}\delta\hat{\beta}\left(A\frac{(\delta'\hat{\beta}A)^{1-\sigma/\sigma}}{\left(1+(\delta'\hat{\beta}A)^{1/\sigma}\right)^{1-\sigma}}\frac{(Rm)^{1-\sigma}}{1-\sigma} + B\right) \\
&= (1+\hat{\gamma})\frac{1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{1/\sigma}}{\left(1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{1/\sigma}\right)^{1-\sigma}}\frac{m^{1-\sigma}}{1-\sigma} - \hat{\gamma}\frac{1+(\delta'\hat{\beta}A)^{1/\sigma}}{\left(1+(\delta'\hat{\beta}A)^{1/\sigma}\right)^{1-\sigma}}\frac{m^{1-\sigma}}{1-\sigma} + \delta B
\end{aligned}$$

for all $m > 0$, and thus A is a solution to the equation

$$A = (1+\hat{\gamma})\left(1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{1/\sigma}\right)^{\sigma} - \hat{\gamma}\left(1+(\delta'\hat{\beta}A)^{1/\sigma}\right)^{\sigma}, \tag{14}$$

and $B = 0$. Let $g(A)$ denote the righthand side of (14).

The case of $\sigma = 1$ can be solved analogously to obtain the equation (14), and the value of B is uniquely obtained from the value of A.

Claim 7 *Equation (14) has a unique solution $A^* \in \mathbb{R}_{++}$.*

Proof: The derivative of g is calculated as

$$g'(A) = (1+\hat{\gamma})\Big(1+\Big(\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}\delta'A\Big)^{-1/\sigma}\Big)^{\sigma-1}\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}} - \hat{\gamma}\Big(1+(\hat{\beta}\delta'A)^{-1/\sigma}\Big)^{\sigma-1}\delta'\hat{\beta}. \tag{15}$$

Note that under $\sigma < 1$, the first (resp. second) term of the righthand side of (15) is increasing (resp. decreasing) in A. Thus an upper-bound of $g'(A)$ under $\sigma < 1$ is given by

$$\lim_{A\to\infty}\left[\Big(1+\big(\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}\delta' A\big)^{-1/\sigma}\Big)^{\sigma-1}\delta'(1+\hat{\gamma}\hat{\beta})\right]-\lim_{A\to 0}\left[\Big(1+(\hat{\beta}\delta' A)^{-1/\sigma}\Big)^{\sigma-1}\delta'\hat{\gamma}\hat{\beta}\right]$$
$$=\delta'(1+\hat{\gamma}\hat{\beta})<1.$$

The second-order derivative of g is

$$g''(A)=(1+\hat{\gamma})\Big(1+\big(\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}\delta' A\big)^{-1/\sigma}\Big)^{\sigma-2}\big(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}\big)^{\frac{\sigma-1}{\sigma}}A^{-\frac{1+\sigma}{\sigma}}\frac{1-\sigma}{\sigma}$$
$$-\hat{\gamma}\Big(1+(\delta'\hat{\beta}A)^{-1/\sigma}\Big)^{\sigma-2}(\delta'\hat{\beta})^{\frac{\sigma-1}{\sigma}}A^{-\frac{1+\sigma}{\sigma}}\frac{1-\sigma}{\sigma},$$

which is proportional to

$$\left(\frac{1+\hat{\gamma}}{\hat{\gamma}}\left[\frac{1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{\frac{-1}{\sigma}}}{1+(\delta'\hat{\beta}A)^{\frac{-1}{\sigma}}}\right]^{\sigma-2}\left[\frac{\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}}{\hat{\beta}}\right]^{\frac{\sigma-1}{\sigma}}-1\right)\frac{1-\sigma}{\sigma}A^{-\frac{1+\sigma}{\sigma}}. \tag{16}$$

When $\sigma = 1$, (16) is equal to 0. When $\sigma > 1$, the sign of (16) is negative, as it can be written as

$$\left(\frac{1+\hat{\gamma}}{\hat{\gamma}}\underbrace{\left[\frac{(\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}})^{\frac{1}{\sigma}}+(\delta' A)^{\frac{-1}{\sigma}}}{\hat{\beta}^{\frac{1}{\sigma}}+(\delta' A)^{\frac{-1}{\sigma}}}\right]^{\sigma-1}}_{\geq 1}\underbrace{\left[\frac{1+(\delta'\hat{\beta}A)^{\frac{-1}{\sigma}}}{1+(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A)^{\frac{-1}{\sigma}}}\right]}_{\geq 1}-1\right)\frac{1-\sigma}{\sigma}A^{-\frac{1+\sigma}{\sigma}}.$$

Thus g is concave under $\sigma > 1$.

We now prove the unique existence of A. First observe that $\lim_{A\to\infty} g'(A) = \delta' < 1$ by (15). Thus, under any σ, there exist $\epsilon > 0$ and $\bar{A}$ such that $g'(A) \leq 1-\epsilon$ at all $A \geq \bar{A}$. This implies that $A > g(A)$ for all A sufficiently large. Given that $g(0) = 1$, the existence of a solution A^* is guaranteed by continuity of g.

If $\sigma < 1$, since $g'(A) < 1$ for all A, there cannot be another solution. If $\sigma \geq 1$, as g is concave, $g'(A^*) < 1$ at the smallest solution A^*. By concavity $g'(A) < 1$ for all $A \geq A^*$ as well, and thus there cannot be another solution. ■

The above observation implies that there exists a unique value function $\hat{W}$ that has the form of $\hat{W}(m) = Au(m) + B$ such that $A > 0$ and $B \in \mathbb{R}$. Below we prove the comparative statics results.

Claim 8 *The unique solution A^* to (14) is increasing in $\hat{\beta}$ if $\sigma < 1$, decreasing in $\hat{\beta}$ if $\sigma > 1$, and constant in $\hat{\beta}$ if $\sigma = 1$.*

Proof: As we have shown in the proof of the previous claim, $g'(A^*) < 1$. Thus, by the implicit function theorem, the unique solution A^* is increasing (resp. decreasing) in $\hat{\beta}$ if the value of $g(A)$ at each $A > 0$ is increasing (resp. decreasing) in $\hat{\beta}$. The derivative of $g(A)$ with respect to $\hat{\beta}$ can be calculated as

$$\hat{\gamma}\delta' A\left[\left(1+\left(\delta'\frac{1+\hat{\gamma}\hat{\beta}}{1+\hat{\gamma}}A\right)^{-1/\sigma}\right)^{\sigma-1} - \left(1+(\delta'\hat{\beta}A)^{-1/\sigma}\right)^{\sigma-1}\right],$$

which is positive if $\sigma < 1$, negative if $\sigma > 1$, and zero if $\sigma = 1$. ■

The actual consumption level is given by

$$c(m) = \operatorname*{argmax}_{c\in[0,m]}\left[\frac{c^{1-\sigma}}{1-\sigma} + \delta\frac{1+\gamma\beta}{1+\gamma}\hat{W}(R(m-c))\right] = \frac{1}{1+(\delta'\frac{1+\gamma\beta}{1+\gamma}A^*)^{1/\sigma}}m.$$

Thus $\lambda = \frac{1}{1+(\delta'\frac{1+\gamma\beta}{1+\gamma}A^*)^{1/\sigma}}$, which is decreasing in $\hat{\beta}$ under $\sigma < 1$, increasing in $\hat{\beta}$ under $\sigma > 1$, and constant in $\hat{\beta}$ under $\sigma = 1$. Furthermore, it is decreasing in β under any σ.

A.8 Proof of Proposition 3

For any lotteries p and q

$$\begin{aligned}\big[u(p) > u(q) \text{ and } (u+v)(p) > (u+v)(q)\big] &\implies \mathcal{C}(\{p,q\}) = \{p\} \succ \{q\} \\ &\implies \{p,q\} \succsim \{p\} \succ \{q\} \qquad \text{(Strotz naiveté)} \\ &\implies \hat{v}(p) > \hat{v}(q).\end{aligned}$$

Regularity requires that u and $u+v$ not be ordinally opposed. Therefore, Lemma 1 implies $\hat{v} \gg_u u+v$. To show necessity, note that $\hat{v} \gg_u u+v$ implies

$$\max_{p\in x}[u(p)+\hat{v}(p)] - \max_{q\in x}\hat{v}(q) \geq \max_{p\in B_{\hat{v}}(x)} u(p) \geq \max_{p\in B_{u+v}(x)} u(p),$$

which implies $x \succsim \{p\}$ for all $p \in \mathcal{C}(x)$.

A.9 Proof of Proposition 4

To show that (1) implies (3), suppose that the individual is naive and take any lotteries p, q. Then

$$\begin{aligned}
\big[u(p) > u(q) \text{ and } v(p) > v(q)\big] \quad &\Longrightarrow \quad \mathcal{C}(\{p,q\}) = \{p\} \succ \{q\} \\
&\Longrightarrow \quad u(p) > \big[u(q) \text{ and } \hat{v}(p) \geq \hat{v}(q)\big] \\
&\Longrightarrow \quad \{p,q\} \succ \{q\} \\
&\Longrightarrow \quad \hat{v}(p) \geq \hat{v}(q),
\end{aligned}$$

which implies $\hat{v} \gg_u v$ since the regularity of $(\succsim, \mathcal{C})$ ensures that u and v are not ordinary opposed. If in addition the individual is sophisticated, we can likewise show

$$\big[u(p) > u(q) \text{ and } \hat{v}(p) > \hat{v}(q)\big] \implies v(p) \geq v(q)$$

and thus $\hat{v} \approx v$.

To show that (3) implies (1), suppose that $\hat{v} \approx \alpha u + (1-\alpha)v$ for some $\alpha \in [0,1]$ and take any lotteries p, q. Then

$$\begin{aligned}
\mathcal{C}(\{p,q\}) = \{p\} \succ \{q\} \quad &\Longrightarrow \quad \big[u(p) > u(q) \text{ and } v(p) \geq v(q)\big] \\
&\Longrightarrow \quad \hat{v}(p) \geq \hat{v}(q) \\
&\Longrightarrow \quad \{p,q\} \succ \{q\},
\end{aligned}$$

and thus the individual is naive. If in addition $\hat{v} \approx v$, then we can show

$$\{p,q\} \succ \{q\} \implies \mathcal{C}(\{p,q\}) = \{p\} \succ \{q\}$$

so that the individual is sophisticated.

The equivalence between (2) and (3) follows as in Ahn, Iijima, Le Yaouanq, and Sarver (2016). (While the definition of $\hat{v} \gg_u v$ in Ahn, Iijima, Le Yaouanq, and Sarver (2016) allows for the case $v \approx -u$, this is ruled out by the regularity).

References

Ahn, D. S., R. Iijima, Y. Le Yaouanq, and T. Sarver (2016): "Behavioral Characterizations of of Naiveté for Time-Inconsistent Preferences," Working paper.

Ahn, D. S., and T. Sarver (2013): "Preference for Flexibility and Random Choice," *Econometrica*, 81, 341–361.

Ali, S. N. (2011): "Learning Self-Control," *Quarterly Journal of Economics*, 126, 857–893.

Aliprantis, C., and K. Border (2006): *Infinite Dimensional Analysis*, 3rd edition. Berlin, Germany: Springer-Verlag.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarlery Journal of Economics*, 130, 1067–1115.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

DellaVigna, S., and U. Malmendier (2006): "Paying Not to Go to the Gym," *American Economic Review*, 96, 694–719.

Dekel, E., and B. L. Lipman (2012): "Costly Self-Control and Random Self-Indulgence," *Econometrica*, 80, 1271–1302.

Freeman, D. (2016): "Revealing Naïveté and Sophistication from Procrastination and Preproperation," Working paper, Simon Fraser University.

Giné, X., D. Karlan, and J. Zinman (2010): "Put your Money where your Butt is: a Commitment Contract for Smoking Cessation," *American Economic Journal: Applied Economics*, 2, 213-235.

Gul, F., and W. Pesendorfer (2001): "Temptation and Self-Control,"*Econometrica*, 69, 1403–1435.

Gul, F., and W. Pesendorfer (2004): "Self-Control and the Theory of Consumption," *Econometrica*, 72, 119–158.

Gul, F., and W. Pesendorfer (2005): "The Revealed Preference Theory of Changing Tastes," *Review of Economic Studies*, 72, 429–448.

Harris, C., and D. Laibson (2001): "Dynamic Choices of Hyperbolic Consumers,"*Econometrica*, 69, 935–957.

Harris, C., and D. Laibson (2013): "Instantaneous Gratification," *Quarterly Journal of Economics*, 128, 205–248.

Heidhues, P., and B. Kőszegi (2009): "Futile Attempts at Self-Control," *Journal of the European Economic Association*, 7, 423–434.

Kaur, S., M. Kremer, and S. Mullainathn (2015): "Self-Control at Work," *Journal of Political Economy*, 123, 1227–1277.

Koszegi, B. (2014): "Behavioral Contract Theory," *Journal of Economic Literature*, 52, 1075–1118.

Kopylov, I. (2012): "Perfectionism and Choice," *Econometrica*, 80, 1819–1943.

Krusell, P., B. Kuruşçu, and A. Smith (2002): "Time Orientation and Asset Prices," *Journal of Monetary Economics*, 49, 107–135.

Krusell, P., B. Kuruşçu, and A. Smith (2010): "Temptation and Taxation," *Econometrica*, 78, 2063–2084.

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

Le Yaouanq, Y. (2015): "Anticipating Preference Reversal," Working paper number TSE-585, Toulouse School of Economics.

Lipman, B., and W. Pesendorfer (2013): "Temptation," in Acemoglu, Arellano, and Dekel, eds., *Advances in Economics and Econometrics: Tenth World Congress*, Volume 1, Cambridge University Press.

Noor, J. (2011): "Temptation and Revealed Preference," *Econometrica*, 79, 601–644.

O'Donoghue, T., and M. Rabin (2001): "Choice and Procrastination," *Quarterly Journal of Economics*, 116, 121–160.

Peleg, M., and M. E. Yaari (1973): "On the Existence of a Consistent Course of Action when Tastes are Changing," *Review of Economic Studies*, 40, 391–401.

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

Shui, H., and L. M. Ausubel (2005): "Time Inconsistency in the Credit Card Market," Working paper, University of Maryland.

Spiegler, R. (2011): *Bounded Rationality and Industrial Organization.* New York, NY: Oxford University Press.

Stovall, J. (2010): "Multiple Temptations," *Econometrica*, 78, 349–376.

Toussaert, S. (2016): "Eliciting Temptation and Self-Control through Menu Choices: A Lab Experiment," Working paper.