Comparative Measures of Naiveté[*]

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

June 4, 2015

Abstract

We propose nonparametric definitions of absolute and comparative naiveté for dynamically inconsistent preferences. These definitions leverage ex-ante choice of menu to identify predictions of future behavior and ex-post choice from menus to identify actual behavior. Naifs prefer the flexibility of a menu to committing to their eventual choices, mistakenly anticipating more virtuous behavior. More naive individuals are more optimistic about their future behavior and demand more flexibility, yet are less virtuous in their actual choices. For Strotzian preferences, our definitions impose linear restrictions relating anticipated and actual temptation utilities to the virtuous utility, and yield further intuitive parametric restrictions in particular specifications such as quasi-hyperbolic discounting. We provide suitable definitions for random choice. Finally, we apply our definitions to understand the welfare implications of commitment devices.

Keywords: Naive, sophisticated, time inconsistent, comparative statics

JEL Classification: D03, D81, D84

[*] Ahn and Sarver acknowledge the financial support of the National Science Foundation through Grants SES-1357719 and SES-1357955. We also thank Ned Augenblick, Drew Fudenberg, David Laibson, Bart Lipman, Jawwad Noor, Matthew Rabin, Philipp Sadowski, Tomasz Strzalecki, Norio Takeoka, and seminar participants at Arizona State, Berkeley, Caltech, Columbia, Duke, Harvard, London School of Economics, Princeton, Queen Mary, UC Riverside, Rochester, and Washington University in St. Louis for helpful comments and discussions.

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

[‡] Department of Economics, Harvard University, Littauer Center G20, Cambridge, MA 02138. Email: riijima@fas.harvard.edu

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

1 Introduction

Models of dynamic inconsistency play an important role in a wide-ranging set of applications in economics. Of particular recent interest are the implications of naiveté when individuals mispredict their future behavior.[1] While naiveté often yields surprising and significant consequences, its effects are usually understood within the context of a specific utility representation of behavior. For example, O'Donoghue and Rabin (2001) generalize the standard (β, δ) model of sophisticated quasi-hyperbolic discounting by adding an additional present-bias parameter $\hat{\beta} \geq \beta$ to capture the individual's possibly naive belief about her future present-bias. The resulting $(\beta, \hat{\beta}, \delta)$ model suggests natural comparisons of naiveté through its parameters, such as larger values of $\hat{\beta}$ intuitively corresponding to more naive individuals.

This parametric evaluation of naiveté relies on a particular utility function. In this paper, we provide general nonparametric definitions of naiveté and sophistication that are divorced from functional form assumptions. Our definitions are worded directly on choice primitives, and not on the components of a utility function. To understand the usefulness of our contribution, as an analogy consider the definition of risk aversion as having a concave utility index for wealth $v'' < 0$. This definition makes sense only under the expected-utility hypothesis, as otherwise the invoked v does not exist. But the notion of risk aversion seems more fundamental, since many people might not maximize expected utility but still consider themselves risk-averse. A more basic and satisfying criterion for risk aversion is an individual's preference for certainly receiving the expected value of a monetary lottery rather than exposing herself to its uncertainty. This definition is workable without assuming any utility representation and can be directly tested without estimating the parameters of a structured model. Understanding naiveté through the parameter $\hat{\beta}$ of the quasi-hyperbolic discounting model has similar limitations relative to a definition of naiveté phrased on choices. Our paper takes steps to overcome these limitations.

Beyond improved theoretical foundations, model-free definitions of naiveté provide relevant substantive benefits. They permit an examination of which positive predictions in applications rely on functional-form assumptions and which predictions are inherent features of naiveté. For example, a more risk-accepting investor will always choose a risky equity position over a risk-free bond whenever a more risk-averse investor does.

[1] A recent survey of empirical applications can be found in Section 2.1 of DellaVigna (2009) and a survey of some theoretical applications in contract theory can be found in Koszegi (2014).

Similarly, we can ask whether predictions regarding savings or procrastination are artifacts of an assumed utility or are robust implications of naiveté. In turn, a deeper understanding of the mechanics of naive choice also improves normative analysis. In particular, effective design of commitment devices can hinge crucially on the assumed level of sophistication. Duflo, Kremer, and Robinson (2011) examine a theoretical model where the optimal timing of when to offer a commitment depends on whether individuals are sophisticated or naive regarding the degree of their present bias, and provide evidence from Kenyan fertilizer adoption that individuals are naive and would benefit from earlier and time-limited commitments. More general definitions of naiveté provide a language broad enough to understand the consequence of policy interventions when citizens have qualitatively different forms of naiveté and are best approximated by a variety of formal models, and which policies work for which assumed models.

We propose definitions both for testing the naiveté of a single agent and for comparing levels of naiveté across agents. We leverage two pieces of choice data. First, we use preference for commitment to measure *anticipated* behavior from an ex-ante perspective before the realization of temptation. Formally, the individual's preferences over different menus captures her demand for commitment and allows an inference of her beliefs regarding her future behavior. Second, fixing a level of commitment, we use choices from availabilities to measure *actual* behavior from an ex-post perspective under the influence of temptation.

To test naiveté and sophistication, our definition compares the agent's predicted value for a set x of different options against her actual ex-post choice $\mathcal{C}(x)$ from that menu x. An individual is sophisticated if she is indifferent between maintaining the flexibility to choose from x later or committing to her eventual choice $\mathcal{C}(x)$ now, i.e., if $x \sim \{\mathcal{C}(x)\}$ from her ex-ante perspective. If she is naive, she believes that she will make a more virtuous choice, so prefers to maintain the flexibility in x, i.e., if $x \succsim \{\mathcal{C}(x)\}$. For example, a naive gym member believes she will work out more than she actually will, and mistakenly pays for a membership that provides the flexibility to visit the gym numerous times.

To compare naiveté across agents, our definition again evaluates ex-post and ex-ante behavior. Comparative naiveté has two components. First, the more naive agent, say individual 1, is more optimistic about her future virtue than the more sophisticated agent, say individual 2. This is measured by comparing her relative demand for commitment: the naive agent's optimism means she is less willing to commit to single options now, while the more sophisticated (and less optimistic) agent is more eager to make commitments to avoid future temptation. Formally, $\{p\} \succ_2 x$ whenever $\{p\} \succ_1 x$, where $\{p\}$ is a commitment to consume p in the future. Second, the more naive agent is actually less virtuous. This is measured by her choice from x. Formally, $\{p\} \succ_1 \{\mathcal{C}_1(x)\}$ whenever $\{p\} \succ_2 \{\mathcal{C}_2(x)\}$. A more naive gym member believes she will work out more often, but actually works out less, than the more sophisticated gym member.

The proposed definitions characterize sharp and intuitive functional inequalities in a variety of special models. Consider the Strotzian model of dynamic inconsistency, where ex-ante normative behavior is dictated by one utility function u while ex-post behavior under temptation is dictated by another utility function v. To allow for naiveté, the individual believes that $\hat{v}$, possibly different than v, governs her future behavior. Our definition of absolute naiveté implies that the believed $\hat{v}$ is a linear combination of the virtuous utility u and the actual temptation utility v. Our definition of comparative naiveté implies that more naive individuals have anticipated temptation utilities that put more weight on u, but actual temptation utilities that put less weight on u. For more structured specifications of the Strotz utilities, the definitions continue to yield interpretable restrictions. In the case of consumption over time dictated by the $(\beta, \hat{\beta}, \delta)$ model of O'Donoghue and Rabin (2001), our definition of absolute naiveté implies that $\hat{\beta} \geq \beta$ and our definition of comparative naiveté implies that if individual 1 with parameters $(\beta_1, \hat{\beta}_2, \delta_1)$ is more naive than individual 2 with parameters $(\beta_1, \hat{\beta}_2, \delta_2)$, then $\delta_1 = \delta_2$ and $\hat{\beta}_1 \geq \hat{\beta}_2 \geq \beta_2 \geq \beta_1$.

We present suitable generalizations of these definitions for the case of random choice under uncertain temptations. The only required variation from the deterministic case is that the determinate choice from a menu is replaced with the average choice under a random choice rule. The random Strotz model of Dekel and Lipman (2012) allows for future temptation to be stochastically realized from a set of multiple temptations. For example, an individual on a diet might be tempted to eat salty snacks or might be tempted to eat sweet desserts. For this model, the random analogs of naiveté and sophistication yield suitable generalized implications: the more naive individual's probability over her future temptations is more optimistic, in the sense that it puts more likelihood on less intense temptations. Note that allowing for randomization is important, even if actual ex-post choice is deterministic, as long as individuals are uncertain about their future behavior. The generalization to random choice has substantive significance, as models of naiveté with uncertainty are common in applications. For example, the predictions in Duflo, Kremer, and Robinson (2011) depend on farmers being uncertain, as well as naive, of their future present-bias.

Our use of ex-ante and ex-post behavior has several precedents. In fact, empirical studies of naiveté also invoke ex-ante and ex-post observations. Purchases of gym memberships in DellaVigna and Malmendier (2006) are taken to occur ex ante and before the experience of temptation, while the visits to the gym are taken to occur ex post when facing the temptation to shirk from exercise. Shui and Ausubel (2004) observe consumers' choices of credit card contracts (assumed ex ante) and their subsequent borrowing behavior (assumed ex post). In perhaps the closest existing match to our primitives, Augenblick, Niederle, and Sprenger (2013) report an experiment where subjects can choose to commit to levels of work effort ex ante and then exert actual effort ex post. These two-tiered observations allow structural estimation of the parameters within specific models of naive choice. With our proposed definitions, the hypothesis that one person is naive or is more naive than another can be rejected from a few choice observations without the need for calibrated parameters.[2]

There are also papers in decision theory that use behavior at different time periods to capture sophistication. Lipman and Pesendorfer (2013) provide a survey of these papers. Noor (2011) considers preferences over a recursive domain that includes ex-ante and ex-post choice preferences as projections and pioneered the approach of using temporal choice as a domain for explicitly testing the sophistication implicitly assumed in most ex-ante axiomatic models of temptation. Kopylov (2012) relaxes Noor's sophistication condition and considers agents who choose flexibility ex ante that is subsequently unused ex post. Kopylov exchews mistaken or naive beliefs, but rather interprets the relaxation of sophistication as reflecting a direct psychic benefit of maintaining positive self-image. These papers are discussed more specifically after we introduce our definitions. Finally, Dekel and Lipman (2012) observe that ex-ante and ex-post choice can be combined to empirically distinguish random Strotz representations from others that involve costly self-control. Much of the technical apparatus from Dekel and Lipman (2012) ends up being useful in studying naiveté, as we will explain in the body of the paper. Finally, a recent independent paper by Le Yaouanq (2015), conceived and executed without awareness of our work, studies similar primitives but proposes a different definition of naiveté, with a more explicit eye towards experimental

[2] A recent experiment by Augenblick and Rabin (2015) incentivized direct reports of subjects' predictions of future behavior.

data collection.

The next section introduces our formal primitives. Section 3 considers the specialized case of deterministic choice and introduces appropriate definitions of naiveté and sophistication. Their implications are characterized for the Strotz model of dynamic inconsistency. As specific applications, we consider the Strotzian versions of quasi-hyperbolic discounting and of more general diminishing impatience. Section 4 considers the case of random choice and characterizes the implications of naiveté for the random Strotz model of Dekel and Lipman (2012). There, we consider as applications the model of Eliaz and Spiegler (2006) and the quasi-hyperbolic discounting model with uncertain present bias. Section 5 examines a model where the individual is offered a selection of commitment devices, and the consequent welfare implications of naiveté.

2 Primitives

We study a two-stage model with an agent who initially decides a menu of several options, and subsequently selects a particular option from that menu.

Let C denote a compact and metrizable space of outcomes. Let $\Delta(C)$ denote the set of lotteries (countably-additive Borel probability measures) over C, with typical elements $p, q, \ldots \in \Delta(C)$. When it causes no confusion, we will slightly abuse notation and write c in place of the degenerate lottery $\delta_c \in \Delta(C)$. Finally, let $\mathcal{K}(\Delta(C))$ denote the family of nonempty compact subsets of $\Delta(C)$ with typical elements $x, y, \ldots \in \mathcal{K}(\Delta(C))$. So $\mathcal{K}(\Delta(C))$ is a family of menus of lotteries. An *expected-utility function* is a continuous function $u : \Delta(C) \to \mathbb{R}$ such that $u(\alpha p + (1-\alpha)q) = \alpha u(p) + (1-\alpha)u(q)$ for all lotteries p, q. A function is *nontrivial* if it is not constant. We write $u \approx v$ when u and v are expected-utility functions and u is a positive affine transformation of v. For fixed expected-utility function u and menu x, let $B_u(x) \equiv \operatorname{argmax}_{p \in x} u(p)$.

We consider a pair of behavioral primitives. The first primitive is a preference relation $\succsim$ on $\mathcal{K}(\Delta(C))$, with indifference $\sim$ and strict preference $\succ$ defined in the standard manner. The behavior encoded in $\succsim$ is taken before the direct experience of temptation but while (possibly incorrectly) anticipating its future occurrence. The second primitive is a random choice rule $\lambda : \mathcal{K}(\Delta(C)) \to \Delta(\Delta((C))$ such that $\lambda^x(x) = 1$, where $\Delta(\Delta(C))$ denotes the space of lotteries over $\Delta(C)$. The behavior encoded in λ is taken while experiencing temptation. For each $x \in \mathcal{K}(\Delta(C))$, λ^x is a probability measure over lotteries, with $\lambda^x(y)$ denoting the probability of choosing a lottery in the set $y \subset x$ when the choice set is the menu x. We refer to the first stage of choice of a menu as occurring "ex ante" and the second stage of choice from a menu as occurring "ex post," that is, before and after the realization of temptation. For example, the purchased gym contracts in DellaVigna and Malmendier (2006) correspond to an ex-ante choice of a menu while the observed number of gym visits corresponds to an ex-post choice from that menu.

At points, we will specialize to choice functions without randomization for their substantive importance and expositional clarity. A random choice function λ is *deterministic* if λ^x is degenerate for all menus x, that is, $\lambda^x = \delta_p$ where δ_p is the Dirac measure supported on p. Identifying the Dirac measure δ_p with p itself, we can notate λ as a standard choice function $\mathcal{C} : \mathcal{K}(\Delta(C)) \to \Delta(C)$.[3] In that case, $\mathcal{C}(x) = p$ for $\delta_p = \lambda^x$.

These primitives echo prior work by Ahn and Sarver (2013) on unforeseen contingencies. That paper inferred unawareness of future taste contingencies by comparing choices before and after the realization of those contingencies. Observing ex-ante demand for flexibility and ex-post exercise of flexibility can reveal unawareness and provide positive foundations for the measurement of an unforeseen contingency, while the standard approach of using only ex-ante preferences cannot. Similarly, here we use demand for commitment in the first stage and then indulgence of temptation in the second stage to infer naiveté. Under-demand for flexibility reveals unawareness of future taste contingencies, while under-demand for commitment reveals naiveté about future temptations.

3 Deterministic choice

We begin with the relatively more straightforward case of choice without randomization before proceeding to the general case with random choices in the next section. Throughout this section we assume a deterministic choice function $\mathcal{C}$. We begin by proposing definitions of sophistication and naiveté for a single individual.

[3] Recall the final outcomes are themselves lotteries. The determinacy here is in the sense that the decision maker does not randomize her selection among these lotteries.

Definition 1. *An individual is* sophisticated *if, for all menus x,*

$$x \sim \{\mathcal{C}(x)\}.$$

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

$$x \succsim \{\mathcal{C}(x)\}.$$

We will say the individual is *strictly naive* if she is naive and unsophisticated.

A sophisticated individual correctly anticipates her choice $\mathcal{C}(x)$ from x. A naive individual erroneously values the option to make more virtuous choices, envisioning her virtuous selections. Inferring sophistication from $x \sim \{\mathcal{C}(x)\}$ assumes that the individual evaluates menus in a consequentialist manner, that is, the individual is indifferent between committing to her (correctly) anticipated choice $\mathcal{C}(x)$ from x at the ex-ante stage or selecting the menu x with the belief that she will choose $\mathcal{C}(x)$ ex post. Put differently, adding an option to a menu is important only if the individual anticipates choosing that option. This assumption is violated if unchosen options from a menu affect well-being. For example, an individual who exerts costly willpower to avoid choosing tempting options as in Gul and Pesendorfer (2001) does not evaluate a menu only by its choice consequences. In this case, she may strictly prefer to remove these unchosen temptations.[4]

In principle, an opposite violation of sophistication where $\{\mathcal{C}(x)\} \succ x$ and individuals over-estimate their future self-control problems is also possible.[5] Many of the following results have analogous statements for this case, and Appendix C records some of those analogs. This case receives less attention and seems less empirically relevant, so we restrict attention in the main paper to traditional naiveté.

We now compare naiveté across individuals. This comparison invokes two conceptually distinct parts. The first considers individuals' ex-ante views of their future behavior. In environments with temptation, an ex-ante desire for commitment is often interpreted as a signal of anticipated temptation. The following comparison of desire for commitment is a slight variation of the comparison introduced by Dekel and Lipman

[4] In a companion paper, we explore alternative definitions of sophistication and naiveté that can be applied to individuals who anticipate exerting costly self-control to resist tempting options.

[5] Ali (2011) shows that such a pessimistic belief can arise and persist in a model of Bayesian experimentation.

(2012).[6]

Definition 2. *Individual* 2 *is* more temptation averse *than individual* 1 *if, for all menus x and lotteries p,*

$$\{p\} \succ_1 x \implies \{p\} \succ_2 x.$$

This definition compares individuals' ex-ante demand for commitment to singletons, with more temptation averse individuals exhibiting higher demand for commitment. That is, if a less temptation averse person strictly prefers to commit to consuming the lottery $\{p\}$, then the more temptation averse person also prefers to commit to $\{p\}$. The more temptation averse individual therefore anticipates a higher value for commitment, while the less temptation averse individual has a more optimistic view of her future behavior. Note that comparability of temptation aversion guarantees that both agents share common preferences over singleton menus, so their virtuous tastes are identical.

The second component of the comparison considers ex-post behavior after the realization of temptation. The following comparison concerns individuals' ex-post choices from menus.

Definition 3. *Individual* 2 *is* more virtuous *than individual* 1 *if, for all menus x and lotteries p,*

$$\{p\} \succ_2 \{\mathcal{C}_2(x)\} \implies \{p\} \succ_1 \{\mathcal{C}_1(x)\}.$$

The more virtuous individual makes better choices from all menus: if the less virtuous agent makes choices from the menu x that are normatively superior to p, as reflected in her ex-ante commitment preference, then the more virtuous agent also makes normatively superior choices from x. Also note that Definition 3 implies both individuals share common preferences over singleton commitments, that is, their virtuous preferences are identical.

Absolute naiveté for a single individual involves both the ex-ante and ex-post perspectives, since a naive individual's ex-ante belief diverges from her ex-post behavior. Correspondingly, comparisons of naiveté across different individuals involves both perspectives as expressed in the prior two definitions: a more naive individual is more optimistic about her future behavior ex ante (that is, she is less temptation averse) yet less disciplined in her actual behavior ex post (that is, she is less virtuous).

[6] The formal definition also appears with different interpretations in Ahn (2007) and Sarver (2008).

Definition 4. *Suppose individuals* 1 *and* 2 *are naive. Individual* 1 *is* more naive *than individual* 2 *if individual* 2 *is more temptation averse and more virtuous than individual* 1*.*

The ubiquitous Strotz model of dynamic inconsistency offers a general application for these concepts. A sophisticated Strotz individual is specified by two preferences. The first is her ex-ante commitment preference over future consumption, as represented by the utility function u. The second is her temptation preference that governs her actual consumption choices at the ex-post stage, as represented by the utility function v. Naivete requires divergence between believed and actual consumption. Specification of a naive Strotzian consumer therefore requires a third preference to capture her possibly erroneous beliefs about her future behavior, as represented by the utility function $\hat{v}$.

Definition 5. *A* Strotz representation *of* $\succsim$ *is a pair* $(u, \hat{v})$ *of nontrivial expected-utility functions such that the function* $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ *defined by*

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

is a utility representation of $\succsim$*.*[7]

While she anticipates she will maximize $\hat{v}$, a naive Strotzian agent's ex-post behavior $\mathcal{C}$ actually maximizes v.

Definition 6. *A* Strotz representation *of* $\mathcal{C}$ *is a pair* (u, v) *of nontrivial expected-utility functions such that*

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

For convenience, we will often combine these two representations and refer to the unified representation for both ex-ante and ex-post choice.

Definition 7. *A* Strotz representation *of* $(\succsim, \mathcal{C})$ *is a triple* $(u, v, \hat{v})$ *such that* $(u, \hat{v})$ *is a Strotz representation of* $\succsim$ *and* (u, v) *is a Strotz representation of* $\mathcal{C}$*.*

The following results demonstrate that the basic definition of naiveté characterizes sharp parametric restrictions on $\hat{v}$ and v. A naive individual believes that her future behavior will be more virtuous than it actually is. Converting this intuition to the parameters of the Strotz model, this means that the anticipated utility $\hat{v}$ is more

[7] Recall $B_{\hat{v}}(x)$ was defined as $\operatorname{argmax}_{q \in x} \hat{v}(q)$.

aligned with the commitment utility u than the actual utility v that will govern future consumption. The structure of the alignment is particular: $\hat{v}$ is a linear combination of u and v, that is, $\hat{v} = \alpha u + (1 - \alpha)v$. The belief $\hat{v}$ puts additional unjustified weight on the normative utility u, but aggregates u with v in a linear manner. This excludes the case where the believed temptation is orthogonal to the actual temptation; for example, this excludes the case where the individual will be tempted to indulge in sweet treats but believes she will be tempted to indulge in salty treats. This structure also relies crucially on the linear structure of the domain of lotteries and the assumed expected-utility functions.

A natural parameterization of comparative naiveté evaluates agents' weights on the normative utility u and the temptation utility v.

Definition 8. *Let u, v, v' be expected-utility functions. Then v is* more u-aligned *than v', written as $v \gg_u v'$, if either $v \approx \alpha u + (1 - \alpha)v'$ for some $\alpha \in [0, 1]$ or $v' \approx -u$.*

An expected-utility function v is more u-aligned than v' if it puts additional weight on u that is not included in v'. It is also more u-aligned than v' if v' is maximally misaligned with u, that is, if v' is exactly $-u$.[8]

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

As mentioned, naiveté corresponds with unwarranted linear alignment of the believed utility $\hat{v}$ with the virtuous utility u. The individual believes her future choices will maximize a convex combination of her actual utility v and her virtuous utility u. Note that the linear structure is a consequence of the definition of naiveté, and is not assumed a priori.

Then comparative naiveté imposes linear restrictions across the agents' ex-ante predictions as captured by $\hat{v}_1, \hat{v}_2$ and ex-post behaviors as captured by v_1, v_2.

Theorem 2. *Suppose $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ have Strotz 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 $u_1 \approx u_2 \equiv u$ and*

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

[8] The special exception for this boundary case is to avoid tedious exceptions in the following characterization theorems.

While they share common normative preferences over singleton commitments, individual 1 is more optimistic about her future behavior than individual 2. This is reflected in the requirement $\hat{v}_1 \gg_u \hat{v}_2$. However, individual 1's actual ex-post choices are even less virtuous than person 2's choices, as reflected in the requirement $v_1 \ll_u v_2$. So to be more naive, an individual must simultaneously be more optimistic about her future virtuous behavior while actually exercising less virtue. As illustrated in Figure 1, comparative naiveté implies that all of the believed and actual temptation utilities are convex combinations of the shared commitment utility u and the more naive individual's actual temptation v_1. As in the case of absolute naiveté, the linear relationship between the parameters is a consequence of the definition of comparative naiveté.

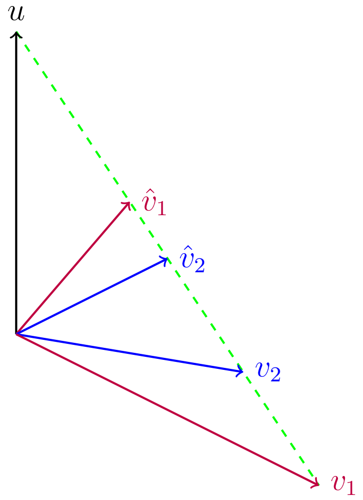

Figure 1: Alignment of believed and actual utilities implied by comparative naiveté

Noor (2011) introduces the following definition of sophistication: if $x \cup \{p\} \succ x$, then $\mathcal{C}(\{p, q\}) = p$ for all $q \in x$.[9] While Noor's definition is generally distinct from Definition 1, the two definitions are equivalent in the Strotzian case. Noor (2011) is interested in providing foundations for pure sophistication, so does not explore naiveté or comparative statics. Kopylov (2012) proposes the following relaxation of Noor's sophistication axiom: if $x \cup \{p\} \succ x$, then $\mathcal{C}(\{p, q\}) = p$ for all $q \in x$ or $\{p\} \succ \{q\}$ for all $q \in x$. That is, even if p is not chosen over other elements ex post, p may still add value to the menu x because it is a more virtuous option. Kopylov's interpretation is not in terms of mistaken predictions about future behavior, but rather in terms of the

[9] This is an equivalent formulation of Noor's axiom used by Lipman and Pesendorfer (2013).

direct non-consequentialist welfare effects of maintaining the availability of unchosen but virtuous options. The difference in interpretation is perhaps most salient when considering Kopylov's notion of comparative perfectionism. His comparison exclusively invokes ex-ante preferences, whereas we argue that comparative naiveté must invoke both ex-ante and ex-post choice. Finally, both Noor (2011) and Kopylov (2012) are interested in the more general self-control preferences of Gul and Pesendorfer (2001), whereas we restrict attention to the Strotzian case.[10]

3.1 Application: Quasi-Hyperbolic Preferences

As a specific application of the previous equivalences, we consider the quasi-hyperbolic model of time inconsistency parameterized by the present-bias factor β. Let $C = [a,b]^{\mathbb{N}}$ be a set of infinite-horizon consumption streams, with elements $c = (c_1, c_2, \dots) \in C$.[11] A lottery $p \in \Delta(C)$ resolves immediately and yields a consumption stream. We focus on the simple case with one-shot resolution of uncertainty for expositional parsimony, but all of the following results generalize to richer settings that incorporate temporal lotteries or true dynamic choice.[12] In more general dynamic environments, simple atemporal lotteries over consumption streams provide sufficient choice observations to generate the following comparative statics.

Suppose the commitment preference over random consumption streams is represented by an expected-utility function whose values $U(c) = U(\delta_c)$ over deterministic streams (that is, whose Bernoulli utility indices) comply with exponential discounting

$$U(c) = \sum_{t=1}^{\infty} \delta^{t-1} u(c_t) \tag{1}$$

for some instantaneous utility function $u : [a,b] \to \mathbb{R}$. The decision-maker understands that she will suffer present bias as parameterized in the quasi-hyperbolic discounting model, but underestimates its magnitude. Specifically, her choice from a menu of

[10] We do explore models with costly self-control in a companion paper.

[11] The product topology on C is compact and metrizable.

[12] Kreps and Porteus (1978) were the first to provide a complete analysis of dynamic choice with uncertainty that resolves gradually through time (i.e., temporal lotteries). The models of temptation in Gul and Pesendorfer (2004) and Noor (2011) used an infinite horizon version of such a setting.

consumption streams maximizes the expected-utility function defined by

$$V(c) = u(c_1) + \beta \sum_{t=2}^{\infty} \delta^{t-1} u(c_t). \tag{2}$$

On the other hand, she anticipates her choice from a menu of consumption streams will maximize the expected-utility function defined by

$$\hat{V}(c) = u(c_1) + \hat{\beta} \sum_{t=2}^{\infty} \delta^{t-1} u(c_t). \tag{3}$$

The individual's ex-ante behavior reflects an optimistic belief that her future present-bias parameter is $\hat{\beta}$, yet her ex-post behavior actually uses the present-bias parameter β.

Definition 9. *A* quasi-hyperbolic (QH) representation *of* $(\succsim, \mathcal{C})$ *is a quadruple* $(u, \beta, \hat{\beta}, \delta)$ *of a continuous and nontrivial function* $u : [a, b] \to \mathbb{R}$ *and scalars* $\beta, \hat{\beta} \in (0, 1]$ *and* $\delta \in (0, 1)$*, such that* $(U, V, \hat{V})$ *is a Strotz representation for* $(\succsim, \mathcal{C})$*, where* U*,* V*, and* $\hat{V}$ *satisfy Equations* (1)*,* (2)*, and* (3)*.*

The behavioral definition of naiveté implies the intuitive restriction that $\hat{\beta} \geq \beta$.

Corollary 1. *Suppose* $(\succsim, \mathcal{C})$ *has a QH representation* $(u, \beta, \hat{\beta}, \delta)$*. Then the individual is naive if and only if* $\hat{\beta} \geq \beta$ *(and is sophisticated if and only if* $\hat{\beta} = \beta$*).*

Proof. By Theorem 1, the individual is naive if and only if $\hat{V} \approx \alpha U + (1 - \alpha)V$ for some $\alpha \in [0, 1]$, where the functions U, V, and $\hat{V}$ satisfy Equations (1), (2), and (3), respectively. Note that

$$\alpha U(c) + (1 - \alpha)V(c) = u(c_1) + (\alpha + (1 - \alpha)\beta) \sum_{t=2}^{\infty} \delta^{t-1} u(c_t).$$

Since the term $u(c_1)$ appears in this expression without any scalar multiple, conclude that $\hat{V} = \alpha U + (1 - \alpha)V$ and hence $\hat{\beta} = \alpha + (1 - \alpha)\beta \in [\beta, 1]$. ■

A similar application of the characterization of comparative naiveté in Theorem 2 provides another set of intuitive comparative restrictions. First, the more naive individual has more optimistic beliefs about her future patience: $\hat{\beta}_1 \geq \hat{\beta}_2$. Second, the more naive individual's behavior is more present-biased: $\beta_1 \leq \beta_2$. These restrictions are also necessary, so the comparison of alternative statistics such as $\hat{\beta}_i - \beta_i$ or $\hat{\beta}_i / \beta_i$ will not work. For example, $\hat{\beta}_1 - \beta_1 \geq \hat{\beta}_2 - \beta_2$ is generally insufficient to guarantee individual 1 is more naive than individual 2 without knowing that $\hat{\beta}_1 \geq \hat{\beta}_2$ and $\beta_1 \leq \beta_2$.

Corollary 2. *Suppose $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ have QH representations $(u_1, \beta_1, \hat{\beta}_1, \delta_1)$ and $(u_2, \beta_2, \hat{\beta}_2, \delta_2)$. Then individual 1 is more naive than individual 2 if and only if $u_1 \approx u_2$, $\delta_1 = \delta_2$, and $\hat{\beta}_1 \geq \hat{\beta}_2 \geq \beta_2 \geq \beta_1$.*

The proof is entirely analogous to that of Corollary 1, hence omitted.

3.2 Application: Diminishing Impatience

The analysis of the quasi-hyperbolic representation in the previous section extends to more general patterns of discounting, such as true hyperbolic discounting. We now relate several properties of discount functions to properties of the perceived discount functions for individuals who satisfy our definition of naiveté. While the prior section corroborates the existing parameter restriction $\hat{\beta} \geq \beta$ for naiveté with quasi-hyerbolic discounting, the analogous formulation for general diminishing impatience is less understood.[13] This section introduces the appropriate restrictions for general discounting while elucidating a common theme through our nonparametric notion of naiveté. That is, the definition of naiveté is useful not only in verifying existing parametric formulations of naiveté, but also in generating novel formulations for less-studied models. As a side benefit, the analysis also uncovers structural relationships between temporal rates of substitution in the anticipated discounting and the impatience of the actual discounting, providing new observationally equivalent tests of diminishing impatience in consumption over time.

Say that $D : \mathbb{N} \cup \{0\} \to (0, 1]$ is a *discount function* if $D(0) = 1$ and $\sum_{t=0}^{\infty} D(t) < \infty$. Suppose as before that consumption in periods $t = 1, 2, \ldots$ is given by $(c_1, c_2, \ldots) \in C = [a, b]^{\mathbb{N}}$. Period 0 commitment preferences over deterministic consumption streams starting in period 1 are represented by

$$U(c) = \sum_{t=1}^{\infty} D(t) u(c_t). \tag{4}$$

[13] Prelec (2004) studies the degree of time inconsistency for a single discount function D, as captured by log-concavity. He suggests this as a criterion for evaluating sophistication, but this approach is clearly conceptually remote from our notion of sophistication that relies on comparing D with a believed discount funtion $\hat{D}$.

Suppose that in period 0 the individual believes that she will apply the discount function $\hat{D}$ in the subsequent period, which yields the following anticipated temptation utility for deterministic consumption streams:

$$\hat{V}(c) = \sum_{t=1}^{\infty} \hat{D}(t-1)u(c_t). \tag{5}$$

In reality, suppose preferences over consumption streams are stationary, and period 1 choices actually maximize

$$V(c) = \sum_{t=1}^{\infty} D(t-1)u(c_t). \tag{6}$$

Definition 10. *A* discounting representation *of* $(\succsim, \mathcal{C})$ *is a triple* $(u, D, \hat{D})$ *of a continuous and nontrivial function* $u : [a,b] \to \mathbb{R}$ *and discount functions* D *and* $\hat{D}$*, such that* $(U, V, \hat{V})$ *is a Strotz representation for* $(\succsim, \mathcal{C})$*, where* U*,* $\hat{V}$*, and* V *satisfy Equations* (4)*,* (5)*, and* (6)*.*

The quasi-hyperbolic representations discussed in the previous section are special cases of the discounting representations where

$$D(t) = \begin{cases} 1 & \text{if } t = 0 \\ \beta\delta^t & \text{if } t > 0. \end{cases}$$

and

$$\hat{D}(t) = \begin{cases} 1 & \text{if } t = 0 \\ \hat{\beta}\delta^t & \text{if } t > 0. \end{cases}$$

Two general properties of discount functions will be important.

Definition 11. *A discount function* $D : \mathbb{N} \cup \{0\} \to (0,1]$ *exhibits* diminishing impatience *if*

$$\frac{D(0)}{D(1)} > \frac{D(t)}{D(t+1)} \quad (\forall t \in \mathbb{N}),$$

and exhibits strong diminishing impatience *if*

$$\frac{D(t)}{D(t+1)} > \frac{D(t+1)}{D(t+2)} \quad (\forall t \in \mathbb{N} \cup \{0\}).$$

Diminishing impatience requires that the intertemporal rate of substitution for any pair of successive periods in the future is strictly more balanced than the intertemporal rate of substitution between today and tomorrow. Strong diminishing impatience further requires that the intertemporal rate of substitution between successive periods is strictly declining over time. Quasi-hyperbolic discount functions exhibit diminishing impatience but not strong diminishing impatience because the discount factor between t and $t+1$ is constant after $t = 1$, whereas true hyperbolic discounting, on the other hand, exhibits strong diminishing impatience.

The following corollary of Theorem 1 uncovers the implications of diminishing and strong diminishing impatience on the perceived future impatience of a naive individual. The individual believes that her ex-post intertemporal rate of substitution between period 1 and period $t+1$ will be governed by the discount factor $\hat{D}(t)$. This discount factor is a convex combination of the virtuous ex-ante discount factor $D(t{+}1)/D(1)$ and the actual tempting discount factor $D(t)$ that governs the intertemporal consumption that the individual will actually choose tomorrow.

Corollary 3. *Suppose $(\succsim, \mathcal{C})$ has a discounting representation $(u, D, \hat{D})$. Then the individual is naive if and only if there exists $\alpha \in [0,1]$ such that*

$$\hat{D}(t) = \alpha \frac{D(t+1)}{D(1)} + (1-\alpha) D(t), \quad \forall t \in \mathbb{N} \cup \{0\}, \tag{7}$$

and the individual is sophisticated if and only if $\alpha = 0$. In addition, if the individual is strictly naive (i.e., $\alpha > 0$), then

1. *The discount function D exhibits diminishing impatience if and only if*

$$\frac{D(0)}{D(t)} > \frac{\hat{D}(0)}{\hat{D}(t)} \quad (\forall t \in \mathbb{N}).$$

2. *The discount function D exhibits strong diminishing impatience if and only if*

$$\frac{D(t)}{D(t+1)} > \frac{\hat{D}(t)}{\hat{D}(t+1)} \quad (\forall t \in \mathbb{N} \cup \{0\}).$$

The two equivalences under naiveté are surprising because they relate (strong) diminishing impatience of the actual temptation, as captured in D, with the intertemporal rate of substitution in the believed temptation, as captured in $\hat{D}$. The first claim says that for a naive individual, diminishing impatience is equivalent to beliefs being biased toward saving desirable consumption for a later date t rather than in the present period 0, as reflected in $\hat{D}(0)/\hat{D}(t) < D(0)/D(t)$. In other words, under-appreciating the temptation for immediate consumption versus later consumption is an inherent feature of naiveté with diminishing impatience. If beliefs are ever biased in the opposite direction (with projected undersaving) then the individual cannot exhibit diminishing impatience in her virtuous utility. Similarly, under-appreciation of the temptation to shift good consumption to immediately prior time periods is an inherent feature of strong diminishing impatience with naiveté. Note that the results do not suggest a relationship between the diminishing impatience of the actual temptation and the diminishing impatience of the believed temptation.

4 Random choice

We now extend the analysis from the prior section to random choice rules. For any $\lambda^x \in \Delta(\Delta(C))$, its average choice $m(\lambda^x)$ is the expectation of the identity function under λ^x, or formally $m(\lambda^x) = \int p \, d\lambda^x \in \Delta(C)$. That is, $m(\lambda^x)$ reduces the compound lottery λ^x into a single lottery. The definitions in the prior section for deterministic choice then generalize by considering reductions of random choices. This reduction from a distribution over multiple lotteries to a single lottery does not assume any attitude towards risk, such as risk neutrality, over deterministic outcomes in C.[14]

Definition 12. *An individual is* sophisticated *if, for all menus x,*

$$x \sim \{m(\lambda^x)\}.$$

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

$$x \succsim \{m(\lambda^x)\}.$$

A sophisticate is indifferent between choosing from a menu x tomorrow and committing to the average choice $m(\lambda^x)$ from that menu. A naif anticipates making more virtuous choices, on average, than she actually will make. The proviso from the prior

[14] Our analysis implicitly assumes indifference to compounding. Indifference to compounding can be relaxed by considering appropriate certain equivalents rather than assuming indifference between λ^x and $m(\lambda^x)$.

subsection regarding the assumption of consequentialism and lack of indirect menu effects applies here as well. When λ^x involves no randomness, this definition reduces to the prior one for deterministic choice: If $\lambda^x = \delta_p$ then the corresponding deterministic choice function takes $\mathcal{C}(x) = p$. Since $m(\delta_p) = p$, this implies $m(\lambda^x) = \mathcal{C}(x)$.

We use the same definition of more temptation averse for random choice as for deterministic choice, but we must extend our previous definition of more virtuous to account for randomness in choice.

Definition 13. *Individual* 2 *is* more virtuous *than individual* 1 *if, for all menus x and lotteries p,*

$$\{p\} \succ_2 \{m(\lambda_2^x)\} \implies \{p\} \succ_1 \{m(\lambda_1^x)\}.$$

The average choice of the more sophisticated individual is more virtuous.

Definition 14. *Suppose individuals* 1 *and* 2 *are naive. Individual* 1 *is* more naive *than individual* 2 *if individual* 2 *is more temptation-averse and more virtuous than individual* 1*.*

A generalization of the classic Strotz model is the random Strotz model that admits uncertainty about future temptations. For example, a dieter might crave potato chips at some times and chocolate cake at other times. Dekel and Lipman (2012) provide a thorough analysis of the random Strotz model. Since a single temptation is parametrized as a single utility vector, a random temptation is parametrized as a probability measure over utility vectors. Note that even if actual choices are degenerate, the random Strotz model is important because it allows an individual to be mistakenly uncertain about her future behavior, while such uncertainty is excluded by the standard Strotz representation.

In what follows, let $\mathcal{V}$ denote the set of all continuous functions $v : C \to \mathbb{R}$, and endow $\mathcal{V}$ with the supremum norm and corresponding Borel σ-algebra. We can identify $\mathcal{V}$ with the set of all expected-utility functions on $\Delta(C)$ by letting $v(p) \equiv \int_C v(c)\,dp$. As before, we say u is a nontrivial expected-utility function if it is not constant. We say $\hat{\mu}$ is a nontrivial measure on $\mathcal{V}$ if it assigns probability zero to the set of constant functions.[15]

[15] Note that the restriction to nontrivial measures in the definitions of the random Strotz representations is without loss of generality since any weight assigned to constant functions can be moved to the commitment utility u without altering the ex-ante preference or ex-post random choice rule.

Definition 15. *A* random Strotz representation *of $\succsim$ is a pair $(u, \hat{\mu})$ of a nontrivial expected-utility function u and a nontrivial probability measure $\hat{\mu}$ over $\mathcal{V}$ such that the function $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ defined by*

$$U(x) = \int_{\mathcal{V}} \max_{p \in B_v(x)} u(p)\, d\hat{\mu}(v)$$

is a utility representation of $\succsim$.

The analogous representation of random choice is more cumbersome because the maximizing Strotz choice set $B_u(B_v(x))$ has potentially multiple elements for a fixed temptation utility $v \in \mathcal{V}$, in turn generating multiple choice probabilities over x for a fixed probability measure over $\mathcal{V}$. A random choice rule maximizes a random Strotz representation if it is generated by some selection function from the correspondence mapping temptations v to possible choices $B_u(B_v(x))$.[16]

Definition 16. *A* random Strotz representation *of λ is a pair (u, μ) of a nontrivial expected-utility function u and a nontrivial probability measure μ over $\mathcal{V}$ such that, for all menus x and all measurable $y \subset x$,*

$$\lambda^x(y) = \mu(p_x^{-1}(y))$$

for some measurable selection function $p_x : \mathcal{V} \to x$ with $p_x(v) \in B_u(B_v(x))$ for all $v \in \mathcal{V}$.[17]

Definition 17. *A* random Strotz representation *of $(\succsim, \lambda)$ is a triple $(u, \mu, \hat{\mu})$ such that $(u, \hat{\mu})$ is a random Strotz representation of $\succsim$ and (u, μ) is a random Strotz representation of λ.*

The definition of naiveté for random Strotz is the stochastic generalization of the definition for deterministic Strotz. In the degenerate case, naiveté implies the believed $\hat{v}$ is more u-aligned than v: $\hat{v} \gg_u v$. In the random case, the believed distribution over all possible temptations stochastically dominates the actual distribution of temptations,

[16] Just as there can be a multiple choice functions that maximize the same deterministic Strotz representation, there can be multiple random choice rules that maximize the same random Strotz representation. However, this multiplicity is not important for our results since observing any maximizing random choice rule provides sufficient information for our comparatives.

[17] Since λ is taken as primitive, the selection function p_x is identified almost surely with respect to μ.

where stochastic dominance is with respect to the $\gg_u$ order. The following relation formalizes the stochastic order generated by $\gg_u$. As is standard, a stochastically dominant measure puts more weight on the upper contour sets of the basic ordering $\gg_u$ over the state space. The following definitions adapt the technology developed by Dekel and Lipman (2012).

Definition 18. *Let u be an expected-utility function. A measurable set $\mathcal{U} \subset \mathcal{V}$ is a* u-upper set *if, for any $v \in \mathcal{U}$ and $v' \in \mathcal{V}$, if $v' \gg_u v$ then $v' \in \mathcal{U}$.*

We let $\gg_u$ notate both the basic ordering over expected-utility functions and the induced stochastic order over measures on expected-utility functions. Note that $v \gg_u v'$ (in the determinate sense) is equivalent to $\delta_v \gg_u \delta_{v'}$ (in the stochastic sense).

Definition 19. *Let u be an expected-utility function, and let μ, μ' be probability measures over $\mathcal{V}$. Then μ is* more u-aligned *than μ', written as $\mu \gg_u \mu'$, if $\mu(\mathcal{U}) \geq \mu'(\mathcal{U})$ for all u-upper sets $\mathcal{U}$.*

Generalizing the earlier result, absolute naiveté is equivalent to $\hat{\mu}$ dominating μ in the stochastic order generated by $\gg_u$.

Theorem 3. *Suppose C has finite cardinality, and suppose $(\succsim, \lambda)$ has a random Strotz representation $(u, \mu, \hat{\mu})$. Then the individual is naive if and only if $\hat{\mu} \gg_u \mu$ (and is sophisticated if and only if $\hat{\mu} \gg_u \mu$ and $\mu \gg_u \hat{\mu}$).*

Analogously, comparative naiveté induces a similar ordering over individuals' predictions and behaviors but in a stochastic sense.

Theorem 4. *Suppose C has finite cardinality, and suppose $(\succsim_1, \lambda_1)$ and $(\succsim_2, \lambda_2)$ have random Strotz representations $(u_1, \mu_1, \hat{\mu}_1)$ and $(u_2, \mu_2, \hat{\mu}_2)$. Then individual 1 is more naive than individual 2 if and only if $u_1 \approx u_2 \equiv u$ and*

$$\hat{\mu}_1 \gg_u \hat{\mu}_2 \gg_u \mu_2 \gg_u \mu_1.$$

To aid intuition further, we will highlight the corollaries of these characterizations for the case where the uncertainty over future behavior is only over the magnitude of the future temptation, and not in its basic direction. For example, the individual knows that she will crave sweet snacks (but not salty snacks) ex post, but is uncertain of how strong her craving for sweets will be.

In what follows, say two expected-utility functions u and v are *independent* if they are nontrivial and it is not the case that $v \approx u$ or $v \approx -u$.

Definition 20. *An* uncertain intensity Strotz representation *of $\succsim$ is a triple $(u, v, \hat{F})$ of two independent expected-utility functions u, v and a cumulative distribution function $\hat{F}$ on $[0, 1]$ such that the function $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ defined by*

$$U(x) = \int_0^1 \max\{u(p) : p \in B_{\alpha u+(1-\alpha)v}(x)\}\, d\hat{F}(\alpha)$$

is a utility representation of $\succsim$.

Definition 21. *An* uncertain intensity Strotz representation *of λ is a triple (u, v, F) of two independent expected-utility functions u, v and a cumulative distribution function F on $[0, 1]$ such that, for all menus x and all measurable $y \subset x$,*

$$\lambda^x(y) = \int_0^1 \mathbf{1}_{[p_x(\alpha)\in y]}\, dF(\alpha)$$

for some measurable selection function $p_x : [0, 1] \to x$ with $p_x(\alpha) \in B_u(B_{\alpha u+(1-\alpha)v}(x))$ for all $\alpha \in [0, 1]$.

Definition 22. *An* uncertain intensity Strotz representation *of $(\succsim, \lambda)$ is a quadruple $(u, v, F, \hat{F})$ such that $(u, v, \hat{F})$ is an uncertain intensity Strotz representation of $\succsim$ and (u, v, F) is an uncertain intensity Strotz representation of λ.*

For the case of an uncertain intensity Strotz representation, the direction of the temptation is known to be v, but the magnitude of that temptation relative to the virtuous utility u is uncertain. The individual underestimates the influence of v, and this bias is reflected in her belief $\hat{F}$ over the intensities in $[0, 1]$ putting more likelihood on larger weighting of u (hence lower weighting of v) in the ex-post stage of choice. Let $\geq_{FOSD}$ denote the usual first-order stochastic dominance order, with $\hat{F} \geq_{FOSD} F$ if $\hat{F}(\alpha) \leq F(\alpha)$ for all $\alpha \in [0, 1]$.

Theorem 5. *Suppose $(\succsim, \lambda)$ has a uncertain intensity Strotz representation $(u, v, F, \hat{F})$. Then the individual is naive if and only if $\hat{F} \geq_{FOSD} F$ (and is sophisticated if and only if $\hat{F} = F$).*

Theorem 6. *Suppose $(\succsim_1, \lambda_1)$ and $(\succsim_2, \lambda_2)$ have uncertain intensity Strotz representations $(u, v, F_1, \hat{F}_1)$ and $(u, v, F_2, \hat{F}_2)$. Then individual 1 is more naive than individual 2 if and only if*

$$\hat{F}_1 \geq_{FOSD} \hat{F}_2 \geq_{FOSD} F_2 \geq_{FOSD} F_1.$$

4.1 Application: Eliaz and Spiegler (2006)

We now apply the general results for random Strotz to an important specialization proposed by Eliaz and Spiegler (2006). In their model, an individual's ex-post choice will be governed by v as in the standard Strotz model. However, an individual's ex-ante assessment is that she will choose virtuously according to u with probability θ and will face temptation and choose according to v with probability $1-\theta$. This is arguably the simplest possible formula that features naive random temptation.

Definition 23. *An* Eliaz–Spiegler representation *of* $\succsim$ *is a triple* (u, v, θ) *of two independent expected-utility functions* u, v *and a scalar* $\theta \in [0,1]$ *such that the function* $U : \mathcal{K}(\Delta(C)) \to \mathbb{R}$ *defined by*

$$U(x) = \theta \max_{p \in x} u(p) + (1-\theta) \max_{p \in B_v(x)} u(p)$$

is a utility representation of $\succsim$.

The implications of the general characterizations for random Strotz are immediate for this special case. The individual mistakenly believes there is some $\theta \geq 0$ probability that she will act virtuously and maximize u rather than v ex post, whereas she will certainly maximize v in reality. The more naive individual believes her ex-post behavior is more likely to be virtuous ($\theta_1 \geq \theta_2$) while her actual (and nonrandom) behavior is identical.

Corollary 4. *Suppose* $\succsim$ *has an Eliaz–Spiegler representation* (u, v, θ)*, and* $\mathcal{C}$ *has a (deterministic) Strotz representation* (u, v)*. Then the individual is naive (and is sophisticated if and only if* $\theta = 0$*).*

Proof. The deterministic ex-post choice of the individual can be expressed as a (degenerate) random choice rule λ that satisfies $\lambda^x = \delta_p$ for $p = \mathcal{C}(x)$. Define a cumulative distribution function F on $[0,1]$ by $F(\alpha) = 1$ for all $\alpha \in [0,1]$, and define a cumulative distribution function $\hat{F}$ by $\hat{F}(\alpha) = 1-\theta$ for $0 \leq \alpha < 1$ and $\hat{F}(1) = 1$. Note that $(u, v, F, \hat{F})$ is an uncertain intensity Strotz representation for $(\succsim, \lambda)$. Then $\hat{F} \geq_{FOSD} F$, and $\hat{F} = F$ if and only if $\theta = 0$. The result then follows from Theorem 5. ■

An analogous application of Theorem 6 yields the following corollary.

Corollary 5. *Suppose* $\succsim_1$ *and* $\succsim_2$ *have the Eliaz–Spiegler representations* (u, v, θ_1) *and* (u, v, θ_2)*, and* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *both have the Strotz representation* (u, v)*. Then individual* 1 *is more naive than individual* 2 *if and only if* $\theta_1 \geq \theta_2$*.*

4.2 Application: Random Quasi-Hyperbolic Discounting

In this section, we describe a generalization of the quasi-hyperbolic discounting model from Section 3.1 that permits uncertainty about the present-bias factor β. This random quasi-hyperbolic discounting representation is a special case of the uncertain intensity Strotz representation where the uncertainty about future intensity is parametrized as uncertainty about the future present-bias parameter β. Several applications in different areas employ naive uncertainty about future present bias. Section 4 of Heidhues and Koszegi (2010) employs random quasi-hyperbolic discounting to explain the structure of credit markets and the consequent welfare implications for consumers. Duflo, Kremer, and Robinson (2011) use the Eliaz and Spiegler (2006) specification of random quasi-hyperbolic discounting, where naiveté is limited to a mistaken belief of some chance of virtuous exponential discounting in all future periods, in their study of fertilizer adoption decisions by Kenyan farmers. Admitting uncertainty about intertemporal substitution in economic models often usefully serves as a reduced-form proxy for a shock in the economy, like wage uncertainty, or for heterogeneity across agents in an aggregate economy, like the distribution of wealth. Similarly, random present-bias can provide a parsimonious channel within the utility function for capturing uncertainty about external factors that affect present-bias.

As in Section 3.1, maintain that consumption in periods $t = 1, 2, \dots$ is given by $(c_1, c_2, \dots) \in C = [a, b]^{\mathbb{N}}$. Suppose the commitment preference over random consumption streams is represented by an expected-utility function whose values $U(c) = U(\delta_c)$ over deterministic streams satisfy exponential discounting, which is equivalent to the extreme case where $\beta = 1$:

$$U(c) = \sum_{t=1}^{\infty} \delta^{t-1} u(c_t). \tag{8}$$

The maximally present-biased individual will value only immediate consumption in period 1 and will ignore consumption in later periods, which is equivalent to the opposite extreme case where $\beta = 0$:

$$V(c) = u(c_1). \tag{9}$$

For any intensity β on the virtuous utility U, we have

$$\beta U(c) + (1 - \beta) V(c) = u(c_1) + \beta \sum_{t=2}^{\infty} \delta^{t-1} u(c_t).$$

Therefore, uncertainty about the present-bias parameter β is equivalent to uncertainty about the intensity of U relative to V, and the present-bias factor is exactly the relative weighting of exponential discounting versus extreme impatience.

Definition 24. *A* random quasi-hyperbolic (RQH) representation *of* $(\succsim, \lambda)$ *is a quadruple* $(u, F, \hat{F}, \delta)$ *of a continuous and nontrivial function* $u : [a, b] \to \mathbb{R}$*, a scalar* $\delta \in (0, 1)$*, and cumulative distribution functions* F *and* $\hat{F}$ *on* $[0, 1]$ *such that* $(U, V, F, \hat{F})$ *is an uncertain intensity Strotz representation for* $(\succsim, \lambda)$*, where* U *and* V *satisfy Equations* (8) *and* (9)*.*

Applying Theorem 5 to the random quasi-hyperbolic representation yields the following corollary: naive random quasi-hyperbolic individuals shift weight in their predictions away from low realizations of the present-bias factor β and towards high realizations of β.

Corollary 6. *Suppose* $(\succsim, \lambda)$ *has a RQH representation* $(u, F, \hat{F}, \delta)$*. Then the individual is naive if and only if* $\hat{F} \geq_{FOSD} F$ *(and is sophisticated if and only if* $\hat{F} = F$*).*

Similarly, the more naive random quasi-hyperbolic individual puts more unwarranted mass on higher realizations of β than the more sophisticated individual. The following corollary follows directly from Theorem 6.[18]

Corollary 7. *Suppose* $(\succsim_1, \lambda_1)$ *and* $(\succsim_2, \lambda_2)$ *have RQH representations* $(u_1, F_1, \hat{F}_1, \delta_1)$ *and* $(u_2, F_2, \hat{F}_2, \delta_2)$*. Then individual 1 is more naive than individual 2 if and only if* $u_1 \approx u_2$*,* $\delta_1 = \delta_2$*, and*

$$\hat{F}_1 \geq_{FOSD} \hat{F}_2 \geq_{FOSD} F_2 \geq_{FOSD} F_1.$$

5 Welfare

We close by considering a setup that explores the welfare implications of policies that introduce new commitment devices to naive consumers. Suppose the government contemplates whether to provide an illiquid forced-savings device. This is equivalent to

[18] Note that Corollary 7 states that $u_1 \approx u_2$ and $\delta_1 = \delta_2$ as part of the implication of individual 1 being more naive than individual 2. This follows from our previous observations that individual 2 being either more temptation averse or more virtuous than individual 1 implies that both have the same commitment preference. The relationship between the distribution functions $\hat{F}_1, \hat{F}_2, F_1, F_2$ then follows from Theorem 6.

introducing an additional commitment device or menu x to the family of existing available menus; the new menu excludes immediate consumption beyond a certain level. A pervasive finding is that the take-up of new commitment devices is minimal under naiveté.[19] Beyond mitigating the effectiveness of new commitment devices, we find that new commitment devices can strictly decrease welfare when consumers are naive. In fact, the existence of such strictly deleterious commitment devices characterizes naiveté. Moreover, the marginal welfare effects of such interventions fail to be monotone in sophistication.

Formally, we consider families of menus to understand the effects of introducing additional commitment devices. For finite $X \subset \mathcal{K}(\Delta(C))$, let $x^*(X) = \{x \in X : x \succsim y \text{ for all } y \in X\}$ denote the set of $\succsim$-maximal menus from the family X. Let $\mathfrak{C}(X) \in \mathcal{C}(x^*(X)) \equiv \{\mathcal{C}(x) : x \in x^*(X)\}$. That is, $\mathfrak{C}(X)$ is the final consumption from the family of menus X when the individual adheres to the following protocol: first, she selects a $\succsim$-maximal menu $x \in x^*(X)$, and second, she consumes $\mathcal{C}(x)$. This allows us to compare the final welfare from different families of commitment devices by comparing their induced final choices, that is, X is better for an individual than Y if $\{\mathfrak{C}(X)\} \succsim \{\mathfrak{C}(Y)\}$.[20]

The next result makes the straightforward but important observation that adding additional commitment devices always makes sophisticated individuals better off. The converse result, that strictly naive individuals can always be made strictly worse off by introducing available commitments devices, requires that singleton menus are dense in the ex-ante preference as they are, for example, whenever a Strotz representation exists. The literature already observed many specific situations where providing flexibility to naive individuals makes them worse off. Our point is that this is a general phenomenon: the possibility of such welfare loss is a necessary consequence of naiveté.

Recall that an individual is strictly naive if she is naive and not sophisticated.

Theorem 7. *If an individual is sophisticated, then $\{\mathfrak{C}(X)\} \succsim \{\mathfrak{C}(Y)\}$ whenever $X \supset Y$. If singleton menus are $\succsim$-dense and the individual is strictly naive, then there exist*

[19] Several studies in this line are surveyed by Bryan, Karlan, and Nelson (2010).

[20] We follow the commonly employed approach of using ex-ante commitment preferences over singletons as the welfare criterion over final consumption $\Delta(C)$. Another established benchmark is the Pareto welfare (partial) order based on improvements with respect to both ex-ante and ex-post preferences. Since Theorems 7 and 9 involve changing ex-ante utility u with possible reciprocal changes to ex-post utility v, they are no longer valid with respect to the Pareto welfare criterion. Theorems 8 and 10 involve losses to both ex-ante and ex-post utility, and therefore hold with respect to either welfare criterion.

$X \supset Y$ with $\{\mathfrak{C}(X)\} \prec \{\mathfrak{C}(Y)\}$.

Proof. By the standard revealed-preference argument, $X \supset Y$ implies $x \succsim y$ for any $x \in x^*(X)$ and $y \in x^*(Y)$. Under sophistication, $\{\mathcal{C}(x)\} \sim x \succsim y \sim \{\mathcal{C}(y)\}$. But $\mathfrak{C}(X) = \mathcal{C}(x)$ for some $x \in x^*(X)$ and $\mathfrak{C}(Y) = \mathcal{C}(y)$ for some $y \in x^*(Y)$, so in particular $\{\mathfrak{C}(X)\} \succsim \{\mathfrak{C}(Y)\}$.

Now assume the individual is strictly naive: there exists a menu x with $x \succ \{\mathcal{C}(x)\}$. By $\succsim$-denseness of the singletons, there exists some lottery p such that $x \succ \{p\} \succ \{\mathcal{C}(x)\}$. Let $X = \{x, \{p\}\}$ and $Y = \{\{p\}\}$. Then $\mathfrak{C}(Y) = p$ and $\mathfrak{C}(X) = \mathcal{C}(x)$, so $\{\mathfrak{C}(Y)\} \succ \{\mathfrak{C}(X)\}$. ■

Example 1. As an example to illustrate the approach in Theorem 7, consider the stopping problems studied by O'Donoghue and Rabin (1999) and O'Donoghue and Rabin (2001). They demonstrate that extreme naifs (where $\hat{\beta} = 1$) will always procrastinate longer than sophisticates on a task that requires immediate costs to yield a flow of benefits and that partial naifs (where $\hat{\beta} > \beta$) always have some task that leads them to procrastinate. We now show this prediction is not an artifact of quasi-hyperbolic discounting but an almost inherent feature of general naiveté in these environments. Suppose $\{d_2\} \succ \{d_1\} \succ \{d_3\}$. Doing it now (in period 1) is tantamount to committing to $\{d_1\}$, while delaying means leaving the menu $\{d_2, d_3\}$ available tomorrow. The only way for the definition of naiveté to have bite in this domain is if $\mathcal{C}(\{d_2, d_3\}) = d_3$ and $\{d_2, d_3\} \succ \{d_3\}$. If $Y = \{\{d_1\}\}$, then the individual must do it now (in period 1). In contrast, $X = \{\{d_1\}, \{d_2, d_3\}\}$ contains the option of doing it now, $\{d_1\}$, or delaying and having the option in the next period whether to do it or not, $\{d_2, d_3\}$. Then $\mathfrak{C}(Y) = d_1$ and $\mathfrak{C}(X) = d_3$, so $\{\mathfrak{C}(Y)\} \succ \{\mathfrak{C}(X)\}$. That is, the naive individual delays longer and gets less welfare than the sophisticated individual if she is given the option X to delay past the first period. The actual choices are worse with X because the individual delays under the (incorrect) belief that she will do it in the following period. The nature of procrastination for naive individuals is therefore neither an artifact of $(\beta, \hat{\beta}, \delta)$ preferences nor even really connected to time discounting. Rather, it is an inherent consequence of the structure of the domain into possible menus (that is, only menus consisting of remaining time periods are observed) and the force of the basic definition of naiveté on this structured domain.

The prior theorem is intuitive because the additional menu that leads to a less virtuous final selection is possibly a superset of an already available menu. Clearly, increasing flexibility for individuals who mistakenly believe they will virtuously exercise that flexibility can decrease their welfare. The next result is sharper, but requires the additional structure of the Strotz model. Under Strotz preferences, there always exists a subset of an existing menu that leaves the individual worse off when added to the family of commitments. That is, there exists a scenario where welfare is harmed by adding stronger commitments (rather than more flexibility) that exclude choices which are otherwise available in some existing forward plan in the status quo. The basic intuition is as follows. Suppose that carrots are more virtuous than pretzels, but pretzels are more virtuous than potato chips. When all three snacks are available, the individual will choose pretzels. If she is naive, she might believe that removing the availability of pretzels will induce her to eat carrots. Anticipating this, she throws the pretzels away. Unfortunately, her prediction is mistaken and she ends up eating potato chips, leaving herself worse off than she was before.

Say an individual has a *preference for commitment* if there exist menus y and $x \subset y$ such that $x \succ y$. If an individual has a preference for commitment, then she is not fully naive in the sense of believing that her future tastes will be identical to her current commitment preference.

Theorem 8. *Suppose $(\succsim, \mathcal{C})$ admits a Strotz representation $(u, v, \hat{v})$, where u and v are independent.*[21] *If the individual is strictly naive and has a preference for commitment, then there exist menus y and $x \subset y$ such that $\{\mathfrak{C}(\{x, y\})\} \prec \{\mathfrak{C}(\{y\})\}$.*

Proof. Under the assumptions of the theorem, it can be shown that there exist lotteries p^1, p^2, p^3 such that[22]

$$\begin{aligned} &u(p^1) > u(p^2) > u(p^3) \\ &\hat{v}(p^2) > \hat{v}(p^1) > \hat{v}(p^3) \\ &v(p^2) > v(p^3) > v(p^1). \end{aligned}$$

Let $y = \{p^1, p^2, p^3\}$ and $x = \{p^1, p^3\}$. The rankings of the lotteries according to u and

[21] That is, it is not that case that $v \approx u$ or $v \approx -u$.

[22] Proof: By Theorem 1, $\hat{v} \gg_u v$. Since it is not the case that $v \approx -u$, this implies $\hat{v} \approx \alpha u + (1-\alpha)v$. Note that $\alpha > 0$ since the individual is strictly naive, and $\alpha < 1$ since $\succsim$ has preference for commitment. Hence, it is not the case that $\hat{v} \approx u$, so there exist lotteries p, q such that $\hat{v}(p) = \hat{v}(q)$ and $u(p) > u(q)$. Since $\hat{v} \approx \alpha u + (1-\alpha)v$ for $\alpha \in (0, 1)$, this also implies that $v(p) < v(q)$. Since it is not the case that $v \approx -u$, there exist lotteries r, s such that $u(r) > u(s)$ and $v(r) > v(s)$, which also implies $\hat{v}(r) > \hat{v}(s)$. It is easy to show that the lotteries $p^1 = (1-\varepsilon)p + \varepsilon[(1/2)s + (1/2)r]$, $p^2 = (1-\varepsilon)q + \varepsilon r$, $p^3 = (1-\varepsilon)q + \varepsilon s$ have the desired properties for $\varepsilon > 0$ sufficiently small.

$\hat{v}$ imply that $x \sim \{p^1\} \succ \{p^2\} \sim y$. The ranking according to v implies that $\mathcal{C}(x) = p^3$ and $\mathcal{C}(y) = p^2$. Therefore, $\{\mathfrak{C}(\{x, y\})\} = \{p^3\} \prec \{p^2\} = \{\mathfrak{C}(\{y\})\}$. ■

Example 2. For a concrete illustration of Theorem 8, consider an individual facing a three-period consumption-savings problem. In period 0, she initially chooses to invest money in a liquid savings account or in a retirement account. In period 1, she then decides whether to make a withdrawal from her savings. If she initially invested in the retirement account, then her early withdrawal results in a tax penalty. In period 2, she finally consumes the remaining balance of the savings or retirement account. For simplicity of exposition, we assume linear utility over static consumption and focus on deterministic consumption streams.

Suppose the individual's period 0 preference $\succsim$ has a Strotz representation $(U, \hat{V})$ where

$$U(c_1, c_2) = c_1 + c_2 \quad \text{and} \quad \hat{V}(c_1, c_2) = c_1 + \hat{\beta} c_2,$$

and suppose the individual's period 1 choice function $\mathcal{C}$ has a Strotz representation (U, V) where

$$V(c_1, c_2) = c_1 + \beta c_2.$$

As standard, assume $0 < \beta < \hat{\beta} \leq 1$.

If the gross interest rate is $R > 1$ and the individual initially has unit wealth, then investing in the liquid savings account is equivalent to choosing the menu

$$y = \{(c_1, c_2) \in \mathbb{R}^2_+ : c_1 + c_2/R = 1\}.$$

The retirement account has a proportional early withdrawal penalty τc_1 associated with a withdrawal of c_1 in period 1, where $\tau \geq 0$. Then investing in the retirement account is equivalent to choosing the menu

$$x^\tau = \{(c_1, c_2) \in \mathbb{R}^2_+ : (1 + \tau)c_1 + c_2/R = 1\}.$$

Note that $x^\tau \subset y$, and that $x^\tau = y$ if $\tau = 0$. The actual choices from y and x^τ are

$$\mathcal{C}(y) = \begin{cases} (1, 0) & \text{if } 1 > \beta R \\ (0, R) & \text{if } 1 \leq \beta R. \end{cases}$$

and

$$\mathcal{C}(x^\tau) = \begin{cases} (1/(1+\tau), 0) & \text{if } 1 > (1+\tau)\beta R \\ (0, R) & \text{if } 1 \leq (1+\tau)\beta R. \end{cases}$$

Suppose $1 > \hat{\beta}R$ and $(1+\tau)\hat{\beta}R > 1 > (1+\tau)\beta R$. In this case, the individual correctly anticipates choosing $(1, 0)$ from the menu y. However, she incorrectly anticipates choosing $(0, R)$ from x^τ, when in fact she will choose $(1/(1+\tau), 0)$. She believes that the tax penalty associated with the retirement account is high enough to deter her from making early withdrawals in period 1, but in reality it is not. Since $U(0, R) > U(1, 0)$, this incorrect belief will lead the individual to initially invest in the illiquid retirement account x^τ over the liquid savings account y in period 0. Therefore, the availability of the retirement account as a savings instrument is strictly detrimental, since

$$\{\mathfrak{C}(\{x^\tau, y\})\} = \{(1/(1+\tau), 0)\} \prec \{(1, 0)\} = \{\mathfrak{C}(\{y\})\}.$$

Note that the perverse welfare effect associated with offering the individual the commitment device x^τ depends crucially on the level of the early withdrawal penalty: τ is high enough that the individual thinks it will deter early withdrawals, but it is low enough that it actually does not. Increasing τ until $(1+\tau)\beta R > 1$ makes the retirement account a strong enough commitment device to increase welfare. For τ is this region,

$$\{\mathfrak{C}(\{x^\tau, y\})\} = \{(0, R)\} \succ \{(1, 0)\} = \{\mathfrak{C}(\{y\})\}.$$

While some forms of partial commitment can make naive individuals worse off, some classes of commitments can unambiguously improve welfare. In Example 2, increasing the early withdrawal penalty magnifies the strength of the commitment device x^τ. When τ is small and the commitment device is weak, the welfare effect of this commitment device is neutral or negative. Once τ exceeds some threshold—the exact value of which depends on the preference parameters—the commitment device becomes strong enough to deliver a positive welfare impact. Expanding on this observation, the following result shows there is a class of commitment devices that will unambiguously improve welfare for any preference: complete commitments to a single outcome. This holds for both sophisticated and strictly naive individuals.

Theorem 9. *Assume $\mathfrak{C}(X \cup \{x\}) \notin x$ implies $\mathfrak{C}(X \cup \{x\}) = \mathfrak{C}(X)$.*[23] *If an individual*

[23] This property simply implies that the tie-breaking procedure used in the selection function $\mathfrak{C}$

is naive, then $\{\mathfrak{C}(X \cup \{p\})\} \succsim \{\mathfrak{C}(X)\}$ *for all lotteries* p.

Proof. By the assumed properties of $\mathfrak{C}$, either $\mathfrak{C}(X \cup \{p\}) = \mathfrak{C}(X)$, in which case the results holds trivially, or $\mathfrak{C}(X \cup \{p\}) = p$. In the latter case, we must have $\{p\} \succsim x$ for all $x \in X$. Since $\mathfrak{C}(X) = \mathcal{C}(y)$ for some $y \in x^*(X)$, naiveté then implies $\{p\} \succsim y \succsim \{\mathcal{C}(y)\}$, and hence $\{\mathfrak{C}(X \cup \{p\})\} \succsim \{\mathfrak{C}(X)\}$. ■

When menu choice is driven purely by temptation, adding extreme commitment devices is never harmful. Of course, such extreme commitment is detrimental if individuals have uncertain virtuous tastes that lead to demand for flexibility. Optimal design of commitment devices for naive individuals with some demand for flexibility remains an important open question, as do suitable comparative statics that isolate naiveté about future temptations and unawareness of future virtuous taste contingencies in a unified manner.[24]

Finally, a natural question is whether a comparative analog of Theorems 7 and 8 holds. Specifically, consider the following conjecture: If an individual is made better off by the introduction of a commitment device, then any other individual who is more sophisticated is also made better off. Excluding the extreme cases of full sophistication or full naiveté, the following theorem shows that this conjecture fails. In fact, for a generic pair of comparable individuals, there exists a new commitment device that leaves the more sophisticated individual strictly worse off while having no effect on the more naive individual.

Theorem 10. *Suppose* $(\succsim_1, \mathcal{C}_1)$ *and* $(\succsim_2, \mathcal{C}_2)$ *admit Strotz representations* $(u_1, v_1, \hat{v}_1)$ *and* $(u_2, v_2, \hat{v}_2)$. *Suppose individual 2 is strictly naive and has a preference for commitment, and that* u_2 *and* v_2 *are independent. If individual 1 is strictly more naive than individual 2,*[25] *then there exist menus* y *and* $x \subset y$ *such that* $\{\mathfrak{C}_1(\{x, y\})\} \sim_1 \{\mathfrak{C}_1(\{y\})\}$ *and* $\{\mathfrak{C}_2(\{x, y\})\} \prec_2 \{\mathfrak{C}_2(\{y\})\}$.

The failure of a monotone relationship between welfare and sophistication resonates with earlier findings, e.g., Heidhues and Koszegi (2009) examine a setting where individuals can pay an up-front cost to impose a penalty on indulging future temptations.

does not change when unchosen options are added. This avoids spurious welfare conclusions that are artifacts of the tie-breaking protocol.

[24] Amador, Werning, and Angeletos (2006) study a consumption-savings problem combining flexibility and temptation, but under the assumption of full sophistication.

[25] That is, individual 1 is more naive than individual 2, but it is not the case that individual 2 is also more naive than individual 1. This restriction still permits a shared ex-ante or a shared ex-post Strotz representation, but not both.

They show that welfare can fail to be monotonic in beliefs, that is, more accurate values of $\hat{\beta}$ do not guarantee higher welfare.[26] However, this finding is for a fixed set of commitments, or for a fixed policy regime. To understand the distribution of marginal welfare effects across individuals from introducing new commitment devices, the relevant consideration is comparing individuals' differences in welfare across two regimes. Theorem 10 speaks to those changes in welfare, providing a negative result that suggests caution when considering the distribution of welfare effects induced by policy changes affecting individuals with heterogeneous levels of naiveté.

[26] In contrast, holding beliefs fixed, welfare is monotonically increasing with more virtuous actual behavior. Trivially, if two individuals share the same ex-ante preference, then the more virtuous individual is better off in any fixed two-stage decision problem X than the less virtuous individual.

A A Comparative from Dekel and Lipman (2012)

In this section, we summarize a relevant result from Dekel and Lipman (2012) that will play a central role in our proofs of Theorems 3 and 4.

Theorem 11 (Dekel and Lipman (2012))**.** *Suppose C has finite cardinality. Suppose $\succsim_1$ and $\succsim_2$ have random Strotz representations (u_1, μ_1) and (u_2, μ_2). Then $\succsim_2$ is more temptation averse than $\succsim_1$ if and only if $u_1 \approx u_2 \equiv u$ and $\mu_1 \gg_u \mu_2$.*

Theorem 4 in Dekel and Lipman (2012) establishes the equivalence of $\succsim_2$ being more temptation averse than $\succsim_1$ and another condition on the representations that they refer to as conditional dominance. However, they also establish that $\mu_1 \gg_u \mu_2$ as an intermediate step in their proof.[27] The equivalence asserted in Theorem 11 is also stated explicitly in Theorem 4 of their working paper, Dekel and Lipman (2010).[28]

B Proofs

This section contains proofs omitted from the main text. The proofs of Theorems 5, 6, and 10 are further relegated to the Online Appendix.

B.1 Preliminaries

Lemmas 1 and 2 below are used in the proofs of Theorems 1 and 2. In the case of finite C, it is easy to show that these lemmas are equivalent to Lemma 3 in Dekel and Lipman (2012), who also noted the connection to the Harsanyi Aggregation Theorem. Since we allow compact C, we include the short proofs of these results in the Online Appendix to show that no technical problems arise in our more general domain.

Lemma 1. *Let u, v, v' be nontrivial expected-utility functions defined on $\Delta(C)$. If for all lotteries p and q we have*

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

[27] To show that $\succsim_2$ being more temptation averse that $\succsim_1$ implies $\mu_1 \gg_u \mu_2$, the relevant results in Dekel and Lipman (2012) are the following: Lemma 3 shows that a partial order vC_uv' used in their paper is equivalent to our order $v \gg_u v'$ (ignoring their normalization of utility functions). Lemmas 4, 5, and 6 and the arguments on page 1296 show that for any set W that is closed under C_u (is a u-upper set in our terminology), $\mu_1(W) \geq \mu_2(W)$.

[28] Dekel and Lipman (2010) impose a normalization on the set of utility functions used in their result. However, by the uniqueness properties of the random Strotz representation established in Theorem 3 of Dekel and Lipman (2012), the probability of any u-upper set is the same for any random Strotz representation of the same preference. Therefore, their normalization of utilities is inconsequential for the result.

then $v \gg_u v'$.

Lemma 2. *Let u, v, v' be expected-utility functions defined on $\Delta(C)$, and suppose $v \gg_u v'$. Then for any menu x,*

$$\max_{p \in B_v(x)} u(p) \geq \max_{q \in B_{v'}(x)} u(q).$$

The following lemma shows that the more temptation averse comparative is implied by $\mu_1 \gg_u \mu_2$. In the case of finite C, this is precisely the necessity part of Theorem 11. Since we wish to include several applications where C is compact but not finite (e.g., dynamic consumption problems where $C = [a, b]^{\mathbb{N}}$), some of our results only require compact C (e.g., Theorems 1, 2, 5, and 6). Lemma 3 is used in the proofs of those results, and is itself proved in the Online Appendix.

Lemma 3. *Suppose C is compact. Suppose $\succsim_1$ and $\succsim_2$ have random Strotz representations (u_1, μ_1) and (u_2, μ_2). If $u_1 \approx u_2 \equiv u$ and $\mu_1 \gg_u \mu_2$, then $\succsim_2$ is more temptation averse than $\succsim_1$.*

B.2 Proof of Theorem 1

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

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

By Lemma 1, this implies that $\hat{v} \gg_u v$.

To establish necessity, suppose $\hat{v} \gg_u v$. By Lemma 2, this implies that for any menu x,

$$U(x) = \max_{p \in B_{\hat{v}}(x)} u(p) \geq \max_{q \in B_v(x)} u(q) = u(\mathcal{C}(x)).$$

Thus $x \succsim \{\mathcal{C}(x)\}$, so the individual is naive.

B.3 Proof of Theorem 2

Theorem 2 follows from the two lemmas below together with Theorem 1.

Lemma 4. *Suppose $\succsim_1$ and $\succsim_2$ have Strotz representations $(u_1, \hat{v}_1)$ and $(u_2, \hat{v}_2)$. Then individual 2 is more temptation averse than individual 1 if and only if $u_1 \approx u_2 \equiv u$ and $\hat{v}_1 \gg_u \hat{v}_2$.*

Proof. Necessity follows as a special case of Lemma 3, since $\hat{v}_1 \gg_u \hat{v}_2$ is equivalent to $\delta_{\hat{v}_1} \gg_u \delta_{\hat{v}_2}$. To establish sufficiency, suppose individual 2 is more temptation averse than individual 1. First, observe that taking $x = \{q\}$ in Definition 2 yields $\{p\} \succ_1 \{q\} \implies \{p\} \succ_2 \{q\}$. Since u_1 and u_2 are nontrivial, it is a standard result that this implies $u_1 \approx u_2$. Let $u \equiv u_2$ in what follows. Then for any lotteries p and q,

$$\begin{aligned}
&\big[u(p) > u(q) \text{ and } \hat{v}_2(p) > \hat{v}_2(q)\big] \\
&\implies \{p, q\} \sim_2 \{p\} \succ_2 \{q\} \\
&\implies \{p, q\} \succsim_1 \{p\} \succ_1 \{q\} \quad \text{(2 more temptation averse than 1)} \\
&\implies \hat{v}_1(p) \geq \hat{v}_1(q).
\end{aligned}$$

By Lemma 1, this implies that $\hat{v}_1 \gg_u \hat{v}_2$. ■

Lemma 5. *Suppose $\succsim_1$ and $\succsim_2$ have Strotz representations $(u_1, \hat{v}_1)$ and $(u_2, \hat{v}_2)$, and $\mathcal{C}_1$ and $\mathcal{C}_2$ have Strotz representations (u_1, v_1) and (u_2, v_2). Then individual 2 is more virtuous than individual 1 if and only if $u_1 \approx u_2 \equiv u$ and $v_2 \gg_u v_1$.*

Proof. To establish sufficiency, suppose individual 2 is more virtuous than individual 1. First, observe that taking $x = \{q\}$ in Definition 3 yields $\{p\} \succ_2 \{q\} \implies \{p\} \succ_1 \{q\}$. Since u_1 and u_2 are nontrivial, it is a standard result that this implies $u_1 \approx u_2$. Let $u \equiv u_2$ in what follows. Then for all menus x and lotteries p,

$$u(p) > u(\mathcal{C}_2(x)) \implies u(p) > u(\mathcal{C}_1(x)). \tag{10}$$

Therefore, for any lotteries p and q,

$$\begin{aligned}
&\big[u(p) > u(q) \text{ and } v_1(p) > v_1(q)\big] \\
&\implies u(\mathcal{C}_1(\{p, q\})) = u(p) > u(q) \\
&\implies u(\mathcal{C}_2(\{p, q\})) \geq u(p) > u(q) \quad \text{(contrapositive of Equation (10))} \\
&\implies v_2(p) \geq v_2(q).
\end{aligned}$$

By Lemma 1, this implies that $v_2 \gg_u v_1$.

To establish necessity, suppose $u_1 \approx u_2 \equiv u$ and $v_2 \gg_u v_1$. By Lemma 2, $v_2 \gg_u v_1$ implies that for any menu x,

$$u(\mathcal{C}_2(x)) = \max_{p \in B_{v_2}(x)} u(p) \geq \max_{q \in B_{v_1}(x)} u(q) = u(\mathcal{C}_1(x)).$$

Thus $u(p) > u(\mathcal{C}_2(x)) \implies u(p) > u(\mathcal{C}_1(x))$, and hence individual 2 is more virtuous than individual 1. ■

B.4 Proof of Corollary 3

The proof of the characterization of naiveté in Equation (7) is a simple generalization of the arguments in the proof of Corollary 1 and is therefore omitted. To prove claim 1, note that by Equation (7),

$$\hat{D}(t) > D(t) \iff \frac{D(t+1)}{D(1)} > D(t) = \frac{D(t)}{D(0)}.$$

The latter holds for all $t \in \mathbb{N}$ if and only if D exhibits diminishing impatience.

To prove claim 2, note first that

$$\begin{aligned} \frac{\hat{D}(t)}{\hat{D}(t+1)} &= \frac{\alpha\frac{D(t+1)}{D(1)} + (1-\alpha)D(t)}{\alpha\frac{D(t+2)}{D(1)} + (1-\alpha)D(t+1)} \\ &= \frac{\alpha\frac{D(t+1)}{D(t)} + (1-\alpha)D(1)}{\alpha\frac{D(t+2)}{D(t+1)} + (1-\alpha)D(1)} \cdot \frac{D(t)}{D(t+1)}. \end{aligned}$$

Therefore,

$$\frac{\hat{D}(t)}{\hat{D}(t+1)} < \frac{D(t)}{D(t+1)} \iff \frac{D(t+1)}{D(t)} < \frac{D(t+2)}{D(t+1)}.$$

The latter holds for all $t \in \mathbb{N} \cup \{0\}$ if and only if D exhibits strong diminishing impatience.

B.5 Proof of Theorem 3

Suppose the random choice rule λ has a random Strotz representation (u, μ). The ex ante preference of a sophisticated individual—which may differ from the individual's actual ex ante preference if she is naive—must also be represented by (u, μ). The following lemma shows how this hypothetical sophisticated preference can be determined from λ and u.

Lemma 6. *Suppose λ has a random Strotz representation (u, μ). Define a binary relation $\succsim^*$ on $\mathcal{K}(\Delta(C))$ by*

$$x \succsim^* y \iff u(m(\lambda^x)) \geq u(m(\lambda^y)).$$

Then (u, μ) is a random Strotz representation for $\succsim^*$.

Proof. Since (u, μ) represents λ, by definition there exists, for all menus x, a measurable selection function $p_x : \mathcal{V} \to x$ with $p_x(v) \in B_u(B_v(x))$ such that

$$\lambda^x(y) = \mu(p_x^{-1}(y))$$

for all measurable $y \subset x$. Thus λ^x is the distribution on x induced by the random variable p_x defined on the measure space $(\mathcal{V}, \mu)$. Therefore, the standard change of variables formula together with the linearity and continuity of u imply

$$\begin{aligned}\int_{\mathcal{V}} \max_{p \in B_v(x)} u(p)\, d\mu(v) &= \int_{\mathcal{V}} u(p_x(v))\, d\mu(v) \\ &= \int_x u(p)\, d\lambda^x(p) = u\Big(\int_x p\, d\lambda^x(p)\Big) = u(m(\lambda^x)),\end{aligned}$$

as desired. ■

Turning now to the proof of Theorem 3, fix random Strotz representations $(u, \hat{\mu})$ and (u, μ) for $\succsim$ and λ, respectively, and define $\succsim^*$ as in Lemma 6. To establish sufficiency, suppose the individual is naive. Then for all menus x and lotteries p,

$$\begin{aligned}\{p\} \succ x &\implies \{p\} \succ \{m(\lambda^x)\} \qquad \text{(naiveté)} \\ &\implies u\big(m\big(\lambda^{\{p\}}\big)\big) = u(p) > u(m(\lambda^x)) \\ &\implies \{p\} \succ^* x.\end{aligned}$$

Thus $\succsim^*$ is more temptation averse than $\succsim$. Since (u, μ) represents $\succsim^*$ by Lemma 6, Theorem 11 implies that $\hat{\mu} \gg_u \mu$. If the individual is sophisticated, then a similar argument shows that the converse also holds: $\succsim$ is also more temptation averse than $\succsim^*$ (in particular, $\succsim = \succsim^*$) and hence $\mu \gg_u \hat{\mu}$.

The following lemma establishes necessity and shows, in particular, that the restriction to finite C is not needed for this direction.

Lemma 7. *Suppose C is compact. Suppose $\succsim$ has a random Strotz representation $(u, \hat{\mu})$, and λ has a random Strotz representation (u, μ). If $\hat{\mu} \gg_u \mu$ then the individual is naive (and if $\hat{\mu} \gg_u \mu$ and $\mu \gg_u \hat{\mu}$ then the individual is sophisticated).*

Proof. Suppose $\hat{\mu} \gg_u \mu$, and define $\succsim^*$ as in Lemma 6. By Lemma 3, $\succsim^*$ is more temptation averse than $\succsim$. By contrapositive, this is equivalent to the condition

$$x \succsim^* \{p\} \implies x \succsim \{p\}.$$

Note that for any menu x, if we take $p = m(\lambda^x)$ then

$$u(m(\lambda^x)) = u(p) = u\big(m\big(\lambda^{\{p\}}\big)\big)$$

and hence $x \sim^* \{p\} = \{m(\lambda^x)\}$. Since $\succsim^*$ is more temptation averse than $\succsim$, this implies $x \succsim \{m(\lambda^x)\}$. Thus the individual is naive. If we also have $\mu \gg_u \hat{\mu}$ then an analogous argument can be used to show that the condition above can be strengthened to $x \succsim^* \{p\} \iff x \succsim \{p\}$. In this case, $x \sim \{m(\lambda^x)\}$ and hence the individual is sophisticated. ■

B.6 Proof of Theorem 4

This result can be separated into three components, two of which have already been proved. First, note that by Theorem 3, individual 2 is naive if and only if $\hat{\mu}_2 \gg_u \mu_2$, where $u \equiv u_2$. Second, by Theorem 11, individual 2 is more temptation averse than individual 1 if and only if $u_1 \approx u_2 \equiv u$ and $\hat{\mu}_1 \gg_u \hat{\mu}_2$. The final step is completed in the following lemma. Note that the restriction to finite C is not needed for necessity; part 1 of the lemma can therefore be used later in the proof of Theorem 6.

Lemma 8. *Suppose $\succsim_1$ and $\succsim_2$ have random Strotz representations $(u_1, \hat{\mu}_1)$ and $(u_2, \hat{\mu}_2)$, and λ_1 and λ_2 have random Strotz representations (u_1, μ_1) and (u_2, μ_2).*

1. *If $u_1 \approx u_2 \equiv u$ and $\mu_2 \gg_u \mu_1$ then individual 2 is more virtuous than individual 1.*
2. *If C has finite cardinality and individual 2 is more virtuous than individual 1, then $u_1 \approx u_2 \equiv u$ and $\mu_2 \gg_u \mu_1$.*

Proof. Define $\succsim_1^*$ and $\succsim_2^*$ as in Lemma 6 for λ_1 and λ_2, respectively. Then (u_1, μ_1) and (u_2, μ_2) represent $\succsim_1^*$ and $\succsim_2^*$. Note that for all menus x and lotteries p,

$$\{p\} \succ_i \{m(\lambda_i^x)\} \iff u_i(p) > u_i(m(\lambda_i^x)) \iff \{p\} \succ_i^* x, \quad i = 1, 2.$$

Therefore, individual 2 is more virtuous than individual 1 if and only if $\succsim_1^*$ is more temptation averse than $\succsim_2^*$. If C has finite cardinality, then by Theorem 11, this is true if and only if $u_1 \approx u_2 \equiv u$ and $\mu_2 \gg_u \mu_1$. For any compact C (not necessarily finite), Lemma 3 shows that if $u_1 \approx u_2 \equiv u$ and $\mu_2 \gg_u \mu_1$, then $\succsim_1^*$ is more temptation averse than $\succsim_2^*$. ■

References

Ahn, D. S. (2007): "Ambiguity Without a State Space," *Review of Economic Studies*, 75, 3–28.

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

Ali, 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.

Amador, M., I. Werning, and G.-M. Angeletos. "Commitment vs. Flexibility," *Econometrica*, 74, 365–396.

Augenblick, N., M. Niederle, and C. Sprenger (2013): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarlery Journal of Economics*, forthcoming.

Augenblick, N., and M. Rabin (2015): "An Experiment on Time Preference and Misprediction in Unpleasant Tasks," Working paper, Haas School of Business and Harvard University.

Bryan, G., D. Karlan, and S. Nelson (2010): "Commitment Devices," *Annual Review of Economics*, 2, 671–698.

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 (2010): "Costly Self-Control and Random Self-Indulgence," Working paper.

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

Dekel, E., B. L. Lipman, and A. Rustichini (2009): "Temptation-Driven Preferences," *Review of Economic Studies*, 76, 9371–971

Duflo, E., M. Kremer, and J. Robinson (2011): "Nudging Farmers to Use Fertilizer: Evidence from Kenya," *American Economic Review*, 101, 2350–2390.

Eliaz, K., and R. Spiegler (2006): "Contracting with Diversely Naive Agents," *Review of Economic Studies*, 72, 689–714.

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.

Gul, F., and W. Pesendorfer (2006): "Random Expected Utility," *Econometrica*, 74, 121–146.

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

Heidhues, P., and B. Koszegi (2010): "Exploiting Naivete about Self-Control in the Credit Market," *American Economic Review*, 100, 2279–2303.

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

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

Kreps, D., and E. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

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

Le Yaouanq, Y. (2015): "Anticipating Temptation," Working paper, 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. (2007): "Commitment and Self-Control," *Journal of Economic Theory*, 135, 1-34.

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

O'Donoghue, T., and M. Rabin (1999): "Doing It Now or Later," *American Economic Review*, 89, 103–124

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.

Prelec, D. (2004): "Decreasing Impatience: A Criterion for Non-stationary Time Preference and 'Hyperbolic' Discounting," *Scandinavian Journal of Economics*, 106, 511–532.

Sarver, T. (2008): "Anticipating Regret: Why Fewer Options May Be Better," *Econometrica*, 76, 263–305.

Shui, H., and L. M. Ausubel (2004): "Time Inconsistency in the Credit Card Market," Working paper, University of Maryland.

Additional Appendices for Online Publication

C Over-Estimation of Self-Control Problems

While our main focus is on naiveté in the traditional sense of underestimation of future temptations, simple variations of our definitions can be used to model an individual who overestimates her future temptations and is therefore overly cautious. In this section, we summarize the implications of such pessimistic violations of sophistication. Formal results are stated for the case the deterministic Strotz representation for simplicity, but the analogous results for random choice are also true.

Definition 25. *An individual is* pessimistic *if, for all menus x, $\{\mathcal{C}(x)\} \succsim x$.*

An individual who is pessimistic has an actual temptation utility than is more aligned with her normative utility than her anticipated temptation utility.

Theorem 12. *Suppose $(\succsim, \mathcal{C})$ has a Strotz representation $(u, v, \hat{v})$. Then the individual is pessimistic if and only if $v \gg_u \hat{v}$.*

The proof of this result is similar to that of Theorem 1 and is omitted.

Definition 26. *Suppose that individuals 1 and 2 are pessimistic. Individual 1 is* more pessimistic *than individual 2 if individual 1 is more temptation averse and more virtuous than individual 2.*

In contrast to the case of individual 1 being more naive than individual 2, now individual 1 is both more cautious than individual 2 in the sense of being more temptation averse, and also more virtuous. This comparative corresponds to a reversal of the ordering of temptation utilities obtained in Theorem 2.

Theorem 13. *Suppose $(\succsim_1, \mathcal{C}_1)$ and $(\succsim_2, \mathcal{C}_2)$ have the Strotz representations $(u_1, v_1, \hat{v}_1)$ and $(u_2, v_2, \hat{v}_2)$. Then individual 1 is more pessimistic than individual 2 if and only if $u_1 \approx u_2 \equiv u$ and*

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

D Additional Proofs

D.1 Proof of Lemma 1

The main step in proving Lemma 1 is the following slight variation of Farkas' Lemma.[29]

Lemma 9. *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.*[30] *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. It is immediate that 2 implies 1. To show 1 implies 2, we first argue that the 1 implies the same implication holds when the strict inequalities are replaced with weak inequalities:

$$[f_1(p) \geq f_1(q) \text{ and } f_2(p) \geq f_2(q)] \implies g(p) \geq g(q). \tag{11}$$

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 $f_2(p^\alpha) > f_2(q^\alpha)$ 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 (11).

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 (11) 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 $p, q \in ca(C)$:

$$[F_1(p) \geq F_1(q) \text{ and } F_2(p) \geq F_2(q)] \implies G(p) \geq G(q). \tag{12}$$

[29] There are two small distinctions between this result and the classic version of Farkas' Lemma. First, Farkas's 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.

[30] That is, there exist $p, q \in \Delta(C)$ such that both $f_1(p) > f_1(q)$ and $f_2(p) > f_2(q)$.

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 (11) 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 (12).

By the Convex Cone Alternative Theorem (an infinite-dimensional version of Farkas' Lemma) (Aliprantis and Border (2006, Corollary 5.84)), Equation (12) 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 $c = g(\delta_{\bar{c}}) - af_1(\delta_{\bar{c}}) - bf_2(\delta_{\bar{c}})$. ■

Turning now to the proof of Lemma 1, if $v' \approx -u$, then by definition $v \gg_u v'$. Alternatively, if it is not the case that $v' \approx -u$, then u and v' are not ordinally opposed. In this case, Lemma 9 implies there exist $a, b \geq 0$ and $c \in \mathbb{R}$ such that

$$v = au + bv' + c.$$

Since v is nontrivial, it must be that $a+b > 0$. Thus $v \approx \alpha u + (1-\alpha)v'$ for $\alpha = a/(a+b) \in [0, 1]$, and hence $v \gg_u v'$.

D.2 Proof of Lemma 2

If $v' \approx -u$, then for any menu x,

$$\max_{q \in B_{v'}(x)} u(q) = \min_{q \in x} u(q) \leq u(p), \ \forall p \in x.$$

In particular,

$$\max_{q \in B_{v'}(x)} u(q) \leq \max_{p \in B_v(x)} u(p).$$

If we do not have $v' \approx -u$, then $v \gg_u v'$ implies $v \approx \alpha u + (1-\alpha)v'$ for some $\alpha \in [0, 1]$. First, consider $\alpha = 0$. In this case, $v \approx v'$. Therefore $B_v(x) = B_{v'}(x)$, which implies

$$\max_{p \in B_v(x)} u(p) = \max_{q \in B_{v'}(x)} u(q).$$

Finally, consider the case of $\alpha > 0$. Note that for any menu x and any $p \in B_v(x)$ and $q \in B_{v'}(x)$,

$$\alpha u(p) + (1-\alpha)v'(p) \geq \alpha u(q) + (1-\alpha)v'(q) \quad \text{and} \quad v'(q) \geq v'(p).$$

Since $\alpha > 0$, these inequalities imply $u(p) \geq u(q)$. Therefore,

$$\max_{p \in B_v(x)} u(p) \geq \max_{q \in B_{v'}(x)} u(q),$$

as claimed.

D.3 Proof of Lemma 3

Suppose (u_1, μ_1) and (u_2, μ_2) are random Strotz representations of $\succsim_1$ and $\succsim_2$ such that $u_1 \approx u_2 \equiv u$ and $\mu_1 \gg_u \mu_2$. Since positive affine transformations of the functions u_i do not change the preferences, it is without loss of generality to assume $u_1 = u_2 \equiv u$. Now fix any menu x, and let $[a, b] = u(x)$. Define $f_x : \mathcal{V} \to [a, b]$ by

$$f_x(v) = \max_{p \in B_v(x)} u(p).$$

By Lemma 2, $v \gg_u v'$ implies $f_x(v) \geq f_x(v')$. Therefore, for any $\alpha \in [a, b]$ and $v \gg_u v'$,

$$v' \in f_x^{-1}([\alpha, b]) \iff f_x(v') \geq \alpha \implies f_x(v) \geq \alpha \iff v \in f_x^{-1}([\alpha, b]).$$

Thus $f_x^{-1}([\alpha, b])$ is a u-upper set. Therefore,

$$\mu_1(f_x^{-1}([\alpha, b])) \geq \mu_2(f_x^{-1}([\alpha, b])).$$

Define distributions $\eta_i^x \equiv \mu_i \circ f_x^{-1}$ on $[a, b]$ for $i = 1, 2$. By the preceding arguments, η_1^x first-order stochastically dominates η_2^x. Therefore, by the change of variables formula,

$$U_1(x) = \int_{\mathcal{V}} f_x(v)\, d\mu_1(v) = \int_a^b \alpha\, d\eta_1^x(\alpha) \geq \int_a^b \alpha\, d\eta_2^x(\alpha) = \int_{\mathcal{V}} f_x(v)\, d\mu_2(v) = U_2(x).$$

Since this is true for every x, it follows immediately that $\succsim_2$ is more temptation averse than $\succsim_1$.

D.4 Proof of Theorem 5

Lemma 10. *Suppose (u, v, F) is an uncertain intensity Strotz representation. Define a function $g : [0, 1] \to \mathcal{V}$ by $g(\alpha) = \alpha u + (1 - \alpha) v$. Define a probability measure μ on $\mathcal{V}$ by $\mu \equiv F \circ g^{-1}$.*[31] *Then the following statements hold.*

[31] We are abusing notation slightly and using F to also denote the probability measure on $[0, 1]$ that has F as its distribution function. That is, for any measurable set $A \subset [0, 1]$, we write $F(A)$ to denote $\int_A dF(\alpha)$. Thus $\mu(E) = \int_{\{\alpha' : g(\alpha') \in E\}} dF(\alpha)$ for any measurable $E \subset \mathcal{V}$.

1. *If (u, v, F) is an uncertain intensity Strotz representation of a preference $\succsim$, then (u, μ) is a random Strotz representation of $\succsim$.*

2. *If (u, v, F) is an uncertain intensity Strotz representation of a random choice rule λ, then (u, μ) is a random Strotz representation of λ.*

3. *Suppose (u, v, F_i) are uncertain intensity Strotz representations for $i = 1, 2$ and define $\mu_i \equiv F_i \circ g^{-1}$. Then $\mu_1 \gg_u \mu_2$ if and only if $F_1 \geq_{FOSD} F_2$.*

Proof. To prove statement 1, note that by assumption $\succsim$ is represented by

$$U(x) = \int_0^1 \max\{u(p) : p \in B_{g(\alpha)}(x)\}\, dF(\alpha).$$

By the standard change of variables formula, this implies

$$\begin{aligned} U(x) &= \int_{\mathcal{V}} \max\{u(p) : p \in B_{\tilde{v}}(x)\}\, d(F \circ g^{-1})(\tilde{v}) \\ &= \int_{\mathcal{V}} \max\{u(p) : p \in B_{\tilde{v}}(x)\}\, d\mu(\tilde{v}), \end{aligned}$$

and hence (u, μ) is a random Strotz representation of $\succsim$.

To prove statement 2, note that by assumption there exists, for all menus x, a measurable selection function $p_x : [0, 1] \to x$ with $p_x(\alpha) \in B_u(B_{g(\alpha)}(x))$ for all $\alpha \in [0, 1]$ such that

$$\lambda^x(y) = \int_0^1 \mathbf{1}_{[p_x(\alpha) \in y]}\, dF(\alpha)$$

for all measurable $y \subset x$. Take any measurable selection function $\tilde{p}_x : \mathcal{V} \to x$ with $\tilde{p}_x(\tilde{v}) \in B_u(B_{\tilde{v}}(x))$ for all $\tilde{v} \in \mathcal{V}$ that also satisfies $p_x(\alpha) = \tilde{p}_x(g(\alpha))$ for all $\alpha \in [0, 1]$.[32] Therefore, for any measurable $y \subset x$,

$$\begin{aligned} \lambda^x(y) &= \int_0^1 \mathbf{1}_{[\tilde{p}_x(g(\alpha)) \in y]}\, dF(\alpha) \\ &= \int_{\mathcal{V}} \mathbf{1}_{[\tilde{p}_x(\tilde{v}) \in y]}\, d(F \circ g^{-1})(\tilde{v}) \\ &= \mu(\tilde{p}_x^{-1}(y)), \end{aligned}$$

and hence (u, μ) is a random Strotz representation of λ.

[32] To see that such a selection function $\tilde{p}_x$ exists, fix any measurable selection function $\hat{p}_x : \mathcal{V} \to x$ with $\hat{p}_x(\tilde{v}) \in B_u(B_{\tilde{v}}(x))$ for all $\tilde{v} \in \mathcal{V}$. Let $\bar{\mathcal{V}} = g([0, 1]) \subset \mathcal{V}$. When the codomain of g is restricted to $\bar{\mathcal{V}}$, i.e., $g : [0, 1] \to \bar{\mathcal{V}}$, this function is a bijection. Now define $\tilde{p}_x(\tilde{v}) = p_x(g^{-1}(\tilde{v}))$ for $\tilde{v} \in \bar{\mathcal{V}}$ and $\tilde{p}_x(\tilde{v}) = \hat{p}_x(\tilde{v})$ for $\tilde{v} \notin \bar{\mathcal{V}}$.

To prove statement 3, suppose $\mu_i \equiv F_i \circ g^{-1}$ for $i = 1, 2$ and $\mu_1 \gg_u \mu_2$. Fix any $\alpha \in [0, 1]$, and let $\mathcal{U} = \{v' \in \mathcal{V} : v' \gg_u \alpha u + (1 - \alpha)v\}$. By construction, $\mathcal{U}$ is a u-upper set, so $\mu_1(\mathcal{U}) \geq \mu_2(\mathcal{U})$. In addition, $g^{-1}(\mathcal{U}) = [\alpha, 1]$. Therefore,

$$F_1([\alpha, 1]) = \mu_1(\mathcal{U}) \geq \mu_2(\mathcal{U}) = F_2([\alpha, 1]).$$

Since this is true for all $\alpha \in [0, 1]$, $F_1 \geq_{FOSD} F_2$.

Conversely, suppose $F_1 \geq_{FOSD} F_2$. Fix any u-upper set $\mathcal{U}$. Note that for any $0 \leq \alpha \leq \alpha' \leq 1$, we have $g(\alpha') \gg_u g(\alpha)$ and hence

$$g(\alpha) \in \mathcal{U} \implies g(\alpha') \in \mathcal{U}.$$

This implies that the set $g^{-1}(\mathcal{U})$ is an interval from some $\alpha^* \in [0, 1]$ to 1.[33] Therefore,

$$\mu_1(\mathcal{U}) = F_1(g^{-1}(\mathcal{U})) \geq F_2(g^{-1}(\mathcal{U})) = \mu_2(\mathcal{U}).$$

Since this is true for all u-upper sets, $\mu_1 \gg_u \mu_2$. ■

Turning now to the proof of Theorem 5, suppose $\succsim$ has an uncertain intensity Strotz representation $(u, v, \hat{F})$, and λ has an uncertain intensity Strotz representation (u, v, F). To establish sufficiency, suppose the individual is naive. The conclusion that $\hat{F} \geq_{FOSD} F$ follows immediately from Theorem 3 and Lemma 10 in the case of finite C. A small additional step is needed in the case where C does not have finite cardinality.

Specifically, there is a finite subset $C^* \subset C$ such that the restrictions of u and v to $\Delta(C^*)$ are also independent expected-utility functions, that is, neither one a positive or negative affine transformation of the other. We abuse notation slightly and also denote these restrictions by u, v. Note also that the restrictions of $\succsim$ and λ to $\mathcal{K}(\Delta(C^*))$ must satisfy naiveté. Let $\mathcal{V}^*$ denote the set of all continuous functions $\tilde{v} : C^* \to \mathbb{R}$. Define measures $\hat{\mu} \equiv \hat{F} \circ g^{-1}$ and $\mu \equiv F \circ g^{-1}$ on $\mathcal{V}^*$, where $g : [0, 1] \to \mathcal{V}^*$ is defined by $g(\alpha) = \alpha u + (1 - \alpha)v$. By Lemma 10, $(u, \hat{\mu})$ and (u, μ) are random Strotz representations of the restrictions of $\succsim$ and λ to $\mathcal{K}(\Delta(C^*))$. Therefore, Theorem 3 implies $\hat{\mu} \gg_u \mu$. By Lemma 10, this implies $\hat{F} \geq_{FOSD} F$.

To establish necessity, suppose $\hat{F} \geq_{FOSD} F$. Define measures $\hat{\mu} \equiv \hat{F} \circ g^{-1}$ and $\mu \equiv F \circ g^{-1}$ on $\mathcal{V}$. Lemma 10 implies that $(u, \hat{\mu})$ and (u, μ) are random Strotz representations for $\succsim$ and λ, and it implies that $\hat{\mu} \gg_u \mu$. By Lemma 7, the individual is naive.

[33] That is, it is equal to either $(\alpha^*, 1]$ or $[\alpha^*, 1]$, where $\alpha^* = \inf(g^{-1}(\mathcal{U}))$.

D.5 Proof of Theorem 6

To establish sufficiency, suppose individual 1 is more naive than individual 2. As argued in the proof of Theorem 5, there exists a finite subset $C^* \subset C$ such that the restrictions of u and v to $\Delta(C^*)$ are also independent expected-utility functions. Note that when the preferences and random choice rules of the individuals are restricted to the domain $\mathcal{K}(\Delta(C^*))$, individual 1 is still more naive than individual 2. Define $\mathcal{V}^*$ and g as in the proof of Theorem 5, and define measures $\hat{\mu}_i \equiv \hat{F}_i \circ g^{-1}$ and $\mu_i \equiv F_i \circ g^{-1}$ on $\mathcal{V}^*$ for $i = 1, 2$. By Lemma 10, $(u, \hat{\mu}_i)$ and (u, μ_i) are random Strotz representations for the restrictions of $\succsim_i$ and λ_i to $\mathcal{K}(\Delta(C^*))$ for $i = 1, 2$. Therefore, Theorem 4 implies

$$\hat{\mu}_1 \gg_u \hat{\mu}_2 \gg_u \mu_2 \gg_u \mu_1.$$

By Lemma 10, this implies

$$\hat{F}_1 \geq_{FOSD} \hat{F}_2 \geq_{FOSD} F_2 \geq_{FOSD} F_1.$$

To establish necessity, suppose

$$\hat{F}_1 \geq_{FOSD} \hat{F}_2 \geq_{FOSD} F_2 \geq_{FOSD} F_1.$$

Define measures $\hat{\mu}_i \equiv \hat{F}_i \circ g^{-1}$ and $\mu_i \equiv F_i \circ g^{-1}$ on $\mathcal{V}$ for $i = 1, 2$. Lemma 10 implies that $(u, \hat{\mu}_i)$ and (u, μ_i) are random Strotz representations for $\succsim$ and λ, and it implies that

$$\hat{\mu}_1 \gg_u \hat{\mu}_2 \gg_u \mu_2 \gg_u \mu_1.$$

By Lemma 7, the individuals are naive. By Lemma 8, individual 2 is more virtuous than individual 1. By Lemma 3, individual 2 is more temptation averse than individual 1. Therefore, individual 1 is more naive than individual 2.

D.6 Proof of Theorem 10

By Theorem 2, $u_1 \approx u_2 \equiv u$ and $\hat{v}_1 \gg_u \hat{v}_2 \gg_u v_2 \gg_u v_1$. There are two cases to consider, depending on whether $v_1 \approx v_2$ or not.

Case 1—$v_1 \approx v_2$: Let $v \equiv v_2 \approx v_1$. Since individual 1 is strictly more naive than individual 2, in this case we must have $\hat{v}_1 \gg_u \hat{v}_2$, but not $\hat{v}_1 \approx \hat{v}_2$. Therefore, it can be shown that there exist lotteries p^1, p^2, p^3, p^4 such that[34]

$$\begin{aligned}
&u(p^1) > u(p^2) > u(p^3) > u(p^4)\\
&\hat{v}_1(p^1) > \hat{v}_1(p^2) > \hat{v}_1(p^3) > \hat{v}_1(p^4)\\
&\hat{v}_2(p^3) > \hat{v}_2(p^1), \hat{v}_2(p^2) > \hat{v}_2(p^4)\\
&v(p^3) > v(p^4) > v(p^1), v(p^2).
\end{aligned}$$

Let $y = \{p^1, p^2, p^3, p^4\}$ and $x = \{p^2, p^4\}$. The rankings of the lotteries according to u and $\hat{v}_1, \hat{v}_2$ imply that $y \sim_1 \{p^1\} \succ_1 \{p^2\} \sim_1 x$ and $y \sim_2 \{p^3\} \prec_2 \{p^2\} \sim_2 x$. The ranking according to v implies that $\mathcal{C}_i(y) = p^3$ and $\mathcal{C}_i(x) = p^4$ for $i = 1, 2$. Therefore, $\{\mathfrak{C}_1(\{x, y\})\} = \{p^3\} = \{\mathfrak{C}_1(\{y\})\}$ and $\{\mathfrak{C}_2(\{x, y\})\} = \{p^4\} \prec_2 \{p^3\} = \{\mathfrak{C}_2(\{y\})\}$.

Case 2—v_1 is not an affine transformation of v_2: Under these assumptions, it can be shown that there exist lotteries p^1, p^2, p^3 such that

$$\begin{aligned}
&u(p^1) > u(p^2) > u(p^3)\\
&\hat{v}_2(p^2) > \hat{v}_2(p^1) > \hat{v}_2(p^3)\\
&v_2(p^2) > v_2(p^3) > v_2(p^1)\\
&v_1(p^3) > v_1(p^2) > v_1(p^1).
\end{aligned}$$

The ranking of these lotteries according to $\hat{v}_1$ is not important for the result, although it is true that the above rankings and $\hat{v}_1 \gg_u \hat{v}_2$ imply $\hat{v}_1(p^1), \hat{v}_1(p^2) > \hat{v}_1(p^3)$. Let $y = \{p^1, p^2, p^3\}$ and $x = \{p^1, p^3\}$. The ranking according to v_1 implies $\mathcal{C}_1(y) = \mathcal{C}_1(x) = p^3$, so $\{\mathfrak{C}_1(\{x, y\})\} = \{p^3\} = \{\mathfrak{C}_1(\{y\})\}$. The rankings according to u and $\hat{v}_2$ imply that $y \sim_2 \{p^2\} \prec_2 \{p^1\} \sim_2 x$. The ranking according to v_2 implies that $\mathcal{C}_2(y) = p^2$ and $\mathcal{C}_2(x) = p^3$. Thus $\{\mathfrak{C}_2(\{x, y\})\} = \{p^3\} \prec_2 \{p^2\} = \{\mathfrak{C}_2(\{y\})\}$.

[34] The arguments needed to prove this claim are similar to those in the proof of Theorem 8 and are omitted.