Theoretical Economics 5 (2010), 127–158 1555-7561/20100127

Uphill self-control

Jawwad Noor
Department of Economics, Boston University

Norio Takeoka
Faculty of Economics, Yokohama National University

This paper extends the theory of temptation and self-control introduced by Gul and Pesendorfer (2001) to allow for increasing marginal costs of resisting temptation, that is, convex self-control costs. It also proves a representation theorem that admits a general class of self-control cost functions. Both models maintain the Order, Continuity, and Set Betweenness axioms but violate Independence.

Keywords. Temptation, self-control, nonlinear self-control costs.

JEL classification. D11.

1. Introduction

Gul and Pesendorfer (2001) (henceforth GP) introduce a theory of choice under temptation. They model an agent who experiences temptation at the moment of choice from a menu and anticipates this in an ex ante period where he selects what menu to face. In this ex ante period he has a particular perspective on what he should choose from menus, embodied in a "normative preference." He understands that his choice from menus will not necessarily respect normative preference, but rather will seek to balance his normative preference with the cost of resisting temptation.

GP axiomatize the model described by (1) and (2) below. Denote the space of alternatives (lotteries) by Δ and the space of menus (nonempty subsets of Δ) by Z. The primitive is a preference $\succsim$ over menus Z, and reflects the ex ante choice between menus prior to ex post (unmodelled) choice from a menu. A general class of representations for $\succsim$ that captures an abstract version of the story in GP is

$$W(x) = \max_{\mu \in x}\{u(\mu) - c(\mu, x)\}, \quad x \in Z, \tag{1}$$

where $u : \Delta \to \mathbb{R}$ represents a von Neumann–Morgenstern (vNM) normative preference and $c(\mu, x)$ reflects the self-control cost of choosing μ from the menu x. The representation suggests that the utility of a menu is its indirect utility: the maximum of normative

Jawwad Noor: jnoor@bu.edu
Norio Takeoka: takeoka@ynu.ac.jp
We thank Larry Epstein, Bart Lipman, the audiences at Boston, Hitotsubashi, Tokyo, and Toyama Universities, and at the First Annual UECE Lisbon Meetings, and also Ed Green (the editor) and two referees for helpful comments. Takeoka gratefully acknowledges financial support from a Grant-in-Aid for Young Scientists (B). The usual disclaimer applies.

DOI: 10.3982/TE536

utility less self-control costs. GP's model is a specialization that tells a very specific story about the self-control cost function c. Their model identifies a vNM function $v:\Delta \to \mathbb{R}$ that represents the temptation perspective, and measures self-control costs c in terms of the difference between the maximum temptation utility achievable in a menu x and the actual temptation utility achieved by a choice $\mu \in x$:

$$c(\mu, x) = \max_{\eta \in x} v(\eta) - v(\mu). \tag{2}$$

That is, the self-control cost of choosing μ from x is identified with the corresponding "temptation opportunity cost." Note that the representation suggests that the agent's ex post choice from a menu x is

$$\mathcal{C}(x) = \arg\max_{\mu \in x} \Big\{ u(\mu) - \Big(\max_{\eta \in x} v(\eta) - v(\mu) \Big) \Big\} = \arg\max_{\mu \in x} \{ u(\mu) + v(\mu) \}.$$

That is, the GP agent's anticipated choices maximize a utility function $w = u + v$, a compromise between normative and temptation utility.

Observe that the GP agent's anticipated choices $\mathcal{C}$ must satisfy the Weak Axiom of Revealed Preference (WARP). In a companion paper (Noor and Takeoka 2010) we argue that the plausible outcome of an agent's internal struggle with temptation is that choice behavior may be inconsistent across choice problems, in the sense of violating WARP. For instance, a dieter may resist temptation and choose to have no dessert from the menu {no dessert, small dessert}, but the presence of a large dessert in {no dessert, small dessert, large dessert} may make it harder for him to skip dessert, and as a compromise he may end up choosing the small dessert. In Noor and Takeoka (2010) we consider the idea that the presence of the large dessert may trigger a craving that weakens his self-control. In this paper we consider an alternative hypothesis: the exertion of self-control may involve an *uphill battle*. That is, small deviations from temptation may be easier to accomplish than larger ones. In the context of the example, the dieter may be able to choose no dessert over the small dessert in the first menu, but may find it substantially harder to choose no dessert over the large dessert in the second menu. He may end up choosing the small dessert in the second menu simply because it would be easier to do so.

The idea of uphill self-control is considered in Takeoka (2006) and Fudenberg and Levine (2006, 2009), where the agent is modelled as having increasing marginal costs of exerting self-control.[1] These papers show that such *convex self-control costs* can generate interesting implications for choice under risk and over time: Takeoka (2006) notes that convex self-control costs can generate the version of the Allais paradox known as the common ratio effect, and also the findings of Keren and Roelofsma (1995) in an intertemporal choice context. We describe these in the concluding section of this paper. In the context of a dual-self model, Fudenberg and Levine (2006, 2009) independently make some of the observations in Takeoka (2006), but show in addition that their model with convex costs can explain a range of behavioral anomalies including Rabin's paradox

[1]Fudenberg and Levine (2006, 2009) cite psychology research that demonstrates that self-control is a limited resource, and suggest that this supports the idea of increasing marginal cost of self-control.

and the observed relationship between cognitive load and risk preferences. They show that plausible parameter values allow them to quantitatively fit their model to data on a range of behaviors.

This paper explores the foundations of the notion of uphill self-control and of models with nonlinear costs of self-control. We axiomatize two models.

General self-control representation: The first model takes the form

$$W(x) = \max_{\mu \in x}\Big\{u(\mu) - c\Big(\mu, \max_{\eta \in x} v(\eta)\Big)\Big\}, \quad x \in Z,$$

where u and v are linear and c satisfies some minimal regularity properties that support its interpretation as the cost of self-control. This expunges from GP's model all but the basic linearity required for the existence of linear normative and temptation utilities, without departing from the basic qualitative story underlying GP's model. Thus, the agent maximizes normative utility net of self-control costs, and the cost $c(\mu, \cdot)$ is increasing for any given possible choice μ. The peculiar axiom associated with this model states that if the menu $\{\mu, \eta\}$ is such that the agent is tempted by η but is able to resist it, then replacing η with a less tempting alternative can only make him better off.

Convex self-control representation: The second model is a nonlinear generalization of GP's model given by

$$W(x) = \max_{\mu \in x}\Big\{u(\mu) - \varphi\Big(\max_{\eta \in x} v(\eta) - v(\mu)\Big)\Big\}, \quad x \in Z,$$

for some increasing convex function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$. This model enriches the general model by requiring self-control costs to depend on the temptation opportunity cost of choice, as in the GP model, but without forcing this dependence to be linear. The peculiar axioms for this model express how the mixing of elements of a menu with others affects the agent's preference for the menu.

For perspective, we point out where our paper stands relative to the current development of the axiomatic literature on temptation. GP's axiomatization of their model makes use of four axioms: Order, Continuity, Independence, and Set Betweenness. The first three are natural extensions of the von Neumann–Morgenstern axioms to a sets-of-lotteries setting, and the fourth expresses the agent's temptation and anticipated choice from menus. Of the four axioms, Set Betweenness is clearly a substantive axiom for a model of decision under temptation. Indeed, existing generalizations of GP's model (Chatterjee and Krishna 2009, Dekel et al. 2009, Kopylov 2009, Stovall 2010) focus on relaxing Set Betweenness while maintaining Independence. This paper (and the companion paper (Noor and Takeoka 2010)) seeks to understand what is potentially lost if one maintains Independence, a convenient and standard axiom in the literature that seems less important than Set Betweenness from the point of view of decision under temptation. We show that the axiom is not auxiliary in nature in that it rules out intuitive qualitative stories about decision under temptation, even if Set Betweenness is retained. In fact, both models axiomatized in this paper satisfy Set Betweenness.

The remainder of the paper is organized as follows. This introduction concludes with a mention of related literature. Sections 2 and 3 axiomatize our general and convex models, respectively. Section 4 concludes with some observations of the convex model's implied properties for ex post choice. All proofs are contained in the Appendices.

1.1 *Related literature*

In a game-theoretic setting, Fudenberg and Levine (2006, 2009) study a long-run self that seeks to maximize the discounted utility of a sequence of short-run impulsive selves and may intervene in the choices of a short-run self but at a cost. They show that the equilibria of the game played by those selves can be regarded as the solution to a maximization problem analogous to (1). Their general setup allows for cases where the cost function might be convex, which would then correspond to specializations of (1) that include the convex self-control model. In Fudenberg and Levine (2009) they construct and analyze a model with convex self-control costs that explains and quantitatively fits a range of experimental findings.

In the temptation literature, Chatterjee and Krishna (2009), Dekel et al. (2009), Kopylov (2009), and Stovall (2010) generalize GP's model. They model agents who are uncertain about temptation (e.g., uncertainty about the temptation preference itself, or uncertainty regarding the strength of self-control, etc.), and Dekel et al. also axiomatize a model where multiple temptations are experienced by the agent. These models relax Set Betweenness but maintain Independence. This paper explores an alternative direction where Set Betweenness is maintained and Independence relaxed. We interpret violations of Independence in terms of nonlinear self-control costs. In a companion paper (Noor and Takeoka 2010), we focus on another possible source of violations of Independence, specifically changes in the agent's self-control ability triggered by the contents of menus.

Nehring (2006) is interested in a more careful description of the notion of self-control, which he interprets in terms of a preference over preferences (second order preferences). Olszewski (forthcoming) relaxes the single dimensionality of temptation in GP's model by permitting different alternatives in a menu to be tempted by different alternatives in the menu. Though not specifically motivated by the idea of uphill self-control, these authors provide foundations for functional forms that can accommodate uphill self-control. On a technical level these papers differ substantially from ours in that they focus on *discrete settings*, whereas we provide an axiomatic generalization of GP's model in a sets-of-lotteries setting.

Finally, we mention Gul and Pesendorfer (2005), who, also in a discrete setting, axiomatize a general model. Their representation for preference over menus is of the form

$$W(x) = f\Big(\max_{\mu\in x} w(\mu), \max_{\eta\in x} v(\eta)\Big),$$

which admits the interpretation that the agent is tempted to maximize some temptation utility v but choice is determined by the maximization of some function w. The two utilities are then aggregated by the function f. To compare, we note that our general model corresponds to the form

$$W(x) = \max_{\mu\in x} f\Big(\mu, \max_{\eta\in x} v(\eta)\Big).$$

Thus, while ex post choice in the Gul and Pesendorfer (2005) model maximizes a utility w, in our model ex post choice from x maximizes the *menu-dependent* utility $f(\mu, \max_{\eta\in x} v(\eta))$. Indeed, ex post choice in their model satisfies the Weak Axiom of Revealed Preference, and this in turn suggests that the model is not suitably interpreted as one involving nonlinear self-control.

2. General model

For any compact metric space C, $\Delta(C)$ denotes the set of all probability measures on the Borel σ-algebra of C, endowed with the weak convergence topology; $\Delta(C)$ is compact and metrizable (Aliprantis and Border 1999, Theorem 14.11), and we often write it simply as Δ. Let $Z = \mathcal{K}(\Delta)$ denote the set of all nonempty compact subsets of Δ. When endowed with the Hausdorff topology, Z is a compact metric space (Aliprantis and Border 1999, Theorem 3.71(3)). An element $x \in Z$ is referred to as a menu. Generic elements of Z are x, y, z, whereas generic elements of Δ are μ, η, ν. For $\alpha \in [0,1]$, $\mu\,\alpha\,\eta \in \Delta$ is the α-mixture that assigns $\alpha\mu(A) + (1-\alpha)\eta(A)$ to each A in the Borel σ-algebra of C. Similarly,

$$x\,\alpha\,y := \alpha x + (1-\alpha) y := \{\mu\,\alpha\,\eta \mid \mu \in x, \eta \in y\} \in Z$$

is an α-mixture of menus x and y.

As in GP, the primitive is a preference $\succsim$ over Z.

2.1 *Axioms*

The first three axioms are familiar from GP.

Axiom 1 (Order). *The preference $\succsim$ is complete and transitive.*

Axiom 2 (Continuity). *The sets $\{y \in Z \mid y \succsim x\}$ and $\{y \in Z \mid x \succsim y\}$ are closed for each $x \in Z$.*

Axiom 3 (Set Betweenness). *For all $x, y \in Z$,*

$$x \succsim y \Longrightarrow x \succsim x \cup y \succsim y.$$

For a complete discussion of Set Betweenness, we refer the reader to GP, who argue that the axiom is consistent with the story where the agent's ranking of menus is sensitive only to anticipated choice from menus and the level of temptation contained in it. What needs to be noted for the purpose of this paper is that the interpretation of Set Betweenness does not hinge on any precise properties of how exertion of self-control in the menu $x \cup y$ affects its desirability. This suggests that Set Betweenness is not inconsistent with generalizations of GP that relax the structure on self-control costs. The interpretations of the following rankings, all consistent with Set Betweenness, should also be noted.

- The ranking of singletons $\{\mu\} \succsim \{\eta\}$ reveals the ex ante preference over the ex post consumption of μ versus η. This "commitment ranking" of lotteries may be interpreted as the agent's normative preference—it is his perspective from a distance on what *should* be consumed ex post.

- The ranking $x \succ x \cup y$ is referred to as a *preference for commitment*, and suggests that some element in y is a source of temptation in $x \cup y$. Note that by this interpretation, $\{\mu\} \succ \{\mu, \eta\}$ reveals directly that η tempts μ. The lack of a preference for commitment in $\{\mu\} \sim \{\mu, \eta\} \succ \{\eta\}$ suggests that η does not tempt μ. Indeed, here the agent anticipates no internal conflict (temptation) when facing $\{\mu, \eta\}$, and therefore it must be that μ is preferred to η under both normative and temptation rankings.

- While the previous bullet points discussed how normative and temptation preferences are revealed by $\succsim$, now we discuss how anticipated ex post choice from menus may be revealed by $\succsim$. The ranking $\{\mu, \eta\} \succ \{\eta\}$ suggests that μ is the anticipated choice from $\{\mu, \eta\}$, since otherwise there would be no reason to value $\{\mu, \eta\}$ over $\{\eta\}$. When $\{\mu\} \succ \{\eta\}$ and $\{\mu, \eta\} \sim \{\eta\}$, the agent is indifferent between $\{\mu, \eta\}$ and $\{\eta\}$ although the former contains the normatively superior alternative μ, which suggests that η is an anticipated choice in $\{\mu, \eta\}$.[2]

GP's fourth axiom formulates the standard vNM independence axiom in the menus setting: for all x, y, z and $\alpha \in (0, 1)$,

$$x \succ y \Longrightarrow \alpha x + (1-\alpha)z \succ \alpha y + (1-\alpha)z.$$

We refer the reader to GP for a discussion of the axiom. We relax Independence so as to impose vNM structure on commitment preference and temptation preference *only*.

Axiom 4 (Commitment Independence). *For any μ, η, ν and $\alpha \in (0, 1)$,*

$$\{\mu\} \succ \{\eta\} \Longrightarrow \{\mu \,\alpha\, \nu\} \succ \{\eta \,\alpha\, \nu\}.$$

Axiom 5 (Temptation Independence). *For any μ, η, ν and $\alpha \in (0, 1)$ such that $\{\mu\} \succ \{\eta\}$,*

$$\{\mu\} \succsim \{\mu, \eta\} \Longleftrightarrow \{\mu \,\alpha\, \nu\} \succsim \{\mu \,\alpha\, \nu, \eta \,\alpha\, \nu\}.$$

Moreover, for any μ, η, η' and $\alpha \in (0, 1)$ such that $\{\mu\} \succ \{\eta\}, \{\eta'\}$,

$$\{\mu\} \succ \{\mu, \eta\} \text{ and } \{\mu\} \succ \{\mu, \eta'\} \Longrightarrow \{\mu\} \succ \{\mu, \eta \,\alpha\, \eta'\}$$
$$\{\mu\} \sim \{\mu, \eta\} \text{ and } \{\mu\} \sim \{\mu, \eta'\} \Longrightarrow \{\mu\} \sim \{\mu, \eta \,\alpha\, \eta'\}.$$

Commitment Independence is readily interpreted. The first part of Temptation Independence states that η tempts (resp. does not tempt) μ if and only if $\eta \,\alpha\, \nu$ tempts

[2] When the agent is ex ante indifferent between whether μ is consumed ex post or η, then ex ante he would be indifferent between the three menus $\{\mu\}$, $\{\mu, \eta\}$, and $\{\eta\}$. Thus, when $\{\mu\} \sim \{\eta\}$, we cannot infer the agent's temptation preference over μ, η and neither can we infer anticipated choice from $\{\mu, \eta\}$.

(resp. does not tempt) $\mu\,\alpha\,\nu$. The second part states that if η and η' both tempt (resp. do not tempt) μ, then the mixture $\eta\,\alpha\,\eta'$ tempts (resp. does not tempt) μ. These are properties that would be expected from a vNM temptation preference.

To introduce the next axiom, write $\eta \succsim_T \mu$ if either

$$\{\mu\} \succ \{\mu, \eta\} \quad \text{or} \quad \{\eta\} \sim \{\mu, \eta\} \succ \{\mu\}.$$

As in the previous discussion, the first condition says that η is more tempting than μ, and the second condition says that μ is not more tempting than η. Thus in either case, η is at least as tempting as μ or, equivalently, we can say that η *weakly tempts* μ.

We adopt the following key axiom for our general model.

Axiom 6 (Temptation Aversion). *If* $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ *for some* μ, η, *then for any* ν,

$$\eta \succsim_T \nu \Longrightarrow \{\mu, \nu\} \succsim \{\mu, \eta\}.$$

Suppose the menu $\{\mu, \eta\}$ is such that η is tempting, though resistible. The axiom simply says that if we replace η with something less tempting, then the agent is not worse off. Intuitively, the lower the temptation in a menu, the lower the self-control costs associated with resisting temptation. It is interesting to note that the axiom also covers the possibility that if ν is less tempting than η and also normatively superior, then ex post the agent may optimally choose to submit to temptation rather than incur any self-control cost. However, even in this case, the agent would be better off with $\{\mu, \nu\}$ than $\{\mu, \eta\}$. Observe that, because of the lower self-control cost, choosing μ in $\{\mu, \nu\}$ is better than choosing μ in $\{\mu, \eta\}$. But then the optimal choice from $\{\mu, \nu\}$ must leave him better off compared to choosing μ in $\{\mu, \eta\}$. The optimal choice may well be to submit to temptation in $\{\mu, \nu\}$, since the normative cost of doing so may be smaller than the self-control cost of resisting.

2.2 *Representation theorem*

The general representation result in this paper is given by the following theorem.

Theorem 1. *A preference* $\succsim$ *satisfies Order, Continuity, Set Betweenness, Commitment Independence, Temptation Independence, and Temptation Aversion if and only if there exists a representation* $W : Z \to \mathbb{R}$ *for* $\succsim$ *defined by*

$$W(x) = \max_{\mu \in x} \Big\{ u(\mu) - c\Big(\mu, \max_{\eta \in x} v(\eta)\Big) \Big\},$$

where $u, v : \Delta \to \mathbb{R}_+$ *are continuous linear functions, and* $c : \Delta \times v(\Delta) \to \mathbb{R}_+$ *is a continuous function that is weakly increasing in its second argument and satisfies*

(i) if $c(\mu, v(\eta)) > 0$, *then* $v(\mu) < v(\eta)$

(ii) if $u(\mu) > u(\eta)$ *and* $v(\mu) < v(\eta)$, *then* $c(\mu, v(\eta)) > 0$.

A preference $\succsim$ that satisfies the noted axiom is referred to as a *general self-control preference*. The representation is often identified with the tuple (u, v, c). The function c possesses minimal properties required to interpret it as a self-control cost function. Monotonicity in its second argument reflects the fact that choosing any given alternative μ is more costly from menus with greater temptation. Condition (i) says that the self-control cost is positive only when self-control is exerted. Indeed, the condition implies that when η is the tempting alternative in the menu, then choosing η is associated with the cost $c(\eta, v(\eta)) = 0$, that is, the cost of submitting to temptation is zero. Condition (ii) says that the self-control cost of resisting temptation is strictly positive.

As the following uniqueness theorem reveals, the cost function c is not pinned down over the entire domain, specifically for (μ, l) such that $v(\mu) < l$ and there is no η with $v(\eta) = l$ such that

$$\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}.$$

The intuition is that alternatives that are always dominated in both normative and temptation terms never have any bearing on the ex ante preference $\succsim$, which is sensitive only to each menu's chosen alternative and most tempting alternative. Dominated alternatives are never chosen and are never most tempting, and thus have no impact on $\succsim$. Nevertheless, the fact that the cost of choosing such unchosen alternatives is not pinned down by $\succsim$ is arguably only a minimal detraction, if at all: alternatives that are never most tempting and never chosen are also not of interest from either a descriptive or a normative standpoint.

We now state the uniqueness result. Say that $\succsim$ is *nondegenerate* if there exists μ, η such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$.

Theorem 2. *Suppose that (u, v, c) and (u', v', c') are both representations of a nondegenerate general self-control preference. Then there exist constants $\alpha_u, \alpha_v > 0$ and β_u, β_v such that $u' = \alpha_u u + \beta_u$ and $v' = \alpha_v v + \beta_v$. Moreover, $c'(\mu, l) = \alpha_u c(\mu, (l - \beta_v)/\alpha_v)$ on the set*

$$\{(\mu, l) \mid v'(\mu) \geq l \text{ or } \{\mu\} \succ \{\mu, \eta\} \succ \{\eta\} \text{ for some } \eta \text{ with } v'(\eta) = l\}.$$

The straightforward proof is omitted.

2.3 *Proof outline for Theorem 1*

Order, Continuity, and Commitment Independence straightforwardly ensure that there exists a continuous mixture linear representation $u : \Delta \to \mathbb{R}$ for the commitment preference. Regarding the temptation preference, Noor and Takeoka (2010, Theorem 2) show that under Temptation Independence there exists a continuous mixture linear function $v : \Delta \to \mathbb{R}$ such that for any μ, η for which $u(\mu) > u(\eta)$,

$$\{\mu\} \succ \{\mu, \eta\} \Longleftrightarrow v(\eta) > v(\mu).$$

That is, v conflicts with u whenever the agent exhibits a preference for commitment. Under the interpretation that a preference for commitment is associated with the experience of temptation—the conflict of underlying normative and temptation preference—the function v takes on its interpretation as a representation for temptation preference.

Next, we show that there exists a representation $W : Z \to \mathbb{R}$ for $\succsim$. Since u is continuous on Δ, there exist a maximal and a minimal lottery $\mu^\Delta, \mu_\Delta \in \Delta$ with respect to u. Continuity and Set Betweenness imply that $\{\mu^\Delta\} \succsim x \succsim \{\mu_\Delta\}$ for all $x \in Z$. By a standard argument, for all $x \in Z$ there exists a unique number $\alpha_x \in [0, 1]$ such that $x \sim \{\mu^\Delta\, \alpha_x\, \mu_\Delta\}$. Thus,

$$W(x) \equiv u(\mu^\Delta\, \alpha_x\, \mu_\Delta)$$

is a representation of $\succsim$. Note that $W(\{\mu\}) = u(\mu)$ for all $\mu \in \Delta$.

The remainder of the proof concerns how to define the self-control cost function c and how to convert W to the desired form. Consider any μ, η satisfying $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$, that is, μ is chosen from $\{\mu, \eta\}$ with the exertion of self-control. The difference $u(\mu) - W(\{\mu, \eta\})$ is naturally interpreted as a measure of the cost of self-control at $\{\mu, \eta\}$, and thus our cost function c must satisfy

$$c(\mu, v(\eta)) = u(\mu) - W(\{\mu, \eta\}) > 0. \tag{3}$$

We define the self-control cost function c as follows. Consider the correspondence $L : v(\Delta) \rightsquigarrow \Delta$ given by

$$L(l) \equiv \{\eta \mid v(\eta) \leq l\}.$$

The set $L(l)$ is a lower contour set defined by temptation utility. Define c by

$$c(\mu, l) \equiv \max\Big[0, \max_{\nu \in L(l)} \big\{u(\mu) - W(\{\mu, \nu\})\big\}\Big].$$

Thus, fixing a choice μ and temptation level l, the cost $c(\mu, l)$ is identified with the maximum value that $u(\mu) - W(\{\mu, \nu\})$ can take as ν varies over the lower contour set $L(l)$; but if this maximum value is negative ($\{\mu, \nu\} \succ \{\mu\}$ for all ν), then $c(\mu, l)$ is set to 0.

This cost function has the desired properties. Temptation Aversion implies that if $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$, then $\{\mu, \nu\} \succsim \{\mu, \eta\}$ whenever $v(\eta) \geq v(\nu)$. Therefore, for such μ, η, it must be that for all $\nu \in L(v(\eta))$,

$$u(\mu) - W(\{\mu, \eta\}) \geq u(\mu) - W(\{\mu, \nu\}),$$

and so we must have (3). On the other hand, the ranking $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$ suggests that η is tempting and also chosen from $\{\mu, \eta\}$, that is, self-control is not exerted. In this case, by definition,

$$c(\eta, v(\eta)) = 0,$$

that is, there is no cost of self-control.

By using (u, v, c) as defined above, we show that the desired functional form for W obtains for all binary menus. The remaining argument is more or less the same as in Gul and Pesendorfer (2001). Specifically, since $\succsim$ satisfies Set Betweenness, the representation can be extended to the set of all finite menus, and by Continuity in the Hausdorff metric topology, the representation can be extended to all of Z, as desired.

3. Convex model

In this section we present a specialization of our general model that has two properties. First, the cost of self-control depends on the temptation opportunity cost of choice, that is, the difference between the realized temptation utility from choice and the maximum temptation utility possible in the menu. Second, the cost of self-control is convex, thereby capturing the idea of uphill self-control. We first formally describe the functional form and then present its axiomatization.

3.1 *Functional form*

As in the general model, let u be a normative utility function and let v be a temptation utility function over lotteries. Both functions are continuous and mixture linear. If μ is chosen in $\{\mu, \eta\}$ with the exertion of self-control, then the temptation opportunity cost is $w = v(\eta) - v(\mu)$. We now describe a functional form where the cost of self-control is a convex transformation of w. This requires us to define the maximum normative benefit from self-control among binary menus where the temptation opportunity cost equals w: for all $w > 0$, let[3]

$$F(w) \equiv \sup\big\{u(\mu) - u(\eta) \mid w = v(\eta) - v(\mu) \text{ for some } \mu, \eta \text{ such that } v(\eta) - v(\mu) > 0 > u(\eta) - u(\mu)\big\}.$$

The inequalities $v(\eta) - v(\mu) > 0 > u(\eta) - u(\mu)$ are equivalent to $\{\mu\} \succ \{\mu, \eta\}$, that is, μ is normatively better but η is more tempting. Since $u(\mu) - u(\eta)$ is the normative benefit from self-control, $F(w)$ is the maximum possible value that the normative benefit from self-control can take among menus $\{\mu, \eta\}$ where the temptation opportunity cost $v(\eta) - v(\mu)$ equals w.

Definition 1 (Convex Self-Control Preference). A preference $\succsim$ is a convex self-control preference if there exists a representation $W : Z \to \mathbb{R}$ for $\succsim$ defined by

$$W(x) = \max_{\mu \in x}\Big\{u(\mu) - \varphi\Big(\max_{\mu' \in x} v(\mu') - v(\mu)\Big)\Big\}, \tag{4}$$

where $u, v : \Delta \to \mathbb{R}_+$ are continuous mixture linear functions and $\varphi : [0, \max_\Delta v - \min_\Delta v] \to \mathbb{R}_+$ is a continuous strictly increasing function such that $\varphi(0) = 0$ and φ is convex on a nondegenerate interval $[0, \overline{w}]$ and satisfies $\varphi(w) \geq F(w)$ for $w > \overline{w}$.

[3]Use the convention that the sup over an empty set is negative infinity. However, for the model, F need not be defined outside the set $B \equiv \{w \mid w = v(\eta) - v(\mu)$ for some μ, η such that $v(\eta) - v(\mu) > 0 > u(\eta) - u(\mu)\}$.

This model describes an agent for whom the costs of self-control increase at an *increasing rate* as more self-control is exerted.

The following remarks explain the restrictions on φ. Self-control costs are revealed only in those menus where self-control is exerted. Self-control is exerted only when the temptation opportunity cost of doing so is not too high. In general, there is a nondegenerate interval $[0, \overline{w}]$ such that the agent never exerts self-control in any menu where the temptation opportunity cost exceeds the threshold $\overline{w}$. Below the threshold level, the shape of the self-control cost function can be pinned down, and here the representation requires that φ should be convex. Above the threshold level, the property $\varphi(w) \geq F(w)$ must be satisfied. To understand this, observe that since F defines an upper bound on the benefit of self-control, it must be that whenever η tempts μ and $v(\eta) - v(\mu) = w > \overline{w}$, the normative benefit $u(\mu) - u(\eta)$ of self-control is always less than the self-control cost $\varphi(w)$:

$$\varphi(w) \geq F(w) \geq u(\mu) - u(\eta).$$

Hence the property "$\varphi(w) \geq F(w)$ for $w > \overline{w}$" is an expression of the fact that self-control is never exerted outside the interval $[0, \overline{w}]$.

Note that outside the interval, only the position of φ relative to F matters, and convexity has no behavioral meaning. Thus the representation requires convexity where it is *meaningful*. See the remarks after Theorem 3 below for when assuming convexity of φ on the full domain is without loss of generality.

3.2 *Axioms*

We augment the general model with three axioms.

Axiom 7 (Weak Binary Independence). *For all μ, μ', η, η', ν, ν' and all $\alpha \in (0, 1)$,*

$$\alpha\{\mu, \mu'\} + (1-\alpha)\{\nu\} \succsim \alpha\{\eta, \eta'\} + (1-\alpha)\{\nu\}$$
$$\Longrightarrow \alpha\{\mu, \mu'\} + (1-\alpha)\{\nu'\} \succsim \alpha\{\eta, \eta'\} + (1-\alpha)\{\nu'\}.$$

The axiom states that the ranking of binary menus $\{\mu, \mu'\}$ and $\{\eta, \eta'\}$ when mixed with a common singleton $\{\nu\}$ is independent of the singleton. This is a weakening of an axiom introduced by Ergin and Sarver (2009), where the axiom holds for all menus rather than just binary menus. Independence (together with Continuity) implies that for all $\mu, \mu', \eta, \eta', \nu, \nu' \in \Delta$ and all $\alpha, \beta \in (0, 1)$,

$$\alpha\{\mu, \mu'\} + (1-\alpha)\{\nu\} \succsim \alpha\{\eta, \eta'\} + (1-\alpha)\{\nu\}$$
$$\Longrightarrow \{\mu, \mu'\} \succsim \{\eta, \eta'\}$$
$$\Longrightarrow \beta\{\mu, \mu'\} + (1-\beta)\{\nu'\} \succsim \beta\{\eta, \eta'\} + (1-\beta)\{\nu'\}.$$

Weak Binary Independence is the weaker implication where the mixing coefficients α, β are equal.

To see the intuition behind the axiom, observe that, by definition,

$$\alpha\{\mu, \mu'\} + (1-\alpha)\{\nu\} = \{\mu\,\alpha\,\nu, \mu'\,\alpha\,\nu\},$$

and consider the interesting case where the agent chooses $\mu\,\alpha\,\nu$ from $\{\mu\,\alpha\,\nu, \mu'\,\alpha\,\nu\}$ after exerting self-control. The self-control cost depends on the temptation opportunity cost borne by choosing $\mu\,\alpha\,\nu$ over $\mu'\,\alpha\,\nu$. The key observation is that this temptation opportunity cost is independent of ν when temptation preferences are vNM.[4] Thus, replacing ν with ν' does not change the agent's temptation opportunity cost, which in turns leaves his self-control cost and his propensity for self-control unchanged. It follows that the effect on the ranking of menus $\{\mu\,\alpha\,\nu, \mu'\,\alpha\,\nu\}$ and $\{\eta\,\alpha\,\nu, \eta'\,\alpha\,\nu\}$ when ν is replaced with ν' comes not from any change in possible self-control costs or self-control ability in either menu, but from the fact that the chosen alternative in each menu is now a mixture with ν' rather than ν. However, ex ante preference over final consumption (revealed by commitment preferences) is also vNM. Since all aspects of the evaluation of a menu—namely anticipated choice and self-control costs—are independent of common mixtures, it follows that ranking of menus mixed with $\{\nu\}$ for a given α is invariant with respect to ν, as required by Weak Binary Independence.

For any menu x, define its *singleton equivalent* $e_x \in \Delta$ by $\{e_x\} \sim x$. To ease notation, we write $e_{\{\mu,\eta\}}$ as $e_{\mu\eta}$. Under Order, Continuity, Set Betweenness, and Commitment Independence, every menu has a singleton equivalent.

Axiom 8 (Self-Control Concavity-1). *For all μ, η and $\alpha \in (0,1)$, if $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$, then for any ν,*

$$\{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\} \succsim \alpha\{e_{\mu\eta}\} + (1-\alpha)\{\nu\}.$$

Assume the hypothesis and recall that by definition $\{\mu, \eta\}$ is as good as $\{e_{\mu\eta}\}$. Suppose both menus are mixed with a common singleton $\{\nu\}$. Observe that since $\mu\,\alpha\,\nu$ and $\eta\,\alpha\,\nu$ get closer to each other as α reduces, the temptation opportunity cost of choosing $\mu\,\alpha\,\nu$ from $\{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\}$ also reduces. In fact this implies that the cost of self-control in $\{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\}$ reduces monotonically as well, because self-control costs are proportional to the temptation opportunity cost. In a story of uphill self-control, the marginal costs of self-control should also fall. The axiom expresses this by suggesting that self-control costs fall "quickly" after mixing: While $\{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\}$ is as good as $\{e_{\mu\eta}\,\alpha\,\nu\}$ when $\alpha = 1$ (by definition of singleton equivalents), the axiom says that it becomes more attractive as α decreases from 1. Observe that the axiom also says that as α increases from 0, self-control costs increase "slowly": While $\{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\}$ is as good as $\{e_{\mu\eta}\,\alpha\,\nu\}$ when $\alpha = 0$, it becomes more attractive as α increases from 0. This is consistent with convexity of self-control costs.

The final axiom is similar in spirit, but applies the idea of the previous axiom to new cases.

[4]That is, $v(\mu'\,\alpha\,\nu) - v(\mu\,\alpha\,\nu) = \alpha[v(\mu') - v(\mu)]$.

Axiom 9 (Self-Control Concavity-2). *For all* μ_1, μ_2, η_1, η_2 *and* $\alpha \in (0,1)$, *if* $\{\mu_i\} \succ \{\mu_i, \eta_i\} \succ \{\eta_i\}$ *for* $i = 1, 2$, *then*

$$\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\} \succsim \alpha\{e_{\mu_1\eta_1}\} + (1-\alpha)\{e_{\mu_2\eta_2}\}.$$

This axiom is interpreted similarly. Suppose the agent exerts self-control at $\{\mu_i, \eta_i\}$ for $i = 1, 2$. Since η_i tempts μ_i, it is intuitive (and implied by the axioms of the general model) that $\eta_1 \, \alpha \, \eta_2$ tempts $\mu_1 \, \alpha \, \mu_2$. Moreover, since the agent can resist η_i and choose μ_i for $i = 1, 2$, it is also intuitive in a story of uphill self-control that he can resist $\eta_1 \, \alpha \, \eta_2$ and choose $\mu_1 \, \alpha \, \mu_2$: observe that when temptation preference is vNM, the temptation opportunity cost of choosing $\mu_1 \, \alpha \, \mu_2$ over $\eta_1 \, \alpha \, \eta_2$ is an α-weighted *average* of the temptation opportunity cost of choosing μ_i over η_i for $i = 1, 2$. Thus, the marginal cost of exerting self-control in $\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\}$ is not higher than that in both $\{\mu_i, \eta_i\}$, $i = 1, 2$. Since the agent exhibits self-control in each of the latter, we may therefore understand that he exerts self-control in $\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\}$.

Turning to the statement of the axiom, assume the hypothesis and suppose that the higher temptation opportunity cost is at $\{\mu_1, \eta_1\}$. As noted, the temptation opportunity cost of choosing $\mu_1 \, \alpha \, \mu_2$ from $\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\}$ is an α-weighted average of that of both $\{\mu_i, \eta_i\}$, $i = 1, 2$. Therefore the temptation opportunity cost—and in turn the self-control cost—of choosing $\mu_1 \, \alpha \, \mu_2$ in $\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\}$ falls monotonically as α goes from 1 to 0. While $\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\}$ is as good as $\{e_{\mu_1\eta_1} \, \alpha \, e_{\mu_2\eta_2}\}$ when $\alpha = 1$, the axiom says that it becomes more attractive as α decreases from 1, suggesting that the cost of self-control falls quickly as α decreases from 1. On the other hand, while $\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\}$ is as good as $\{e_{\mu_1\eta_1} \, \alpha \, e_{\mu_2\eta_2}\}$ when $\alpha = 0$, the axiom says that it becomes more attractive as α increases from 0, suggesting that the cost of self-control increases slowly as α increases from 0. This expresses the convexity of self-control costs.

3.3 *Representation theorem*

Say that $\succsim$ is a *nondegenerate preference* if there exist $\mu, \mu' \in \Delta$ with $\{\mu\} \succ \{\mu, \mu'\} \succ \{\mu'\}$. The following theorem is the main result of this section.

Theorem 3. *A nondegenerate preference* $\succsim$ *satisfies all the axioms of Theorem 1, Weak Binary Independence, and Self-Control Concavity-1 and -2 if and only if* $\succsim$ *is a convex self-control preference.*

This establishes the behavioral foundations of the convex self-control model. A discussion of the proof of the result is deferred to the next subsection.

Recall that the convex self-control representation requires φ to be convex only on an interval $[0, \overline{w}]$. Our axioms do not guarantee that a convex extension to the full domain $[0, \max_\Delta v - \min_\Delta v]$ exists. The issue is technical: for the extension to exist, it is necessary that φ be Lipschitz continuous on $[0, \overline{w}]$. It is possible to describe restrictions on preferences that guarantee this, but we omit them because of their lack of transparency.

As a corollary of the theorem, we obtain the GP model when Self-Control Concavity-2 is strengthened to a self-control linearity condition.

Corollary 1. *A convex self-control preference $\succsim$ admits a GP representation if and only if it satisfies self-control linearity: For all $\mu_1, \mu_2, \eta_1, \eta_2$ and $\alpha \in (0,1)$, if $\{\mu_i\} \succ \{\mu_i, \eta_i\} \succ \{\eta_i\}$ for $i = 1, 2$, then*

$$\{\mu_1 \, \alpha \, \mu_2, \eta_1 \, \alpha \, \eta_2\} \sim \alpha\{e_{\mu_1\eta_1}\} + (1-\alpha)\{e_{\mu_2\eta_2}\}.$$

Thus, self-control linearity is the behavioral expression of the linear self-control costs property in GP's model. This alternative axiomatization of GP's model provides perspective on the behavioral foundations of their model by highlighting the various implications of Independence in the presence of Order, Continuity, and Set-Betweenness. Indeed, this permits a more transparent evaluation of that axiom and, in turn, of the model.

Now turn to the uniqueness properties of the convex self-control representation. Given a representation (u, v, φ), the *self-control subdomain* is defined as

$$R = \Big\{ w \in \Big[0, \max_{\Delta} v - \min_{\Delta} v\Big] \,\Big|\, w = v(\eta) - v(\mu) \text{ for some } \mu, \eta \text{ such that } \{\mu\} \succ \{\mu, \eta\} \succ \{\eta\} \Big\}.$$

If $v(\eta) - v(\mu) \notin R$, self-control is never exerted at $\{\mu, \eta\}$. Thus, the actual shape of φ outside R is immaterial in the description of choice behavior. Note that since preference satisfy the Self-Control Concavity-1 axiom, R is an interval with $\inf R = 0$.[5] Notice also that the threshold level $\overline{w}$ associated with the representation must satisfy $R \subset [0, \overline{w}]$ because self-control is never exerted when $v(\eta) - v(\mu) > \overline{w}$.

Identify any convex self-control representation (4) with the corresponding tuple (u, v, φ). The uniqueness properties of the representation mirror those of the general representation.

Theorem 4. *Suppose that (u, v, φ) and $(\tilde{u}, \tilde{v}, \tilde{\varphi})$ are both representations of a convex self-control preference. Then there exist constants $\alpha_u, \alpha_v > 0$ and β_u, β_v such that $\tilde{u} = \alpha_u u + \beta_u$ and $\tilde{v} = \alpha_v v + \beta_v$. Moreover, when R and $\tilde{R}$ are the self-control subdomains for (u, v, φ) and $(\tilde{u}, \tilde{v}, \tilde{\varphi})$, respectively, $\tilde{R} = \alpha_v R$ and $\tilde{\varphi}(\alpha_v w) = \alpha_u \varphi(w)$ for all $w \in R$.*

The theorem states that u and v are unique up to positive affine transformation. When φ and $\tilde{\varphi}$ are differentiable, the stated condition on φ and $\tilde{\varphi}$ implies that for $\tilde{w} = \alpha_v w$ and $w \in R$,

$$\frac{\tilde{w}\tilde{\varphi}''(\tilde{w})}{\tilde{\varphi}'(\tilde{w})} = \frac{w\varphi''(w)}{\varphi'(w)},$$

where f' and f'' denote the first and the second derivatives of f, respectively. Thus the curvature of φ is uniquely determined within the self-control subdomain.

[5] Under Self-Control Concavity-1, we have $\{\mu\} \succ \{\mu, \eta \, \alpha \, \mu\} \succ \{\eta \, \alpha \, \mu\}$ for all $\alpha \in (0,1)$ if $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. See Lemma 7 in the Appendix.

3.4 *Proof outline for Theorem 3*

The main technical difficulty lies in showing that the self-control cost function in the general model is a function of temptation opportunity cost:

$$c\Big(\mu, \max_{\eta\in x} v(\eta)\Big) = \varphi\Big(\max_{\eta\in x} v(\eta) - v(\mu)\Big).$$

Convexity of the function φ then follows readily from Self-Control Concavity-2.

The functions u, v, and W are determined as in the general model. Take any μ, η satisfying $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. This ranking suggests that μ is chosen from $\{\mu, \eta\}$ with self-control, and the difference $u(\mu) - W(\{\mu, \eta\})$ expresses the cost of self-control at $\{\mu, \eta\}$. The corresponding temptation opportunity cost is $w = v(\eta) - v(\mu)$. Define the self-control cost function by

$$\varphi(v(\eta) - v(\mu)) \equiv u(\mu) - W(\{\mu, \eta\}). \tag{5}$$

The key step in the proof is to show that φ is indeed well defined. This is demonstrated by establishing that for all μ, μ', η, η' such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $\{\mu'\} \succ \{\mu', \eta'\} \succ \{\eta'\}$,

$$v(\eta) - v(\mu) \geq v(\eta') - v(\mu') \Longrightarrow u(\mu) - W(\{\mu, \eta\}) \geq u(\mu') - W(\{\mu', \eta'\}). \tag{6}$$

To see how (6) is obtained, begin by noting that by Continuity and the fact that the set of all lotteries with finite supports is dense in Δ under the weak convergence topology, we can focus on μ, μ', η, η' that have finite supports. Hence these lotteries can be viewed as vectors in the interior of the unit simplex of a finite dimensional space, $\mathbb{R}^n$. An important preliminary observation is that Weak Binary Independence implies the property (7) below. Recall that the axiom states that the ranking of two menus $\{a, a'\}$ and $\{b, b'\}$ when mixed with a common singleton $\{\nu\}$ is independent of the singleton. Intuitively, if a common "translation" is applied to the elements of both menus $\{a, a'\}$ and $\{b, b'\}$, then the ranking of the menus is uneffected. Formally, a translation is a vector $\theta \in \mathbb{R}^n$ such that $\sum_{i=1}^n \theta(i) = 0$. We say that a translation θ is admissible for a menu x if $\mu + \theta \in \Delta$ for all $\mu \in x$. Weak Binary Independence implies that if θ is admissible for $\{a, a'\}$, then

$$W(\{a + \theta, a' + \theta\}) = W(\{a, a'\}) + u(\theta). \tag{7}$$

This property is used in the main argument for showing (6), described next.

Assume the hypothesis of (6). For simplicity, suppose that the translation $\theta \equiv \mu' - \mu$ is admissible for $\{\mu, \eta\}$. By (7), $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ implies $\{\mu+\theta\} \succ \{\mu+\theta, \eta+\theta\} \succ \{\eta+\theta\}$ and so $\{\mu'\} \succ \{\mu', \eta + \theta\} \succ \{\eta + \theta\}$. Note that

$$v(\eta) - v(\mu) \geq v(\eta') - v(\mu') \Longrightarrow v(\eta + \theta) \geq v(\eta').$$

By Temptation Aversion, $v(\eta+\theta) \geq v(\eta')$ implies $W(\{\mu', \eta'\}) \geq W(\{\mu', \eta+\theta\})$ (recall that the axiom states that reducing temptation in a menu makes it better). Then, again by (7),

$$W(\{\mu', \eta'\}) \geq W(\{\mu', \eta + \theta\}) = W(\{\mu, \eta\}) + u(\theta) = W(\{\mu, \eta\}) + u(\mu') - u(\mu).$$

Rearrange this to see that (6) is proved.

The above argument relies on the assumption that $\theta = \mu' - \mu$ is admissible for $\{\mu, \eta\}$. For the interested reader, here is the outline of the proof for (6) when $\theta = \mu' - \mu$ is not admissible for $\{\mu, \eta\}$. Assuming the hypothesis for (6), Self-Control Concavity-2 together with the linearity of temptation utility ensures that for all $\alpha \in [0, 1]$,[6]

$$\{\mu \,\alpha\, \mu'\} \succ \{\mu \,\alpha\, \mu', \eta \,\alpha\, \eta'\} \succ \{\eta \,\alpha\, \eta'\}.$$

Notice that by assumption,

$$1 \geq \beta \geq \gamma \geq 0 \Longleftrightarrow v(\eta \,\beta\, \eta') - v(\mu \,\beta\, \mu') \geq v(\eta \,\gamma\, \eta') - v(\mu \,\gamma\, \mu').$$

Moreover, by Continuity, there exists a small neighborhood of $\{\mu \,\alpha\, \mu', \eta \,\alpha\, \eta'\}$ in which a small translation $\theta \equiv \mu \,\gamma\, \mu' - \mu \,\beta\, \mu'$ is admissible for $\{\mu \,\beta\, \mu', \eta \,\beta\, \eta'\}$. This allows us to proceed by applying our earlier argument in this small neighborhood of $\{\mu \,\alpha\, \mu', \eta \,\alpha\, \eta'\}$. Then, since $[0, 1]$ is compact, there exists a finite number of mixing coefficients $1 = \alpha^0 \geq \alpha^1 \geq \cdots \geq \alpha^I = 0$ such that

$$\begin{aligned} u(\mu) - W(\{\mu, \eta\}) &= u(\mu \,\alpha^0\, \mu') - W(\{\mu \,\alpha^0\, \mu', \eta \,\alpha^0\, \eta'\}) \geq \cdots \\ &\geq u(\mu \,\alpha^I\, \mu') - W(\{\mu \,\alpha^I\, \mu', \eta \,\alpha^I\, \eta'\}) = u(\mu') - W(\{\mu', \eta'\}). \end{aligned}$$

This "chain" linking $u(\mu) - W(\{\mu, \eta\})$ and $u(\mu') - W(\{\mu', \eta'\})$ then proves (6).

To conclude, having established (6), it follows that

$$v(\eta) - v(\mu) = v(\eta') - v(\mu') \Longrightarrow u(\mu) - W(\{\mu, \eta\}) = u(\mu') - W(\{\mu', \eta'\}).$$

It then follows that φ as defined in (5) is indeed well defined. By (6), φ is increasing. By Self-Control Concavity-1, if $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$, then $\{\mu\} \succ \{\mu, \eta \,\alpha\, \mu\} \succ \{\eta \,\alpha\, \mu\}$ and, hence, φ is defined on an interval $(0, \overline{w})$. Convexity of φ is a consequence of Self-Control Concavity-2.

4. Concluding remarks: Ex post choice

While the convex self-control model is a representation for an ex ante preference over menus, it suggests that ex post choice is given by the choice correspondence defined by

$$\mathcal{C}_\varphi(x) = \arg\max_{\mu \in x} \Big\{ u(\mu) - \varphi\Big(\max_{\mu' \in x} v(\mu') - v(\mu)\Big)\Big\}.$$

We conclude this paper with some observations about this choice correspondence in the context of choice under risk. These reproduce the observations in Takeoka (2006).

An immediate observation is that $\mathcal{C}_\varphi$ is menu-dependent via its dependence on the most tempting alternative in the menu. If φ is convex, for instance, this would imply that while an agent can pick a "good" alternative over a moderately tempting alternative, adding an even more tempting alternative to the menu may induce the agent to

[6] The first ranking comes from linearity of v. The second ranking is implied by Self-Control Concavity-2 and Commitment Independence because $\{\mu \,\alpha\, \mu', \eta \,\alpha\, \eta'\} \succsim \{e_{\mu\eta} \,\alpha\, e_{\mu'\eta'}\} \succ \{\eta \,\alpha\, \eta'\}$.

choose the moderately tempting alternative, thereby violating the Weak Axiom of Revealed Preference. The intuition for such choice is that the loss of self-control ability due to the presence of a great temptation may make the agent unable to choose the good alternative, but he may nevertheless have enough self-control to resist the great temptation. He chooses the moderately tempting alternative as a compromise. An analysis of the notion of menu-dependent self-control can be found in the companion paper (Noor and Takeoka 2010).

Next consider implications for choice under risk and over time. As noted in our discussion of Self-Control Concavity-1, mixing two alternatives with a third can only enhance the propensity for self-control. An implication is the possibility that for some lotteries r, s, ℓ,

$$\{r\} \succ \{r, s\} \sim \{s\} \quad \text{and} \quad \{r\,\alpha\,\ell\} \succ \{r\,\alpha\,\ell, s\,\alpha\,\ell\} \succ \{s\,\alpha\,\ell\}$$

for all sufficiently small $\alpha \in (0, 1)$. That is, s may overwhelmingly tempt r, but the agent may be able to exhibit self-control if both are mixed with a third common lottery ℓ. Ex post choice would thus exhibit the reversal

$$\mathcal{C}_\varphi(\{r, s\}) = \{s\} \quad \text{and} \quad \mathcal{C}_\varphi(\{r\,\alpha\,\ell, s\,\alpha\,\ell\}) = \{r\,\alpha\,\ell\}.$$

The idea that mixing can induce reversals is reminiscent of the following two experimental findings.

The first is the *common ratio effect* (Allais 1953, Kahneman and Tversky 1979). Subjects in experiments are observed to prefer, for instance, \$3,000 for sure over a 0.8 chance of \$4,000, but also prefer a 0.2 chance of \$4,000 to a 0.25 chance of \$3,000. If we take s to denote a safe alternative and r to denote a risky prospect, and if we let ℓ denote the zero outcome, then the convex model generates such choices (here $\alpha = 0.25$). Observe that this involves a temptation to be risk averse, a hypothesis that is reminiscent of Loewenstein et al. (2001), who survey evidence suggesting that the emotional response to risk is that of aversion (anxiety, dread, etc.).

A second experimental finding is that of Keren and Roelofsma (1995). Experiments on time preference show that subjects who prefer a small immediate reward over a larger later reward (scenario 1) tend to reverse preferences when both rewards are delayed by a common number of periods (scenario 2). Such reversals have been explained in terms of temptation by immediate gratification (see GP (2004) for instance), which sways choice in the first scenario but is irrelevant in the second. Keren and Roelofsma (1995) find that if the rewards are made probabilistic, so that each is received with a common probability, then reversals tend to disappear: subjects prefer the later larger-outcome lottery in both the first and the second scenario. This is generated by the convex model since the α-mixing with the zero reward increases self-control, and thus the temptation by immediate gratification that made the immediate sure reward overwhelming is now resistible. Thus, delaying a pair of rewards has a similar effect on choice as proportionately reducing the probability of receiving either reward.[7]

[7] The convex model lends itself to an infinite horizon extension in the spirit of Gul and Pesendorfer (2004) and Noor (2007). In this setting, further implications of convexity for the interaction of risk and time preferences and also for the timing of resolution of risk can be studied.

While convex self-control costs have been shown to generate various behaviors observed in experiments, further experiments are necessary before it can be claimed that convexity *explains* them. The reader is referred to the companion paper (Noor and Takeoka 2010) for further remarks and a more general discussion of menu-dependent self-control.

Appendices

A. Proof of Theorem 1

The proof of necessity of the axioms is routine. For Temptation Aversion, observe that since $v(\nu) \leq v(\eta)$ and $c(\mu, \cdot)$ is weakly increasing, $W(\{\mu, \eta\}) = u(\mu) - c(\mu, v(\eta)) \leq u(\mu) - c(\mu, v(\nu)) \leq W(\{\mu, \nu\})$.

Sufficiency of the axioms is established in a sequence of lemmas.

Lemma 1. *(i) There exists a continuous linear function $u : \Delta \to \mathbb{R}_+$ such that*

$$\{\mu\} \succsim \{\eta\} \Longleftrightarrow u(\mu) \geq u(\eta).$$

(ii) There exists a continuous function $W : Z \to \mathbb{R}_+$ that represents $\succsim$ and satisfies $W(\{\mu\}) = u(\mu)$ for all $\mu \in \Delta$.

(iii) There exists a continuous linear function $v : \Delta \to \mathbb{R}_+$ such that if $\{\mu\} \succ \{\eta\}$, then

$$\{\mu\} \succ \{\mu, \eta\} \Longleftrightarrow v(\eta) > v(\mu).$$

Proof. (i) The first assertion follows from Order, Continuity, Commitment Independence, and the mixture space theorem.

(ii) Since u is continuous on Δ, there exist a maximal and a minimal lottery $\mu^\Delta, \mu_\Delta \in \Delta$ with respect to u. Without loss of generality, we can assume $u(\mu^\Delta) = 1$ and $u(\mu_\Delta) = 0$. From Continuity and Set Betweenness, $\{\mu^\Delta\} \succsim x \succsim \{\mu_\Delta\}$ for all $x \in Z$. By a standard argument, for all $x \in Z$, there exists a unique number $\alpha(x) \in [0, 1]$ such that $x \sim \{\mu^\Delta \, \alpha(x) \, \mu_\Delta\}$. Define

$$W(x) \equiv u(\mu^\Delta \, \alpha(x) \, \mu_\Delta) \in [0, 1].$$

Then W represents $\succsim$. Moreover, $W(\{\mu\}) = u(\mu)$ for all $\mu \in \Delta$.

To show continuity of W, let $x^n \to x$. Since $u(\mu^\Delta) = 1$ and $u(\mu_\Delta) = 0$, $W(x) = \alpha(x)$. So we want to show $\alpha(x^n) \to \alpha(x)$. By contradiction, suppose otherwise. Then there exists a neighborhood $B(\alpha(x))$ of $\alpha(x)$ such that $\alpha(x^m) \notin B(\alpha(x))$ for infinitely many m. Let $\{x^m\}$ denote the corresponding subsequence of $\{x^n\}$. Since $x^n \to x$, $\{x^m\}$ also converges to x. Since $\{\alpha(x^m)\}$ is a sequence in $[0, 1]$, there exists a convergent subsequence $\{\alpha(x^\ell)\}$ with a limit $\alpha \neq \alpha(x)$. On the other hand, since $x^\ell \to x$ and $x^\ell \sim \{\mu^\Delta \, \alpha(x^\ell) \, \mu_\Delta\}$, Continuity implies $x \sim \{\mu^\Delta \, \alpha \, \mu_\Delta\}$. Since $\alpha(x)$ is unique, $\alpha(x) = \alpha$, which is a contradiction.

(iii) See Noor and Takeoka (2010, Theorem 2). □

Without loss of generality, assume that $v(\Delta) = [0, 1]$. By construction, if $\{\mu\} \succ \{\mu, \eta\}$, then $v(\eta) > v(\mu)$. If $\{\mu\} \sim \{\mu, \eta\} \succ \{\eta\}$, then $v(\mu) \geq v(\eta)$.

LEMMA 2. *For all* $\mu, \eta, \nu \in \Delta$, *if* $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ *and* $v(\nu) \le v(\eta)$, *then* $\{\mu, \nu\} \succsim \{\mu, \eta\}$.

PROOF. The first case is where $\{\nu\} \succsim \{\eta\}$. Since $\{\mu\} \succ \{\mu, \eta\}$, we know $u(\mu) > u(\eta)$ and $v(\mu) < v(\eta)$. For all $\alpha \in (0, 1)$, $v(\eta) > v(\nu\ \alpha\ \mu)$ and $u(\nu\ \alpha\ \mu) > u(\eta)$. Thus $\{\nu\ \alpha\ \mu\} \succ \{\nu\ \alpha\ \mu, \eta\}$. By Temptation Aversion, $\{\mu, \nu\ \alpha\ \mu\} \succsim \{\mu, \eta\}$. By Continuity, we have $\{\mu, \nu\} \succsim \{\mu, \eta\}$ as $\alpha \to 1$.

Next suppose $\{\eta\} \succ \{\nu\}$. If $\{\eta\} \succ \{\eta, \nu\}$, we have $v(\nu) > v(\eta)$, which contradicts the assumption. Hence Set Betweenness implies $\{\eta\} \sim \{\eta, \nu\} \succ \{\nu\}$. By Temptation Aversion, $\{\mu, \nu\} \succsim \{\mu, \eta\}$. □

Define the correspondence $L: v(\Delta) \rightsquigarrow \Delta$ by

$$L(l) := \{\eta \mid v(\eta) \le l\}.$$

By continuity and linearity of v, it is clear that $L(l)$ is a nonempty compact convex set for each l. Define the self-control cost function by

$$c(\mu, l) = \max\Big[0, \max_{\nu \in L(l)} \big\{u(\mu) - W(\{\mu, \nu\})\big\}\Big].$$

The following lemma clarifies various properties of c. Properties (iii)–(vi) correspond to the properties in the statement of the theorem.

LEMMA 3. *(i) For any* μ, l, *if* $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ *for some* η *with* $v(\eta) = l$, *then* $c(\mu, l) = u(\mu) - W(\{\mu, \eta\}) > 0$.

(ii) For any μ, l, *if* $\{\mu\} \succ \{\mu, \eta\}$ *for some* $\eta \in L(l)$, *then* $c(\mu, l) > 0$.

(iii) For any μ, l, *if* $l \le v(\mu)$, *then* $c(\mu, l) = 0$.

(iv) If $u(\mu) > u(\eta)$ *and* $l = \max_{\mu,\eta} v$, *then*

$$v(\mu) < v(\eta) \Longleftrightarrow c(\mu, l) > 0.$$

(v) For any μ, $c(\mu, \cdot)$ *is weakly increasing.*

(vi) The function c *is continuous.*

PROOF. (i) For any $\nu \in L(l)$, $v(\nu) \le v(\eta)$, and thus by Lemma 2, $u(\mu) - W(\{\mu, \nu\}) \le u(\mu) - W(\{\mu, \eta\})$. Since $\eta \in L(l)$, it follows that $\max_{\nu \in L(l)}\{u(\mu) - W(\{\mu, \nu\})\} = u(\mu) - W(\{\mu, \eta\}) > 0$ and thus $c(\mu, l) = u(\mu) - W(\{\mu, \eta\})$.

(ii) Obvious from the definition of c.

(iii) Under the hypothesis, $\{\mu\} \not\succ \{\mu, \eta\}$ for all $\eta \in L(l)$. Consequently,

$$\max_{\nu \in L(l)} \{u(\mu) - W(\{\mu, \nu\})\} \le 0 \quad \text{and so} \quad c(\mu, l) = 0.$$

(iv) Sufficiency obtains from part (ii). For the converse, note that if $v(\mu) \ge v(\eta)$, then $l = v(\mu)$, and thus part (iii) implies $c(\mu, l) = 0$.

(v) For any $l, l' \in v(\Delta)$,

$$\begin{aligned} l' < l &\Longrightarrow L(l') \subset L(l) \\ &\Longrightarrow \max_{\nu \in L(l')} \{u(\mu) - W(\{\mu, \nu\})\} \le \max_{\nu \in L(l)} \{u(\mu) - W(\{\mu, \nu\})\} \\ &\Longrightarrow c(\mu, l') \le c(\mu, l). \end{aligned}$$

(vi) We show below that $L : v(\Delta) \rightsquigarrow \Delta$ is a continuous correspondence. The assertion then follows from the following argument: Since u and W are continuous, the maximum theorem implies that $(\mu, l) \mapsto \max_{\nu \in L(l)}\{u(\mu) - W(\{\mu, \nu\})\}$ is continuous. Moreover, since the upper envelope of two continuous functions is continuous, the function c is continuous.

To show that L is upper hemicontinuous, take any sequence $\{l_n\} \subset v(\Delta)$ that converges to some $l \in v(\Delta)$, and suppose that $\eta_n \in L(l_n)$ for each n. We must show that there is a subsequence $\{l_{n(m)}\}$ such that $\eta_{n(m)} \to \eta$ for some $\eta \in L(l)$. Since $\{\eta_n\}$ is a sequence in a compact set Δ, it has a convergent subsequence $\eta_{n(m)} \to \eta$ for some η. Since $v(\eta_{n(m)}) \le l_{n(m)}$ for each m and since v is continuous, it follows that $v(\eta) \le l$ and thus $\eta \in L(l)$, as desired.

To show that L is lower hemicontinuous, take any sequence $\{l_n\} \subset v(\Delta)$ that converges to some $l \in v(\Delta)$ and suppose that $\eta \in L(l)$. We must show that there exists a subsequence $\{l_{n(m)}\}$ such that $\eta_{n(m)} \to \eta$, where $\eta_{n(m)} \in L(l_{n(m)})$ for each m. Consider two possibilities:

– There exists N such that $l_n \ge v(\eta)$ for all $n \ge N$. Then $\eta \in L(l_n)$ for each $n \ge N$. In particular, lower hemicontinuity is established by taking the subsequence $\{l_N, l_{N+1}, \ldots\}$ and the corresponding trivial sequence $\{\eta\}$ that converges to η.

– For all N there exists $n_N \ge N$ such that $l_{n_N} < v(\eta)$. Take the subsequence $\{l_{n(m)}\}$ satisfying $l_{n(m)} < v(\eta)$ for all m. Construct $\{\eta_{n(m)}\}$ as follows: Let η_* be the minimizer of v over Δ (normalized so that $v(\eta_*) = 0$) and let $\alpha_{n(m)}$ satisfy $v(\eta\, \alpha_{n(m)}\, \eta_*) = l_{n(m)}(v(\eta)/l) \le l_{n(m)}$.[8] Then $\eta_{n(m)} \in L(l_{n(m)})$, where $\eta_{n(m)} := \eta\, \alpha_{n(m)}\, \eta_*$ for each m. To see that $\eta_{n(m)} \to \eta$, observe that

$$\begin{aligned} v(\eta\, \alpha_{n(m)}\, \eta_*) &= l_{n(m)} \frac{v(\eta)}{l} \\ &\Longrightarrow \alpha_{n(m)} v(\eta) = l_{n(m)} \frac{v(\eta)}{l} \text{ (since } v \text{ is linear and } v(\eta_*) = 0) \\ &\Longrightarrow \alpha_{n(m)} = \frac{l_{n(m)}}{l} \text{ (note that } v(\eta) > 0 \text{ since } v(\eta) > l_{n(m)} \ge 0, \text{ and also} \\ &\qquad l_{n(m)}/l < 1 \text{ since } l_{n_m} < v(\eta) \le l). \end{aligned}$$

Since $l_{n(m)} \to l$, it follows that $\alpha_{n(m)} \to 1$ and, in turn, $\eta_{n(m)} \to \eta$, as desired. $\square$

[8]Note that $l > 0$; otherwise $0 \le l_{n(m)} < v(\eta) \le l = 0$ is a contradiction. Recall also that $\eta \in L(l)$ implies $v(\eta) \le l$, and thus $l_{n(m)}(v(\eta)/l) \le l_{n(m)}$.

Lemma 4. *For all* $\mu, \eta \in \Delta$,

$$W(\{\mu, \eta\}) = \max_{\nu \in \{\mu,\eta\}} \Big\{ u(\nu) - c\Big(\nu, \max_{\{\mu,\eta\}} v\Big)\Big\}.$$

Proof. Consider the various cases. In each case, let $l = \max_{\{\mu,\eta\}} v$.

(i) $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. Since $v(\mu) < v(\eta) = l$, Lemma 3(i) implies $W(\{\mu, \eta\}) = u(\mu) - c(\mu, l) = u(\mu) - c(\mu, \max_{\{\mu,\eta\}} v)$. Since $c(\eta, l) = 0$, we have $u(\eta) - c(\eta, \max_{\{\mu,\eta\}} v) = u(\eta)$, and since $\{\mu, \eta\} \succ \{\eta\}$, it follows that

$$u(\mu) - c\Big(\mu, \max_{\{\mu,\eta\}} v\Big) > u(\eta) - c\Big(\eta, \max_{\{\mu,\eta\}} v\Big).$$

Indeed, $W(\{\mu, \eta\}) = \max_{\nu \in \{\mu,\eta\}} u(\nu) - c(\nu, \max_{\{\mu,\eta\}} v)$, as desired.

(ii) $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$. By definition of $c(\mu, l)$,

$$c(\mu, l) \geq \max_{\nu \in L(l)} \big\{u(\mu) - W(\{\mu, \nu\})\big\} \geq u(\mu) - W(\{\mu, \eta\}).$$

In particular, $W(\{\mu, \eta\}) \geq u(\mu) - c(\mu, l) = u(\mu) - c(\mu, \max_{\{\mu,\eta\}} v)$. Then

$$u(\eta) - c\Big(\eta, \max_{\{\mu,\eta\}} v\Big) = u(\eta) = W(\{\mu, \eta\}) \geq u(\mu) - c\Big(\mu, \max_{\{\mu,\eta\}} v\Big)$$

and, hence, $W(\{\mu, \eta\}) = \max_{\nu \in \{\mu,\eta\}} u(\nu) - c(\nu, \max_{\{\mu,\eta\}} v)$.

(iii) $\{\mu\} \sim \{\mu, \eta\} \succ \{\eta\}$ or $\{\eta\} \sim \{\eta, \mu\} \succsim \{\mu\}$. Suppose $\{\mu\} \sim \{\mu, \eta\} \succ \{\eta\}$. Then $l = v(\mu) \geq v(\eta)$ and, in particular, $c(\mu, l) = 0$. Since $c(\eta, l) \geq 0$,

$$\begin{aligned} W(\{\mu, \eta\}) &= u(\mu) \\ &= u(\mu) - c\Big(\mu, \max_{\{\mu,\eta\}} v\Big) \\ &= u(\eta) \text{ since } \{\mu\} \succ \{\eta\} \text{ and } c(\mu, l) = 0 \\ &\geq u(\eta) - c(\eta, l) \text{ since } c(\eta, l) \geq 0. \end{aligned}$$

This establishes the result.

For the case where $\{\eta\} \sim \{\eta, \mu\} \succsim \{\mu\}$, we have $l = v(\eta) \geq v(\mu)$ (this is without loss of generality when $\{\mu\} \sim \{\mu, \eta\} \sim \{\eta\}$), $c(\eta, l) = 0$, and $c(\mu, l) \geq 0$. Arguing as above yields the result.

(iv) $\{\eta\} \succ \{\eta, \mu\} \succsim \{\mu\}$. The argument is analogous to that in cases (i) and (ii). □

Lemma 5. *For all finite menus* $x \in Z$,

$$W(x) = \max_{\nu \in x} \Big\{ u(\nu) - c\Big(\nu, \max_{x} v\Big)\Big\}.$$

Proof. The argument is similar to that used in the conclusion of the proof in Gul and Pesendorfer (2001, Theorem 1). Gul and Pesendorfer (2001, Lemma 2) show that if $\succsim$ satisfies Set Betweenness, for all finite menus $x \in Z$,

$$W(x) = \max_{\mu \in x} \min_{\eta \in x} W(\{\mu, \eta\}) = \min_{\eta \in x} \max_{\mu \in x} W(\{\mu, \eta\}). \tag{8}$$

Fix $\mu \in x$ arbitrarily. Since $c(\nu, \cdot)$ is weakly increasing for all ν,

$$\min_{\eta \in x} W(\{\mu, \eta\}) = \min_{\eta \in x} \max_{\nu \in \{\mu, \eta\}} u(\nu) - c\Big(\nu, \max_{\{\mu,\eta\}} v\Big) \geq \min_{\eta \in x} \max_{\nu \in \{\mu, \eta\}} u(\nu) - c\Big(\nu, \max_{x} v\Big)$$
$$= \max_{\nu \in \{\mu, \eta^\mu\}} u(\nu) - c\Big(\nu, \max_{x} v\Big),$$

where η^μ is a minimizer of the associated minimization problem. Since the above inequality holds for all $\mu \in x$, if follows from (8) that

$$W(x) \geq \max_{\mu \in x} \max_{\nu \in \{\mu, \eta^\mu\}} u(\nu) - c\Big(\nu, \max_{x} v\Big) = \max_{\nu \in x}\Big\{u(\nu) - c\Big(\nu, \max_{x} v\Big)\Big\}. \tag{9}$$

On the other hand, fix $\eta \in x$ arbitrarily. Since $c(\nu, \cdot)$ is weakly increasing,

$$\max_{\mu \in x} W(\{\mu, \eta\}) = \max_{\mu \in x} \max_{\nu \in \{\mu, \eta\}} u(\nu) - c\Big(\nu, \max_{\{\mu,\eta\}} v\Big) \leq \max_{\mu \in x} \max_{\nu \in \{\mu, \eta\}} u(\nu) - c\Big(\nu, \min_{\mu \in x} \max_{\{\mu,\eta\}} v\Big)$$
$$= \max_{\nu \in x} u(\nu) - c\Big(\nu, \min_{\mu \in x} \max_{\{\mu,\eta\}} v\Big) = \max_{\nu \in x} u(\nu) - c\Big(\nu, \max_{\{\mu^\eta,\eta\}} v\Big),$$

where μ^η is a minimizer of the associated minimization problem. Since $c(\nu, \cdot)$ is weakly increasing and the above inequality holds for all $\eta \in x$, if follows from (8) that

$$\begin{aligned} W(x) &\leq \min_{\eta \in x} \max_{\nu \in x}\Big\{u(\nu) - c\Big(\nu, \max_{\{\mu^\eta,\eta\}} v\Big)\Big\} = \max_{\nu \in x}\Big\{u(\nu) - \max_{\eta \in x} c\Big(\nu, \max_{\{\mu^\eta,\eta\}} v\Big)\Big\} \\ &= \max_{\nu \in x}\Big\{u(\nu) - c\Big(\nu, \max_{x} v\Big)\Big\}. \end{aligned} \tag{10}$$

Taking (9) and (10) together, the desired result holds. □

Lemma 6. *For all $x \in Z$, W can be written as the desired form.*

Proof. By Lemma 0 of Gul and Pesendorfer (2001, p. 1421), there exists a sequence of subsets x^n of x such that each x^n is finite and $x^n \to x$ in the Hausdorff metric. By Lemma 5,

$$W(x^n) = \max_{\nu \in x^n}\Big\{u(\nu) - c\Big(\nu, \max_{x^n} v\Big)\Big\}. \tag{11}$$

Since c is continuous by Lemma 3(vi), the maximum theorem implies that the right-hand side of (11) converges to

$$\max_{\nu \in x}\Big\{u(\nu) - c\Big(\nu, \max_{x} v\Big)\Big\}.$$

On the other hand, by Lemma 1(ii), $W(x^n) \to W(x)$. This completes the proof. □

B. Proof of Theorem 3

Proof of necessity of axioms is omitted. The proof of sufficiency follows.

Let (u, v, W) be the objects guaranteed by Lemma 1. Since u and v are mixture linear, assume that $u(\Delta) = v(\Delta) = [0, 1]$.

Let

$$A \equiv \big\{w \in [0,1] \mid w = v(\eta) - v(\mu) \text{ for some } \mu, \eta \text{ such that } \{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}\big\}.$$

Since $\succsim$ is a nondegenerate preference, A is nonempty.

We show the following preliminary lemma.

Lemma 7. *For all μ, η, ν and $\alpha \in (0,1)$,*

$$\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\} \Longrightarrow \{\mu\,\alpha\,\nu\} \succ \{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\} \succ \{\eta\,\alpha\,\nu\}.$$

Proof. If $\{\mu\} \succ \{\mu, \eta\}$, we have $u(\mu) > u(\eta)$ and $v(\eta) > v(\mu)$. Since u and v are mixture linear, $u(\mu\,\alpha\,\nu) > u(\eta\,\alpha\,\nu)$ and $v(\eta\,\alpha\,\nu) > v(\mu\,\alpha\,\nu)$. Thus, $\{\mu\,\alpha\,\nu\} \succ \{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\}$.

As shown in Lemma 1(ii), there exists $e_{\mu\eta} \in \Delta$ such that $\{\mu, \eta\} \sim \{e_{\mu\eta}\}$. By Self-Control Concavity-1, $\{\mu, \eta\}\,\alpha\,\{\nu\} \succsim \{e_{\mu\eta}\,\alpha\,\nu\}$. Since $\{\mu, \eta\} \sim \{e_{\mu\eta}\} \succ \{\eta\}$, by Commitment Independence, $\{e_{\mu\eta}\,\alpha\,\nu\} \succ \{\eta\,\alpha\,\nu\}$. Thus, we have $\{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\} \succ \{\eta\,\alpha\,\nu\}$ as desired. □

Lemma 8. *(i) The set A is an interval with* $\inf A = 0$, *and (ii) if* $\sup A \in A$, *then* $\sup A = 1$.

Proof. (i) It suffices to show that for all $w \in A$, $\alpha w \in A$ for all $\alpha \in (0,1)$. Let $w \in A$. There exist μ, η such that $w = v(\eta) - v(\mu)$ and $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. By Lemma 7, $\{\mu\} \succ \{\mu, \eta\,\alpha\,\mu\} \succ \{\eta\,\alpha\,\mu\}$. Thus $\alpha w = \alpha(v(\eta) - v(\mu)) = v(\eta\,\alpha\,\mu) - v(\mu) \in A$.

(ii) Since $\sup A \in A$, there exist μ, η such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $v(\eta) - v(\mu) = \sup A$. By contradiction, suppose $\sup A < 1$. Then either $\max_\Delta v > v(\eta)$ or $\min_\Delta v < v(\mu)$. In case of the former, Continuity implies that there exists ν sufficiently close to η such that $\{\mu\} \succ \{\mu, \nu\} \succ \{\nu\}$ and $v(\nu) > v(\eta)$. Thus $\sup A < v(\nu) - v(\mu) \in A$, which is a contradiction. The symmetric argument can be applied to the latter case. □

Define $\varphi : A \to (0,1]$ by

$$\varphi(w) \equiv u(\mu) - W(\{\mu, \eta\}),$$

where μ, η satisfy $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $w = v(\eta) - v(\mu)$.

The lemmas below establish that φ is well defined.

Lemma 9. *For all $\mu, \eta, \mu', \eta' \in \Delta$ and $\alpha \in (0,1)$, if $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $\{\mu'\} \succ \{\mu', \eta'\} \succ \{\eta'\}$, then $\{\mu\,\alpha\,\mu'\} \succ \{\mu\,\alpha\,\mu', \eta\,\alpha\,\eta'\} \succ \{\eta\,\alpha\,\eta'\}$ for all $\alpha \in [0,1]$.*

Proof. Since $\{\mu\} \succ \{\mu, \eta\}$ and $\{\mu'\} \succ \{\mu', \eta'\}$, we have $u(\mu) > u(\eta)$, $v(\eta) > v(\mu)$, $u(\mu') > u(\eta')$, and $v(\eta') > v(\mu')$. Since u and v are mixture linear, $u(\mu\,\alpha\,\mu') > u(\eta\,\alpha\,\eta')$ and $v(\eta\,\alpha\,\eta') > v(\mu\,\alpha\,\mu')$, and, hence, $\{\mu\,\alpha\,\mu'\} \succ \{\mu\,\alpha\,\mu', \eta\,\alpha\,\eta'\}$. As shown in Lemma 1(ii), there exist $\nu, \nu' \in \Delta$ such that $\{\mu, \eta\} \sim \{\nu\}$ and $\{\mu', \eta'\} \sim \{\nu'\}$. By Self-Control Concavity-2, $\{\mu\,\alpha\,\mu', \eta\,\alpha\,\eta'\} \succsim \{\nu\,\alpha\,\nu'\}$. Since $\{\nu\} \succ \{\eta\}$ and $\{\nu'\} \succ \{\eta'\}$, Commitment Independence implies that $\{\nu\,\alpha\,\nu'\} \succ \{\eta\,\alpha\,\eta'\}$ for all $\alpha \in [0,1]$. Therefore, we have $\{\mu\,\alpha\,\mu'\} \succ \{\mu\,\alpha\,\mu', \eta\,\alpha\,\eta'\} \succ \{\eta\,\alpha\,\eta'\}$. □

Take any finite subset $\mathbf{c} = \{c_1, \ldots, c_N\} \subset C$. Define

$$\Delta_{(N,\mathbf{c})} \equiv \left\{ \nu \in \mathbb{R}^N_+ \;\middle|\; \sum_{i=1}^{N} \nu(c_i) = 1 \right\} \subset \Delta, \qquad \Theta_{(N,\mathbf{c})} \equiv \left\{ \theta \in \mathbb{R}^N \;\middle|\; \sum_{i=1}^{N} \theta(c_i) = 0 \right\}.$$

For all $\mu \in \Delta_{(N,\mathbf{c})}$ and $\theta \in \Theta_{(N,\mathbf{c})}$, if $\mu + \theta \in \Delta_{(N,\mathbf{c})}$, we can view $\mu + \theta$ as the lottery obtained by shifting μ toward θ. For all $\mu \in \Delta_{(N,\mathbf{c})}$, say that $\theta \in \Theta_{(N,\mathbf{c})}$ is admissible for μ if $\mu + \theta \in \Delta_{(N,\mathbf{c})}$.

Lemma 10. *Given any two menus $x, y \subset \Delta_{(N,\mathbf{c})}$, the following statements are equivalent:*

(i) For all $\alpha \in [0,1]$ and $\mu, \eta \in \Delta_{(N,\mathbf{c})}$, $x \,\alpha\, \{\mu\} \succsim y \,\alpha\, \{\mu\} \Longrightarrow x \,\alpha\, \{\eta\} \succsim y \,\alpha\, \{\eta\}$.

(ii) For all $\theta \in \Theta_{(N,\mathbf{c})}$ that are admissible for x, y, $x \succsim y \Longleftrightarrow x + \theta \succsim y + \theta$.

Proof. Inspecting the proof of Ergin and Sarver (2009, Lemma 4) reveals that the proof works for any two fixed menus x, y. □

The preceding lemma yields that Weak Binary Independence is equivalent to the condition that for all $\mu, \mu', \eta, \eta' \in \Delta_{(N,\mathbf{c})}$ and admissible translations $\theta \in \Theta_{(N,\mathbf{c})}$ for these lotteries,

$$\{\mu, \mu'\} \succsim \{\eta, \eta'\} \Longrightarrow \{\mu + \theta, \mu' + \theta\} \succsim \{\eta + \theta, \eta' + \theta\}, \tag{12}$$

which is referred to as *translation invariance.*

For all $\theta \in \Theta_{(N,\mathbf{c})}$, let $u(\theta)$ denote $\sum_i u(c_i)\theta(c_i)$.

Lemma 11. *For all $\mu, \mu' \in \Delta_{(N,\mathbf{c})}$ and $\theta \in \Theta_{(N,\mathbf{c})}$, if $\mu + \theta, \mu' + \theta \in \Delta_{(N,\mathbf{c})}$, then $W(\{\mu + \theta, \mu' + \theta\}) = W(\{\mu, \mu'\}) + u(\theta)$.*

Proof. By Set Betweenness, assume that $\{\mu\} \succsim \{\mu, \mu'\} \succsim \{\mu'\}$. Since u is continuous, there exists $\alpha \in [0,1]$ such that $W(\{\mu, \mu'\}) = u(\mu \,\alpha\, \mu')$. If $\mu + \theta, \mu' + \theta \in \Delta_{(N,\mathbf{c})}$, then $\mu \,\alpha\, \mu' + \theta = (\mu + \theta) \,\alpha\, (\mu' + \theta) \in \Delta_{(N,\mathbf{c})}$. Hence translation invariance implies that

$$W(\{\mu + \theta, \mu' + \theta\}) = u(\mu \,\alpha\, \mu' + \theta) = u(\mu \,\alpha\, \mu') + u(\theta) = W(\{\mu, \mu'\}) + u(\theta). \qquad \square$$

Lemma 12. *Take all $\mu, \mu', \eta, \eta' \in \Delta$ with finite supports. Assume that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $\{\mu'\} \succ \{\mu', \eta'\} \succ \{\eta'\}$. Then*

$$v(\eta) - v(\mu) \geq v(\eta') - v(\mu') \Rightarrow u(\mu) - W(\{\mu, \eta\}) \geq u(\mu') - W(\{\mu', \eta'\}).$$

Proof. Let $\mathbf{c} \equiv \{c_1, \ldots, c_N\} \subset C$ be the union of the supports of μ, μ', η, η'. Hence, these lotteries belong to $\Delta_{(N,\mathbf{c})}$.

Step 1. We claim that if $\theta \equiv \mu' - \mu \in \Theta_{(N,\mathbf{c})}$ is admissible for η, then $u(\mu) - W(\{\mu, \eta\}) \geq u(\mu') - W(\{\mu', \eta'\})$. Since v is mixture linear,

$$v(\eta + \theta) - v(\mu') = v(\eta + \theta) - v(\mu + \theta) = v(\eta) - v(\mu) \geq v(\eta') - v(\mu').$$

Thus $v(\eta+\theta) \geq v(\eta')$. Furthermore, by translation invariance as given in (12), $\{\mu+\theta\} \succ \{\mu+\theta, \eta+\theta\} \succ \{\eta+\theta\}$, that is, $\{\mu'\} \succ \{\mu', \eta+\theta\} \succ \{\eta+\theta\}$. By Lemma 2, $\{\mu', \eta'\} \succsim \{\mu', \eta+\theta\}$. Thus, from Lemma 11,

$$\begin{aligned} u(\mu') - W(\{\mu', \eta'\}) &\leq u(\mu') - W(\{\mu', \eta+\theta\}) = u(\mu+\theta) - W(\{\mu+\theta, \eta+\theta\}) \\ &= u(\mu) + u(\theta) - W(\{\mu, \eta\}) - u(\theta) = u(\mu) - W(\{\mu, \eta\}), \end{aligned}$$

as desired.

We now turn to the general case where $\theta \equiv \mu' - \mu \in \Theta_{(N,\mathbf{c})}$ is not admissible for η. Take a lottery ν in the interior of $\Delta_{(N,\mathbf{c})}$. For all $\alpha \in (0,1)$ sufficiently close to 1, let $a \equiv \mu\,\alpha\,\nu$, $b \equiv \eta\,\alpha\,\nu$, $a' \equiv \mu'\,\alpha\,\nu$, and $b' \equiv \eta'\,\alpha\,\nu \in \Delta_{(N,\mathbf{c})}$. Continuity implies $\{a\} \succ \{a,b\} \succ \{b\}$ and $\{a'\} \succ \{a',b'\} \succ \{b'\}$. Furthermore, $v(b) - v(a) = v(\eta\,\alpha\,\nu) - v(\mu\,\alpha\,\nu) = \alpha(v(\eta) - v(\mu)) \geq \alpha(v(\eta') - v(\mu')) = v(b') - v(a')$. From Lemma 9, for all $\beta \in [0,1]$, $\{a\,\beta\,a'\} \succ \{a\,\beta\,a', b\,\beta\,b'\} \succ \{b\,\beta\,b'\}$. Notice also that $a\,\beta\,a', b\,\beta\,b' \in \Delta_{(N,\mathbf{c})}$ for all $\beta \in [0,1]$.

Step 2. We claim that for all $\beta \in [0,1]$, there exists a relative open interval $O(\beta)$ containing β such that for all $\tilde{\beta} \in O(\beta)$,

$$\tilde{\beta} \geq \beta \Longleftrightarrow u(a\,\tilde{\beta}\,a') - W(\{a\,\tilde{\beta}\,a', b\,\tilde{\beta}\,b'\}) \geq u(a\,\beta\,a') - W(\{a\,\beta\,a', b\,\beta\,b'\}). \tag{13}$$

Since $v(b) - v(a) \geq v(b') - v(a')$, we have, for all $\tilde{\beta} \in (0,1)$ with $\tilde{\beta} \geq \beta$,

$$\begin{aligned} v(b\,\beta\,b') - v(a\,\beta\,a') &= \beta(v(b) - v(a)) + (1-\beta)(v(b') - v(a')) \\ &\leq \tilde{\beta}(v(b) - v(a)) + (1-\tilde{\beta})(v(b') - v(a')) = v(b\,\tilde{\beta}\,b') - v(a\,\tilde{\beta}\,a'). \end{aligned}$$

Let $\theta \equiv a\,\beta\,a' - a\,\tilde{\beta}\,a' \in \Theta_{(N,\mathbf{c})}$. Notice that

$$b\,\tilde{\beta}\,b' + \theta = (\eta\,\tilde{\beta}\,\eta')\,\alpha\,\nu + (\beta - \tilde{\beta})(a - a').$$

Since $(\eta\,\beta\,\eta')\,\alpha\,\nu$ is in the interior of $\Delta_{(N,\mathbf{c})}$, there exists a relative open interval $O(\beta)$ containing β such that $(\eta\,\tilde{\beta}\,\eta')\,\alpha\,\nu + (\beta - \tilde{\beta})(a - a') \in \Delta_{(N,\mathbf{c})}$ for all $\tilde{\beta} \in O(\beta)$. That is, for all $\tilde{\beta} \in O(\beta)$, θ is admissible for $b\,\tilde{\beta}\,b'$. Thus, by Step 1, we have (13).

Step 3. We claim that $u(a) - W(\{a,b\}) \geq u(a') - W(\{a',b'\})$. Let $O(\beta)$ be an open interval containing $\beta \in [0,1]$ guaranteed by Step 2. Since $\{O(\beta) \mid \beta \in [0,1]\}$ is an open cover of $[0,1]$, there exists a finite subcover, denoted by $\{O(\beta^i) \mid i = 1, \ldots, I\}$. Without loss of generality, assume $\beta^i \leq \beta^{i+1}$. Define $\beta^0 = 0$ and $\beta^{I+1} = 1$. Since $\beta^0 \in O(\beta^1)$ and $\beta^{I+1} \in O(\beta^I)$, from Step 2,

$$\begin{aligned} u(a') - W(\{a',b'\}) &\leq u(a\,\beta^1\,a') - W(\{a\,\beta^1\,a', b\,\beta^1\,b'\}) \leq \cdots \\ &\leq u(a\,\beta^I\,a') - W(\{a\,\beta^I\,a', b\,\beta^I\,b'\}) = u(a) - W(\{a,b\}). \end{aligned}$$

From Step 3, for all $\alpha \in (0,1)$ sufficiently close to 1,

$$u(\mu\,\alpha\,\nu) - W(\{\mu\,\alpha\,\nu, \eta\,\alpha\,\nu\}) \geq u(\mu'\,\alpha\,\nu) - W(\{\mu'\,\alpha\,\nu, \eta'\,\alpha\,\nu\}).$$

Continuity ensures that $u(\mu) - W(\{\mu,\eta\}) \geq u(\mu') - W(\{\mu',\eta'\})$ as $\alpha \to 1$. □

Lemma 13. *For all $\mu, \mu', \eta, \eta' \in \Delta$ such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $\{\mu'\} \succ \{\mu', \eta'\} \succ \{\eta'\}$,*

$$v(\eta) - v(\mu) \geq v(\eta') - v(\mu') \Rightarrow u(\mu) - W(\{\mu, \eta\}) \geq u(\mu') - W(\{\mu', \eta'\}).$$

Proof. Let μ^+ and μ^- be a maximal and a minimal lottery in Δ with respect to v. By continuity and mixture linearity of v, for all α sufficiently close to 1, $v(\eta\,\alpha\,\mu^+) - v(\mu) > v(\eta'\,\alpha\,\mu^-) - v(\mu')$. Since the set of lotteries with finite supports is dense in Δ under the weak convergence topology (Aliprantis and Border 1999, p. 513, Theorem 15.10), there exist sequences $\{\mu_n\}$, $\{\eta_n\}$, $\{\mu'_n\}$, and $\{\eta'_n\}$ with finite supports such that $\mu_n \to \mu$, $\eta_n \to \eta\,\alpha\,\mu^+$, $\mu'_n \to \mu'$, and $\eta'_n \to \eta'\,\alpha\,\mu^-$. Moreover, by continuity of W, $W(\{\mu_n\}) > W(\{\mu_n, \eta_n\}) > W(\{\eta_n\})$ and $W(\{\mu'_n\}) > W(\{\mu'_n, \eta'_n\}) > W(\{\eta'_n\})$, and by continuity of v, $v(\eta_n) - v(\mu_n) > v(\eta'_n) - v(\mu'_n)$. By Lemma 12, for all n, we have $u(\mu_n) - W(\{\mu_n, \eta_n\}) \geq u(\mu'_n) - W(\{\mu'_n, \eta'_n\})$. Continuity of u and W implies that $u(\mu) - W(\{\mu, \eta\,\alpha\,\mu^+\}) \geq u(\mu') - W(\{\mu', \eta'\,\alpha\,\mu^-\})$ as $n \to \infty$. Again, by continuity, $u(\mu) - W(\{\mu, \eta\}) \geq u(\mu') - W(\{\mu', \eta'\})$ as $\alpha \to 1$. □

Lemma 14. *The function $\varphi : A \to (0, 1]$ is (i) well defined, (ii) weakly increasing, and (iii) continuous.*

Proof. (i) Take any μ, μ', η, η' such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $\{\mu'\} \succ \{\mu', \eta'\} \succ \{\eta'\}$. From Lemma 13, if $v(\eta) - v(\mu) = v(\eta') - v(\mu')$, $u(\mu) - W(\{\mu, \eta\}) = u(\mu') - W(\{\mu', \eta'\})$. Hence, φ is well defined.

(ii) Take $w, w' \in A$ such that $w' < w$. There exist μ, μ', η, η' such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$, $\{\mu'\} \succ \{\mu', \eta'\} \succ \{\eta'\}$, $w = v(\eta) - v(\mu)$, and $w' = v(\eta') - v(\mu')$. By Lemma 13, $\varphi(w) = u(\mu) - W(\{\mu, \eta\}) \geq u(\mu') - W(\{\mu', \eta'\}) = \varphi(w')$.

(iii) Take any $w^0 \in A$. For any sequence $w^n \to w^0$, $n = 1, 2, \ldots$, we want to show that $\varphi(w^n) \to \varphi(w^0)$. First suppose $w^0 < \sup A$. Take any $w \in (w^0, \sup A)$. There exist μ, η such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $w = v(\eta) - v(\mu)$. Since $w^n \to w$, $w^n < w$ for all sufficiently large n. Define $\alpha^n \equiv w^n/w$ for $n = 0$ and all sufficiently large n. By Lemma 7, $\{\mu\} \succ \{\mu, \eta\,\alpha^n\,\mu\} \succ \{\eta\,\alpha^n\,\mu\}$ and $w^n = v(\eta\,\alpha^n\,\mu) - v(\mu)$. By continuity of W,

$$\lim_{n\to\infty} \varphi(w^n) = \lim_{n\to\infty} u(\mu) - W(\{\mu, \eta\,\alpha^n\,\mu\}) = u(\mu) - W(\{\mu, \eta\,\alpha^0\,\mu\}) = \varphi(w^0).$$

Next suppose $w^0 = \sup A$. Since $w^0 \in A$, there exist μ, η such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $w^0 = v(\eta) - v(\mu)$. Define $\alpha^n \equiv w^n/w^0 \in (0, 1]$. By Lemma 7, $\{\mu\} \succ \{\mu, \eta\,\alpha^n\,\mu\} \succ \{\eta\,\alpha^n\,\mu\}$. Moreover, $w^n = v(\eta\,\alpha^n\,\mu) - v(\mu)$. By continuity of W,

$$\lim_{n\to\infty} \varphi(w^n) = \lim_{n\to\infty} u(\mu) - W(\{\mu, \eta\,\alpha^n\,\mu\}) = u(\mu) - W(\{\mu, \eta\}) = \varphi(w^0).$$ □

Denote the closure of A by $\overline{A}$. By Lemma 8(i), $\overline{A}$ is a closed nondegenerate interval including 0. Let $\overline{w} = \sup A$. Define $\varphi(0) = \inf\{\varphi(w) \mid w \in A\}$ and $\varphi(\overline{w}) = \sup\{\varphi(w) \mid w \in A\}$.

Lemma 15. *The function $\varphi : \overline{A} \to [0, 1]$ is a unique continuous and weakly increasing extension of φ. Moreover, (i) $\varphi(0) = 0$, (ii) φ is weakly convex, and (iii) strictly increasing.*

Proof. Since φ is continuous and weakly increasing, the former statement holds.

(i) We show that $\varphi(0)=0$. Take any $w \in A$. There exist μ, η such that $w = v(\eta) - v(\mu)$ and $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. By Lemma 7, $\{\mu\} \succ \{\mu, \eta\,\alpha\,\mu\} \succ \{\eta\,\alpha\,\mu\}$ for all $\alpha \in (0,1)$. Thus,

$$\varphi(0) = \lim_{\alpha \to 0} \varphi(\alpha w) = \lim_{\alpha \to 0} u(\mu) - W(\{\mu, \eta\,\alpha\,\mu\}) = 0.$$

(ii) We show that φ is convex on A. Then, by continuity, φ is convex on $\overline{A}$. Take any $w_i \in (0, \overline{w})$, $i = 1, 2$. Without loss of generality, assume $w_1 < w_2$. There exists $\mu, \eta_2 \in \Delta$ such that $\{\mu\} \succ \{\mu, \eta_2\} \succ \{\eta_2\}$ and $w_2 = v(\eta_2) - v(\mu)$. Let $\eta_1 = \eta_2\,(w_1/w_2)\,\mu$. Then, $w_1 = v(\eta_1) - v(\mu)$. Moreover, by Lemma 7, $\{\mu\} \succ \{\mu, \eta_1\} \succ \{\eta_1\}$. Since v is mixture linear, $\alpha w_1 + (1-\alpha) w_2 = v(\eta_1\,\alpha\,\eta_2) - v(\mu)$ for all $\alpha \in (0,1)$. By Lemma 7, $\{\mu\} \succ \{\mu, \eta_1\,\alpha\,\eta_2\} \succ \{\eta_1\,\alpha\,\eta_2\}$. In the proof of Lemma 1(ii), we show that for all $x \in Z$, there exists $\nu \in \Delta$ such that $\{\nu\} \sim x$. Let $\nu_i \in \Delta$ satisfy $\{\nu_i\} \sim \{\mu, \eta_i\}$. By Self-Control Concavity-2, $\{\mu, \eta_1\,\alpha\,\eta_2\} \succsim \{\nu_1\,\alpha\,\nu_2\}$. Thus, we have

$$\begin{aligned}
\varphi(\alpha w_1 + (1-\alpha) w_2) &= u(\mu) - W(\{\mu, \eta_1\,\alpha\,\eta_2\}) \\
&\le u(\mu) - u(\nu_1\,\alpha\,\nu_2) = \alpha(u(\mu) - u(\nu_1)) + (1-\alpha)(u(\mu) - u(\nu_2)) \\
&= \alpha\big(u(\mu) - W(\{\mu, \eta_1\})\big) + (1-\alpha)\big(u(\mu) - W(\{\mu, \eta_2\})\big) \\
&= \alpha\varphi(w_1) + (1-\alpha)\varphi(w_2).
\end{aligned}$$

(iii) First of all, since $\varphi(0) = 0$, $\varphi(0) < \varphi(w)$ for all $w \neq 0$. Next, take $w, w' \in \overline{A}$ such that $w' > w > 0$. There exists $\alpha \in (0,1)$ with $w = \alpha w'$. Since φ is convex,

$$\varphi(w) = \varphi(\alpha w') \le \alpha\varphi(w') + (1-\alpha)\varphi(0) < \varphi(w'),$$

as desired. □

Lemma 16. *Let* $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$. *If* $v(\eta) - v(\mu) \in A$, *then* $u(\eta) \ge u(\mu) - \varphi(v(\eta) - v(\mu))$.

Proof. There exist μ', η' such that $\{\mu'\} \succ \{\mu', \eta'\} \succ \{\eta'\}$ and $v(\eta') - v(\mu') = v(\eta) - v(\mu)$. Since $\varphi(v(\eta) - v(\mu)) = \varphi(v(\eta') - v(\mu')) = u(\mu') - W(\{\mu', \eta'\})$, it suffices to show that $u(\mu') - W(\{\mu', \eta'\}) \ge u(\mu) - u(\eta)$.

We claim that $u(\mu') - u(\eta') > u(\mu) - u(\eta)$. Suppose otherwise, that is, $u(\mu) - u(\eta) \ge u(\mu') - u(\eta')$. Let

$$L \equiv \big\{\alpha \in [0,1] \mid \{\mu\,\alpha\,\mu'\} \succ \{\mu\,\alpha\,\mu', \eta\,\alpha\,\eta'\} \succ \{\eta\,\alpha\,\eta'\}\big\}.$$

By assumption, $0 \in L$ and $1 \notin L$. Moreover, by Continuity, L is open in $[0,1]$. Let $\bar{\alpha} \equiv \sup L \in (0,1]$. By Continuity, $\bar{\alpha} \notin L$, and hence

$$\{\mu\,\bar{\alpha}\,\mu'\} \succ \{\mu\,\bar{\alpha}\,\mu', \eta\,\bar{\alpha}\,\eta'\} \sim \{\eta\,\bar{\alpha}\,\eta'\}. \tag{14}$$

Since $u(\mu) - u(\eta) \ge u(\mu') - u(\eta') > \varphi(v(\eta') - v(\mu'))$ and $v(\eta') - v(\mu') = v(\eta) - v(\mu)$,

$$u(\mu\,\alpha\,\mu') - u(\eta\,\alpha\,\eta') > \varphi(v(\eta\,\alpha\,\eta') - v(\mu\,\alpha\,\mu')) \tag{15}$$

for all $\alpha \in [0,1]$. On the other hand, since $\bar{\alpha}$ is a supremum of L, there exists a sequence $\{\alpha^n\}$ in L converging to $\bar{\alpha}$. We have $\{\mu\, \alpha^n\, \mu'\} \succ \{\mu\, \alpha^n\, \mu', \eta\, \alpha^n\, \eta'\} \succ \{\eta\, \alpha^n\, \eta'\}$, and hence $u(\mu\, \alpha^n\, \mu') - u(\eta\, \alpha^n\, \eta') > \varphi(v(\eta\, \alpha^n\, \eta') - v(\mu\, \alpha^n\, \mu')) = u(\mu\, \alpha^n\, \mu') - W(\{\mu\, \alpha^n\, \mu', \eta\, \alpha^n\, \eta'\})$. Continuity and (15) imply $u(\mu\, \bar{\alpha}\, \mu') - u(\eta\, \bar{\alpha}\, \eta') > \varphi(v(\eta\, \bar{\alpha}\, \eta') - v(\mu\, \bar{\alpha}\, \mu')) = u(\mu\, \bar{\alpha}\, \mu') - W(\{\mu\, \bar{\alpha}\, \mu', \eta\, \bar{\alpha}\, \eta'\})$, that is, $W(\{\mu\, \bar{\alpha}\, \mu', \eta\, \bar{\alpha}\, \eta'\}) > u(\eta\, \bar{\alpha}\, \eta')$, which contradicts (14).

Since $v(\eta') - v(\mu') = v(\eta\, \alpha\, \eta') - v(\mu\, \alpha\, \mu')$ for all $\alpha \in L$, by Lemma 14(i), $u(\mu') - W(\{\mu', \eta'\}) = u(\mu \alpha \mu') - W(\{\mu \alpha \mu', \eta \alpha \eta'\})$. Thus taking Continuity and the above claims together yields

$$
\begin{aligned}
u(\mu') - W(\{\mu', \eta'\}) &= u(\mu\, \bar{\alpha}\, \mu') - W(\{\mu\, \bar{\alpha}\, \mu', \eta\, \bar{\alpha}\, \eta'\}) = u(\mu\, \bar{\alpha}\, \mu') - u(\eta\, \bar{\alpha}\, \eta') \\
&= \bar{\alpha}(u(\mu) - u(\eta)) + (1-\bar{\alpha})(u(\mu') - u(\eta')) \geq u(\mu) - u(\eta),
\end{aligned}
$$

as desired. □

Let

$$
B \equiv \left\{ w \in [0,1] \mid w = v(\eta) - v(\mu) \text{ for some } \{\mu\} \succ \{\mu, \eta\} \right\}.
$$

By Continuity, B is open in $[0,1]$. If $\{\mu\} \succ \{\mu, \eta\}$, then $\{\mu\} \succ \{\mu, \eta\, \alpha\, \mu\}$ for all $\alpha \in (0,1)$. Hence, B is an interval satisfying $\inf B = 0$. Moreover, by definition, $A \subset B$, or $\sup A \leq \sup B$.

Define $F : B \to \mathbb{R}_+$ by

$$
F(w) \equiv \sup \left\{ u(\mu) - u(\eta) \mid w = v(\eta) - v(\mu) \text{ for some } \{\mu\} \succ \{\mu, \eta\} \right\}.
$$

Lemma 17. *The function F is weakly concave.*

Proof. Take $w_i \in B$, $i = 1, 2$, and $\alpha \in (0,1)$. There exist $\mu_i^n, \eta_i^n \in \Delta$ such that $\{\mu_i^n\} \succ \{\mu_i^n, \eta_i^n\}$, $v(\eta_i^n) - v(\mu_i^n) = w_i$, and $u(\mu_i^n) - u(\eta_i^n) \to F(w_i)$. Since $v(\eta_i^n) > v(\mu_i^n)$ and $u(\mu_i^n) > u(\eta_i^n)$, we have $v(\eta_1^n\, \alpha\, \eta_2^n) > v(\mu_1^n\, \alpha\, \mu_2^n)$ and $u(\mu_1^n\, \alpha\, \mu_2^n) > u(\eta_1^n\, \alpha\, \eta_2^n)$. Thus $\{\mu_1^n\, \alpha\, \mu_2^n\} \succ \{\mu_1^n\, \alpha\, \mu_2^n, \eta_1^n\, \alpha\, \eta_2^n\}$. Since

$$
\alpha w_1 + (1-\alpha) w_2 = \alpha(v(\eta_1^n) - v(\mu_1^n)) + (1-\alpha)(v(\eta_2^n) - v(\mu_2^n)) = v(\eta_1^n\, \alpha\, \eta_2^n) - v(\mu_1^n\, \alpha\, \mu_2^n),
$$

then

$$
\begin{aligned}
F(\alpha w_1 + (1-\alpha) w_2) &\geq \limsup u(\mu_1^n\, \alpha\, \mu_2^n) - u(\eta_1^n\, \alpha\, \eta_2^n) \\
&= \limsup \alpha(u(\mu_1^n) - u(\eta_1^n)) + (1-\alpha)(u(\mu_2^n) - u(\eta_2^n)) \\
&= \alpha F(w_1) + (1-\alpha) F(w_2).
\end{aligned}
$$

□

By Theorem 10.3 (Rockafellar 1970, p. 85), F can be uniquely extended to the closure of B in a continuous and concave way.

Lemma 18. *The function F satisfies: (i) $F(w) > \varphi(w)$ for all $w \in A$, (ii) $F(\overline{w}) \geq \varphi(\overline{w})$, and (iii) if $\overline{w} \in A$, $F(\overline{w}) = \varphi(\overline{w})$.*

Proof. (i) There exist μ, η such that $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$ and $w = v(\eta) - v(\mu)$. By definition,

$$F(w) \geq u(\mu) - u(\eta) > u(\mu) - W(\{\mu, \eta\}) = \varphi(w).$$

(ii) By Lemma 14(iii), φ is continuous. Moreover, since F is concave, F is continuous. For any sequence $w^n \to \overline{w}$, by part (i), $F(w^n) > \varphi(w^n)$. By continuity of F and φ, $F(\overline{w}) \geq \varphi(\overline{w})$.

(iii) Take any sequence $w^n \in A$ satisfying $w^n \to \overline{w}$. By part (ii), $F(\overline{w}) \geq \varphi(\overline{w})$. By way of contradiction, suppose $F(\overline{w}) > \varphi(\overline{w}) = \sup\{\varphi(w) \mid w \in A\}$. For all $w \in A$ and $\mu, \eta \in \Delta$ such that $w = v(\eta) - v(\mu)$ and $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$, by Lemma 16, we must have $\varphi(w) \geq u(\mu) - u(\eta)$. Thus there exist sequences $w^n \to \overline{w}$, $\{\mu^n\}_{n=1}^{\infty}$ and $\{\eta^n\}_{n=1}^{\infty}$ such that $w^n = v(\eta^n) - v(\mu^n) \in A$, $\{\mu^n\} \succ \{\mu^n, \eta^n\} \succ \{\eta^n\}$, and $u(\mu^n) - u(\eta^n) > c > \sup\{\varphi(w) \mid w \in A\}$, where $c > 0$ is a constant number. Since $\{\mu^n\}_{n=1}^{\infty}$ and $\{\eta^n\}_{n=1}^{\infty}$ are sequences in Δ, we can assume $\mu^n \to \mu^0$ and $\eta^n \to \eta^0$ without loss of generality. Since

$$u(\mu^n) - u(\eta^n) > c > \varphi(v(\eta^n) - v(\mu^n)) = u(\mu^n) - W(\{\mu^n, \eta^n\}),$$

continuity implies $u(\mu^0) - u(\eta^0) > u(\mu^0) - W(\{\mu^0, \eta^0\})$, that is, $W(\{\mu^0, \eta^0\}) > u(\eta^0)$. On the other hand, since $\overline{w} = v(\eta^0) - v(\mu^0) > 0$ and $u(\mu^0) > u(\eta^0)$, we have $\{\mu^0\} \succ \{\mu^0, \eta^0\}$. Hence $\{\mu^0\} \succ \{\mu^0, \eta^0\} \succ \{\eta^0\}$, which contradicts $\overline{w} \notin A$. □

Since $v(\Delta) = [0, 1]$, $\max_x v - v(\mu) \in [0, 1]$ for all $x \in Z$ and $\mu \in x$. Now we define a function $\overline{\varphi} : [0, 1] \to \mathbb{R}_+$ as

$$\overline{\varphi}(w) \equiv \begin{cases} \varphi(w) & \text{if } w \in [0, \overline{w}] \\ \dfrac{\varphi(\overline{w})}{\overline{w}} w & \text{if } w \in (\overline{w}, 1]. \end{cases} \tag{16}$$

Lemma 19. *The function $\overline{\varphi}$ is continuous, strictly increasing, and satisfies*

$$\overline{\varphi}(w) \begin{cases} = \varphi(w) & \text{if } w \in \overline{A} \\ \geq F(w) & \text{elsewhere.} \end{cases}$$

Proof. By Lemma 14, φ is continuous and strictly increasing on $[0, \overline{w}]$. Moreover, since $\varphi(\overline{w})/\overline{w} > 0$, $(\varphi(\overline{w})/\overline{w})w$ is continuous and increasing on $(\overline{w}, 1]$. Since $\overline{\varphi}(\overline{w}) = \varphi(\overline{w})$, $\overline{\varphi}$ is continuous and strictly increasing on $[0, 1]$.

If $\overline{w} \in A$, $\overline{w} = 1$ by Lemma 8(ii). Assume $\overline{w} \notin A$. Since F is concave, there exists a supporting affine function L at $(\overline{w}, F(\overline{w}))$. That is, L satisfies that $L(w) \geq F(w)$ for all w and $L(\overline{w}) = F(\overline{w})$. Since L is an affine function, $L(w)$ can be written as $aw + b$ for some $a, b \in \mathbb{R}$. If $b < 0$, for small w, $L(w) < 0$ and hence $\varphi(w) < F(w) \leq L(w) < 0$, which is a contradiction. Thus, we must have $b \geq 0$. Since $F(\overline{w}) = L(\overline{w})$,

$$\frac{F(\overline{w})}{\overline{w}} = a + \frac{b}{\overline{w}} \geq a.$$

Moreover, by Lemma 18(iii), $\varphi(\overline{w}) = F(\overline{w})$. Thus, we have $\varphi(\overline{w})/\overline{w} \geq a$. Therefore,

$$\frac{\varphi(\overline{w})}{\overline{w}} w - L(w) \begin{cases} = 0 & \text{if } w = \overline{w} \\ \geq 0 & \text{if } w > \overline{w} \\ \leq 0 & \text{if } w < \overline{w}. \end{cases} \tag{17}$$

Now take any $w \in (\overline{w}, 1]$. By (17),

$$\overline{\varphi}(w) = \frac{\varphi(\overline{w})}{\overline{w}} w \geq L(w) \geq F(w),$$

as desired. □

Lemma 20. *For all* $\mu, \eta \in \Delta$,

$$W(\{\mu, \eta\}) = \max_{\nu \in \{\mu, \eta\}} \Big\{ u(\nu) - \overline{\varphi}\Big(\max_{\{\mu,\eta\}} v - v(\nu)\Big)\Big\}.$$

Proof. Without loss of generality, assume $\{\mu\} \succsim \{\eta\}$. By Set Betweenness, $\{\mu\} \succsim \{\mu, \eta\} \succsim \{\eta\}$. There are four cases:

Case (i): $\{\mu\} \succ \{\mu, \eta\} \succ \{\eta\}$. In this case, $v(\eta) > v(\mu)$. By definition of φ, $W(\{\mu, \eta\}) = u(\mu) - \varphi(v(\eta) - v(\mu)) > u(\eta)$. Thus $W(\{\mu, \eta\})$ can be expressed as the desired form.

Case (ii): $\{\mu\} \succ \{\mu, \eta\} \sim \{\eta\}$. We have $v(\eta) > v(\mu)$. If $v(\eta) - v(\mu) \in A$, by Lemma 16, $W(\{\mu, \eta\}) = u(\eta) \geq u(\mu) - \varphi(v(\eta) - v(\mu))$ as desired. If $v(\eta) - v(\mu) \notin A$, we have either $v(\eta) - v(\mu) = \sup A$ or $v(\eta) - v(\mu) \notin \overline{A}$. The former case implies $\varphi(v(\eta) - v(\mu)) = F(v(\eta) - v(\mu))$ by Lemma 18(iii). For the latter case, by Lemma 19, $\overline{\varphi}(v(\eta) - v(\mu)) \geq F(v(\eta) - v(\mu))$. Thus, in each case,

$$\overline{\varphi}(v(\eta) - v(\mu)) \geq F(v(\eta) - v(\mu)) \geq u(\mu) - u(\eta).$$

Thus, $W(\{\mu, \eta\}) = u(\eta) \geq u(\mu) - \overline{\varphi}(v(\eta) - v(\mu))$.

Case (iii): $\{\mu\} \sim \{\mu, \eta\} \succ \{\eta\}$. By construction of v, $v(\mu) \geq v(\eta)$. Since $W(\{\mu, \eta\}) = u(\mu) > u(\eta) - \overline{\varphi}(v(\mu) - v(\eta))$, $W(\{\mu, \eta\})$ is represented by the desired form.

Case (iv): $\{\mu\} \sim \{\mu, \eta\} \sim \{\eta\}$. If $v(\eta) \geq v(\mu)$, $W(\{\mu, \eta\}) = u(\eta) \geq u(\mu) - \overline{\varphi}(v(\eta) - v(\mu))$. If $v(\mu) \geq v(\eta)$, we have $W(\{\mu, \eta\}) = u(\mu) \geq u(\eta) - \overline{\varphi}(v(\mu) - v(\eta))$. In either case, $W(\{\mu, \eta\})$ is represented by the desired form. □

Finally, we can show that the representation extends to the entire domain. The argument is similar to that used in the conclusion of the proof in Gul and Pesendorfer (2001, Theorem 1). Briefly, by Gul and Pesendorfer (2001, Lemma 2), Set Betweenness implies that the representation extends to all finite menus. Then, given that the set of finite menus is dense in Z in the Hausdorff topology, the continuity of the representation permits the representation to extend to all menus. For a more detailed argument, see Lemmas 5 and 6 in the proof of Theorem 1.

C. Proof of Corollary 1

Let $(u, v, \overline{\varphi})$ be a representation constructed as in the proof of Theorem 3. If $\succsim$ satisfies Self-Control Linearity, we can show a counterpart of Lemma 15(ii) as follows. A proof is omitted.

Lemma 21. *The function* $\varphi : \overline{A} \to [0, 1]$ *satisfies* $\varphi(\alpha w + (1-\alpha)w') = \alpha\varphi(w) + (1-\alpha)\varphi(w')$ *for all* $\alpha \in [0, 1]$.

Since $\varphi(0)=0$, we have $\varphi(\alpha w)=\varphi(\alpha w+(1-\alpha)0)=\alpha\varphi(w)$ for all $w\in[0,\overline{w}]$ and $\alpha\in[0,1]$. Thus, φ is a linear function on $[0,\overline{w}]$. Define $K\equiv\overline{\varphi}(\overline{w})/\overline{w}$. By Lemma 21, for all $w\in[0,\overline{w}]$,

$$\overline{\varphi}(w)=\overline{\varphi}\left(\frac{w}{\overline{w}}\overline{w}\right)=\overline{\varphi}(\overline{w})\frac{w}{\overline{w}}=Kw.$$

Moreover, by (16), $\overline{\varphi}(w)=Kw$ for all $w\in(\overline{w},1]$. That is, $\overline{\varphi}$ is written as a linear function with a positive slope K. Redefine v as Kv. Then (u,v) is a GP representation.

References

Aliprantis, Charalambos D. and Kim C. Border (1999), *Infinite Dimensional Analysis*, second edition. Springer, Berlin. [131, 152]

Allais, Maurice (1953), "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Americaine." *Econometrica*, 21, 503–546. [143]

Chatterjee, Kalyan and R. Vijay Krishna (2009), "A 'dual self' representation for stochastic temptation." *American Economic Journal: Microeconomics*, 1, 148–167. [129, 130]

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini (2009), "Temptation-driven preferences." *Review of Economic Studies*, 76, 937–971. [129, 130]

Ergin, Haluk and Todd Sarver (2009), "A unique costly contemplation representation." Unpublished paper, Department of Economics, Washington University in St. Louis. [137, 150]

Fudenberg, Drew and David K. Levine (2006), "A dual-self model of impulse control." *American Economic Review*, 96, 1449–1476. [128, 130]

Fudenberg, Drew and David K. Levine (2009), "Self-control, risk aversion, and the Allais paradox." Unpublished paper, Department of Economics, Washington University in St. Louis. [128, 130]

Gul, Faruk and Wolfgang Pesendorfer (2001), "Temptation and self-control." *Econometrica*, 69, 1403–1435. [127, 135, 136, 147, 148, 156]

Gul, Faruk and Wolfgang Pesendorfer (2004), "Self-control and the theory of consumption." *Econometrica*, 72, 119–158. [143]

Gul, Faruk and Wolfgang Pesendorfer (2005), "The simple theory of temptation and self-control." Unpublished paper, Department of Economics, Princeton University. [130, 131]

Kahneman, Daniel and Amos Tversky (1979), "Prospect theory: An analysis of decision under risk." *Econometrica*, 47, 263–291. [143]

Keren, Gideon and Peter Roelofsma (1995), "Immediacy and certainty in intertemporal choice." *Organizational Behavior and Human Decision Processes*, 63, 287–297. [128, 143]

Kopylov, Igor (2009), "Perfectionism and choice." Unpublished paper, Department of Economics, University of California, Irvine. [129, 130]

Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch (2001), "Risk as feelings." *Psychological Bulletin*, 127, 267–286. [143]

Nehring, Klaus (2006), "Self-control through second-order preferences." Unpublished paper, Department of Economics, University of California, Davis. [130]

Noor, Jawwad (2007), "Commitment and self-control." *Journal of Economic Theory*, 135, 1–34. [143]

Noor, Jawwad and Norio Takeoka (2010), "Menu-dependent self-controlfe." Unpublished paper, Department of Economics, Boston University. [128, 129, 130, 134, 143, 144]

Olszewski, Wojciech (forthcoming), "A model of consumption-dependent temptation." *Theory and Decision*. [130]

Rockafellar, R. Tyrrell (1970), *Convex Analysis*. Princeton University Press, Princeton, New Jersey. [154]

Stovall, John E. (2010), "Multiple temptations." *Econometrica*, 78, 349–376. [129, 130]

Takeoka, Norio (2006), "Temptation, certainty effect, and diminishing self-control." Unpublished paper, Faculty of Economics, Yokohama National University. [128, 142]

Submitted 2009-3-6. Final version accepted 2009-12-12. Available online 2009-12-13.