A Model of Self-Discipline*

Maximilian Mihm† Kemal Ozbek‡

September 3, 2016

Abstract

Self-discipline is the effort an individual exerts to regulate her own moods, thereby reducing internal conflicts between her normative preferences and temptations. In this paper, we propose a model of self-discipline where a decision-maker balances the benefits of regulating her moods against a cost of self-discipline effort. We provide an axiomatic characterization of the model in a menu-choice framework, and show how costs of self-discipline can be elicited and compared across individuals. Our model generalizes well-known models of temptation-driven behavior by viewing temptations as the endogenous outcome of a self-discipline choice problem.

Keywords: menu-choice, preference for commitment, self-discipline, temptation.

"In reading the lives of great men, I found that the first victory they won was over themselves ... self-discipline with all of them came first." Harry S. Truman

1 Introduction

Self-discipline is the effort an individual exerts to regulate her own moods. For example, someone who wants to follow a diet may exercise self-discipline to suppress

*We would like to thank Nageeb Ali, Eddie Dekel, Bart Lipman, Paola Manzini and conference participants at RUD 2016 for comments.

†New York University Abu Dhabi. Email: max.mihm@nyu.edu.

‡University of St Andrews. Email: kemal.ozbek@st-andrews.ac.uk.

her appetite; someone who wants to save for retirement may exercise self-discipline to curb her spending-mood; and someone who wants to gain a promotion may exercise self-discipline to overcome her laziness. By regulating their moods, individuals reduce the chances that they will make "bad" choices (e.g., overeat, overspend, or shirk at work). However, the fact that dieters sometimes overeat, savers sometimes overspend, and careerist sometimes shirk, suggests that self-discipline is not always easy.

In this paper, we propose an axiomatic model of self-discipline. We view self-discipline as an effort that a decision-maker (DM) exerts to regulate her own moods, thereby reducing internal conflicts between her normative preferences and temptations. Normative preferences reflect the DM's long-term goals and objectives (e.g., to lose weight, save for retirement, or gain a promotion). Temptations, on the other hand, are temporary preferences that can strike at random and reflect the DM's immediate desires, urges and cravings (e.g., overeating, overspending, or shirking). By regulating her moods (e.g., suppressing her appetite, curbing her spending-mood, or overcoming her laziness), the DM reduces the chances of being struck by temptations that lead her to make normatively bad choices. On the other hand, self-discipline may be hard, and so the DM must balance the benefit of regulating her moods against the costs of self-discipline effort.

In section 2, we provide a formal model of a self-discipline choice problem. Although self-discipline can represent unobservable efforts, the self-discipline choice problem induces a preference relation over menus that is – in principle – revealed by choice behavior. We therefore study the behavioral implications of self-discipline in a two-period framework, where the DM chooses a menu in the first period anticipating her self-discipline problem before she selects an alternative in the second period.

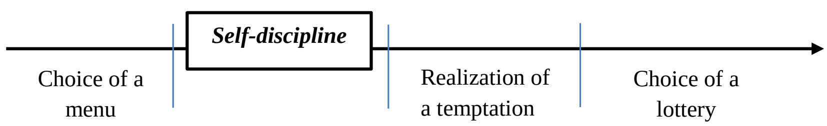

Figure 1: Timeline

Our main result (Theorem 1) provides an axiomatic characterization of self-discipline preferences. In particular, two novel axioms identify distinctive features of the self-discipline model. *Temptation Monotonicity* reflects the idea that if the ex-post choice from menu A is normatively better than the ex-post choice from menu B for *every* temptation preference that could strike in the second period, then the DM prefers menu A to menu B. *Aversion to Randomization* reflects the idea that if DM has the opportunity to adjust her self-discipline effort, she prefers an early resolution of uncertainty in order to better respond to self-discipline incentives. Together with four standard axioms from the menu-choice literature, these behavioral conditions characterize the testable implications of the self-discipline model in terms of menu-choice data.

We also show that model parameters can be elicited from preferences over menus (Theorem 2). In particular, the DM's normative preferences are uniquely identified by her commitment ranking over singleton menus,[1] and we provide an explicit formula that shows how the minimal costs of self-discipline can be recovered from data on singleton equivalents. Finally, we show that a decision-maker with higher costs of self-discipline values commitment more than one with lower costs, establishing a comparative measure of self-discipline in terms of behavioral data (Theorem 3).

Our self-discipline model generalizes well-known models of temptation-driven behavior by viewing temptations as the endogenous outcome of a self-discipline choice problem. For instance, we show in Proposition 1 that the random Strotz model in Dekel and Lipman [2012] corresponds to the special case of self-discipline preferences that satisfy a Set Independence axiom (which implies Aversion to Randomization). Dekel and Lipman [2012] show that random Strotz preferences generalize the Strotz and self-control models in Gul and Pesendorfer [2001], as well as the multiple temptation model in Stovall [2010] and dual-self models in Chatterjee and Krishna [2009] and Olszewski [2007]. As a result, each of these models represent important special cases of self-discipline preferences.

Other models of temptation-driven behavior overlap with self-discipline preferences.

[1]It is customary to interpret the commitment ranking as the DM's normative preference. Noor [2011] provides a critical discussion of this interpretation.

Chandrasekher [2014] proposes a generalization of random Strotz, weakening the assumption that preferences are complete while retaining Set Independence. The overlap with self-discipline preferences is the random Strotz model. Noor and Takeoka [2010] characterize convex self-control preferences, which can violate Aversion to Randomization but satisfy a Set Betweenness axiom that self-discipline preferences can violate. The overlap with self-discipline preferences is the self-control model. Dekel et al. [2009] characterize a model with multi-dimensional temptations that can violate Weak Set Betweenness (which is implied by Temptation Monotonicity), but satisfies Set Independence. The overlap with self-discipline preferences is the multiple temptation model. Hence, each of these models overlap with self-discipline preferences, neither one nesting the other.

In a discrete choice setting, Nehring [2006] proposes a model of preferences over second-order preferences, where a DM exercises control over her future choice dispositions. The interpretation of his model is similar to ours. However, we consider a framework with menus of lotteries, and this extended setting is important to interpret, characterize, and identify our self-discipline model.

Finally, self-discipline preferences share some common features with variational preferences (first introduced in Maccheroni et al. [2006]). The closest model in the menu-choice literature with a similar structure is the costly contemplation model in Ergin and Sarver [2010]. Costly contemplation can also induce a desire for early resolution of uncertainty, generating systematic violations of Set Independence. However, costly contemplation preferences exhibit a desire for flexibility, and our preferences satisfy this condition if and only if self-discipline is irrelevant (i.e., there is no conflict between normative preferences and temptations).

The remainder of the paper is organized as follows. In Section 2, we describe the framework and formally define self-discipline preferences. Section 3 introduces our axioms. In Section 4, we provide our main results: a representation theorem, identification result, and comparative statics. We also highlight some novel implications of the self-discipline model for ex-post choice behavior. Section 5 concludes. Proofs are given in a separate Appendix.

2 Model

In this section, we describe our framework and formally define self-discipline preferences.

2.1 Framework

In the following, X is a finite set of n alternatives, with typical elements $x, y, z \in X$ called *outcomes*; P is the set of all probability distributions on X, with typical elements $p, q, r \in P$ called *lotteries*; and $\mathcal{A}$ is the set of non-empty closed subsets of P, with typical elements $A, B, C \in \mathcal{A}$ called *menus*.

Our primitive is a binary relation $\succsim$ on the set of menus, with asymmetric part denoted $\succ$ and symmetric part denoted $\sim$. We interpret the binary relation $\succsim$ as the preference relation of a DM who chooses a menu in period 1, anticipating that she will choose a lottery from the menu in period 2. We call the restriction of $\succsim$ to singleton menus the *commitment ranking*. A functional $U : \mathcal{A} \to \mathbb{R}$ represents $\succsim$ if, for all menus A and B,

$$A \succsim B \quad \Leftrightarrow \quad U(A) \geq U(B).$$

Let $co(A)$ denote the convex hull of menu A; since the set of alternatives is finite $co(A)$ is also a menu.

For menus $A, B \in \mathcal{A}$ and $\alpha \in [0, 1]$, let

$$\alpha A + (1 - \alpha)B = \{\alpha p + (1 - \alpha)q \, : \, p \in A, \; q \in B\}.$$

The menu $\alpha A + (1 - \alpha)B$ can be interpreted as a randomization over menus A and B that is resolved *after* the DM has chosen a lottery, i.e., the DM chooses lottery $\alpha p + (1 - \alpha)q \in \alpha A + (1 - \alpha)B$ not knowing whether the lottery $p \in A$ or $q \in B$ will determine her final outcome.

Given a binary relation $\succsim$ on menus, let $A(p) = \{q \in A : \; \{q\} \sim \{p\}\}$ denote the intersection of menu $A \in \mathcal{A}$ and the "indifference curve" of the commitment ranking corresponding to lottery $p \in P$. If the commitment ranking is continuous, $A(p)$ is a menu whenever it is non-empty.

2.2 Temptations

For any two vectors $v, w \in \mathbb{R}^n$, let $v \cdot w$ denote the dot product of v and w. An expected utility function on P can be identified with an element of $\mathbb{R}^n$; hence, if $v \in \mathbb{R}^n$ and $p \in P$, we use $v \cdot p$ and $v(p)$ interchangeably.

Let $\mathcal{V} = \{v \in \mathbb{R}^n : \sum_{i=1}^n v_i = 0,\ v \cdot v = 1\}$, with typical elements $u, v, w \in \mathcal{V}$ called *(expected) utilities*. For any non-constant $w \in \mathbb{R}^n$, there exists a unique $v_w \in \mathcal{V}$ such that, for all $p, q \in P$, $w(p) \geq w(q)$ if and only if $v_w(p) \geq v_w(q)$. Hence, any non-trivial expected utility preference on P can be represented by a unique utility in $\mathcal{V}$.

Let $\Delta(\mathcal{V})$ be the set of all finitely-additive Borel probability distributions on $\mathcal{V}$, with typical elements $\pi, \rho, \mu \in \Delta(\mathcal{V})$ called *distributions (over temptations)*.[2] For a utility $v \in \mathcal{V}$, denote by $\delta_v \in \Delta(\mathcal{V})$ the distribution that assigns probability 1 to v.

For $u \in \mathcal{V}$, define a partial order over $\mathcal{V}$ by $v\, C_u\, w$ (read "v is closer to u than w") if, whenever $u(p) > u(q)$,

$$w(p) \geq w(q) \quad \Rightarrow \quad v(p) \geq v(q).$$

Hence, the utility v is closer to u than w if, whenever w agrees with u, v also agrees with u. A subset $\mathcal{W} \subseteq \mathcal{V}$ is *closed under* C_u if $w \in \mathcal{W}$ and $v\, C_u\, w$ implies $v \in \mathcal{W}$.

Dekel and Lipman [2012] propose an extension of C_u from $\mathcal{V}$ to $\Delta(\mathcal{V})$, which formalizes the intuition that one distribution over temptations is more closely aligned (in a stochastic sense) with specific utility $u \in \mathcal{V}$ than another distribution. Specifically, define $\pi \trianglerighteq_u \rho$ (read "π is stochastically closer to u than ρ") if $\pi(\mathcal{W}) \geq \rho(\mathcal{W})$ for every $\mathcal{W} \subseteq \mathcal{V}$ that is (i) closed in $\mathcal{V}$, and (ii) closed under C_u. Hence, the distribution π is stochastically closer to u than ρ if the utilities realized under π

[2] In the literature, a distribution over temptations is also sometimes called a random utility. See, e.g., Luce and Suppes [1965, Chapter 5.3] or Gul and Pesendorfer [2006] for detailed discussions.

agree more often with u than the utilities realized under ρ.[3]

For menu $A \in \mathcal{A}$ and utility $v \in \mathcal{V}$, let

$$M_v(A) = \arg\max_{p \in A} v(p),$$

denote the set of lotteries in menu A that maximize the utility v. Since A is non-empty and closed, and v is a continuous function on P, $M_v(A)$ is also a menu by the Maximum theorem.

For menu $A \in \mathcal{A}$ and utility $u \in \mathcal{V}$, define the function $\varphi_A^u : \mathcal{V} \to \mathbb{R}$ by

$$\varphi_A^u(v) = \max_{p \in M_v(A)} u(p),$$

i.e., for any $v \in \mathcal{V}$, $\varphi_A^u(v)$ is the maximum of u over $M_v(A)$. Since u is a continuous function on P and $M_v(A)$ is a menu for all $v \in \mathcal{V}$, φ_A^u is well-defined.

2.3 Self-discipline

Dekel and Lipman [2012] propose a random Strotz model to describe the behavior of a DM concerned about temptations:

Definition 1. [Random Strotz] A binary relation $\succsim$ on menus is a *random Strotz preference* if there exists $u \in \mathcal{V}$ and $\pi \in \Delta(\mathcal{V})$ such that the functional $U : \mathcal{A} \to \mathbb{R}$, defined by

$$U(A) = \int_{\mathcal{V}} \varphi_A^u(v)\, \pi(dv),$$

represents $\succsim$.

The random Strotz model can be interpreted as follows. The utility function u represents the DM *normative preference* over lotteries, which reveals her long-term goals and objectives (e.g., to lose weight, save for retirement, or gain a promotion). However, when the DM chooses a lottery in the second period, she has temporary preferences – *temptations* – that reveal her immediate desires, cravings and urges

[3] Dekel and Lipman [2012] give a number of equivalent definitions of the partial order $\trianglerighteq_u$, which allow for a direct interpretation in terms of first order stochastic dominance.

(to eat a hamburger, purchase a new phone, or take an extended lunch break). Temptations are overwhelming: if the DM's temptation at the moment of choice is represented by the utility $v \in \mathcal{V}$, she is only able to choose one of the tempting lotteries in $M_v(A)$ from menu A. Until the moment of choice, the DM is uncertain about the nature of the temptation that will strike. When uncertainty about temptations is described by the distribution $\pi \in \Delta(\mathcal{V})$, the DM prefers menu A over menu B if and only if

$$\int_{\mathcal{V}} \varphi_A^u(v)\, \pi(dv) \geq \int_{\mathcal{V}} \varphi_B^u(v)\, \pi(dv).$$

To formalize the idea that individuals exercise effort to regulate their own behavior, we generalize the random Strotz model by viewing temptations as the endogenous outcome of a self-discipline choice problem.

Our model formalizes the following idea. Temptations represent the DM immediate desires, urges and cravings when she is called upon to choose a lottery. Until the moment of choice, the DM is uncertain about the nature of temptations that will strike. We view self-discipline as an effort the DM can exert to regulate her own moods before a temptation is realized. By regulating her moods, the DM reduces the chances that she will be struck by a temptation the leads her to make normatively bad choices. For example, by suppressing her appetite, the dieter is less likely to be struck by a temptation that leads her to overeat; by curbing her mood to spend, the saver is less likely to be struck by a temptation that leads her to overspend; and by overcoming her laziness, the careerist is less likely to be struck by a temptation that leads her to shirk at work. However, self-discipline can be hard, and so the DM must balance the benefits of regulating her moods against the cost of self-discipline effort. These costs are represented by a function $c : \Delta(\mathcal{V}) \to \mathbb{R}$, where $c(\pi)$ is a behavioral measure of the effort required to induce the distribution π over temptations.

A self-discipline preference reflects the behavior of a DM who acts "as if" she anticipates a costly self-discipline effort in the future:

Definition 2. [Self-discipline] A binary relation $\succsim$ on menus is a *self-discipline preference* if there exists $u \in \mathcal{V}$ and a proper lower-semicontinuous function

$c : \Delta(\mathcal{V}) \to [0, \infty]$ such that the functional $U : \mathcal{A} \to \mathbb{R}$, defined by

$$U(A) = \max_{\pi \in \Delta(\mathcal{V})} \left(\int_{\mathcal{V}} \varphi^u_A(v)\, \pi(dv) - c(\pi) \right), \tag{1}$$

represents $\succsim$. In this case, we say that $\succsim$ is represented by (u, c).

The self-discipline model admits a natural multi-system interpretation, in which the utility u represents cognitive processes responsible for setting normative objectives and regulating (at a cost) other cognitive processes responsible for desires, urges and cravings. This approximates the way many psychologists view temptation and self-control problems (see, e.g., Normann and Shallice [2000]). It also has a connection with dual-self models studied in many applications, where one self (the planner) is responsible for setting long-term goals and objectives, and can control (at a cost) the behavior of another self (the doer) who executes short-term decision-making (see, e.g., Thaler and Shefrin [1981], Bénabou and Tirole [2004], Benhabib and Bisin [2005], Fudenberg and Levine [2006], and Ali [2011]).

Properness and lower-semicontinuity are minimal properties of a cost function to ensure that the self-discipline choice problem is well-defined.[4] We impose no other a priori restrictions on the distribution over temptations that could be induced by self-discipline effort. Instead, we assume that the DM anticipates her self-discipline choice problem, and show that menu-choice reveals her self-discipline costs.

The following examples illustrate some special cases, which generalize existing models of temptation-driven behavior. Let $\succsim$ be a self-discipline preference represented by (u, c).

Example 1. [Random Strotz] For some distribution over temptations $\pi \in \Delta(\mathcal{V})$, let c be defined on $\Delta(\mathcal{V})$ by

$$c(\rho) = \begin{cases} 0 & \text{if } \rho = \pi \\ \infty & \text{otherwise} \end{cases}.$$

[4]The function $c : \Delta(\mathcal{V}) \to [0, \infty]$ is proper if $c(\pi) < \infty$ for some $\pi \in \Delta(\mathcal{V})$. Lower semicontinuity is defined with respect to the weak* topology.

Then $U(A) = \int_{\mathcal{V}} \varphi^u_A(v)\, \pi(dv)$ for all $A \in \mathcal{A}$. Hence, random Strotz preferences can be viewed as the special case of self-discipline preferences where there is a common solution to the self-discipline choice problem for all menus (in this case, the distribution π). $\square$

Example 2. [Costly Strotz] For some function $k : \mathcal{V} \to [0, \infty]$, which is proper and lower semicontinuous, let c be defined on $\Delta(\mathcal{V})$ by $c(\pi) = \int_{\mathcal{V}} k(v)\, \pi(dv)$. Then $U(A) = \max_{v \in \mathcal{V}} \left(\varphi^u_A(v) - k(v)\right)$ for all $A \in \mathcal{A}$. Costly Strotz preferences represent the special case of our model where ex-post choice is non-stochastic. They can be viewed as a generalization of the Strotz model in Gul and Pesendorfer [2001] where – instead of a fixed temptation in period 2 – the DM can exercise self-discipline to align her temptations more closely with her normative utility. $\square$

Example 3. [Costly dual-self] For some $v \in \mathcal{V}$ and a proper lower-semicontinuous function $k : [0,1] \to [0, \infty]$, let c be defined on $\Delta(\mathcal{V})$ by

$$c(\pi) = \begin{cases} k(\beta) & \text{if } \pi = \beta\delta_u + (1-\beta)\delta_v \\ \infty & \text{otherwise} \end{cases}.$$

Then $U(A) = \max_{\rho \in [0,1]} \left(\beta\varphi^u_A(u) + (1-\beta)\varphi^u_A(v)\right)$ for all $A \in \mathcal{A}$. Costly dual-self preferences represent a special case of our general model where self-discipline effort is "one-dimensional". They can be viewed as a generalization of the dual-self model in Chatterjee and Krishna [2009] where – instead of a fixed chance that an "alter ego" (utility function v) will select a lottery in period 2 – the DM can exercise self-discipline to increase the chances that her normative utility u will prevail. $\square$

3 Axioms

In this Section, we present our axioms. The first three axioms are standard:

Axiom 1. *[Non-trivial Weak Order] For all menus $A, B, C \in \mathcal{A}$, (i) $A \succsim B$ or $B \succsim A$, and (ii) $A \succsim B$ and $B \succsim C$ implies $A \succsim C$. Moreover, there exist menus $A, B \in \mathcal{A}$ such that $A \succ B$.*

Axiom 2. *[Mixture-Continuity] For all menus $A, B, C \in \mathcal{A}$, the following sets are closed:*

$$\{\alpha \in [0,1] \,:\, \alpha A + (1-\alpha)B \succsim C\} \quad \textit{and} \quad \{\alpha \in [0,1] \,:\, C \succsim \alpha A + (1-\alpha)B\}\,.$$

Axiom 3. *[Indifference to Convexification] For all menus $A \in \mathcal{A}$, $A \sim co(A)$.*

Axiom 1 requires that the preference relation is complete, transitive, and non-trivial. Axiom 2 imposes continuity for mixtures over menus. Axiom 3 follows from the assumption that normative preferences and temptations satisfy the standard von Neumann-Morgenstern (vNM) axioms. As a result, the DM always chooses an extreme point from a menu: since A and $co(A)$ have the same extreme points, $A \sim co(A)$. The remaining axioms identify distinctive features of self-discipline and we discuss them in more detail.

3.1 Monotonicity

Our next axiom is a monotonicity condition based on comparing ex-post choices from menus "temptation-by-temptation". Suppose that *for every* temptation preference $v \in \mathcal{V}$, a lottery is tempting in A that is normatively preferred to every lottery that is tempting in B (i.e., there exists $p \in M_v(A)$ such that $\{p\} \succsim \{q\}$ for all $q \in M_v(B)$). Then the DM could "replicate" the self-discipline effort she would exert for menu B when she faces menu A, and guarantee a normatively better outcome with menu A. Hence, she should prefer A to B.

The following definition formalizes the idea that menu A dominates menu B "temptation-by-temptation":

Definition 3. [Temptation dominance] Menu A *temptation-dominates* menu B (denoted $A \trianglerighteq B$) if, whenever $\{p\} \succ \{q\}$,

$$\frac{1}{2}A(p) + \frac{1}{2}B(q) \supseteq \frac{1}{2}B(p) + \frac{1}{2}A(q).$$

To motivate the definition, suppose $A \trianglerighteq B$. Now fix a specific temptation preference $v \in \mathcal{V}$, and suppose p is tempting in B, q is tempting in A, and $\{p\} \succ \{q\}$. Since

$\{p\} \succ \{q\}$, $A \trianglerighteq B$ implies that there exists $p' \in A(p)$ and $q' \in B(q)$ such that $\frac{1}{2}p' + \frac{1}{2}q' = \frac{1}{2}p + \frac{1}{2}q$. Moreover, since temptations satisfy the standard vNM independence axiom, $\frac{1}{2}p + \frac{1}{2}q$ is tempting in $\frac{1}{2}A + \frac{1}{2}B$, and so $\frac{1}{2}p' + \frac{1}{2}q'$ is tempting in $\frac{1}{2}A + \frac{1}{2}B$. But this implies that p' is tempting in A (by vNM independence). Hence, for any lottery that is tempting in menu B (e.g., p), there is normatively better lottery that is tempting in A (e.g., p'). The same argument applies *for every* temptation, and so the DM should prefer menu A to menu B:

Axiom 4. *[Temptation Monotonicity] For all menus $A, B \in \mathcal{A}$,*

$$A \trianglerighteq B \quad \Rightarrow \quad A \succsim B.$$

3.2 Randomization

Our last two axioms reflect the idea that temptations are not necessarily fixed, but are the endogenous outcome of a self-discipline choice problem. In particular, the opportunity to adjust self-discipline effort in response to incentives induces a desire for early resolution of uncertainty. To illustrate, consider an extension of the DM's preferences to lotteries over menus. In particular, for menus $A, B \in \mathcal{A}$, let $\alpha \circ A + (1-\alpha) \circ B$ denote a randomization over menus A and B that is resolved in period 1 (immediately after the DM has chosen a menu). By contrast, the menu $\alpha A + (1-\alpha)B$ denotes a randomization over A and B that is resolved in period 2 (immediately after the DM has chosen a lottery). When deciding how much self-discipline to exercise, a DM may strictly prefer an early resolution of uncertainty so that she can condition her self-discipline effort on the realization of uncertainty:

Axiom 5. *[Aversion to Randomization] For all menus $A, B \in \mathcal{A}$, lotteries $p, q \in P$, and $\alpha \in (0, 1)$,*

$$A \sim \{p\} \text{ and } B \sim \{q\} \quad \Rightarrow \quad \alpha\{p\} + (1-\alpha)\{q\} \succsim \alpha A + (1-\alpha)B.$$

To motivate the axiom, suppose the DM is indifferent between menus A and $\{p\}$, and between menus B and $\{q\}$. A standard independence argument would imply that the DM should also be indifferent between $\alpha \circ \{p\} + (1-\alpha) \circ \{q\}$ and $\alpha \circ A + (1-\alpha) \circ B$. However, the optimal self-discipline effort for menus A and B may differ. Intuitively, the DM would like to make her choice of self-discipline contingent on the menu that is actually payoff relevant, and so prefers an early resolution of uncertainty when randomizing over menus A and B (i.e., $\alpha \circ A + (1-\alpha) \circ B \succsim \alpha A + (1-\alpha) B$). On the other hand, self-discipline is redundant for singleton menus, and so the timing of the resolution of uncertainty is irrelevant for the singleton menus $\{p\}$ and $\{q\}$ (i.e., $\alpha \circ \{p\} + (1-\alpha) \circ \{q\} \sim \alpha\{p\} + (1-\alpha)\{q\}$). Hence, $\alpha\{p\} + (1-\alpha)\{q\} \succsim \alpha A + (1-\alpha)B$ is implied by a standard independence argument together with a desire for early resolution of uncertainty for non-singleton menus.

Our final axiom excludes intrinsic (psychological) motivations for an early resolution of uncertainty, focusing the desire for early resolution on the role of self-discipline effort:

Axiom 6. *[Independence of Degenerate Decisions] For all menus $A, B \in \mathcal{A}$, lotteries $p, q \in P$, and $\alpha \in (0,1)$,*

$$\alpha A + (1-\alpha)\{p\} \succsim \alpha B + (1-\alpha)\{p\} \quad \Rightarrow \quad \alpha A + (1-\alpha)\{q\} \succsim \alpha B + (1-\alpha)\{q\}.$$

To motivate the axiom, suppose the DM prefers menu $\alpha A + (1-\alpha)\{p\}$ to $\alpha B + (1-\alpha)\{p\}$. While the incentives to exercise self-discipline in $\alpha A + (1-\alpha)\{p\}$ may depend on menu A and probability α, they are independent of the particular lottery p because self-discipline is redundant for singleton menus (where the decision is degenerate). Likewise, the incentives to exercise self-discipline in menu $\alpha B + (1-\alpha)\{p\}$ depend only on B and α. As such, replacing singleton menu $\{p\}$ with menu $\{q\}$ does not alter the incentives to exercise self-discipline. Since self-discipline is the only reason the DM may prefer an early resolution of uncertainty, the DM therefore also prefers $\alpha A + (1-\alpha)\{q\}$ to $\alpha B + (1-\alpha)\{q\}$.

4 Analysis

In this section, we analyze the implications of the self-discipline model for menu-choice.

4.1 Characterization

The following theorem shows that the axioms in Section 3 characterize the behavior of a DM who chooses among menus "as if" she anticipates a self-discipline choice problem:

Theorem 1. *A binary relation on menus $\succsim$ is a self-discipline preference if and only if it satisfies Axioms 1–6.*

Proof sketch: It is straightforward to show that self-discipline preferences satisfy Axioms 1–6. For the converse, we show in the proof of Theorem 1 that Axioms 1–2 and 6 imply that the commitment ranking can be represented by a utility $u \in \mathcal{V}$. Using singleton equivalents, we then define a functional I over the set $\Phi = \{\varphi_A^u : A \in \mathcal{A}\}$ such that, for all menus A and B, $A \succsim B$ if and only if $I(\varphi_A^u) \geq I(\varphi_B^u)$. The key step in the proof is to show that I is monotone, i.e., $\varphi_A^u \geq \varphi_B^u$ implies $I(\varphi_A^u) \geq I(\varphi_B^u)$. To establish this, we show in Lemma 1 that for convex menus A and B, $A \trianglerighteq B$ if and only if $\varphi_A^u \geq \varphi_B^u$. The monotonicity of I then follows from Axioms 3 and 4. The remainder of the proof uses Axioms 5–6 to show that I is a continuous convex functional, and applies a duality argument to establish the desired representation. $\square$

In addition to establishing testable implications of self-discipline, the characterization facilitates a comparison of self-discipline preferences with other models of menu-choice.

Axioms 1–3 are satisfied by most models of menu-choice in a framework with lotteries. Many models satisfy a stronger continuity condition introduced in Dekel et al. [2001]:

Axiom. *[Strong Continuity] For all menus $A \in \mathcal{A}$, the following sets are closed in the Hausdorff topology:*

$$\{B \in \mathcal{A} : B \succsim A\} \quad \textit{and} \quad \{B \in \mathcal{A} : A \succsim B\}.$$

Strong Continuity is a technical condition, which implies Mixture Continuity. However, Strong Continuity is inconsistent with some important special cases of self-discipline preferences, such as the Strotz model in Gul and Pesendorfer [2001] and the dual-self model in Chatterjee and Krishna [2009]. For a weak order that satisfies Strong Continuity, Axiom 3 is implied by Set Independence (see definition below); we require Indifference to Convexification explicitly because self-discipline preferences can violate both Strong Continuity and Set Independence.

Axiom 4 reveals that our DM is concerned that she can be tempted by ex-ante inferior alternatives in the second period. In particular, it implies the Weak Set Betweenness axiom in Dekel et al. [2009]:

Axiom. *[Weak Set Betweenness] For all menus $A, B \in \mathcal{A}$, if $\{p\} \succsim \{q\}$ for all $p \in A$ and $q \in B$, then $A \succsim A \cup B \succsim B$.*

It is straightforward to show that $\{p\} \succsim \{q\}$ for all $p \in A$ and $q \in B$ implies $A \trianglerighteq A \cup B \trianglerighteq B$ (in the sense of Definition 3). Following the discussion in Section 3.1, Weak Set Betweenness can therefore also be interpreted as a monotonicity condition based on comparing menus "temptation-by-temptation". However, the logic behind Weak Set Betweenness uses only the assumption that temptations are continuous weak orders.[5] Temptation Monotonicity extends the set of menus that are comparable by using the additional assumption that temptations satisfy the standard vNM independence axiom (thereby strengthening the axiom).

Axiom 5 is related to the Aversion to Contingent Planning (ACP) axiom in Ergin

[5]To illustrate, suppose $\{p\} \succsim \{q\}$ for all $p \in A$ and $q \in B$. Now fix a specific temptation preference, and suppose p is tempting in A and q is tempting in B. Since we have fixed a temptation preference, either p or q must be tempting in $A \cup B$. In either case, there is a lottery that is tempting in A that is normatively preferred to the lottery that is tempting in $A \cup B$, and a lottery that is tempting in $A \cup B$ that is normatively preferred to the lottery that is tempting in B. The same argument applies for any temptation preference that is representable by a continuous utility function (regardless of whether the temptation satisfies vNM independence or not).

and Sarver [2010], who also introduce Axiom 6 in the context of their costly contemplation model. For a weak order, Aversion to Randomization implies ACP if every menu has a singleton equivalent, while ACP implies Aversion to Randomization if preferences satisfy Strong Continuity. However, singleton equivalents may fail to exist for the costly contemplation preferences in Ergin and Sarver [2010], while self-discipline preferences can violate Strong Continuity. Aversion to Randomization and ACP are therefore not interchangeable.

Together, Axioms 5 and 6 reveal the systematic violations of the Set Independence axiom in Dekel et al. [2001] induced by a self-discipline choice problem:

Axiom. *[Set Independence] For all menus $A, B, C \in \mathcal{A}$ and $\alpha \in (0,1)$,*

$$A \succsim B \quad \Rightarrow \quad \alpha A + (1-\alpha)C \succsim \alpha B + (1-\alpha)C.$$

Set Independence can be viewed as combining a standard independence argument – $A \succsim B$ implies $\alpha \circ A + (1-\alpha) \circ C \succsim \alpha \circ B + (1-\alpha) \circ C$ – with indifference to the timing of the resolution of uncertainty – $\alpha \circ A + (1-\alpha) \circ C \sim \alpha A + (1-\alpha)C$ – (see, e.g., Gul and Pesendorfer [2001, p. 1407]). The standard independence condition seems compelling, but indifference to the timing of the resolution of uncertainty is less compelling when there are hidden actions available to a DM (see, e.g., Kreps and Porteus [1978], and Machina [1984]). Self-discipline effort is a specific example of a hidden action, which induces violations of Set Independence in our framework. In particular, a self-discipline preference will satisfy Set Independence if and only if there is a common self-discipline effort that is optimal *for every* menu:

Proposition 1. *Let $\succsim$ be a self-discipline preference. Then $\succsim$ is a random Strotz preference if and only if it satisfies the Set Independence axiom.*

Dekel and Lipman [2012] characterize the class of continuous-intensity random Strotz preferences. Proposition 1 complements their result with an axiomatic characterization of random Strotz preferences without additional continuity conditions, also providing a novel interpretation of the random Strotz model as the special case of self-discipline where the self-discipline choice problem has a common solution for all menus.

4.2 Identification

Our next result shows how self-discipline costs can be recovered from menu-choice data. The normative utility is identified uniquely from the DM's commitment ranking. However, it is generally not possible to identify a unique cost function. To see why, consider a self-discipline preference $\succsim$ represented by (u, c), and let

$$\mathcal{D}(A|u,c) = \arg\max_{\pi \in \Delta(\mathcal{V})} \left[\int_{\mathcal{V}} \varphi_A(v)\, \pi(dv) - c(\pi) \right]$$

denote the set of optimal distributions for the self-discipline choice problem with menu A. Now suppose $c(\pi) < \infty$ for some $\pi \notin \bigcup_{A \in \mathcal{A}} \mathcal{D}(A|u,c)$. Define another cost function $\tilde{c}$ such that $\tilde{c}(\pi) > c(\pi)$ and $\tilde{c}(\rho) = c(\rho)$ for all $\rho \neq \pi$. Then clearly $(u, \tilde{c})$ also represents $\succsim$. Hence, it is not possible to identify a unique cost for a distribution that is never optimal. However, the following theorem shows that there is a unique minimal cost function. We observe that, for a self-discipline preference $\succsim$, every menu $A \in \mathcal{A}$ has a *singleton equivalent* $p_A \in P$ such that $\{p_A\} \sim A$.

Theorem 2. *Let $\succsim$ be a self-discipline preference represented by (u, c). Then:*

(i) The utility u is the unique function in $\mathcal{V}$ such that, for all $p, q \in P$,

$$\{p\} \succsim \{q\} \quad \Leftrightarrow \quad u(p) \geq u(q).$$

(ii) The function $c^* : \Delta(\mathcal{V}) \to [0, \infty]$, defined by

$$c^*(\pi) = \sup_{A \in \mathcal{A}} \left(\int_{\mathcal{V}} \varphi_A(v)\, \pi(dv) - u(p_A) \right) \quad \forall \pi \in \Delta(\mathcal{V}), \tag{2}$$

is the unique minimal self-discipline cost function such that (u, c^*) represents $\succsim$.

Theorem 2 shows how the parameters of a self-discipline model can be elicited from menu-choice data. First, the normative utility u can be recovered directly from the commitment ranking. Second, the cost function c^* can be constructed from data on singleton equivalents. For instance, suppose we want to determine the cost of

$\pi \in \Delta(\mathcal{V})$ using experimental data. For any menu A, $c^*(\pi) \geq \int \varphi_A(v)\,\pi(dv) - u(p_A)$, where p_A is the singleton equivalent of A. This provides a lower bound on the self-discipline cost of π. Data on singleton equivalents for other menus allows us to estimate a more precise lower bound. The formula in Eq. (2) shows that this procedure approximates $c^*(\pi)$ arbitrarily closely, establishing a direct connection between the self-discipline cost function and menu-choice behavior.

For the menu-choice implications of the self-discipline model, Theorem 2 shows that it is without loss of generality to consider only the minimal cost function c^*. The following corollary shows that cost function c^* also measures the cost of any feasible self-discipline effort.

Corollary 1. *Suppose the self-discipline preference $\succsim$ has canonical representation (u, c^*), and let c be another cost function such that (u, c) also represents $\succsim$. Then:*

(i) $\pi \in \bigcup_{A \in \mathcal{A}} \mathcal{D}(A|u, c)$ *implies* $c(\pi) = c^*(\pi)$, *and*

(ii) $\mathcal{D}(A|u, c) \subseteq \mathcal{D}(A|u, c^*)$ *for all* $A \in \mathcal{A}$.

The DM therefore never exercise a self-discipline effort that would be suboptimal with the minimal cost function, and the cost of any self-discipline effort the DM anticipates exercising is equal to the minimal cost. As a result, we refer to c^* as the *canonical cost function*, and call (u, c^*) the *canonical representation* of $\succsim$. The following Proposition shows that the canonical cost function also satisfies some properties that seem intuitive for a measure of self-discipline costs:

Proposition 2. *Let $\succsim$ be a self-discipline preference with the canonical representation (u, c^*). Then c^* satisfies the following properties:*

(i) Groundedness: $c^*(\pi) = 0$ *for some* $\pi \in \Delta(\mathcal{V})$.

(ii) Convexity: $c^*(\alpha\pi + (1-\alpha)\rho) \leq \alpha c^*(\pi) + (1-\alpha)c^*(\rho)$ *for all* $\pi, \rho \in \Delta(\mathcal{V})$ *and* $\alpha \in [0, 1]$.

(iii) Monotonicity: $\pi \trianglerighteq_u \rho$ *implies* $c^*(\pi) \geq c^*(\rho)$.

Groundedness means that the DM has the option to exercise no self-discipline effort, thereby incurring no cost. In particular, a distribution π with $c^*(\pi) = 0$ can be interpreted as a distribution over temptations the DM can achieve without any self-discipline effort. Convexity means that the DM can randomize over self-discipline effort. For example, to induce the distribution $\alpha\pi + (1 - \alpha)\rho$ the DM could plan to toss a biased coin, exercising the self-discipline needed to induce π when the coin lands on heads and the self-discipline need to induce ρ when the coin lands on tails. The cost of $\alpha\pi + (1 - \alpha)\rho$ should then be bounded above by the expected cost of the randomization (i.e., $\alpha c(\pi) + (1 - \alpha)c(\rho)$); the inequality being strict if there is an alternative, less costly self-discipline effort that would also achieve the distribution $\alpha\pi + (1 - \alpha)\rho$ over temptations (without randomizing over π and ρ). Finally, monotonicity captures the intuitive idea that aligning temptations closer with her normative utility requires more self-discipline, and should therefore be costlier.

Example 4. [Costly dual-self (cont.)] To illustrate, consider the costly dual-self preference in Example 3. It is straightforward to see that the cost function c is grounded if and only if k is grounded; convex if and only if k is convex; and monotone if and only if k is increasing. Hence, Proposition 2 shows that it is without loss of generality to assume (i) the DM incurs no cost when the "alter ego" (utility v) selects a lottery for sure ($k(0) = 0$), (ii) the DM incurs a higher cost to increase the chances that her normative utility will prevail ($k(\beta)$ is increasing in β), and (iii) the "marginal cost" of self-discipline is increasing (k is convex). □

4.3 Comparative statics

As an application of the identification result in Theorem 2, we provide a behavioral measure of comparative self-discipline. Consider two DMs with self-discipline preferences $\succsim_1$ and $\succsim_2$. Intuitively, if self-discipline is costlier for DM2 than for DM1, DM2 should find commitment – which eliminates the need to exercise self-discipline – more valuable. Dekel and Lipman [2012] define a comparative behavior that formalizes when DM2 finds commitment more valuable than DM1 in terms of menu-choice data:

Definition 4. [Desire for commitment] Let $\succsim_1$ and $\succsim_2$ be binary relations on the set of menus $\mathcal{A}$. Then $\succsim_2$ has a *stronger desire for commitment* than $\succsim_1$ if, for all menus A and lotteries p,

$$\{p\} \succsim_1 A \;\Rightarrow\; \{p\} \succsim_2 A.$$

The following theorem shows that, in the context of the self-discipline model, this comparative characterizes when one DM has a higher cost of self-discipline than the other:

Theorem 3. *Let $\succsim_1$ and $\succsim_2$ be self-discipline preferences with canonical representations (u_1, c_1) and (u_2, c_2), respectively. Then the following are equivalent:*

(i) $\succsim_2$ *has a stronger desire for commitment than* $\succsim_1$.

(ii) $u_2 = u_1$ *and* $c_2 \geq c_1$.

Theorem 3 shows that a DM with higher costs of self-discipline values commitment more than an agent with lower self-discipline costs, establishing a comparative measure of self-discipline in terms of behavioral data.[6]

4.4 Ex-post choice

We conclude by highlighting some implications of self-discipline for ex-post choices from menus. To simplify, consider the costly Strotz model in Example 2, which is the special case of self-discipline that induces non-stochastic ex-post choice behavior. Specifically, for a costly Strotz preference represented by (u, k), let $\mathcal{C} : \mathcal{A} \to \mathcal{A}$, defined by

$$\mathcal{C}(A) = \bigcup_{v \in \mathcal{D}(A|u,k)} \arg\max_{p \in M_v(A)} u(p) \quad \forall A \in \mathcal{A},$$

denote the ex-post choice correspondence on menus induced by (u, k). Recall that a choice correspondence $\mathcal{C}$ on menus satisfies the *Weak Axiom of Revealed Preference*

[6]The restriction that $u_2 = u_1$ is required to compare the canonical costs, as these costs are measured in the same units as the normative utility.

(WARP) if

$$p, q \in A \cup B,\ p \in \mathcal{C}(A),\ q \in \mathcal{C}(B) \quad \Rightarrow \quad p \in \mathcal{C}(B)$$

and *vNM independence (IND)* if

$$\mathcal{C}(\alpha A + (1-\alpha)B) = \alpha\mathcal{C}(A) + (1-\alpha)\mathcal{C}(B) \quad \forall A, B \in \mathcal{A} \text{ and } \alpha \in (0,1).$$

The following examples illustrate that the ex-post choices induced by costly Strotz model can violate WARP and IND.

Example 5. [WARP] Consider three utilities $u, w, \tilde{w} \in \mathcal{V}$, where u represents the DM's normative ranking, and define the self-discipline cost function k by

$$k(v) = \begin{cases} 0 & \text{if } v \in \{w, \tilde{w}\} \\ \infty & \text{otherwise} \end{cases},$$

Now consider three lotteries $p, q, r \in P$ such that $u(p) > u(q) > u(r)$, while

$$w(r) > w(p) > w(q) \quad \text{and} \quad \tilde{w}(q) > \tilde{w}(r) > \tilde{w}(p).$$

In menu $\{p, q\}$, w solves the self-discipline choice problem, while in the menu $\{p, q, r\}$, $\tilde{w}$ solves the self-discipline choice problem. Hence, $\mathcal{C}(\{p, q\}) = \{p\}$ but $\mathcal{C}(\{p, q, r\}) = \{q\}$, violating WARP. $\square$

Example 5 illustrates that the costly Strotz model is consistent with well-known violations of WARP (see, e.g., Sippel [1997], or Echenique et al. [2011]).[7] The following example illustrates that the model can also induce violations of IND consistent, for example, with the common ratio effect observed in Allais [1953] and Kahneman and Tversky [1979]:[8]

Example 6. [Independence] Consider the same setting as Example 5, but define

[7]In a discrete-choice framework, a related example is discussed in Nehring [2006].

[8]Let $[\alpha; x]$ denote a lottery in which the DM receives \$$x$ with probability α, and \$0 otherwise, and let $p = [1; 3000]$, $q = [\frac{4}{5}; 4000]$, and $z = [1; 0]$. The common ratio effect is the finding that many subjects in experiments choose p from $\{p, q\}$, but also choose $0.25q + 0.75z$ from $\{0.25p + 0.75q, 0.25q + 0.75z\}$, consistent with the period 2 behavior in Example 6.

the self-discipline cost function k by

$$k(v) = \begin{cases} 0 & \text{if } v = w \\ \frac{1}{2}\left(u(p) - u(q)\right) & \text{if } v = u \\ \infty & \text{otherwise} \end{cases} .$$

In menu $\{p, q\}$, u solves the self-discipline choice problem, while w solves the self-discipline choice problem in menu $\frac{1}{4}\{p, q\} + \frac{3}{4}\{r\}$. Hence, $\mathcal{C}(\{p, q\}) = \{p\}$ but we have $\mathcal{C}\left(\frac{1}{4}\{p, q\} + \frac{3}{4}\{r\}\right) = \left\{\frac{1}{4}q + \frac{3}{4}r\right\}$, violating IND. $\square$

The examples illustrate that the costly Strotz model can provide a simple temptation-driven explanation for well-documented deviations from standard rationality postulates. The following proposition shows that these features of ex-post choice are directly related to violations of Set Independence for preferences over menus.

Proposition 3. *Let $\succsim$ be a costly Strotz preference that induces the choice correspondence $\mathcal{C}$. Then $\mathcal{C}$ satisfies WARP and IND if and only if $\succsim$ satisfies Set Independence.*

The idea that temptations can induce violations of WARP and IND has been observed before. For instance, the convex self-control model in Noor and Takeoka [2010] can induce ex-post violations of WARP and IND. However, the overlap of costly Strotz and convex self-control preferences satisfies Set Independence. Proposition 3 therefore illustrates that the endogenous choice of temptations in the costly Strotz model induces a novel source of deviations from WARP and IND. For instance, in Noor and Takeoka's [2010] example of the common ratio effect, the normative utility must exhibit *more* risk aversion than the temptation utility; while in Example 6 the normative utility exhibits *less* risk aversion than the temptation utility. Hence, the common ratio effect arises in their model when a DM controls her urge to avoid risks, while in our model the common ratio effect arises because the DM regulates her urge to seek risks.

5 Conclusion

In this paper, we propose a model of temptation-driven behavior, in which an agent can exercise costly self-discipline to align her temptations more closely with her normative objectives. In a menu-choice framework, we provide an axiomatic characterization that establishes testable implications of self-discipline. We also show that self-discipline cost can be elicited from choice-data, and provide a behavioral measure of comparative self-discipline.

Self-discipline preferences generalize many well-known models of temptation-driven behavior, such as the random Strotz, self-control, and multiple temptation models. Moreover, by incorporating incentive effects that impact a decision-maker's self-discipline choice problem, the model can accommodate novel behavioral implications of temptation and self-control problems. For example, in period 1, self-discipline induces a desire for early resolution of uncertainty, generating violations of Set Independence. In period 2, self-discipline induces menu-dependent temptation rankings, generating commonly observed violations of standard rationality postulates such as WARP and IND. The self-discipline model is therefore sufficiently general to rationalize a wide-range of temptation-driven behaviors, while having enough structure to identify behavioral meaningful parameters from choice-data.

References

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

Aliprantis, C. D. and K. C. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer.

Allais, M. (1953): "Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Postulats et Axiomes de l'École amÉricaine," *Econometrica: Journal of the Econometric Society*, 503–546.

Bénabou, R. and J. Tirole (2004): "Willpower and Personal Rules," *Journal of Political Economy*, 112, 848–886.

Benhabib, J. and A. Bisin (2005): "Modeling Internal Commitment Mechanisms and Self-control: A Neuroeconomics Approach to Consumption-Saving Decisions," *Games and Economic Behavior*, 52, 460–492.

Billingsley, P. (1995): *Probability and Measure*, John Wiley and Sons, Inc.

Chandrasekher, M. (2014): "Incomplete Menu Preferences and Ambiguity," Mimeo, Arizona State University.

Chatterjee, K. and R. V. Krishna (2009): "A "Dual Self" Representation for Stochastic Temptation," *American Economic Journal: Microeconomics*, 1, 148–167.

De Oliveira, H., T. Denti, M. Mihm, and K. Ozbek (2016): "Rationally Inattentive Preferences and Subjective Informaction Costs," Mimeo.

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

Dekel, E., B. L. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891–934.

——— (2009): "Temptation-Driven Preferences," *The Review of Economic Studies*, 76, 937–971.

Dunford, N. and J. T. Schwartz (1958): *Linear Operators: General Theory*, Interscience Publishers.

Echenique, F., S. Lee, and M. Shum (2011): "The Money Pump as a Measure of Revealed Preference Violations," *Journal of Political Economy*, 119, 1201–1223.

Ergin, H. and T. Sarver (2010): "A Unique Costly Contemplation Representation," *Econometrica*, 78, 1285–1339.

Fudenberg, D. and D. K. Levine (2006): "A Dual-Self Model of Impulse Control," *The American Economic Review*, 1449–1476.

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

——— (2006): "Random Expected Utility," *Econometrica*, 74, 121–146.

HERSTEIN, I. N. AND J. MILNOR (1953): "An Axiomatic Approach to Measurable Utility," *Econometrica*, 291–297.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica: Journal of the Econometric Society*, 263–291.

KREPS, D. AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

LUCE, R. D. AND P. SUPPES (1965): *Preference, utility, and subjective probability*, Wiley and Sons.

MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2004): "Variational Representation of Preferences under Ambiguity," *International Center for Economic Research Working Papers*, 5/2004.

——— (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, 1447–1498.

MACHINA, M. J. (1984): "Temporal Risk and the Nature of Induced Preferences," *Journal of Economic Theory*, 33, 199–231.

NEHRING, K. (2006): "Self-Control through Second Order Preferences," Mimeo.

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

NOOR, J. AND N. TAKEOKA (2010): "Uphill Self-Control," *Theoretical Economics*, 5, 127–158.

NORMANN, D. A. AND T. SHALLICE (2000): "Attention to Action: Willed and Automatic Control of Behavior," in *Cognitive Neuroscience: A Reader*, Oxford: Blackwell.

Olszewski, W. (2007): "Preferences Over Sets of Lotteries," *The Review of Economic Studies*, 74, 567–595.

Sippel, R. (1997): "An Experiment on the Pure Theory of Consumer's Behaviour," *The Economic Journal*, 107, 1431–1444.

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

Thaler, R. H. and H. M. Shefrin (1981): "An Economic Theory of Self-Control," *The Journal of Political Economy*, 392–406.

A Appendix

A.1 Preliminaries

Let $B(\Sigma)$ be the set of bounded Σ-measurable functions mapping $\mathcal{V}$ to $\mathbb{R}$. When endowed with the supnorm, $B(\Sigma)$ is a Banach space. The topological dual of $B(\Sigma)$ is the space $ba(\Sigma)$ of all bounded and finitely-additive set functions $\mu : \Sigma \to \mathbb{R}$, the duality being

$$\langle \varphi, \mu \rangle = \int_{\mathcal{V}} \varphi(v)\, \mu(dv)$$

for all $\varphi \in B(\Sigma)$ and all $\mu \in ba(\Sigma)$ (see, e.g., Dunford and Schwartz [1958, p. 258]). For $\varphi, \psi \in B(\Sigma)$, we write $\varphi \geq \psi$ if $\varphi(v) \geq \psi(v)$ for all $v \in \mathcal{V}$.

Let Φ be a non-empty subset of $B(\Sigma)$, and Φ_c be the constant functions in Φ. Set Φ is called a *tube* if $\Phi = \Phi + \mathbb{R}$. A functional $I : \Phi \to \mathbb{R}$ is

(i) *normalized* if $I(k) = k$ for all $k \in \Phi_c$,[9]

(ii) *monotone* if $\varphi \geq \psi$ implies $I(\varphi) \geq I(\psi)$ for all $\varphi, \psi \in \Phi$,

(iii) *translation invariant* if $I(\alpha\varphi + (1-\alpha)k) = I(\alpha\varphi) + (1-\alpha)k$ for all $\varphi \in \Phi$, $k \in \Phi_c$, and $\alpha \in [0,1]$, such that $\alpha\varphi, \alpha\varphi + (1-\alpha)k \in \Phi$,

[9] We abuse notation by writing k for $k\mathbf{1}$, where it is obvious from the context.

(iv) *vertically invariant* if $I(\varphi + c) = I(\varphi) + c$ for all $\varphi \in \Phi$ and $c \in \mathbb{R}$ such that $\varphi + c \in \Phi$,

(v) a *niveloid* if $I(\varphi) - I(\psi) \leq \sup_{v \in \mathcal{V}}(\varphi(v) - \psi(v))$ for all $\varphi, \psi \in \Phi$.

For notational convenience, we denote $\alpha A + (1-\alpha)B$ by $A[\alpha]B$ for $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$.

Let P^o be the interior of P (i.e., the set of lotteries with full support), and $\mathcal{A}^o \subset \mathcal{A}$ the collection of non-empty closed subsets of P^o. Denote by $\bar{p} = (1/n, ..., 1/n)$ the uniform distribution over X.

For $u \in \mathcal{V}$ and $A \in \mathcal{A}$, define $\varphi_A^u : \mathcal{V} \to \mathbb{R}$ by

$$\varphi_A^u(v) = \max_{p \in M_v(A)} u(p) \quad \forall v \in \mathcal{V}. \tag{3}$$

When u is clear from the context, we omit the superscript u. By the Maximum Theorem (see, e.g., Aliprantis and Border [2006, pp. 569–570]), φ_A is an upper semicontinuous function taking values in $K = [u_*, u^*]$, where $u_* = \min_{p \in P} u(p)$ and $u^* = \max_{p \in P} u(p)$. Upper semicontinuous functions are Σ-measurable (Billingsley [1995, pp. 184–186]). As a result, $\varphi_A \in B(\Sigma, K)$, where $B(\Sigma, K)$ denotes the functions in $B(\Sigma)$ assuming values in K. Let $\Phi = \{\varphi_A : A \in \mathcal{A}\}$ and $\Phi^o = \{\varphi_A : A \in \mathcal{A}^o\}$. Clearly $0 \in \Phi^o$ and $\Phi^o \subseteq \Phi$. Moreover, since $\varphi_{A[\alpha]B} = \alpha\varphi_A + (1-\alpha)\varphi_B$ for any $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$, both Φ and Φ^o are convex sets. It is straightforward to show that$\varphi_A = \varphi_{co(A)}$ for all $A \in \mathcal{A}$.

A.2 Lemmas

In this Section, we state and prove three lemmas that are essential to establish our results. The first lemma characterizes an important dominance relation on menus. The second lemma provides a representation for a binary relation satisfying Axioms 1–4. The third lemma establishes that there is a common solution to the self-discipline choice problem for a collection of menus if and only if the DM is indifferent to the timing of the resolution of uncertainty for these menus.

A.2.1 A dominance relation on menus

In the following, fix some $u \in \mathcal{V}$.

For $A \in \mathcal{A}$ and $p, q \in P$ such that $u(p) \geq u(q)$, let $A_p = \{r \in co(A) : u(r) = u(p)\}$ and $A_{p,q} = \{r \in co(A) : u(p) \geq u(r) \geq u(q)\}$. Now define a partial order $\trianglerighteq_u^*$ on $\mathcal{A}$ by $A \trianglerighteq_u^* B$ if

$$u(p) > u(q) \quad \Rightarrow \quad A_p[\frac{1}{2}]B_q \supseteq B_p[\frac{1}{2}]A_q.$$

There is a "duality" between the order $\trianglerighteq_u$ on $\Delta(\mathcal{V})$ and the order $\trianglerighteq_u^*$ on $\mathcal{A}$:

(i) For $\pi, \rho \in \Delta(\mathcal{V})$, $\pi \trianglerighteq_u \rho$ if and only if $\langle \varphi_A, \pi \rangle \geq \langle \varphi_A, \rho \rangle$ for all $A \in \mathcal{A}$.

(ii) For $A, B \in \mathcal{A}$, $A \trianglerighteq_u^* B$ if and only if $\langle \varphi_A, \pi \rangle \geq \langle \varphi_B, \pi \rangle$ for all $\pi \in \Delta(\mathcal{V})$.

Part (i) follows from Theorem 4 and Lemma 6 in Dekel and Lipman [2012]. Part (ii) follows directly from the following lemma.

Lemma 1. *For all $A, B \in \mathcal{A}$, $A \trianglerighteq_u^* B$ if and only if $\varphi_A \geq \varphi_B$.*

Proof. **[Sufficiency]:** Let $A, B \in \mathcal{A}$ such that $A \trianglerighteq_u^* B$.

First, suppose $A = A_p \cup A_q$ and $B = B_p \cup B_q$ for two lotteries $p, q \in P$ with $u(p) > u(q)$, where A_p, A_q, B_p and B_q are non-empty. Fix any $v \in \mathcal{V}$, and let $r \in M_v(A_p)$, $r' \in M_v(A_q)$, $s \in M_v(B_p)$, and $s' \in M_v(B_q)$. Since $A_p[\frac{1}{2}]B_q \supseteq B_p[\frac{1}{2}]A_q$, $\frac{1}{2}v(r) + \frac{1}{2}v(s') \geq \frac{1}{2}v(s) + \frac{1}{2}v(r')$, and so $v(r) - v(r') \geq v(s) - v(s')$. Hence, $\varphi_A(v) \geq \varphi_B(v)$. Since v was arbitrary, $\varphi_A \geq \varphi_B$.

Now consider arbitrary $A, B \in \mathcal{A}$. Fix $v \in \mathcal{V}$, and let $r \in \arg\max_{p \in M_v(A)} u(p)$ and $s \in M_v(B)$. By way of contradiction, suppose $u(s) > u(r)$. Then $A_s[\frac{1}{2}]B_r \supseteq B_s[\frac{1}{2}]A_r$. Since $r \in A_r$ and $s \in B_s$, both of these sets are non-empty. Since $A_s[\frac{1}{2}]B_r \supseteq B_s[\frac{1}{2}]A_r$, it follows that A_s and B_r are non-empty. Hence, by the argument in the previous paragraph, $\varphi_{A_s \cup A_r} \geq \varphi_{B_s \cup B_r}$, and so there exists $s' \in M_v(A_s \cup A_r) \cap A_s$, implying $s' \in M_v(co(A))$. Since $u(s') > u(r)$, this contradicts $r \in \arg\max_{p \in M_v(A)} u(p)$. Hence, $\varphi_A \geq \varphi_B$.

[Necessity]: Let $A, B \in \mathcal{A}$ such that $\varphi_A \geq \varphi_B$.

Let $p, q \in P$ such that $u(p) > u(q)$. The following steps show $A_p[\frac{1}{2}]B_q \supseteq B_p[\frac{1}{2}]A_q$.

Step 1: If $u(w_A) = \min_{r \in A} u(r) \geq \max_{s \in B} u(s) = u(b_B)$, then the claim is trivially true. Thus, assume that $u(b_B) > u(w_A)$. Since $u, -u \in \mathcal{V}$, we must have $u(b_A) = \max_{r \in A} u(r) \geq u(b_B)$ and $u(w_A) \geq \min_{s \in B} u(s) = u(w_B)$, and so if $u(p) > u(b_B)$ or $u(w_A) > u(q)$, the claim holds easily. Hence, assume $u(b_B) \geq u(p) > u(q) \geq u(w_A)$.

Step 2: We now argue that $\varphi_{A_{p,q}} \geq \varphi_{B_{p,q}}$. To see this, let $v \in \mathcal{V}$ such that $\varphi_B(v) = u(r)$ for some $r \in M_v(B)$. First, if $u(r) > u(p)$, then for any $s \in co(B)$ with $u(p) > u(s)$, there exists some $\alpha \in (0,1)$ satisfying $\alpha r + (1-\alpha)s \in B_p$. Hence, $v \cdot [\alpha r + (1-\alpha)s] \geq v \cdot s$ implying $M_v(B_{p,q}) \subset B_p$. The same argument applies for A yielding $M_v(A_{p,q}) \subseteq A_p$, and so $\varphi_{A_{p,q}}(v) = \varphi_{B_{p,q}}(v)$. Second, if $u(p) \geq u(r) \geq u(q)$, then clearly $\varphi_{B_{p,q}}(v) = \varphi_B(v)$ and $\varphi_{A_{p,q}}(v) = \varphi_A(v)$, and so $\varphi_{A_{p,q}}(v) \geq \varphi_{B_{p,q}}(v)$ since we have $\varphi_A \geq \varphi_B$. Finally, if $u(q) > u(r)$, then for any $s \in B$ with $u(s) > u(q)$, there exists some $\alpha \in (0,1)$ satisfying $\alpha r + (1-\alpha)s \in B_q$. Hence, for any $s \in B$ with $u(s) > u(q)$, $v \cdot [\alpha r + (1-\alpha)s] \geq v \cdot s$ implying $M_v(B_{p,q}) \subset B_q$ and so $\varphi_{A_{p,q}}(v) \geq \varphi_{B_{p,q}}(v)$. Thus, we have $\varphi_{A_{p,q}} \geq \varphi_{B_{p,q}}$.

Step 3: We now show that $\varphi_{A_p \cup A_q} \geq \varphi_{B_p \cup B_q}$. Fix some $w \in \mathcal{W} = \{v \in \mathcal{V} \,:\, v \cdot u = 0\}$ and notice that there is a unique $\bar{\alpha} \in (-1,1)$ such that $\bar{v} = \bar{\alpha} u + \left(\sqrt{1-\bar{\alpha}^2}\right) w \in \mathcal{V}$ satisfies $M_{\bar{v}}(A_p \cup A_q) = M_w(A_p) \cup M_w(A_q)$. That is, $\bar{v}(r) = \bar{v}(s)$ for all $r \in M_w(A_p)$ and $s \in M_w(A_q)$. We claim that $M_{\bar{v}}(B_p \cup B_q) \cap M_w(B_q) \neq \emptyset$. That is, $\bar{v}(r) \leq \bar{v}(s)$ for all $r \in M_w(B_p)$ and $s \in M_w(B_q)$. Assume, for contradiction, that this is not true. For any $r \in P$, let $r_{w,u} = (u \cdot r) \cdot u + (w \cdot r) \cdot w \in \mathbb{R}^n$ denote the projection of r onto the space spanned by u and w in $\mathbb{R}^n$. Note that for any $E \in \mathcal{A}$ and $r \in P$, all points in $M_w(E_r)$ are projected onto the same point $r^E_{w,u} \in \mathbb{R}^n$. Let $M_{w,u}(E_r) = \{r^E_{w,u} \in \mathbb{R}^n\}$ and let $M_{w,u}(E_{s,t}) = \bigcup_{u(s) \geq u(r) \geq u(t)} M_{w,u}(E_r)$ for any $E \in \mathcal{A}$ and $s, t \in P$. Note that by assumption we have $\bar{v}(p^B_{w,u}) > \bar{v}(q^B_{w,u})$ and $\bar{v}(p^A_{w,u}) = \bar{v}(q^A_{w,u})$. Now, without loss of generality, assume that

$$\bar{v}(p^B_{w,u}) > \bar{v}(p^A_{w,u}) = \bar{v}(q^A_{w,u}) > \bar{v}(q^B_{w,u})$$

Define a function $f : [0,1] \to \mathbb{R}$ such that $f(a) = w(r[a]^A_{w,u}) - w(r[a]^B_{w,u})$ where $r[a] = ap + (1-a)q \in P$ for all $a \in [0,1]$. Observe that f is a continuous function from $[0,1]$ into $\mathbb{R}$ such that $f(0) > 0$ and $f(1) < 0$. Hence, by the Intermediate

Value Theorem, there is some $a \in (0,1)$ satisfying $f(a) = 0$. That is, there is some $a \in (0,1)$ such that $w(r[a]^A_{w,u}) - w(r[a]^B_{w,u}) = 0$ implying $r[a]^A_{w,u} = r[a]^B_{w,u}$. Therefore, the set $F = M_{w,u}(A_{p,q}) \cap M_{w,u}(B_{p,q})$ is non-empty. Moreover, it can be easily verified that F is a closed (and convex) set. Let $r[a^*]$ be the unique element of $M_u(F)$. Note that $a^* \in (0,1)$; that is, $u(p) > u(r[a^*]) > u(q)$ since $p^A_{w,u} \neq p^B_{w,u}$ and $q^A_{w,u} \neq q^B_{w,u}$.

For all $a \in (a^*, 1]$, let $\alpha^A(a) \in (-1,1)$ be the unique number such that the vector $v^A_a = \alpha^A(a)u + \left(\sqrt{1-(\alpha^A(a))^2}\right) w \in \mathcal{V}$ satisfies $v^A_a(r[a^*]) = v^A_a(r[a]^A_{w,u})$. Note that $\alpha^A(a)$ is a monotonically decreasing upper semicontinuous function. Hence $\alpha^*_A = \lim_{a \to a^*} \alpha^A(a)$ is well defined. Similarly, for all $a \in (a^*, 1]$ let $\alpha^B(a) \in (-1,1)$ such that $v^B_a = \alpha^B(a)u + \left(\sqrt{1-(\alpha^B(a))^2}\right) w \in \mathcal{V}$ satisfies $v^B_a(r[a^*]) = v^B_a(r[a]^B_{w,u})$ and define $\alpha^*_B = \lim_{a \to a^*} \alpha^B(a)$. Observe that it must be $-1 < \alpha^*_B \leq \alpha^*_A < 1$. Moreover since $r[a^*]$ is the unique element of $M_u(F)$, without loss of generality, we can assume that $\alpha^*_B < \alpha^*_A$.

Pick some $\hat{\alpha} \in (\alpha^*_B, \alpha^*_A)$ and let $\hat{v} = \hat{\alpha}u + \left(\sqrt{1-\hat{\alpha}^2}\right) w \in \mathcal{V}$. Note that since $\alpha^*_B < \hat{\alpha}$ and $\alpha^B(a)$ is decreasing, there must be some $\hat{r}_B \in M_{\hat{v}}(M_{w,u}(B_{p,q}))$ such that $u(\hat{r}_B) > u(r[a^*])$. Moreover, since $\hat{\alpha} < \alpha^*_A \leq \alpha^A(a)$ for all $a \in (a^*, 1]$, we have $u(r[a^*]) \geq u(\hat{r}_A)$ for all $\hat{r}_A \in M_{\hat{v}}(M_{w,u}(A_{p,q}))$. Combining these, we deduce $\varphi_{A_{p,q}}(\hat{v}) < \varphi_{B_{p,q}}(\hat{v})$, a contradiction. Hence, it must be $M_{\bar{v}}(B_p \cup B_q) \cap M_w(B_q) \neq \emptyset$.

Note that any $v \in \mathcal{V} \backslash \{u, -u\}$ can be uniquely given as $v = v^\alpha_{u,w} = \alpha u + \left(\sqrt{1-\alpha^2}\right) w$ for some $w \in \mathcal{W}$ and some $\alpha \in (-1,1)$. Moreover, for any $E \in \mathcal{A}$, and for any $v^\alpha_{u,w}, v^\beta_{u,w} \in \mathcal{V}$ with $w \in W$ and $\alpha, \beta \in (-1,1)$, we have $M_{v^\alpha_{u,w}}(E_s \cup E_t) \cap M_w(E_t) = \emptyset$ implies $M_{v^\beta_{u,w}}(E_s \cup E_t) \cap M_w(E_t) = \emptyset$ whenever $s, t \in P$ with $u(s) > u(t)$ and $\alpha < \beta$. Therefore, given that $M_{\bar{v}}(A_p \cup A_q) \cap M_w(A_q) \neq \emptyset$ implies $M_{\bar{v}}(B_p \cup B_q) \cap M_w(B_q) \neq \emptyset$, for any $v = \alpha u + \left(\sqrt{1-\alpha^2}\right) w \in \mathcal{V}$ with $w \in W$ and $\alpha \in (-1,1)$, we must have $M_v(B_p \cup B_q) \cap M_w(B_q) = \emptyset$ implies $M_v(A_p \cup A_q) \cap M_w(A_q) = \emptyset$. Thus, we obtain $\varphi_{A_p \cup A_q} \geq \varphi_{B_p \cup B_q}$.

Step 4: Finally let $\bar{r} = \frac{1}{2}p + \frac{1}{2}q$, and let $C_{\bar{r}} = A_p[\frac{1}{2}]B_q$, $C_q = A_q[\frac{1}{2}]B_q$, and $D_{\bar{r}} = A_q[\frac{1}{2}]B_p$. We want to show that $C_{\bar{r}} \supseteq D_{\bar{r}}$. Note that since $\varphi_{A_p \cup A_q} \geq \varphi_{B_p \cup B_q}$ we have $\varphi_{C_{\bar{r}} \cup C_q} = \frac{1}{2}\varphi_{A_p \cup A_q} + \frac{1}{2}\varphi_{B_q} \geq \varphi_{B_p \cup B_q} + \frac{1}{2}\varphi_{A_q} = \varphi_{D_{\bar{r}} \cup C_q}$. Assume for contradiction that there exists some $s \in D_{\bar{r}} \setminus C_{\bar{r}}$. Then $C_{\bar{r}}$ and $E = co(\{s\} \cup C_q)$ are both closed and convex sets in $\mathbb{R}^n$ with $C_{\bar{r}} \cap E = \emptyset$. Hence, by a Strong

Separating Hyperplane Theorem (see, e.g., Dunford and Schwartz [1958, p. 417] or Aliprantis and Border [2006, p. 207]), there exists some $v \in \mathcal{V}$ and $k \in \mathbb{R}$ such that $v \cdot e > k$ for all $e \in E$ and $v \cdot f < k$ for all $f \in C_{\bar{r}}$. If $v \cdot s \geq v \cdot e$ for all $e \in C_q$, then we must have $\varphi_{D_{\bar{r}} \cup C_q}(v) = u \cdot s > u \cdot q = \varphi_{D_{\bar{r}} \cup C_q}(v)$, a contradiction. Thus, assume that we have $\max_{e \in C_q} v \cdot e > v \cdot s$ and let $c \in C_q$ such that $v \cdot c \geq v \cdot e$ for all $e \in C_q$. Now define $\alpha = \frac{u \cdot s - u \cdot c}{(v \cdot c - v \cdot s) + (u \cdot s - u \cdot c)} \in (0, 1)$ and $w = \alpha v + (1 - \alpha) u$. Then we have $w \cdot s = w \cdot c \geq w \cdot e$ for all $e \in C_q$. On the other hand, $w \cdot c = w \cdot s > w \cdot f$ for all $f \in C_{\bar{r}}$. Thus, we must have $\varphi_{D_{\bar{r}} \cup C_q}(v_w) = u \cdot s > u \cdot q = \varphi_{D_{\bar{r}} \cup C_q}(v_w)$, a contradiction. To summarize, there cannot be any $s \in D_{\bar{r}} \setminus C_{\bar{r}}$, and so $A_p[\frac{1}{2}]B_q \supseteq B_p[\frac{1}{2}]A_q$. □

A.2.2 Implications of Axioms 1–4

The following lemma shows how the self-discipline representation is obtained from Axioms 1–6.

Lemma 2. *Let $\succsim$ be a binary relation on $\mathcal{A}$ that satisfies Axioms 1–6. Then:*

(i) There exists $u \in \mathcal{V}$ such that, for all $p, q \in P$, $u(p) \geq u(q)$ if and only if $\{p\} \succsim \{q\}$.

(ii) Every menu $A \in \mathcal{A}$ has a singleton equivalent $p_A \in P$.

(iii) The function c^* defined on $\Delta(\mathcal{V})$ by

$$c^*(\pi) = \sup_{A \in \mathcal{A}} \left(\langle \varphi_A^u, \pi \rangle - u(p_A) \right) \quad \forall \pi \in \Delta(\mathcal{V})$$

is non-negative lower-semicontinuous and proper.

(iv) The functional $U : \mathcal{A} \to \mathbb{R}$, defined by

$$U(A) = \max_{\pi \in \Delta(\mathcal{V})} \left(\langle \varphi_A^u, \pi \rangle - c^*(\pi) \right) \quad \forall A \in \mathcal{A},$$

represents $\succsim$.

Proof. Let $\succsim$ be a binary relation on $\mathcal{A}$ that satisfies Axioms 1–6.

[Part (i)]: Let $p, q \in P$ and assume that $\{p\} \sim \{q\}$. By Axiom 5, we have $\{p\} \succsim \{q\}[\frac{1}{2}]\{p\}$. This implies $\{q\}[\frac{1}{2}]\{p\} \succsim \{q\}$ by Axiom 6, and so we must have $\{p\} \sim \{q\}[\frac{1}{2}]\{p\}$. By Axiom 6 again, we have $\{p\}[\frac{1}{2}]\{r\} \sim \{q\}[\frac{1}{2}]\{r\}$ for any $r \in P$. Hence, by Herstein and Milnor [1953], there exists $u \in \mathcal{V}$ representing the commitment ranking.

[Part (ii)]: Since A is non-empty and compact, and u is a continuous function on P, there exist some $b, w \in A$ such that $\{b\} \succsim \{p\} \succsim \{w\}$ for all $p \in A$. Clearly $\{b\} \succsim A \succsim \{w\}$ by Axiom 4, and so by Axiom 2, the following (non-empty) sets, whose union is equal to $[0, 1]$, must be closed:

$$\{\alpha \in [0,1] : \{b\}[\alpha]\{w\} \succsim A\} \quad \text{and} \quad \{\alpha \in [0,1] : A \succsim \{b\}[\alpha]\{w\}\}\,.$$

Since $[0, 1]$ is a connected set, the two subsets above must intersect; that is, there must exist some $\alpha \in [0, 1]$ such that $A \sim \{b\}[\alpha]\{w\}$. Let $p_A \in P$ be equal to $\alpha b + (1 - \alpha)w$. Finally, note that if $A \in \mathcal{A}^o$, then $b, w \in P^o$, and so $p_A \in P^o$.

[Part (iii)]: For any $\pi \in \Delta(\mathcal{V})$, $\left\langle \varphi_{\{p\}}, \pi \right\rangle - u(p) = u(p) - u(p) = 0$, and so c^* is non-negative. Since c^* is the supremum of continuous functions, it is lower semicontinuous. Finally, for any $A \in \mathcal{A}$, $\langle \varphi_A, \delta_{-u} \rangle = u(w_A)$. Since $u(r) \succsim u(w_A)$ for all $r \in A$, it follows that $A_p[\frac{1}{2}]\{w_A\}_q \supseteq \{w_A\}_p[\frac{1}{2}]A_q$ for all $p, q \in P$ such that $\{p\} \succ \{q\}$. To see why, note that $\{w_A\}_p \neq \emptyset$ if and only if $\{p\} \sim \{w_A\}$, but then $A_q = \emptyset$. Hence, by Axiom 4, $A \succsim \{w_A\}$ and so $u(p_A) \geq u(w_A)$. It follows that $\langle \varphi_A, \delta_{-u} \rangle - u(p_A) \leq 0$ for all $A \in \mathcal{A}$, and so $c^*(\delta_{-u}) = 0$. Hence, c^* is proper.

[Part (iv)]: To establish the desired representation, we show that there is a normalized convex niveloid $I : \Phi \to \mathbb{R}$ such that, for all menus A and B, $A \succsim B$ if and only if $I(\varphi_A) \geq I(\varphi_B)$. Following the approach in Maccheroni et al. [2006], an application of Fenchel-Moreau duality then establishes $I(\varphi_A) = \max_{\pi \in \Delta(\mathcal{V})} (\langle \varphi_A, \pi \rangle - c^*(\pi))$ for all $A \in \mathcal{A}$. The key step in the proof is establishing the functional I which satisfies the prerequisite properties. For technical reasons, we start by defining a functional I^o on $\mathcal{A}^o$, and then use Axiom 2 to extend the functional to $\mathcal{A}$.

Define the functional $I^o : \Phi^o \to \mathbb{R}$ by $I^o(\varphi_A) = u(p_A)$ for all $A \in \mathcal{A}$. If p_A and q_A are two singleton equivalents of A, then $u(p_A) = u(q_A)$. Hence, I^o is well-defined. Moreover, for any menus $A, B \in \mathcal{A}^o$ with singleton equivalents p_A and p_B, $A \succsim B$

if and only if $\{p_A\} \succsim \{p_B\}$ if and only if $u(p_A) \geq u(p_B)$. Hence, $I^o(\varphi_A) \geq I^o(\varphi_B)$ if and only if $A \succsim B$.

We start by establishing that I^o is a normalize convex niveloid. We proceed in steps. Step 1 shows that I^o is normalized. Step 2 uses Axiom 5 to show I^o is convex. Step 3 uses Axiom 6 to show I^o is translation invariant. Step 4 uses a result in Maccheroni et al. [2004] to show that I^o is vertically invariant. Step 5 uses Axioms 4 and 3, and Lemma 1 to show that I^o is monotone. Finally, Step 6 applies a number of results in Maccheroni et al. [2004] to establish that I^o is a niveloid.

Step 1: I^o is normalized.

Let $k \in \mathbb{R}$ such that $k \in \Phi^o$. That means there is a lottery $p \in P^o$ such that $k = \varphi_{\{p\}} = u(p)$. Hence, $I^o(k) = I^o(\varphi_{\{p\}}) = u(p) = k$.

Step 2: I^o is convex.

Let $A, B \in \mathcal{A}^o$ and $\alpha \in [0,1]$. Note that $A[\alpha]B \in \mathcal{A}^o$ and so $\varphi_{A[\alpha]B} \in \Phi^o$. By part (ii) there exist some $p_A, p_B \in P^o$ such that $\{p_A\} \sim A$ and $\{p_B\} \sim B$. By Axiom 5, we have $\{p_A\}[\alpha]\{p_B\} \succsim A[\alpha]B$, and so

$$\begin{aligned}
\alpha I^o(\varphi_A) + (1-\alpha) I^o(\varphi_B) &= \alpha u(p_A) + (1-\alpha) u(p_B) \\
&= I^o(\varphi_{\{p_A\}[\alpha]\{p_B\}}) \\
&\geq I^o(\varphi_{A[\alpha]B}) \\
&= I^o(\alpha\varphi_A + (1-\alpha)\varphi_B).
\end{aligned}$$

Step 3: I^o is translation invariant.

Let $A \in \mathcal{A}^o$, $p \in P$ with $u(p) = k$, and $\alpha \in (0,1)$. Let $b, w \in A$ such that $\{b\} \succsim \{q\} \succsim \{w\}$ for all $q \in A$. By part (i), $\{b\}[\alpha]\{\bar{p}\} \succsim \{q\}[\alpha]\{\bar{p}\} \succsim \{w\}[\alpha]\{\bar{p}\}$ for all $q \in A$, and so by Axiom 4 $\{b\}[\alpha]\{\bar{p}\} \succsim A[\alpha]\{\bar{p}\} \succsim \{w\}[\alpha]\{\bar{p}\}$. The argument used in the proof of Claim part (ii) yields a $\beta \in [0,1]$ such that

$$(\{b\}[\alpha]\{\bar{p}\})\,[\beta]\,(\{w\}[\alpha]\{\bar{p}\}) \sim A[\alpha]\{\bar{p}\}.$$

Hence, $r = \beta b + (1-\beta) w \in P^o$ satisfies $\{r\}[\alpha]\{\bar{p}\} \sim A[\alpha]\{\bar{p}\}$. By Axiom 6, it follows that $\{r\}[\alpha]\{p\} \sim A[\alpha]\{p\}$, and so

$$\begin{aligned}
I^o(\alpha\varphi_A + (1-\alpha)k) &= I^o(\varphi_{A[\alpha]\{p\}}) = I^o(\varphi_{\{r\}[\alpha]\{p\}}) \\
&= \alpha u(r) + (1-\alpha)u(p) = \alpha u(r) + (1-\alpha)k \\
&= \alpha u(r) + (1-\alpha)u(\bar{p}) + (1-\alpha)k \\
&= I^o(\varphi_{\{r\}[\alpha]\{\bar{p}\}}) + (1-\alpha)k \\
&= I^o(\varphi_{A[\alpha]\{\bar{p}\}}) + (1-\alpha)k \\
&= I^o(\alpha\varphi_A + (1-\alpha)\varphi_{\{\bar{p}\}}) + (1-\alpha)k \\
&= I^o(\alpha\varphi_A) + (1-\alpha)k,
\end{aligned}$$

establishing that I^o is translation invariant.

Step 4: I^o is vertically invariant.

The result follows from Step 1 of the proof of Lemma 20 in Maccheroni et al. [2004] once we show that for all $A \in \mathcal{A}^o$ and $k \in \mathbb{R}$ such that $\varphi_A + k \in \Phi^o$, there exists some $\alpha \in (0,1)$ satisfying $\frac{\varphi A}{\alpha}, \frac{\varphi_A+k}{\alpha} \in \Phi^o$. To see this, let $p^\theta = \theta p + (1-\theta)\bar{p}$ for any $p \in P$ and $\theta > 0$ and note that for any given $p \in P^o$ there exists some $\theta > 1$ such that $p^\theta \in P^o$. Clearly if $p^\theta \in P^o$, then $p^{\theta'} \in P^o$ for any $\theta' < \theta$.

Since $A \in \mathcal{A}^o$ is a finite set, there must exist some $\theta > 1$ such that $p^\theta \in P^o$ for all $p \in A$. Pick any such $\theta > 1$, and call it θ_A. Since $\varphi_A + k \in \Phi^o$, there must exist some $B \in \mathcal{A}^o$ such that $\varphi_B = \varphi_A + k$. Similarly, $B \in \mathcal{A}^o$ is a finite set, and so there must exist some $\theta > 1$ such that $p^\theta \in P^o$ for all $p \in B$. Pick any such $\theta > 1$, and call it θ_B and let $\theta_* = \min\{\theta_A, \theta_B\}$.

Let $A^{\theta_*} = \{p^{\theta_*} : p \in A\} \in \mathcal{A}^o$ and $B^{\theta_*} = \{p^{\theta_*} : p \in B\} \in \mathcal{A}^o$. Observe that $v \cdot p^\theta = \theta(v \cdot p)$ for any $v \in \mathcal{V}$, $p \in P$, and $\theta > 0$. Therefore, we have $\varphi_{A^{\theta_*}} = \theta_* \cdot \varphi_A \in \Phi^o$ and $\varphi_{B^{\theta_*}} = \theta_* \cdot \varphi_B \in \Phi^o$. Let $\alpha = 1/\theta_*$. We have shown $\frac{\varphi_A}{\alpha}, \frac{\varphi_A+k}{\alpha} \in \Phi^o$ as desired. Hence, by Maccheroni et al. [2004, Lemma 20], I^o is vertically invariant.

Step 5: I^o is monotone.

Let $A, B \in \mathcal{A}^o$ such that $\varphi_A \geq \varphi_B$. Since $\varphi_A = \varphi_{co(A)}$ and $\varphi_B = \varphi_{co(B)}$, it follows that $\varphi_{co(A)} \geq \varphi_{co(B)}$. Moreover, it follows immediately from Lemma 1, that

$\varphi_{co(A)} \geq \varphi_{co(B)}$ implies $A \trianglerighteq B$. Hence, by Axiom 4, $co(A) \succsim co(B)$. As a result, Axiom 3 implies $I^o(\varphi_A) \geq I^o(\varphi_B)$.

Step 6: I^o is a niveloid.

Since I^o is vertically invariant, the functional $I^* : \Phi^o + \mathbb{R} \to \mathbb{R}$, defined by $I^*(\varphi + k) = I^o(\varphi) + k$ for all $\varphi \in \Phi^o$, is the unique vertically invariant extension of I^* to the tube generated by Φ^o (Maccheroni et al. [2004, Lemma 22]). Moreover, since Φ^o is a convex set and I^o is a convex functional, the obvious adaption of the arguments in Maccheroni et al. [2004] establishes that I^* is also convex. We now show that I^* must also be monotone. By the first paragraph in the proof of Lemma 24 in Maccheroni et al. [2004], it is sufficient to show that if $\varphi, \psi \in \Phi^o$ and $\varphi + k \geq \psi$, then $I^*(\varphi + k) \geq I^*(\psi)$.

Let $A, B \in \mathcal{A}^o$ and $k \in \mathbb{R}$ such that $\varphi_A + k \geq \varphi_B$. Clearly there exists $\alpha \in (0, 1)$ such that

$$\alpha(\varphi_A + k) + (1 - \alpha)\varphi_B = \alpha\varphi_A + (1 - \alpha)\varphi_B + \alpha k \in \Phi^o.$$

Moreover, since $\varphi_A + k \geq \varphi_B$, $\alpha(\varphi_A + k) + (1 - \alpha)\varphi_B \geq \varphi_B$. Now assume, for contradiction, that $I^*(\varphi_A + k) < I^*(\varphi_B)$. Since I^* is convex, this would imply

$$\begin{aligned} I^o(\varphi_B) &= \alpha I^*(\varphi_B) + (1 - \alpha)I^*(\varphi_B) \\ &> \alpha I^*(\varphi_A + k) + (1 - \alpha)I^*(\varphi_B) \\ &\geq I^*(\alpha(\varphi_A + k) + (1 - \alpha)\varphi_B). \\ &= I^o(\alpha(\varphi_A + k) + (1 - \alpha)\varphi_B) \end{aligned}$$

which contradicts that I^o is monotone, and thus I^* must be monotone.

Since I^o is vertically invariant, and its unique vertically invariant extension to the tube generated by Φ^o, I^*, is monotone, I^o is a niveloid by Lemma 23 in Maccheroni et al. [2004]. In sum, we have shown that I^o is a normalized convex niveloid.

We now extend I^o to Φ. For any menu $A \in \mathcal{A}$ and number $m \in \mathbb{N}$, define $A^m = A[\frac{m-1}{m}]\{\bar{p}\}$ and denote $\varphi_A^m = \varphi_{A^m}$. Note that for all $A \in \mathcal{A}$ and $m \in \mathbb{N}$, $A^m \in \mathcal{A}^o$ and $\varphi_A^m \to \varphi_A$ uniformly as $m \to \infty$. Define the a functional $I : \Phi \to \mathbb{R}$

by

$$I(\varphi_A) = \lim_{m\to\infty} I^o(\varphi_A^m) \qquad \forall A \in \mathcal{A}.$$

Since I^o is a niveloid, it is a continuous function, and so I^o preserves convergence. Thus, for any menu $A \in \mathcal{A}$, the sequence $\{I^o(\varphi_A^m)\}_{m\in\mathbb{N}}$ converges to a point in $[u_*, u^*]$ showing that I is well-defined. The following arguments show that I preserves the properties of I^o, i.e., it is also a normalized convex niveloid.

Since I^o is a niveloid, we have $I^o(\varphi_A^m) - I^o(\varphi_A^m) \leq \max(\varphi_A^m - \varphi_A^m)$ for any $A, B \in \mathcal{A}$, and $m \in \mathbb{N}$. Thus we obtain,

$$\begin{aligned}
I(\varphi_A) - I(\varphi_B) &= \lim_{m\to\infty}(I^o(\varphi_A^m)) - \lim_{m\to\infty}(I^o(\varphi_B^m)) \\
&= \lim_{m\to\infty}(I^o(\varphi_A^m) - I^o(\varphi_B^m)) \\
&\leq \lim_{m\to\infty}(\max(\varphi_A^m - \varphi_B^m)) \\
&= \lim_{m\to\infty}\frac{m-1}{m}(\max(\varphi_A - \varphi_B)) \\
&= \max(\varphi_A - \varphi_B),
\end{aligned}$$

establishing that I is a niveloid.

Clearly I is normalized. Now let $A, B \in \mathcal{A}$, and $\alpha \in [0,1]$. Since Φ is a convex set, $\alpha\varphi_A + (1-\alpha)\varphi_B \in \Phi$, and so by convexity of I^o we have

$$\begin{aligned}
I(\alpha\varphi_A + (1-\alpha)\varphi_B) &= \lim_{m\to\infty}\left(I^o(\varphi_{A[\alpha]B}^m)\right) \\
&= \lim_{m\to\infty}(I^o(\alpha\varphi_A^m + (1-\alpha)\varphi_B^m) \\
&\leq \lim_{m\to\infty}(\alpha I^o(\varphi_A^m) + (1-\alpha)I^o(\varphi_B^m)) \\
&= \alpha\lim_{m\to\infty} I^o(\varphi_A^m) + (1-\alpha)\lim_{m\to\infty} I^o(\varphi_B^m) \\
&= \alpha I(\varphi_A) + (1-\alpha)I(\varphi_B),
\end{aligned}$$

showing that I is convex. As a result, I is a normalized convex niveloid which assumes values in K.

Since Φ is a convex subset of $B(\Sigma, K)$ and I is a normalized convex niveloid, the obvious adaption of the arguments in the proof of Lemma 27 in Maccheroni et al.

[2004] establishes that

$$I(\varphi) = \max_{\pi \in \Delta(\mathcal{V})} \left(\langle \varphi, \pi \rangle - c^*(\pi) \right) \qquad \forall \varphi \in \Phi.$$

Hence, it remains to show that, for all $A, B \in \mathcal{A}$, $A \succsim B$ if and only if $I(\varphi_A) \geq I(\varphi_B)$. We establish the contrapositive for each direction.

First, suppose that $A \succ B$.

Using parts (i) and (ii), we can find $p, q \in P$ such that $A \succ \{p\} \succ \{q\} \succ B$. Then by Axiom 2, there exists some $M \in \mathbb{N}$ such that for all $m \geq M$,

$$A^m \succsim \{p\} \succ \{q\} \succsim B^m$$

Otherwise, it must be the case that $\{p\} \succsim A$ or $B \succsim \{q\}$, a contradiction. Thus, for all $m \geq M$, we must have

$$I^o(\varphi_A^m) \geq u(p) > u(q) \geq I^o(\varphi_B^m)$$

Since weak inequalities are preserved in the limit, we obtain

$$I(\varphi_A) \geq u(p) > u(q) \geq I(\varphi_B)$$

and so $I(\varphi_A) > I(\varphi_B)$.

For the converse, suppose that $I(\varphi_A) > I(\varphi_B)$.

By construction, $u^* \geq I(\varphi_A) > I(\varphi_B) \geq u_*$. Hence, there exist $p, q \in P$ such that

$$I(\varphi_A) > u(p) > u(q) > I(\varphi_B).$$

Since I is continuous, there exists $M \in \mathbb{N}$ such that for all $m \geq M$,

$$I^o(\varphi_A^m) \geq u(p) > u(q) \geq I^o(\varphi_B^m).$$

implying that for all $m \geq M$,

$$A^m \succsim \{p\} \succ \{q\} \succsim B^m$$

Hence, by Axiom 2, it follows that

$$A \succsim \{p\} \succ \{q\} \succsim B,$$

and so $A \succ B$.

As a result, the function $U : \mathcal{A} \to \mathbb{R}$, defined by $U(A) = I(\varphi_A)$ represents $\succsim$. $\square$

A.2.3 Indifference to the timing of the resolution of uncertainty

Our final lemma is used in the proof of Propositions 1 and 3. For a self-discipline preference $\succsim$ represented by (u, c), it provides a necessary and sufficient condition for there to be a common solution to the self-discipline choice problem for a finite collection of menus. The argument follows the proof of Lemma 1 in De Oliveira et al. [2016].

Let $U : \mathcal{A} \to \mathbb{R}$ and $\mathcal{D} : \mathcal{A} \to \mathbb{R}$ represent, respectively, the value function and policy correspondence of the self-discipline choice problem

$$\max_{\pi \in \Delta(\mathcal{V})} (\langle \varphi_A, \pi \rangle - c(\pi))$$

with parameters (u, c).

For $A_1, ..., A_N \in \mathcal{A}$ and $\alpha_1, ..., \alpha_N \in (0, 1)$ such that $\sum_{i=1}^N \alpha_i = 1$, denote by $\sum_i \alpha_i A_i = \{\sum_i \alpha_i p_i \,:\, p_i \in A_i \; \forall i = 1, ..., N\}$. We observe (without proof) that U has the following convexity property:

$$U(\sum_i \alpha_i A_i) \leq \sum_i \alpha_i U(A_i).$$

Lemma 3. *Let $A_1, ..., A_N \in \mathcal{A}$ and $\alpha_1, ..., \alpha_N \in (0, 1)$ such that $\sum_{i=1}^N \alpha_i = 1$. Then the following statements are equivalent:*

(i) $U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i)$.

(ii) $\bigcap_i \mathcal{D}(A_i) \neq \emptyset$.

Proof. [**(i) implies (ii)**]: Let $A_1, ..., A_N \in \mathcal{A}$ and $\alpha_1, ..., \alpha_N \in (0,1)$ such that $\sum_{i=1}^N \alpha_i = 1$ and $U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i)$. We proceed by induction on N. If $N = 1$, (ii) trivially holds. Now suppose that $N > 1$, and that (i) implies (ii) for $N - 1$.

Without loss of generality, let $\alpha_1 = \min_i \alpha_i$, and set

$$B = \frac{\alpha_2}{1-\alpha_1} A_2 + ... + \frac{\alpha_N}{1-\alpha_1} A_N.$$

Since $\alpha_i/(1-\alpha_1) \leq 1$ for all $i = 2, ..., N$, we have $B \in \mathcal{A}$. By the convexity property of U,

$$U(B) \leq \sum_{i=2}^N \left(\frac{\alpha_i}{1-\alpha_1} \right) U(A_i)$$

and

$$\sum_i \alpha_i U(A_i) = U\left(\sum_i \alpha_i A_i \right) = U\left(A_1[\alpha_1]B\right) \leq \alpha_1 U(A_1) + (1-\alpha_1) U(B).$$

Hence, we get

$$\sum_{i=2}^N \alpha_i U(A_i) = (1-\alpha_1) U(B)$$

and

$$\alpha_1 U(A_1) + (1-\alpha_1) U(B) = U(A_1[\alpha_1]B).$$

Now choose some $\pi \in \mathcal{D}(A_1[\alpha_1]B)$. Then

$$\langle \alpha_1 \varphi_{A_1} + (1-\alpha_1)\varphi_B, \pi \rangle - U\left(A_1[\alpha_1]B\right) = c(\pi) \geq \langle \varphi_{A_1}, \pi \rangle - U\left(A_1\right).$$

Replacing $U(\varphi_{A_1})$ with $\frac{1}{\alpha_1} U\left(A_1[\alpha_1]B\right) - \frac{1-\alpha_1}{\alpha_1} U(B)$, and rearranging, we get

$$(1-\alpha_1)\langle \varphi_B, \pi \rangle - \frac{1-\alpha_1}{\alpha_1} U(B) \geq (1-\alpha_1)\langle \varphi_{A_1}, \pi \rangle - \frac{1-\alpha_1}{\alpha_1} U\left(A_1[\alpha_1]B\right).$$

Multiplying both sides of the inequality by $\alpha_1/(1-\alpha_1)$ and adding $(1-\alpha_1)\langle \varphi_B, \pi \rangle$, we get

$$\langle \varphi_B, \pi \rangle - U(B) \geq \langle \alpha_1 \varphi_{A_1} + (1-\alpha_1)\varphi_B, \pi \rangle - U\left(A_1[\alpha_1]B\right),$$

which implies that $\langle \varphi_B, \pi \rangle - U(B) \geq c(\pi)$. It follows that $\pi \in \mathcal{D}(B)$. An analogous argument shows that $\pi \in \mathcal{D}(A_1)$. As a result,

$$\mathcal{D}\left(A_1[\alpha_1]B\right) \subset \mathcal{D}(A_1) \cap \mathcal{D}(B).$$

Since $\sum_{i=2}^{N} \alpha_i U(A_i) = (1-\alpha_1)U(B)$, by the inductive assumption, $\mathcal{D}(B) \subset \mathcal{D}(A_i)$ for all $i = 2, ..., N$, and so $\mathcal{D}\left(\sum_i \alpha_i A_i\right) \subset \mathcal{D}(A_i)$ for all $i = 1, ..., N$. Since $\mathcal{D}\left(\sum_i \alpha_i A_i\right) \neq \emptyset$, we have $\bigcap_i \mathcal{D}(A_i) \neq \emptyset$.

[**(ii) implies (i)**]: Suppose (ii) holds, and let$\pi \in \bigcap_i \mathcal{D}\left(A_i\right)$. Then

$$\sum_i \alpha_i U(A_i) = \left\langle \sum_i \alpha_i \varphi_{A_i}, \pi \right\rangle - c(\pi) \leq U\left(\sum_i \alpha_i A_i\right).$$

On the other hand, the convexity property of U implies

$$\sum_i \alpha_i U(A_i) \geq U\left(\sum_i \alpha_i A_i\right).$$

Hence, $U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i)$, and so (i) holds. □

A.3 Proofs for the results in the text

Proof of Theorem 1

It is straightforward to show that a self-discipline preference satisfies Axioms 1–6 (in particular, Axiom 4 follows from the sufficiency part of Lemma 1). For the converse, let $\succsim$ be a binary relation that satisfies Axioms 1–6. Then by Lemma 2, (u, c^*) represents $\succsim$ and so $\succsim$ is a self-discipline preference.

Proof of Theorem 2

Let $\succsim$ be a self-discipline preference represented by (u, c).

[**Part (i)**]: It is straightforward to show that u represents the commitment ranking. With the normalization $u \in \mathcal{V}$, uniqueness follows by standard arguments.

[Part (ii)]: By Lemma 2, (u, c^*) also represents $\succsim$. It therefore remains to show that $c \geq c^*$ (establishing c^* as the minimal self-discipline cost function). By way of contradiction, suppose $c(\pi) < c^*(\pi)$ for some $\pi \in \Delta(\mathcal{V})$. Then, by definition of c^*, there exists a menu $A \in \mathcal{A}$ such that $\langle \varphi_A, \pi \rangle - u(p_A) > c(\pi)$, i.e., $\langle \varphi_A, \pi \rangle - c(\pi) > u(p_A)$. Hence, $u(p_A) = \max_{\rho \in \Delta(\mathcal{V})} (\langle \varphi_A, \rho \rangle - c(\rho)) > u(p_A)$, a contradiction.

Proof of Corollary 1

Let $\succsim$ be a self-discipline preference with canonical representation (u, c^*), and let c be another cost function such that (u, c) also represents $\succsim$. By Theorem 2, $c^* \leq c$.

[Part (i)]: Let $\pi \in \mathcal{D}(A|u, c)$ for some $A \in \mathcal{A}$, and let p_A be the singleton equivalent of A. Hence, $\langle \varphi_A, \pi \rangle - c(\pi) = u(p_A)$. Now suppose $c(\pi) > c^*(\pi)$. Then

$$\langle \varphi_A, \pi \rangle - c^*(\pi) > u(p_A) = \max_{\rho \in \Delta(\mathcal{V})} [\langle \varphi_A, \rho \rangle - c^*(\rho)],$$

a contradiction. Hence, $c(\pi) = c^*(\pi)$.

[Part (ii)]: Let $\pi \in \mathcal{D}(A|u, c)$ for some $A \in \mathcal{A}$ with singleton equivalent p_A. From part (ii),

$$\langle \varphi_A, \pi \rangle - c^*(\pi) = \langle \varphi_A, \pi \rangle - c(\pi) = u(p_A),$$

and so $\pi \in \mathcal{D}(A|u, c^*)$.

Proof of Proposition 2

Let (u, c^*) be a canonical representation for a self-discipline preference. In the proof of Lemma 2, part (iii), we show that c^* is grounded. Since c^* is the supremum over linear functions, c^* is convex. Finally, to establish u-monotonicity, let $\pi, \rho \in \Delta(\mathcal{V})$ with $\pi \trianglerighteq_u \rho$. Then

$$\langle \varphi_A, \pi \rangle \geq \langle \varphi_A, \rho \rangle \qquad \forall A \in \mathcal{A}.$$

Hence,

$$\sup_{A \in \mathcal{A}} (\langle \varphi_A, \pi \rangle - u(p_A)) \geq \sup_{A \in \mathcal{A}} (\langle \varphi_A, \rho \rangle - u(p_A)),$$

and so $c^*(\pi) \geq c^*(\rho)$.

Proof of Theorem 3

Let $\succsim_1$ and $\succsim_2$ be self-discipline preferences with canonical representations (u_1, c_1^*) and (u_2, c_2^*), respectively.

[(i) implies (ii)]: Suppose $\succsim_2$ has a stronger preference for commitment than $\succsim_1$. Thus, for any $p, q \in P$, $\{p\} \succsim_1 \{q\}$ implies $\{p\} \succsim_2 \{q\}$. We want to show that also $\{p\} \succsim_2 \{q\}$ implies $\{p\} \succsim_1 \{q\}$ and so $u_2 = u_1$. Suppose, for contradiction, that this is not the case; that is, there exist $p, q \in P$ such that $\{p\} \succsim_2 \{q\}$ and $\{q\} \succ_1 \{p\}$. Since $\{q\} \succ_1 \{p\}$, we must have $\{q\} \succsim_2 \{p\}$ and so $\{p\} \sim_2 \{q\}$. Since u_2 is non-constant, there must exist some $r \in P$ such that either $\{r\} \succ_2 \{p\}$ or $\{p\} \succ_2 \{r\}$. Suppose $\{r\} \succ_2 \{p\}$ (the argument in the opposite case is analogous). Then $\{r\}[\alpha]\{p\} \succ \{p\}$ for all $\alpha \in (0,1)$. On the other hand, since $\{q\} \succ_1 \{p\}$, there exists some $\alpha \in (0,1)$ such that $\{q\} \succ_1 \{r\}[\alpha]\{p\}$. Since $\succsim_2$ has a stronger preference for commitment, it follows that $\{q\} \succsim_2 \{r\}[\alpha]\{p\} \succ_2 \{p\} \sim_2 \{q\}$, which is a contradiction.

Now consider a menu A and let $\{p_A\} \sim_1 A$ and $\{q_A\} \sim_2 A$. Since $\succsim_2$ has a stronger preference for commitment than $\succsim_1$, $\{p_A\} \succsim_2 \{q_A\}$ and so $u_2(p_A) \geq u_2(q_A)$. As a result, for any $\pi \in \Delta(\mathcal{V})$,

$$\begin{aligned} c_2(\pi) &= \sup_{A \in \mathcal{A}} \left(\langle \varphi_A, \pi \rangle - u_2(q_A) \right) \\ &\geq \sup_{A \in \mathcal{A}} \left(\langle \varphi_A, \pi \rangle - u_1(p_A) \right) \\ &= c_1(\pi). \end{aligned}$$

[(ii) implies (i)]: Suppose that $u_1 = u_2$ and $c_1 \leq c_2$. Let $\{p\} \succsim_1 A$ for some $p \in P$ and $A \in \mathcal{A}$. Then we have,

$$\begin{aligned} u_2(p) &= u_1(p) \\ &\geq \max_{\pi \in \Delta(\mathcal{V})} \left(\langle \varphi_A, \pi \rangle - c_1(\pi) \right) \\ &\geq \max_{\pi \in \Delta(\mathcal{V})} \left(\langle \varphi_A, \pi \rangle - c_2(\pi) \right), \end{aligned}$$

and so $\{p\} \succsim_2 A$.

Proof of Proposition 1

Let $\succsim$ be a self-discipline preference with canonical representation (u, c^*). It is straightforward to show that random Strotz preferences satisfy Set Independence. For the converse, define the functional $U : \mathcal{A} \to \mathbb{R}$ as in Lemma 2 part (iv).

Let $A, B \in \mathcal{A}$ and $\alpha \in (0, 1)$, and let p_A and p_B be singleton equivalents of A and B, respectively. By Set Independence, $\{p_A\}[\alpha]\{p_B\} \sim \{p_A\}[\alpha]B \sim A[\alpha]B$. Hence, $U(A[\alpha]B) = \alpha U(A) + (1 - \alpha)U(B)$. By induction, for menus $A_1, ..., A_N \in \mathcal{A}$ and $\alpha_1, ..., \alpha_N \in [0, 1]$ such that $\sum_i \alpha_i = 1$,

$$U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i).$$

By Lemma 3, it follows that $\bigcap_i \mathcal{D}(A_i|u, c^*) \neq \emptyset$. Hence, the collection of closed sets $\{\mathcal{D}(A|u, c^*) \,:\, A \in \mathcal{A}\}$ has the finite intersection property. Since $\Delta(\mathcal{V})$ is compact, it follows that $\bigcap_{A \in \mathcal{A}} \mathcal{D}(A|u, c^*) \neq \emptyset$.

Now let $\pi \in \bigcap_{A \in \mathcal{A}} \mathcal{D}(A|u, c^*)$. Then, for all menus $A \in \mathcal{A}$,

$$U(A) = \langle \varphi_A, \pi \rangle - c(\pi),$$

and so, for all menus $A, B \in \mathcal{A}$,

$$A \succsim B \quad \Leftrightarrow \quad \langle \varphi_A, \pi \rangle \geq \langle \varphi_A, \pi \rangle .$$

Hence, is a random Strotz preference represented by (u, π).

Proof of Proposition 3

Let $u \in \mathcal{V}$, and let $k : \mathcal{V} \to [0, \infty]$ be lower semicontinuous and proper. Without loss of generality, we assume that k is grounded. Denote by $U : \mathcal{A} \to \mathbb{R}$ the value function, and by $\mathcal{D} : \mathcal{A} \rightrightarrows \mathcal{V}$ the policy correspondence, corresponding to the self-discipline choice problem

$$\max_{v \in \mathcal{V}} \left(\varphi_A(v) - k(v)\right).$$

Let $\mathcal{C} : \mathcal{A} \rightrightarrows \mathcal{A}$ denote the choice correspondence induced by (u, k), defined by

$$\mathcal{C}(A) = \bigcup_{v \in \mathcal{D}(A|u,k)} \arg\max_{p \in M_v(A)} u(p) \quad \forall A \in \mathcal{A}.$$

For a menu A and lottery p, denote by $\mathcal{W}_p = \{w \in \mathcal{V} : p \in M_w(A)\} \cap \mathcal{D}(A)$.

[**Sufficiency**]: Consider a menu $A \in \mathcal{A}$ with $p \in \mathcal{C}(A)$. Since $\succsim$ is a costly Strotz preference, $\mathcal{W}_p \neq \emptyset$ and there exists $\kappa \in \mathbb{R}_+$ such that $k(w) = \kappa$ all $w \in \mathcal{W}_p$. By way of contradiction, suppose that $\kappa > 0$. Since k is grounded, it follows that there exists $q \in A$ such that $u(p) > u(q)$. Since $\kappa > 0$, there exists $\bar{\alpha} \in (0,1)$ such that $\bar{\alpha}(u(p) - u(q)) < \kappa$. Hence, $p \notin \mathcal{C}(A[\alpha]\{q\})$, contradicting vNM independence. It follows that, for all menus $A \in \mathcal{A}$ and $p \in \mathcal{C}(A)$, $w \in \mathcal{W}_p$ implies $k(w) = 0$. Hence, for any menu A, if $p, q \in \mathcal{C}(A)$, then $w \in \mathcal{W}_p \cup \mathcal{W}_q$ implies $k(w) = 0$ and so $u(p) = u(q)$.

As a result, with an obvious abuse of notation, $U(A) = u(\mathcal{C}(A))$. Now observe that vNM Independence implies, for any $A, B \in \mathcal{A}$ and $\alpha \in (0,1)$,

$$u\left(\mathcal{C}(A[\alpha]B)\right) = \alpha u\left(\mathcal{C}(A)\right) + (1-\alpha) u\left(\mathcal{C}(B)\right).$$

Hence, $\succsim$ satisfies Set Independence.

[**Necessity**]: Suppose $\succsim$ satisfies Set Independence. By Lemma 3 and the argument in the proof of Proposition 1, there exists $v \in \bigcap_{A \in \mathcal{A}} \mathcal{D}(A)$, and so $U(A) = \varphi_A(v)$ for all $A \in \mathcal{A}$. As a result, for any menu $A \in \mathcal{A}$, $\mathcal{C}(A) = \arg\max_{p \in M_v(A)} u(p)$, and $\mathcal{C}$ therefore satisfies WARP and vNM independence.

A.4 Models of temptation-driven behavior

In this Section, we briefly review some well-known models of temptation-driven behavior referred to in the main text. All of these models have in common that temptations in period 2 can induce a desire for commitment in period 1.

A.4.1 Axioms

We start by reviewing some axioms referred to in the discussion. Let $\mathcal{P}$ denote the set of non-trivial weak orders on $\mathcal{A}$, and let $\succsim \in \mathcal{P}$.

Upper Semicontinuity: For all $A \in \mathcal{A}$, the set $\{B : B \succsim A\}$ is closed is the Hausdorff topology on $\mathcal{A}$.

Finiteness: For all $A \in \mathcal{A}$, there exists a finite $A' \subseteq co(A)$ such that for all B, $A \subseteq co(B) \subseteq co(A)$ implies $B \sim A$.

Set Betweenness: For all $A, B \in \mathcal{A}$, $A \succsim B$ implies $A \succsim A \cup B \succsim B$.

Desire for Commitment: For all $A \in \mathcal{A}$, there exists $p \in A$ such that $\{p\} \succsim A$.

Upper Semicontinuity (which is "one half" of Strong Continuity) and Finiteness are technical axioms. However, Set Betweenness and Desire for Commitment reveal temptation-driven behavior (see Gul and Pesendorfer [2001, Section 2] and Dekel et al. [2009, Section 4])

A.4.2 Temptation-driven preferences

We now review some important models of temptation-driven behavior in the menu-choice literature.

Self-control preferences and their generalizations Gul and Pesendorfer [2001] characterize a model of self-control, where the DM compromises between her temptation and normative rankings in period 2, incurring an opportunity cost.

Self-control: $\succsim$ is a *self-control preference* if there exists $u, v \in \mathbb{R}^n$ such that

$$U(A) = \max_{p \in A} u(p) + v(q) - \max_{q \in A} v(q) \quad \forall A \in \mathcal{A}$$

represents $\succsim$.

Let $\mathcal{P}_{SC}$ denote the class of self-control preferences. Theorem 1 in Gul and Pesendorfer [2001] shows that a preference relation is in $\mathcal{P}_{SC}$ if and only if it satisfies Strong Continuity, Set Independence and Set Betweenness.

Dekel et al. [2009] generalize self-control preferences by allowing the DM to be affected by multiple temptations at once, and by allowing the set of temptations to be random.

Multi-dimensional temptation: $\succsim$ is a *multi-dimensional temptation preference* if there exists $u \in \mathbb{R}^n$, $I, J_1, ..., J_I \in \mathbb{N}$, $(v_1^i, ..., v_{J_i}^i) \in \mathbb{R}^{n \times J_i}$ for all $i = 1, ..., n$, and $(\alpha_1, ..., \alpha_I) \in \mathbb{R}_{++}^I$ with $\sum_i \alpha_i = 1$, such that

$$U(A) = \sum_i \alpha_i \left(\max_{p \in A} u(p) - c_i(p, A) \right) \quad \forall A \in \mathcal{A}$$

represents $\succsim$, where

$$c_i(p, A) = \left(\sum_{j=1}^{J_i} \max_{q \in A} v_j^i(q) \right) - \sum_{j=1}^{J_i} v_j^i(p) \quad \forall p \in P, \ A \in \mathcal{A}.$$

Let $\mathcal{P}_T$ denote the class of multi-dimensional temptation preferences. Theorem 1 in Dekel et al. [2009] shows that a preference relation is in $\mathcal{P}_T$ if and only if it satisfies Strong Continuity, Set Independence, Finiteness, Desire for Commitment and a technical axiom called Approximate Improvements are Chosen (Dekel et al. [2009, Axiom 4]).

Stovall [2010] characterizes the special case of multi-dimensional temptation where only one temptation affects the DM at a time, but temptations can occur at random.

Multiple temptation: $\succsim$ is a *multiple temptation preferences* if there exists $I \in \mathbb{N}$, $u, v_1, ..., v_I \in \mathbb{R}^n$ and $(\alpha_1, ..., \alpha_I) \in \mathbb{R}_{++}^I$ with $\sum_i \alpha_i = 1$, such that

$$U(A) = \sum_{i=1}^{I} \alpha_i \left(\max_{p \in A} u(p) + v_i(p) - \max_{q \in A} v_i(p) \right) \quad \forall A \in \mathcal{A}$$

represents $\succsim$.

Let $\mathcal{P}_{MT}$ denote the class of multiple temptation preferences. Theorem 1 in Stovall

[2010] shows that a preference relation is in $\mathcal{P}_{MT}$ if and only if it satisfies Strong Continuity, Finiteness, Set Independence, and Weak Set Betweenness.

Noor and Takeoka [2010] take a different approach to generalize self-control, by allowing convex costs of self-control.

Convex self-control: $\succsim$ is a convex *self-control preference* if there exist $u, v \in \mathbb{R}^n$ and a continuous strictly increasing convex function $\gamma : [0, u^* - u_*] \to \mathbb{R}_+$ such that

$$U(A) = \max_{p \in A} \left(u(p) - \gamma \left[\max_{q \in A} v(q) - v(p) \right] \right)$$

represents $\succsim$.

Theorem 3 in Noor and Takeoka [2010] provides a characterization of the class of convex self-control preferences $\mathcal{P}_{SCS}$. In particular, a convex self-control preference satisfies Strong Continuity and Set Betweenness, but may violate Set Independence (and Aversion to Randomization).

Strotz preferences and their generalizations Gul and Pesendorfer [2001] also characterize a model of temptation where the DM is unable to exercise self-control: in period 2, she is able to choose only from the tempting alternatives in a menu.

Strotz: $\succsim$ is a *Strotz preference* if there exists $u, v \in \mathcal{V}$ such that

$$U(A) = \varphi_A(v) \quad \forall A \in \mathcal{A}$$

represents $\succsim$.

Let $\mathcal{P}_{St}$ denote the class of Strotz preferences. Theorem 3 in Gul and Pesendorfer [2001] shows that a preference relation is in $\mathcal{P}_{SC} \cup \mathcal{P}_{St}$ if and only if it satisfies Mixture Continuity, Upper Semicontinuity, Set Independence, and Set Betweenness.

Chatterjee and Krishna [2009] generalize Strotz preferences: in period 1 the DM is uncertain if she will be tempted in period 2, but if she is tempted she succumbs.

Dual-self: $\succsim$ is a *dual-self preference* if there exists $u, v \in \mathcal{V}$ and $\alpha \in [0, 1]$ such that

$$U(A) = \alpha\varphi_A(u) + (1 - \alpha)\varphi_A(v) \quad \forall A \in \mathcal{A}$$

represents $\succsim$.

Theorem 1 in Chatterjee and Krishna [2009] provides an axiomatic characterization of the class $\mathcal{P}_{DS}$ of dual-self preferences. Olszewski [2007] characterizes the subclass $\mathcal{P}_{MM} \subset \mathcal{P}_{DS}$ of *maxmin preferences*, where $v = -u$, and so $U(A) = \alpha \max_{p \in A} u(p) + (1-\alpha) \min_{p \in A} u(p)$ for all $A \in \mathcal{A}$.

The class of random Strotz preferences $\mathcal{P}_{RS}$ in Definition 1 generalizes $\mathcal{P}_{DS}$ by allowing for greater uncertainty about the temptation ranking in period. Chandrasekher [2014] generalizes random Strotz by allowing for Knightian uncertainty regarding temptation rankings, inducing incompleteness in the preference relation over menus.

A.4.3 Relation with self-discipline

Theorem 1 in Dekel and Lipman [2012] shows that $\mathcal{P}_{RS} \supseteq \mathcal{P}_{MT} \supseteq \mathcal{P}_{SC}$. Clearly, $\mathcal{P}_{RS} \supseteq \mathcal{P}_{DS} \supseteq (\mathcal{P}_{MM} \cup \mathcal{P}_{St})$. Moreover, it follows from the axiomatic characterizations in Dekel et al. [2009] and Noor and Takeoka [2010] that $\mathcal{P}_{T} \cap \mathcal{P}_{CSC} = \mathcal{P}_{SC}$. It follows from our results that self-discipline preferences have the following relation with these models of temptation-driven behavior.

Corollary 2. *Let $\mathcal{P}_{SD}$ be the class of self-discipline preferences. Then:*

(i) $\mathcal{P}_{SD} \supseteq \mathcal{P}_{RS} \supseteq \mathcal{P}_{MT} \supseteq \mathcal{P}_{SC}$.

(ii) $\mathcal{P}_{SD} \supseteq \mathcal{P}_{RS} \supseteq \mathcal{P}_{DS} \supseteq (\mathcal{P}_{MM} \cup \mathcal{P}_{St})$.

(iii) $(\mathcal{P}_{SD} \cap \mathcal{P}_{CSC}) = (\mathcal{P}_{T} \cap \mathcal{P}_{CSC}) = \mathcal{P}_{SC}$.

(iv) $(\mathcal{P}_{SD} \cap \mathcal{P}_{T}) = \mathcal{P}_{MT}$.