Accepted Manuscript

Mood-driven Choices and Self-regulation

Maximilian Mihm, Kemal Ozbek

PII: S0022-0531(18)30164-9
DOI: https://doi.org/10.1016/j.jet.2018.05.006
Reference: YJETH 4783

To appear in: *Journal of Economic Theory*

Received date: 17 July 2017
Revised date: 23 April 2018
Accepted date: 7 May 2018

Please cite this article in press as: Mihm, M., Ozbek, K. Mood-driven Choices and Self-regulation. *J. Econ. Theory* (2018), https://doi.org/10.1016/j.jet.2018.05.006

Mood-driven Choices and Self-regulation*

Maximilian Mihm† Kemal Ozbek‡

May 9, 2018

Abstract

We model a decision maker who can exert costly effort to regulate herself, thereby reducing internal conflicts between her normative objectives and mood-driven choices. We provide an axiomatic characterization of the model, and show how costs of self-regulation can be elicited and compared across individuals. In a consumption-saving problem we show that self-regulation can generate unintended income effects, which have important implications for public policies on saving behavior. We also provide several examples to illustrate how self-regulation can rationalize many well-known choice anomalies. These behavioral implications follow from a key feature of the model that self-regulation decisions can respond to changes in incentives.

Keywords: choice anomalies, consumption-saving, desire for commitment, internal conflict, random Strotz, self-regulation.

*We would like to thank the editor and two anonymous reviewers, as well as Nageeb Ali, Madhav Chandrasekher, Kalyan Chatterjee, Eddie Dekel, Ryota Iijima, Paola Manzini, Rani Spiegler; conference participants at ESEM 2016 and RUD 2016; seminar audiences at London Business School, Penn State and UC Davis; and especially Bart Lipman for useful comments and suggestions.

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

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

1 Introduction

A standard economic agent considers all alternatives that are available to her, and chooses the one that best satisfies her normative objectives. However, in reality, there are many mood-driven factors–such as addiction, temptation, or inattention–that can cause the agent to make inferior choices leading to a conflict with long-term objectives.[1] As such, an agent who is concerned that moods might jeopardize her objectives has an incentive to take precautionary actions. For instance, temporary cravings for nicotine may lead an agent to smoke–especially when she is feeling stressed, anxious, or irresolute–which conflicts with her long-term goal to quit smoking.[2] As a result, the agent may want to regulate herself by taking precautionary actions, so as to better resist cravings, lessen their frequency, or reduce their intensity.

The objective of this paper is to understand how the possibility of self-regulation can affect choice behavior, and how observed choice behavior can reveal the self-regulation problem. To study these questions, we model a decision maker (DM) who can exert costly effort to reduce conflicts between her normative preferences and mood-driven choices, where these efforts could be physical (e.g., exercise, rest), mental (e.g., meditation, self-motivation), or could take many other forms (e.g., medical, social, or monetary).

A key feature of our model is that the DM responds to incentives when choosing her self-regulation effort. To illustrate, consider a two period consumption-saving problem. In the first period, the DM receives some income $m > 0$, consumes $x_1 \leq m$, and saves the rest for consumption x_2 in the next period. The DM evaluates a consumption bundle (x_1, x_2) in terms of a normative utility $w(x_1) + \delta w(x_2)$, for some function $w(.)$ and discount factor $\delta \in (0, 1)$. However, at the moment of deciding how much to consume in period 1, the DM becomes more impatient and deviates from her normative objectives by discounting her period 2 utility at a rate

[1] Mood-driven factors that can impact choice behavior may also include choice overload, cold feet, habits, present bias, selfishness, and status quo bias among many others (see Section 6).

[2] Alternatively, temporary urges to consume sugar may cause the agent to eat sweets–especially when she is feeling depressed or sad–which conflict with her goal to loose weight. Or temporary lack of concentration may cause the agent to miss important details of an insurance plan–especially when she is feeling tired or restless–which conflict with her goal to purchase the right policy.

$\hat{\beta}\delta$, where $\hat{\beta} \in (0,1)$ measures her impatience mood.[3]

In this context, self-regulation can be viewed as an effort the DM can exert to alter her impatience by way of increasing $\beta \in [\hat{\beta}, 1]$ but at a cost $c(\beta)$. Anticipating that she will choose some bundle (x_1^β, x_2^β) when her future discount rate is $\beta\delta$, the DM then faces the following self-regulation choice problem:

$$\max_{\beta \in [\hat{\beta},1]} \left[w(x_1^\beta) + \delta\, w(x_2^\beta) - c(\beta) \right].$$

By regulating her impatience, the DM is able to more closely align the consumption bundle (x_1^β, x_2^β) she actually chooses with the normatively ideal bundle (x_1^1, x_2^1), thereby increasing her realized normative utility. However, she must balance this benefit against the self-regulation cost $c(\beta)$. In particular, the solution to this self-regulation problem depends on the incentives in her consumption-saving problem. For instance, we show in Section 2, that self-regulation can generate unintended income effects, which impact the way the DM's savings responds to changes in the interest rate. For a wide range of model parameters, the DM *decreases* savings relative to consumption in period 1 in response to an *increase* in the interest rate. This negative income effect, which is driven by the ability of the DM to regulate impatience, can impact the viability of policy instruments aimed at encouraging savings behavior.

Some other economic settings where self-regulation can have interesting implications for choice behavior include portfolio-choice or labor-supply problems. For instance, in a portfolio-choice problem, a change in wealth may alter the DM's incentives to adjust her risk-preferences, thereby affecting the share of wealth she invests in safe assets. Alternatively, in a labor-supply problem, a change in the compensation scheme–such as a change in the share of flat-rate versus piece-rate payments–may alter the DM's incentives to adjust her leisure-preferences, thereby affecting her labor-supply.

In well-known models of internal conflicts, such as the self-control model in Gul and Pesendorfer [2001] and the dual-self model in Chatterjee and Krishna [2009],

[3]This is the two-period version of the beta-delta model introduced in Phelps and Pollak [1968], and later used by Laibson [1997] and O'Donoughe and Rabin [2001] to study present bias.

there is no opportunity for the DM to exert effort to influence her internal conflicts, in order to alter her mood-driven choices. As a result, these models do not generate the type of responses to incentives (e.g., the unintended income effects in the consumption-saving problem) that arise under the self-regulation model. Moreover, our self-regulation model is consistent with many other choice anomalies that have been widely documented in experiments (e.g., the attraction, common ratio and magnitude effects), but that are not consistent with these other models (see Section 7).

Our general model formalizes self-regulation by identifying a novel parameter–the cost of self-regulation–that determines how a DM is able to regulate herself. We view these costs as subjective, and possibly encompassing many forms of self-regulation effort. A challenge in identifying such costs is that self-regulation itself may often be a hidden action, and therefore not observable (for instance, when it represents a purely mental effort). To overcome this challenge, we study the observable implications of self-regulation in a framework where the DM chooses a choice set (or *menu*) today (when she is in a "cool" state), anticipating that she will exercise self-regulation (in an interim period) before selecting an alternative tomorrow (when she is in a "hot" state).[4]

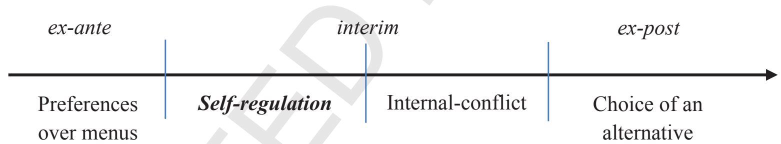

Figure 1: Timeline

Our identifying assumption is that the DM anticipates the benefits and costs of self-regulation when she chooses a menu, so that her preferences over menus–which can be revealed by choice behavior–incorporates her self-regulation problem. In this menu-choice framework, we give an axiomatic characterization of a general model of costly self-regulation (Theorem 1), and show how self-regulation costs can be elicited from choice-data (Theorem 2). We also show that a DM with higher costs

[4]In the literature, a cool state refers to a situation where the DM does not experience internal conflicts, in contrast to a hot state where she faces conflicts (see, e.g., Noor [2007]).

of self-regulation values commitment more, establishing a comparative measure of ability to self-regulate in terms of observable behavior (Theorem 3).

Our model of self-regulation is primarily related to the literature where internal conflicts generate a desire for commitment; that is, the DM would prefer committing to the normatively best alternative in a menu if she had the option to do so. Axiomatic work on this literature was pioneered by the self-control model of Gul and Pesendorfer [2001], while Dekel and Lipman [2012] study the random Strotz model, which subsumes also many other models of internal conflict.[5] A key implication of these models is that the DM, in terms menu choice, will be indifferent between mixing commitment equivalents of menus and mixing the menus themselves since her internal conflicts are fixed, and cannot be changed. We show that these models constitute the special case of our model where the DM does not have the opportunity to alter the conflicts by taking precautionary actions (Proposition 1).

A novel feature of our model is that internal conflicts are not necessarily fixed, and rather can be influenced by costly self-regulation. In terms of preferences for commitment, this feature of the model induces a key axiom, *Increasing Desire for Commitment* (IDC). This axiom reflects the idea that it is better to mix commitment equivalents of menus than the menus themselves since when menus are mixed (i) the utility gap between normatively better and worse alternatives reduces, resulting in a decrease in the benefits of self-regulation, and (ii) there are more inferior alternatives in the mixed menu than there are in each of the original menus, resulting in a more complicated self-regulation problem. As such, relative to the mixture of the commitment equivalents (where self-regulation is not needed), the mixture of the menus both reduces the potential benefit of self-regulation and makes it harder to effectively utilize self-regulation effort.

There are also models in the literature that, in contrast to the self-regulation model, can imply a decrease in desire for commitment with mixtures (e.g., Noor and Takeoka [2010] and Masatlioglu, Nakajima, and Ozdenoren [2017]). More broadly, the self-regulation model is also related to a literature on cognitive control of preferences (e.g., Nehring [2006] and Koida [2017]), and to the literature on

[5]See also the survey article of Lipman and Pesendorfer [2013] on temptations.

variational menu choice preferences (e.g., Ergin and Sarver [2010]). We discuss the formal connection with these other models after presenting our axiomatic characterization of the self-regulation model (see Section 6).

The paper is organized as follows. Section 2 presents the two period consumption-saving problem in more detail, and shows how self-regulation can generate unintended income effects. In Section 3, we introduce the framework and define preferences over menus induced by a general model of self-regulation. Section 4 presents the axioms. Section 5 contains our representation, identification, and comparative statics results. We also characterize two special cases of the model where self-regulation effort is fixed or constrained. In Section 6, we discuss the related literature. Section 7 provides several examples to illustrate some implications of the self-regulation model for ex-post choice from menus. Section 8 concludes. Proofs are given in an Appendix.

2 A consumption-saving problem

To illustrate our general model, we consider a consumption-saving problem with an agent who can exercise self-regulation to adjust her impatience. We show that the ability to regulate impatience can lead to unintended income effects in savings behavior that have important implications for public policies aimed at encouraging saving behavior.

Agent's problem

Consider a two period consumption-saving problem similar to the one described in the Introduction. Specifically, the agent receives income $m > 0$ and consumes $x_1 \leq m$ in period 1. The amount not consumed, $m - x_1$, is saved, and earns interest at a gross rate of return $R > 1$, which is then taxed (at a rate t) in period 2 leading to a net rate of return $R_t > 1$. There are no other sources of income. Hence, the agent consumes $x_2 = (m - x_1)R_t$ in period 2.

The agent obtains utility from consumption in each period according to the iso-elastic utility function $w_\theta(x) = \frac{1}{1-\theta}x^{1-\theta}$ for some $\theta > 0$ (with $w_\theta(x) = \ln(x)$ when

$\theta = 1$), and discounts period 2 utility at some rate $\delta \in (0, 1)$ to aggregate her period utilities. However, at the point of deciding how much to consume today and how much to save for tomorrow, the agent becomes more impatient. As a result, she deviates from her normative objective of discounting with δ, and instead discounts her period 2 utility at a rate $\hat{\beta}\delta$ where $\hat{\beta} \in (0, 1)$ measures her impatience mood.

On the other hand, by regulating herself in advance, the agent can adjust her impatience to $\beta \in [\hat{\beta}, 1]$ at some cost given by a function $c(\beta)$ which is increasing in β. Anticipating that she will consume the bundle (x_1^β, x_2^β), which is optimal when her future discount rate is $\beta\delta$, the agent faces the following self-regulation choice problem, $\max_{\beta \in [\hat{\beta}, 1]} \left[w(x_1^\beta) + \delta\, w(x_2^\beta) - c(\beta) \right]$, as discussed in the Introduction.[6]

Policy-maker's problem

Now consider the problem of a policy-maker who wants to encourage savings. The policy-maker aims to decrease first period consumption relative to second period consumption, or simply decrease the ratio $x_{1,2}^\beta = x_1^\beta / x_2^\beta$. One way of implementing this policy goal is adjusting the net interest rate R_t; for instance, by changing the tax rate t or the interest rate R. Indeed, when the net interest rate increases, the ratio of first period consumption to second period consumption, regardless of β, decreases; that is, $\frac{\partial x_{1,2}^\beta}{\partial R_t} < 0$. Thus, in order to encourage savings, the policy-maker might want to forgo some tax/interest revenue by raising interest rate R_t.

This policy, however, may cause–contrary to its objective–less savings when the agent can adjust her impatience β. The reason is that when R_t increases the agent becomes relatively richer. Thus, the marginal benefit of self-regulation diminishes, and the agent therefore wants to choose a lower level of β which can, in turn, induce her to save relatively less, and consume more in period 1.

In fact, when the optimal impatience β^* is elastic enough, the indirect effect of raising R_t, through β^*, can overcome its direct effect on the relative share of first period consumption $x_{1,2}^{\beta^*}$, and can overall lead to an increase, $\frac{\partial x_{1,2}^{\beta^*}}{\partial R_t} > 0$, in the ratio

[6]In particular, the optimal bundle is $x_1^\beta = \frac{m}{1+(\beta\delta R_t^{1-\theta})^{1/\theta}}$ and $x_2^\beta = \frac{m\,(\beta\delta R_t)^{1/\theta}}{1+(\beta\delta R_t^{1-\theta})^{1/\theta}}$ when the discount rate is $\beta\delta$.

of first period consumption relative to second period consumption. In particular, this holds whenever $\frac{\partial\beta^*/\partial R_t}{\beta^*/R_t} < -1$.

Discussion

To set a viable policy goal, it is therefore important to identify whether the agent can adjust her impatience, and how elastic her adjustments are to policy changes. A main factor which affects the elasticity of impatience adjustment is the agent's cost of self-regulation. Our identification result, Theorem 2, provides a method for eliciting such costs under a general self-regulation model (see Section 5.2).

To avoid incurring self-regulation costs, the agent in the consumption-saving problem would be willing to pay a premium for the option to commit ex-ante to the normatively best two-period allocation of her income. Intuitively, the premium the agent would be willing to pay should depend on her costs of self-regulation, where higher costs should reflect in a willingness to pay a higher premium. In this regard, our comparative statics result, Theorem 3, provides a comparative measure of ability to self-regulate in terms of desire for commitment under a general self-regulation model (see Section 5.3).

Most models of internal conflict, such as the self-control model in Gul and Pesendorfer [2001] and the dual-self model in Chatterjee and Krishna [2009], cannot generate the type of income effects that arise under the self-regulation model.[7] In these models, a change in the net interest rate can only scale up or down the value of the agent's problem without altering her behavior. Unintended income effects arise in our model because the DM exhibits an increasing desire for commitment, and our characterization result, Theorem 1, provides a way of testing such behavioral responses under a general self-regulation model (see Section 5.1).

[7] The convex self-control model in Noor and Takeoka [2010] can also generate income effects in the consumption-saving example. However, in general, the self-regulation model and the convex self-control model induce different ex-ante and ex-post choice behavior. For instance, we show that these models can have different asset choice implications for the consumption-saving problem (see Section 6). We also give an example of ex-post choice patterns, which are compatible with the self-regulation model, but which cannot be generated by the convex self-control model (see Section 7).

3 A general model of self-regulation

In this section, we describe the menu-choice framework and define the preferences induced by a general model of self-regulation.

3.1 Framework

In the following, X is a finite set of n prizes, 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*.[8]

For any $\alpha \in [0,1]$, let $\alpha A + (1-\alpha)B$ denote a *mixed-menu*, which is the mixture of menus A and B, where

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

Our primitive is a binary relation $\succsim$ on the set of menus $\mathcal{A}$, 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 the set of singleton menus the *commitment ranking*.[9] A functional $U : \mathcal{A} \to \mathbb{R}$ represents $\succsim$ when, for all menus A and B, $A \succsim B$ if and only if $U(A) \geq U(B)$.

3.2 Conflicts and moods

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.

[8]Although our menus-of-lotteries framework focuses on choice under risk, our analysis could be adapted for choice over time (respectively, choice under uncertainty), for instance, by replacing the set of lotteries P with a convex subset of $\mathbb{R}^n$ as the set of finite consumption streams (respectively, the set of acts over a finite state space).

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

Let $\mathcal{V} = \{v \in \mathbb{R}^n : \sum_{i=1}^n v_i = 0,\, v \cdot v = 1\}$ be the set of *utilities*, with typical elements $u, v, w \in \mathcal{V}$. 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 function can be represented by a unique utility in $\mathcal{V}$.

We say two utilities $u, v \in \mathcal{V}$ *conflict* on the ranking of $p, q \in P$ whenever $u(p) > u(q)$ and $v(p) < v(q)$, or vice versa. More generally, u and v conflict in menu A when $M_u(A) \cap M_v(A) = \emptyset$, where for any $w \in \mathcal{V}$, $M_w(A) = \arg\max_{p \in A} w(p)$ denotes the lotteries in menu A that maximize w.[10]

We think of *moods* as temporary states of mind that give rise to compositions of conflicts, which can be represented by distributions over utilities. This reflects the idea that, while normative objectives may be stable and representable in terms of a single utility, mood-driven behaviors can be temporary and random.

Formally, let $\Delta(\mathcal{V})$ be the set of all finitely-additive Borel probability distributions on $\mathcal{V}$, with typical elements $\pi, \rho, \sigma \in \Delta(\mathcal{V})$ called *distributions.* With this formalization of a mood in mind, we denote by $b_A^u(\pi)$ the benefit the DM obtains from menu A when her normative utility is u and her mood induces a composition of conflicts given by the distribution π, where

$$b_A^u(\pi) = \int_{\mathcal{V}} [\max_{p \in M_v(A)} u(p)]\, \pi(dv).$$

Benefit $b_A^u(.)$ provides a general way to model how moods can lead to internal conflicts in menu A, and what normative utility the DM anticipates obtaining under these conflicts.[11]

For a fixed $\hat{\pi} \in \Delta(\mathcal{V})$, Dekel and Lipman [2012] call $b_A^u(\hat{\pi})$ the random Strotz model and show that it generalizes many well-known models of internal conflicts that have been studied in the literature:

- Self-Control: When $\hat{\pi} \in \Delta(\mathcal{V})$ is the uniform distribution over the set

[10]Since A is non-empty and closed, and w is a continuous function on P, $M_w(A)$ is also a menu by the Maximum theorem.

[11]Since u is continuous on P and $M_v(A)$ is a menu for all $v \in \mathcal{V}$, b_A^u is well-defined.

$\{w_{v+au} \in \mathcal{V} : a \in [0, \frac{1}{k}],\ v \in \mathcal{V},\ k > 0\}$, then,

$$b^u_A(\hat{\pi}) = \max_{p \in A} (u(p) + k\, v(p)) - \max_{q \in A} k\, v(q). \tag{1}$$

This is the self-control model, where the DM reconciles with a conflicting utility $k\, v$, while experiencing a psychological cost of its maximal forgone utility (Gul and Pesendorfer [2001]). In this model, k measures the DM's capacity of self-control, where a lower k corresponds to a greater ability to resist temptations. When k tends to ∞, temptations become overwhelming, and the model converges to the Strotz model.

Let $\delta_v \in \Delta(\mathcal{V})$ be the distribution that assigns probability 1 to v for any utility $v \in \mathcal{V}$.

- Dual-Self: When $\hat{\pi} \in \Delta(\mathcal{V})$ is such that $\hat{\pi} = (1-\theta)\, \delta_u + \theta\, \delta_v$ for some $\theta \in [0, 1]$ and some $v \in \mathcal{V}$, then,

$$b^u_A(\hat{\pi}) = (1 - \theta) \max_{p \in A} u(p) + \theta \max_{p \in M_v(A)} u(p). \tag{2}$$

This is the dual-self model, where with probability $1 - \theta$ choice is made by the normative self u, and with probability θ choice is made by an alter ego v (Chatterjee and Krishna [2009]).[12] The case $\theta = 0$ is the standard model, while the case $\theta = 1$ is the Strotz model. The case $v = -u$ is the Hurwicz criterion where $\theta \in [0, 1]$ is interpreted as a decision weight (measuring pessimism/optimism) rather than as a probability (Arrow and Hurwicz [1972], Olszewski [2007]).[13]

3.3 Self-regulation

We view self-regulation as an effort the DM can exert to alter her mood-driven behavior so as to make normatively better choices. By regulating herself, the DM

[12] Eliaz and Spiegler [2006] apply the dual-self model to study the effects of sophistication and naivety in optimal contracting problems.

[13] In particular, $b^u_A(\hat{\pi}) = \theta \max_{p \in A} u(p) + (1 - \theta) \min_{p \in A} u(p)$ for $\hat{\pi} = \theta\delta_u + (1 - \theta)\delta_{-u}$.

can (i) improve the way she resolves internal conflicts, (ii) reduce the chances of conflicts, or (iii) weaken the intensity of conflicts.

For instance, self-regulation may allow the DM to decrease k in the self-control model, thereby increasing her capacity to resist the internal conflict. Or self-regulation may decrease θ in the dual-self-model, thereby lowering the chances that the alter-ego realizes leading to a conflict. Or, in general, self-regulation may allow the DM to weaken the intensity of conflicts, for instance, by altering her mood-driven behavior from the utility v to an alternative utility w, that is "less conflicting" with u (see Section 3.4 for a definition of less conflicting utility).

However, self-regulation requires effort, and so the DM must incur the costs in order to exploit the benefits of self-regulation. These costs are represented by a function $c : \Delta(\mathcal{V}) \to [0, \infty]$, where $c(\pi)$ is a behavioral measure of the effort required to induce π. A cost function c is *proper* whenever $c(\pi) < \infty$ for some $\pi \in \Delta(\mathcal{V})$, and *lower semi-continuous* whenever $\liminf_{\pi \to \bar{\pi}} c(\pi) \geq c(\bar{\pi})$ for every $\bar{\pi} \in \Delta(\mathcal{V})$, where convergence is defined with respect to the weak*-topology.

A self-regulation preference reflects the behavior of a DM who acts as if she anticipates exerting an optimal level of self-regulation before making a choice from a menu.

Definition 1. [Self-regulation preference] A binary relation $\succsim$ on menus is a *self-regulation preference* if there exists $u \in \mathcal{V}$ and a proper lower semi-continuous function $c : \Delta(\mathcal{V}) \to [0, \infty]$ such that the functional $U : \mathcal{A} \to \mathbb{R}$, defined by

$$U(A) = \max_{\pi \in \Delta(\mathcal{V})} \left[b_A^u(\pi) - c(\pi) \right],$$

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

The self-regulation model admits a natural multi-system interpretation, in which the utility u represents the cognitive process (i.e., the planner) responsible for setting normative objectives and regulating at a cost c other cognitive processes (i.e., the doers) responsible for cravings, urges, attentiveness, biases and selfish desires.[14]

[14]Related multi-system models have also been applied in the literature; see, for instance, Thaler

The self-regulation model also allows for many possible interpretations for the way a DM's internal conflicts realize and how she reconciles with them ex-post. As illustrated above, the benefit $b^u_A(\pi)$ might correspond to the value of menu A under self-control, or dual-self, or any combination of such models. In particular, for each menu A, the optimal level of self-regulation is determined by weighing the benefit against the cost. As such, self-regulation effort can vary across different menus leading to a potentially different ex-post realization and resolution of internal conflicts for each menu.

In general, the DM may consider all distributions in $\Delta(\mathcal{V})$ as part of her optimization problem. However, the DM could also consider only a subset $\Pi \subset \Delta(\mathcal{V})$. Such constraints can always be incorporated by setting $c(\pi) = \infty$ for any $\pi \notin \Pi$, which is observationally equivalent to excluding π in Definition 1. For instance, the DM might consider only a fixed $\hat{\pi}$, which corresponds to the model in Eq. 1 or the model in Eq. 2. In that case, one can set $c(\pi) = \infty$ for all $\pi \neq \hat{\pi}$, and obtain, respectively, the self-control model in Gul and Pesendorfer [2001] or the dual-self model in Chatterjee and Krishna [2009].

3.4 Canonical cost functions

Properness and lower semi-continuity are minimal properties of a cost function to ensure that the self-regulation problem is well-defined. We impose no other a priori restrictions on the cost function. On the other hand, there are a number of intuitive properties that, without loss of generality, can be imposed on a self-regulation cost function (see Corollary 1 in Section 5.2). To state the properties, we require some additional notation.

For $u \in \mathcal{V}$, define a partial order over $\mathcal{V}$ by $v \geq^*_u w$ (read "v is less conflicting with u than w is") if, whenever $u(p) > u(q)$, then $v(p) < v(q)$ implies $w(p) < w(q)$. Hence, the utility v is less conflicting with u than w is if, whenever v conflicts with u, w also conflicts with u; that is, v is more closely aligned with u than w is. This notion of comparative conflict can be generalized with distributions over utilities by

and Shefrin [1981], Bénabou and Tirole [2004], Benhabib and Bisin [2005], Fudenberg and Levine [2006], and Ali [2011], among many others.

understanding that each distribution represents a possible composition of conflict.

Dekel and Lipman [2012] extend $\geq^*_u$ from $\mathcal{V}$ to $\Delta(\mathcal{V})$ as follows: let $\pi \trianglerighteq^*_u \rho$ (read "π is stochastically less conflicting with u than ρ is") if $\pi(\hat{\mathcal{V}}) \geq \rho(\hat{\mathcal{V}})$ for every $\hat{\mathcal{V}} \subset \mathcal{V}$ that is (i) closed in $\mathcal{V}$, and (ii) closed under $\geq^*_u$.[15] Hence, the distribution π is stochastically less conflicting with u than ρ is if π puts more weight on less conflicting utilities than ρ does; that is, if π is more closely aligned (in a stochastic sense) with u than ρ is.

Definition 2. [Canonical cost functions] We say a cost function c is canonical if it satisfies (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 $\alpha \in (0,1)$, and (iii) *monotonicity:* $c(\pi) \geq c(\rho)$ for all $\pi, \rho \in \Delta(\mathcal{V})$ with $\pi \trianglerighteq^*_u \rho$.

Groundedness implies that the DM has the option not to exercise self-regulation, thereby incurring no cost. When c is grounded, $U(\{p\}) = u(p)$ for every lottery p, and so the ex-ante utility over singleton menus coincides with the normative utility over lotteries.[16]

Convexity means that the cost of a "less focused" composition of conflicts should not exceed the average cost of "more focused" compositions. For example, a distribution $\sigma = \alpha\pi + (1-\alpha)\rho$, where $\pi, \rho \in \Delta(\mathcal{V})$ and $\alpha \in (0,1)$, can be seen as less focused since its support is larger than either distribution π or ρ.

Monotonicity captures the idea that having a distribution less conflicting with normative utility requires more self-regulation effort, and should therefore be costlier. In particular, monotonicity implies that the cost of self-regulation is maximal for δ_u, since u is the least conflicting utility.

[15] A subset $\hat{\mathcal{V}} \subset \mathcal{V}$ is said to be *closed under* $\geq^*_u$ if $w \in \hat{\mathcal{V}}$ and $v \geq^*_u w$ implies $v \in \hat{\mathcal{V}}$.

[16] Since cost function c is lower semi-continuous and its domain $\Delta(\mathcal{V})$ is compact, there is always a $\pi \in \Delta(\mathcal{V})$ that minimizes c, and so in general $U(\{p\}) = u(p) - c(\pi)$ for every lottery $p \in P$. As a result, while the ex-ante utility over singletons in general may differ from the normative utility over lotteries, they both imply the same ordinal ranking over lotteries (see also Corollary 1).

4 Axioms

In this section, we discuss behavioral implications of the self-regulation model for menu-choice. We start by presenting four standard axioms from the literature, and then focus on two further axioms which reflect key features of self-regulation.

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

Axiom 3. *[Weak Set Independence] For all menus $A, B \in \mathcal{A}$ and $\alpha \in (0,1)$, if $\alpha A + (1-\alpha)\{p\} \succsim \alpha B + (1-\alpha)\{p\}$, then $\alpha A + (1-\alpha)\{q\} \succsim \alpha B + (1-\alpha)\{q\}$ for all lotteries $p, q \in P$.*

Let $co(A)$ denote the convex hull of menu A. Since it is closed, $co(A)$ is also a menu.

Axiom 4. *[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 preferences over mixtures of menus. Axiom 3 reflects the idea that normative preferences and conflicting utilities satisfy the standard von Neumann-Morgenstern (vNM) independence axiom, and as a result they are linear over lotteries.[17] Axiom 4 reflects the idea that in a conflict situation only maximal alternatives matter. Since A and $co(A)$ have the same extreme points, and for linear utilities only the extreme points in a menu can be maximal, the DM is indifferent between A and $co(A)$.

[17] Axiom 3 is introduced in Ergin and Sarver [2010] in the context of their costly contemplation model, which is similar in spirit to the Weak Certainty Independence axiom in Maccheroni, Marinacci, and Rustichini [2006] used in their analysis of ambiguity aversion. In our analysis, Axiom 3 reveals that the DM focuses on consequentialist motivations for self-regulation and excludes any psychological motivations to affect her mood-driven choices. In particular, since self-regulation is only relevant for menus A and B, replacing the common commitment menu $\{p\}$ with menu $\{q\}$ does not alter the incentives to self-regulate, and so does not alter ranking of the mixed-menus.

4.1 Dominance and monotonicity

Our next axiom is a monotonicity condition which allows the DM to rank some menus independent of any specific mood she may experience in the future. In this regard, the monotonicity axiom reflects the idea that the DM deems possible only conflicts that generate choices consistent with the following independence condition.

Condition (IND)**.** *For $p \in A$ and $q \in B$, p is chosen in A and q is chosen in B if and only if $\frac{1}{2}p + \frac{1}{2}q$ is chosen in $\frac{1}{2}A + \frac{1}{2}B$.*

In our setting, IND implies the weak axiom of revealed preference (WARP) condition, which can be stated as follows.[18]

Condition (WARP)**.** *For $p, q \in A \cap B$, p is chosen in A and q is chosen in B if and only if p is chosen in B and q is chosen in A.*

A conflict induces choices that satisfy WARP if and only if the choices can be rationalized by a weak order over lotteries. Moreover, the conflict leads to choices that satisfy IND if and only if this weak order satisfies the standard vNM independence condition. The following definition expresses when menu A offers a better ex-post choice situation than menu B for every conflict consistent with IND.

For menu A and lottery p, let $A(p) = \{r \in A : \{r\} \sim \{p\}\}$ denote the set of lotteries in A that are indifferent to p in terms of the DM's commitment rankings.

Definition 3. [Dominance] Menu A *dominates* menu B (denoted $A \trianglerighteq B$) if, for all $p, q \in P$, $\{p\} \succ \{q\}$ implies $\frac{1}{2}A(p) + \frac{1}{2}B(q) \supset \frac{1}{2}B(p) + \frac{1}{2}A(q)$.

To illustrate the idea behind Definition 3, let $A \trianglerighteq B$ and suppose the DM considers a conflict that induces choices consistent with IND. In that case, for each lottery p chosen in B, there must be a lottery p' chosen in A such that $\{p'\} \succsim \{p\}$. To see this, assume that a lottery q is chosen in menu A and a lottery p is chosen in menu B . If $\{q\} \succsim \{p\}$, then we are done; otherwise, we have $\{p\} \succ \{q\}$ and so

[18] To see the implication, let $p, q \in A \cap B$ and assume p is chosen in A and q is chosen in B. By IND, $\frac{1}{2}p + \frac{1}{2}q$ is chosen in $\frac{1}{2}A + \frac{1}{2}B$. Hence, IND implies p is chosen in B and q is chosen in A.

$\frac{1}{2}B(p) + \frac{1}{2}A(q) \subset \frac{1}{2}A(p) + \frac{1}{2}B(q)$. As a result, $\frac{1}{2}q + \frac{1}{2}p \in \frac{1}{2}A(p) + \frac{1}{2}B(q)$; that is, there are lotteries $p' \in A(p)$ and $q' \in B(q)$ such that $\frac{1}{2}p' + \frac{1}{2}q' = \frac{1}{2}q + \frac{1}{2}p$. Moreover, by IND, $\frac{1}{2}q + \frac{1}{2}p$ is chosen in $\frac{1}{2}A + \frac{1}{2}B$; that is, $\frac{1}{2}p' + \frac{1}{2}q'$ is chosen in $\frac{1}{2}A + \frac{1}{2}B$. As such, by IND, p' is chosen in A where $\{p'\} \sim \{p\}$.

As a result, when A dominates B, any conflict consistent with IND always leads to a better choice in menu A. Therefore, self-regulation efforts induce better outcomes with menu A than with menu B, which motivates the following axiom.

Axiom 5. *[Monotonicity] For all menus $A, B \in \mathcal{A}$, if $A \trianglerighteq B$, then $A \succsim B$.*

4.2 Mixtures and desire for commitment

Most models of internal conflict satisfy Axioms 1–5. The DM's ability to self-regulate, which is the key feature of our model, is revealed by the following axiom.

Axiom 6. *[Increasing Desire for Commitment] For all menus $A, B \in \mathcal{A}$, lotteries $p, q \in P$, if $A \sim \{p\}$ and $B \sim \{q\}$, then $\alpha\{p\} + (1-\alpha)\{q\} \succsim \alpha A + (1-\alpha)B$ for all $\alpha \in (0,1)$.*

Increasing Desire for Commitment (IDC) reflects the idea that it is better to mix commitment equivalent of menus rather than the menus themselves. The reason is that, when a menu is mixed, *incentives* for self-regulation change for two reasons:

(i) the utility gap between normatively better and worse alternatives in a mixed menu reduces, resulting in a decrease in the benefits of self-regulation, and

(ii) there are more inferior alternatives in a mixed menu resulting in a more complicated self-regulation problem.

As such, relative to the mixture of commitment equivalents (where self-regulation is redundant), the mixture of menus A and B both (i) reduces the potential benefit of self-regulation and (ii) makes it harder to effectively utilize the self-regulation effort. As a result, mixing menus increases the desire for commitment.[19]

[19] In an extension of our framework, Increasing Desire for Commitment can be also interpreted as a desire for early resolution of uncertainty (see Section 6).

5 Analysis of self-regulation

In this section, we provide our main findings: a representation theorem, as well as identification and comparative statics results. We also characterize two special cases of the self-regulation model, where self-regulation efforts are fixed or constrained.

5.1 Characterization

The following theorem shows that Axioms 1–6 characterize the behavior of a DM who chooses among menus "as if" she anticipates a self-regulation choice problem. As such, Axioms 1–6 are not only necessary, but also sufficient to test the self-regulation model using menu-choice data.

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

Proof sketch: It is straightforward to show that self-regulation preferences satisfy Axioms 1–6. For the converse, Lemma 2 (Appendix A.2) shows that if a binary relation $\succsim$ satisfies Axioms 1–6, then there exist a normative utility $u \in \mathcal{V}$, representing the commitment ranking, and a self-regulation cost function $c : \Delta(\mathcal{V}) \to [0, \infty]$ such that (u, c) represents $\succsim$. In particular, Axioms 1–6 imply that every menu $A \in \mathcal{A}$ has a *commitment equivalent* lottery $p_A \in P$ such that $\{p_A\} \sim A$. Using commitment equivalents, then a functional I over the set $\Phi = \{\varphi^u_A : \mathcal{V} \to \mathbb{R} \,|\, \varphi^u_A(v) = \max_{p \in M_v(A)} u(p),\ v \in \mathcal{V},\ A \in \mathcal{A}\}$ can be defined such that, for all menus A and B, $A \succsim B$ if and only if $I(\varphi^u_A) \geq I(\varphi^u_B)$. The key step is to show that I is monotone, i.e., $\varphi^u_A \geq \varphi^u_B$ implies $I(\varphi^u_A) \geq I(\varphi^u_B)$. To establish this, Lemma 1 (Appendix A.2) shows that for convex menus A and B, $A \trianglerighteq B$ if and only if $\varphi^u_A \geq \varphi^u_B$. The monotonicity of I then follows from Axioms 4 and 5. The remainder of the proof uses Axioms 3–6 to show that I is continuous and convex, and employs duality arguments to establish the desired representation. $\square$

Remark 1. As discussed in Section 4.1, Axiom 5 reflects the idea that the internal conflicts satisfy vNM independence over lotteries. This facilitates comparison with the literature and also plays an important technical role in the proof of Theorem

1. In particular, the restriction to expected utility functions ensures that Φ is a convex space, which is important to apply the duality arguments used to derive the representation.

5.2 Elicitation

The normative utility in the self-regulation model is identified by the DM's commitment ranking. In particular, for a given self-regulation preference $\succsim$, there exists a unique $u \in \mathcal{V}$ such that for all $p, q \in P$, $\{p\} \succsim \{q\}$ if and only if $u(p) \geq u(q)$ (see Lemma 2 in Appendix A.2). Moreover, the following result shows that, for each self-regulation preference, there is a unique minimal cost function.[20]

Theorem 2. *Let $\succsim$ be a self-regulation preference with normative utility $u \in \mathcal{V}$. Then, the function $c^* : \Delta(\mathcal{V}) \to [0, \infty]$, defined by $c^*(\pi) = \sup_{A \in \mathcal{A}} (b_A^u(\pi) - u(p_A))$ for all $\pi \in \Delta(\mathcal{V})$, is the unique minimal cost function where (u, c^*) represents $\succsim$.*

Theorem 2 shows that self-regulation preferences can always be represented by the unique normative utility and minimal cost function (u, c^*). In particular, the minimal cost function c^* can be constructed from data on commitment equivalent menus. For instance, since $c^*(\pi) \geq b_A^u(\pi) - u(p_A)$ for any menu A, the commitment equivalent p_A can be used to determine $b_A^u(\pi) - u(p_A)$ as a lower bound on the cost of π. Using commitment equivalents for other menus then leads to a more precise lower bound. Theorem 2 shows that this procedure approximates $c^*(\pi)$ arbitrarily closely, thereby establishing a direct connection between the self-regulation cost function and menu-choice behavior. Moreover, as the following corollary shows, c^* satisfies the canonical properties in Definition 2.

Corollary 1. *Let c^* be the minimal cost function for a given self-regulation preference $\succsim$. Then c^* satisfies groundedness, convexity, and monotonicity.*

As such, the minimal cost function c^* is canonical, and we refer to (u, c^*) as the *canonical representation* of $\succsim$.

[20] In general, it is not possible to identify a unique cost function. For instance, if a self-regulation preferences $\succsim$ is represented by (u, c), then for any positive constant k, (u, c_k) also represents $\succsim$, where $c_k(\pi) = c(\pi) + k$ for all $\pi \in \Delta(\mathcal{V})$.

Remark 2. Our approach to identification relies on the ex-ante implications of the self-regulation model, with the assumption that the DM acts as if she anticipates solving a costly self-regulation problem before the choice of an alternative in a future period. In this regard, Theorem 2 shows how the DM's anticipated costs of self-regulation can be elicited with menu-choice data. A related issue would be to identify whether the DM is sophisticated enough to correctly anticipate the parameters of her self-regulation problem. To address this question, in addition to menu choice-data, one would also require ex-post choice data from each menu. For instance, using ex-ante data on commitment equivalents and ex-post stochastic choice data, Ahn, Iijima, Le Yaouang, and Sarver [2017] provide a behavioral condition to test sophistication/naivety in the random Strotz model, and one could adapt their approach to also study the sophistication of a DM with more general self-regulation preferences.[21]

5.3 Comparative statics

As an application of our identification result, we consider a behavioral measure of comparative self-regulation. Let $\succsim_1$ and $\succsim_2$ be self-regulation preferences of two DMs. We say that DM2 is *less able to self-regulate* than DM1 when their normative utilities are the same, but self-regulation is costlier for DM2 than DM1; that is, $u_2 = u_1$ and $c_2^* \geq c_1^*$.[22]

Intuitively, when DM2 is less able to self-regulate, she should find the option of commitment–which eliminates the need to exercise self-regulation–more valuable than DM1. Dekel and Lipman [2012] define a comparative menu-choice behavior that formalizes when DM2 finds commitment more valuable than DM1.

Definition 4. [Comparative 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, $\{p\} \succsim_1 A$ implies $\{p\} \succsim_2 A$.

[21] Ahn, Iijima, Le Yaouang, and Sarver [2017]'s key assumption is that the ex-ante data is generated by a random Strotz model. In this regard, our characterization of the ex-ante implications of random Strotz (Proposition 1 in Section 5.4) can be used to test their assumption.

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

The following theorem shows that the comparative in Definition 4 characterizes when DM2 is less able to self-regulate than DM1.

Theorem 3. *Let $\succsim_1$ and $\succsim_2$ be self-regulation preferences with canonical representations (u_1, c_1^*) and (u_2, c_2^*), respectively. Then, $\succsim_2$ has a stronger desire for commitment than $\succsim_1$ if and only if $u_2 = u_1$ and $c_2^* \geq c_1^*$.*

Theorem 3 provides a behavioral measure of comparative ability to self-regulate. In particular, Theorem 3 implies that the utility difference between the normatively best alternative in a menu and the commitment equivalent of the menu is higher for a DM who is less able to self-regulate. As such, the DM will be willing to pay a higher premium for the option to commit ex-ante to the normatively best alternative, thereby avoiding higher self-regulation costs.

5.4 Special cases

We conclude this section by characterizing two special cases of the self-regulation model.

Fixed self-regulation

An important special case of our self-regulation model is one where the DM's self-regulation effort is fixed. This special case is characterized by the following axiom.

Axiom. *[Neutral Desire for Commitment] For all menus $A, B \in \mathcal{A}$ and lotteries $p, q \in P$, if $A \sim \{p\}$ and $B \sim \{q\}$, then $\alpha\{p\} + (1-\alpha)\{q\} \sim \alpha A + (1-\alpha)B$ for all $\alpha \in (0, 1)$.*

Neutral Desire for Commitment (NDC) reflects the idea that the DM's desire for commitment does not change for a menu when it is mixed with another menu. As such, NDC is stronger than IDC, and the following result shows that self-regulation preferences satisfy NDC if and only if there is a common solution to the DM's self-regulation choice problem.

Proposition 1. *Let $\succsim$ be a self-regulation preference. Then $\succsim$ satisfies the Neutral Desire for Commitment axiom if and only if there exists a unique $\pi \in \Delta(\mathcal{V})$ such that for all $A, B \in \mathcal{A}$, $A \succsim B$ if and only if $b^u_A(\pi) \geq b^u_B(\pi)$.*

Dekel and Lipman [2012] study the representation in Proposition 1, which they call random Strotz. In particular, they characterize the class of "continuous-intensity" random Strotz preferences. However, there are also important random Strotz preferences, such as the Strotz and the dual-self preferences, which are not in the continuous-intensity class. Proposition 1 (together with Theorem 1) provides a characterization result for the entire class of random Strotz preferences.

Constrained self-regulation

Another special case of our self-regulation model is one where the DM's self-regulation effort is constrained, rather than costly. This special case is characterized by the following axiom.

Axiom. *[Weak Neutral Desire for Commitment] For each menu $A \in \mathcal{A}$ and lottery $p \in P$, if $A \sim \{p\}$, then $A \sim \alpha\{p\} + (1-\alpha)A$ for all $\alpha \in (0,1)$.*

Weak Neutral Desire for Commitment (WNDC) reflects the idea that the DM's desire for commitment does not change for a menu when it is mixed with its commitment equivalent menu. As such, WNDC is weaker than NDC, and the following result shows that self-regulation preferences satisfy WNDC if and only if the DM's self-regulation choice is costless, but constrained.

Proposition 2. *Let $\succsim$ be a self-regulation preference. Then $\succsim$ satisfies the Weak Neutral Desire for Commitment axiom if and only if there exists a set $\Pi \subset \Delta(\mathcal{V})$ such that for all $A, B \in \mathcal{A}$, $A \succsim B$ if and only if $\max_{\pi \in \Pi} b^u_A(\pi) \geq \max_{\pi \in \Pi} b^u_B(\pi)$.*

As such, relative to the random Strotz model where the self-regulation effort is fixed, for the representation in Proposition 2 self-regulation choices can vary across menus, but are restricted to a constraint set.

6 Related literature

In this section, we review the related literature. Most models of internal conflict extend on Gul and Pesendorfer [2001]'s self-control model, which is characterized by a non-trivial weak order on menus that satisfies the following axioms.

Axiom. *[Strong Continuity] For all menus $A \in \mathcal{A}$, the upper, $\{B \in \mathcal{A} : B \succsim A\}$, and the lower, $\{B \in \mathcal{A} : A \succsim B\}$, contour sets are closed.*

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

Axiom. *[Set Independence] For all menus $A, B, C \in \mathcal{A}$, if $A \succsim B$, then for all $\alpha \in (0, 1)$, $\alpha A + (1-\alpha)C \succsim \alpha B + (1-\alpha)C$.*

Strong Continuity directly implies Mixture Continuity (Axiom 2), while Set Independence directly implies both Weak Set Independence (Axiom 3) and Increasing Desire for Commitment (Axiom 6). On the other hand, while Set Independence and Strong Continuity together imply Indifference to Convexification (Axiom 4), Set Betweenness together with Strong Continuity and Set Independence implies Monotonicity (Axiom 5).

The self-regulation model can induce preferences that violate these three axioms used to characterize the self-control model. In the following, we consider other models in the literature which also relax these three axioms, and we discuss their relations with the self-regulation model.

Strong Continuity

To characterize the Strotz model, Gul and Pesendorfer [2001] weaken the Strong Continuity axiom by dropping Lower Semicontinuity.[23] Similarly, to characterize the dual-self model in Eq. 2, Chatterjee and Krishna [2009] relax the Strong Continuity axiom. Clearly, both models are random Strotz, and as such, they are both special cases of the self-regulation model. Moreover, self-regulation preferences contain many other preferences violating Strong Continuity, such as the class of

[23] Lower Semicontinuity only requires that the lower contour sets are closed.

"non-continuous intensity" random Strotz preferences (see Section 5.4). As such, we use in our analysis Mixture Continuity to permit many plausible models of internal conflict that can lead to preferences violating Strong Continuity.

Set Betweenness

Dekel, Lipman, and Rustichini [2009] argue that Set Betweenness can exclude some natural features (e.g., randomness) of internal conflicts. As an alternative, they propose the Weak Set Betweenness (WSB) axiom, which restricts Set Betweenness to certain menus.

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 $A \succsim A \cup B \succsim B$.*

Axiom 5 directly implies WSB, whereas WSB implies Axiom 5 if Strong Continuity and Set Independence are also satisfied. Although self-regulation preferences imply WSB, this axiom is too weak to ensure that each ex-post choice behavior is rationalized by a weak order satisfying vNM independence.[24]

Stovall [2010] generalizes the self-control model by allowing random internal conflicts, where his axiomatization replaces Set Betweenness with WSB. Stovall [2010]'s model is a special case of the self-regulation model where preferences over menus satisfy Set Independence and Strong Continuity, as well as a finiteness axiom.[25] Dekel, Lipman, and Rustichini [2009] further relax WSB to characterize a model with multi-dimensional conflicts, where preferences over menus satisfy the Strong Continuity and Set Independence axioms. The overlap with the self-regulation model is Stovall [2010]'s model of random self-control.

To capture perfectionism in choice making, Kopylov [2012] restricts Set Betweenness to the sets with common normatively best alternatives, while retaining both Strong Continuity and Set Independence axioms. The overlap of his model with the self-regulation model is the self-control model.

[24]In Appendix A.4, we provide two examples of menu comparisons which illustrate why our monotonicity condition (Axiom 5) cannot be replaced with WSB in our characterization of the self-regulation model.

[25]Stovall [2010] also provides examples demonstrating how Set Betweenness can exclude some intuitive choice behaviors that might occur under the presence of internal conflicts.

Set Independence

Noor and Takeoka [2010] extend on the self-control model by allowing for convex self-control costs. In their axiomatization, they relax Set Independence, while retaining both Strong Continuity and Set Betweenness. In particular, convex self-control preferences satisfy the following axiom.

Axiom. *[Decreasing Desire for Commitment] For all menus $A, B \in \mathcal{A}$ and lotteries $p, q \in P$, if $A \sim \{p\}$ and $B \sim \{q\}$, then $\alpha A + (1-\alpha)B \succsim \alpha\{p\} + (1-\alpha)\{q\}$ for all $\alpha \in (0, 1)$.*

Thus, preferences satisfying DDC also capture systematic violations of Set Independence, but do so in the opposite direction of self-regulation preferences. The overlap of the convex self-control model with the self-regulation model leads to preferences that satisfy the NDC axiom.[26] As such, the overlap is a random Strotz model, and in particular, it coincides with the self-control model since it leads to preferences that also satisfy Strong Continuity and Set Betweenness.

Timing of resolution of uncertainty

By associating menus with degenerate lotteries, preferences over menus can be extended to preferences over "lotteries of menus". In this extended choice setting, mixtures of commitment equivalents can be interpreted as early resolutions of uncertainty, while mixtures of menus can be interpreted as late resolutions (see, e.g., Ergin and Sarver [2015]). In this regard, whether preferences violate NDC or not would reveal differing attitudes towards the timing of the resolution of uncertainty. For instance, while IDC can be interpreted as a desire for early resolution, NDC can be interpreted as an indifference to the timing of resolution, and DDC can be interpreted as a desire for late resolution. In particular, our self-regulation model implies IDC, since early resolution allows the DM to focus her costly self-regulation efforts towards the payoff relevant menu.

To illustrate an economic implication of these differing attitudes, consider the consumption-saving problem in Section 2. Suppose that, in an earlier period 0,

[26] Given our Axioms 1-5, the NDC axiom is equivalent to the Set Independence axiom.

the DM can invest in two risky assets, A and B. In period 1, Asset A returns an income of m_h with probability α and an income of m_l with probability $1-\alpha$. Over periods 1 and 2, Asset B yields consumption vector (x_1^h, x_2^h) with probability α and consumption vector (x_1^l, x_2^l) with probability $1-\alpha$. Now, suppose that for $k = h, l$, the DM is indifferent in period 1 between facing the consumption-saving problem with income m_k or committing to the consumption vector (x_1^k, x_2^k). As such, while asset A can be seen as a mixture of menus, asset B is a mixture of commitment equivalents of these menus. In that case, (i) models satisfying NDC (e.g., self-control or dual-self) induce indifference between assets A and B, (ii) models satisfying DDC (e.g., convex self-control) induce a preference for asset A, and (iii) models satisfying IDC (e.g., self-regulation) induce a preference for asset B, which provides more commitment.

Other related models

There are also other models, which overlap with the self-regulation model. Chandrasekher [2014] proposes a generalization of the random Strotz model, weakening the assumption that preferences are complete while retaining the NDC axiom. The overlap of the model in Chandrasekher [2014] with the self-regulation model is the random Strotz model.

Masatlioglu, Nakajima, and Ozdenoren [2017] generalize the Strotz model by allowing the DM to resist internal conflicts subject to her willpower stock. Similarly, Grant, Hsieh, and Liang [2017] generalize the self-control model by allowing the DM to reconcile with internal conflicts subject to her stock of willpower. Both models lead to preferences which can violate Strong Continuity and IDC (Axiom 6). In particular, the overlap of the willpower model in Masatlioglu, Nakajima, and Ozdenoren [2017] with the self-regulation model is the Strotz model, while the overlap of the willpower model in Grant, Hsieh, and Liang [2017] with the self-regulation model is the self-control model.

Similar in spirit to the discrete choice model in Nehring [2006], Koida [2017] proposes, using lotteries over menus, a model of internal conflict, where the DM can exercise cognitive control over mental states that trigger ex-post choices. Koida

[2017]'s model, which he calls Anticipated Stochastic Choice (ASC), can also lead to preferences that violate Strong Continuity, Set Betweenness, and Set Independence. However, the overlap of ASC preferences and self-regulation preferences satisfy the WNDC axiom and so they can be represented by the constrained self-regulation model in Proposition 2. Koida [2017]'s model extends this representation in two directions: (i) the set of distributions can vary with the menu, and (ii) the support of the distributions can contain non-expected utility functions. As such, ASC preferences can violate Axiom 5; hence, neither the self-regulation nor the ASC model nests the other.

The self-regulation model also shares features with variational models, first introduced in Maccheroni, Marinacci, and Rustichini [2006]. The closest models in the menu-choice literature with a similar structure are the costly contemplation model in Ergin and Sarver [2010] and the rational inattention model in De Oliveira, Denti, Mihm, and Ozbek [2017]. In particular, the IDC axiom, which reveals the hidden action problem in our model, is related to the Aversion to Contingent Planning (ACP) axiom used to characterize these other models.

Axiom. *[Aversion to Contingent Planning] For all menus $A, B \in \mathcal{A}$, if $A \succsim B$, then for all $\alpha \in (0,1)$, $A \succsim \alpha A + (1-\alpha)B$.*

For a preference relation satisfying Axioms 1–5, IDC is equivalent to ACP and Upper Semicontinuity.[27] As such, in our analysis, IDC cannot be replaced with ACP since we only assume Mixture Continuity (Axiom 2). Moreover, these variational models lead to a desire for flexibility, and the self-regulation leads to preferences that violate desire for flexibility except when cost of self-regulation is zero. As a result, the overlap of these models with the self-regulation model is the standard expected utility model.

A DM could also take precautionary actions to regulate many other mood-driven factors which impact choice behavior. For instance, the DM could self-regulate in order to better control addictive or habitual behaviors (Bernheim and Rangel [2004], Gul and Pesendorfer [2007]); or in order to become more attentive to choice alternatives (Masatlioglu, Nakajima, and Ozbay [2012], Manzini and Mariotti

[27] Upper Semicontinuity only requires that the upper contour sets are closed.

[2014], Lleras, Masatlioglu, Nakajima, and Ozbay [2017]); or in order to weaken choice overload (Buturak and Evren [2017]) or status quo bias (Masatlioglu and Ok [2005], Dean, Kibris, and Masatlioglu [2017]); or in order to lessen present bias to avoid procrastination (O'Donoughe and Rabin [2001]); or in order to overcome pessimistic/optimistic biases in payoff relevant beliefs (Epstein and Kopylov [2007]); or in order to reduce selfishness to make more virtuous choices (Dillenberger and Sadowski [2012]).

7 Ex-post choice

In this section, we provide several examples to illustrate some possible implications of the self-regulation model for the ex-post choice of alternatives from menus. In particular, these examples show that, for both deterministic and stochastic choice, self-regulation can lead to systematic violations of well-known consistency principles, which have been widely documented in experiments.

Deterministic choice

We start by considering a simple model that generalizes the self-control model in Eq. (1). In particular, suppose that self-regulation effort allows the DM to adjust her capacity of self-control, k, in order to better resist temptations in her ex-post choice problem. Specifically, let $u \in \mathcal{V}$ be the DM's normative utility, and $v \in \mathcal{V}$ be her temptation utility. When the DM does not self-regulate, her temptation utility v receives equal weight in reconciling with the normative utility u; that is, $k = 1$. However, the DM can exert effort with a cost $\bar{c} > 0$ to lower k to $\frac{1}{2}$, thereby increasing her capacity of self-control to better resist ex-post temptations.

While the self-control model, where k is fixed, leads to ex-post choices that satisfy WARP, the following example illustrates that the ability to self-regulate can lead to choices that violate WARP. In particular, the choice behavior in this example is consistent with the well-documented *attraction effect* (Huber and Puto [1983]).

Example 1. **[WARP]** Consider lotteries $p, q, r \in P$ where $u(p) > u(q) > u(r)$ and

$v(r) > v(q) > v(p)$ and $u(r)+v(r) > u(q)+v(q) > u(p)+v(p)$. In addition, assume (i) $\Big[u(p) + \frac{1}{2}v(p)\Big] - \Big[u(r) + \frac{1}{2}v(r)\Big] > \bar{c}$ and (ii) $\bar{c} > \Big[u(p) + \frac{1}{2}v(p)\Big] - \Big[u(q) + \frac{1}{2}v(q)\Big]$.

Under these conditions, for menu $\{p, q\}$, it is optimal not to self-regulate. In that case, temptation utility v is weighted equally as the normative utility u, and so the DM cannot resist the most tempting alternative and therefore chooses q. On the other hand, for menu $\{p, q, r\}$, it is optimal to self-regulate. In that case, temptation utility v receives half the weight as the normative utility u, and so the DM can resist the most tempting alternative r and chooses p. Since q is chosen in $\{p, q\}$, and p is chosen in $\{p, q, r\}$, this choice behavior violates WARP.[28] □

Moreover, while the self-control model satisfies the IND condition in Section 4.1, the following example illustrates that the ability to self-regulate can lead to choices that violate IND. In particular, the choice behavior in this example is consistent with the well-documented *common ratio effect* (Allais [1953] and Kahneman and Tversky [1979]).

Example 2. **[IND]** Consider the same setting as in Example 2, but now replace assumption (ii) with assumption (ii') $\bar{c} > \frac{1}{4}\Big(\Big[u(p) + \frac{1}{2}v(p)\Big] - \Big[u(r) + \frac{1}{2}v(r)\Big]\Big)$.

Under these conditions, for menu $\{p, r\}$, it is optimal to self-regulate. In that case, temptation utility v receives half the weight of the normative utility u, and so the DM can resist the most tempting alternative r and chooses p. On the other hand, for any lottery s, it is optimal not to self-regulate for menu $\frac{1}{4}\{p, r\} + \frac{3}{4}\{s\}$. In that case, temptation utility v is weighted equally with the normative utility u, and so the DM cannot resist the most tempting alternative and therefore chooses $\frac{1}{4}r + \frac{3}{4}s$. Since p is chosen in $\{p, r\}$, and $\frac{1}{4}r + \frac{3}{4}s$ is chosen in $\frac{1}{4}\{p, r\} + \frac{3}{4}\{s\}$, this choice behavior violates IND.[29] □

Noor and Takeoka [2010] show that the convex self-control model can also lead to ex-post choices that violate WARP and IND. However, the ex-post choice patterns

[28] In general, the attraction effect refers to the increased attractiveness of an alternative (e.g., p) when an inferior alternative (e.g., r) is added to the choice set $\{p, q\}$.

[29] Let $[\alpha; x]$ denote a lottery in which the DM receives \$$x$ with probability α, and \$0 otherwise, and let $p = [1; 3000]$, $r = [\frac{4}{5}; 4000]$, and $s = [1; 0]$. The common ratio effect is the finding, for instance, that many subjects in experiments choose p from $\{p, r\}$, but also choose $\frac{1}{4}r + \frac{3}{4}s = [\frac{1}{5}; 4000]$ from $\{\frac{1}{4}p + \frac{3}{4}s, \frac{1}{4}r + \frac{3}{4}s\} = \{[\frac{1}{4}; 3000], [\frac{1}{5}; 4000]\}$, violating IND as in Example 2.

generated by these models can differ significantly. For instance, consider Example 1, where the DM chooses q from $\{p, q\}$ and p from $\{p, q, r\}$ whilst $u(p) > u(q) > u(r)$. Moreover, assume that $\bar{c} \geq [u(q) + \frac{1}{2}v(q)] - [u(r) + \frac{1}{2}v(r)]$, hence r is chosen from $\{q, r\}$. This choice pattern, where q is chosen from $\{p, q\}$, r is chosen from $\{q, r\}$ and p is chosen from $\{p, q, r\}$, is incompatible with the convex-self control model (see Appendix A.5).

Similarly, the willpower model in Masatlioglu, Nakajima, and Ozdenoren [2017] can also lead to ex-post choices that violate WARP and IND. However, the ex-post choice patterns generated by their model and the self-regulation model can differ significantly. For instance, consider Example 2, where the DM chooses p from $\{p, r\}$ and $\frac{1}{4}r + \frac{3}{4}s$ from $\frac{1}{4}\{p, r\} + \frac{3}{4}\{s\}$ for any s. Now let $s = p$. Then the choice pattern, where p is chosen from $\{p, r\}$ and $\frac{1}{4}r + \frac{3}{4}p$ is chosen from $\frac{1}{4}\{p, r\} + \frac{3}{4}\{p\}$, is incompatible with the model in Masatlioglu, Nakajima, and Ozdenoren [2017] (see Appendix A.5).

There are also other choice anomalies documented in experiments that the self-regulation model can generate. An example is the *magnitude effect*, which has been observed in settings with both choice over time and choice under risk.[30] For choice over time (respectively, under risk), the magnitude effect refers to the finding that people are more impatient (respectively, risk-seeking) when stakes are small, but they become increasingly patient (respectively, risk-averse) when stakes are large (see, e.g., Prelec and Lowenstein [1991]). Such choice patterns can naturally arise with self-regulation since the benefits of self-regulation increase when the magnitudes increase, which generates higher incentives to self-regulate and influence the way ex-post choices are driven.[31]

[30]Motivated by robust findings on the magnitude effect in experimental settings for intertemporal choice, Noor and Takeoka [2018] characterize, by using preferences over consumption streams, a model of discounting utility with endogenous impatience, where the DM exerts effort in empathizing with future utilities by supressing her selfishness.

[31]By contrast, Noor and Takeoka [2010]'s convex self-control model, due to increasingly higher costs of patient choices (respectively, risk-averse choices), induces higher impatience (respectively, risk-seeking) when stakes get larger. For instance, in the consumption-saving problem discussed in Section 2, an increase in income m would induce the DM to make consumption-saving choices as if she becomes more impatient, which is incompatible with findings on the magnitude effect.

Stochastic choice

We now consider a simple model which generalizes the dual-self model in Eq. (2). In particular, self-regulation efforts allow the DM to lower the probability θ that her alter ego realizes, in order to increase the chances of choosing normatively better alternatives. Specifically, let $u \in \mathcal{V}$ be the DM's normative utility, and $v \in \mathcal{V}$ be her alter ego utility. When the DM does not self-regulate, with probability $\frac{1}{3}$ her ex-post choice are made by maximizing u first and with probability $\frac{2}{3}$ by maximizing v first. However, the DM can exert effort with a cost $\bar{c} > 0$ to increase the probability of u to $\frac{1}{2}$ (hence, decrease the probability of v to $\frac{1}{2}$), thereby increasing the chances of choosing normatively better alternatives.

While the ex-post choices induced by the dual-self model satisfy a regularity principle (REG) of stochastic choice, the following example illustrates that the ability to self-regulate can lead to violations of REG.[32] In particular, the choice behavior in this example is consistent with the *asymmetrical dominance effect* (Huber, Payne, and Puto [1982]).

Example 3. **[REG]** Consider lotteries $p, q, r \in P$ such that $u(p) > u(q) > u(r)$ and $v(r) > v(q) > v(p)$, and suppose that $\bar{c}$ satisfies (i) $\frac{1}{6}\left(u(p) - u(r)\right) > \bar{c}$ and (ii) $\bar{c} > \frac{1}{6}\left(u(p) - u(q)\right)$.

Under these conditions, for menu $\{p, q\}$ it is optimal not to self-regulate. In that case, with probability $\frac{2}{3}$ the alter ego v realizes and chooses q instead of p, and with probability $\frac{1}{3}$ the normative self u realizes and chooses p. On the other hand, for menu $\{p, q, r\}$ it is optimal to self-regulate. In that case, with probability $\frac{1}{2}$ the normative self u realizes and chooses p, and with probability $\frac{1}{2}$ the alter ego v realizes and chooses r instead of p. Since p is chosen with probability $\frac{1}{3}$ in $\{p, q\}$, and p is chosen with a higher probability $\frac{1}{2}$ in $\{p, q, r\}$, this stochastic choice behavior violates REG. $\square$

Moreover, while the ex-post choices induced by the dual-self model satisfy the strong stochastic transitivity (SST) principle, the following example illustrates that

[32] The regularity principle asserts that adding a new alternative to a choice set should never increase the probability of selecting an existing alternative (see, e.g., Rieskamp, Busemeyer, and Mellers [2006]).

the ability to self-regulate can lead to violations of SST.[33] In particular, the choice behavior in this example is consistent with the *comparability effect* (Mellers and Biagini [1994]).

Example 4. **[SST]** Consider the same setting as in Example 3, but now with the additional assumption (iii) $\bar{c} > \frac{1}{6}(u(q) - u(r))$.

Under these conditions, for menu $\{p, q\}$ it is optimal not to self-regulate. In that case, with probability $\frac{2}{3}$ the alter ego v realizes and chooses q. Similarly, for menu $\{q, r\}$, it is optimal not to self-regulate. In that case, again with probability $\frac{2}{3}$ the alter ego v realizes and chooses r. On the other hand, for menu $\{p, r\}$, it is optimal to self-regulate. In that case, only with probability $\frac{1}{2}$ the alter ego v realizes and chooses r. Since q is chosen in $\{p, q\}$ with probability $\frac{2}{3} > \frac{1}{2}$, r is chosen in $\{q, r\}$ with probability $\frac{2}{3} > \frac{1}{2}$, but r is chosen in $\{p, r\}$ with probability $\frac{1}{2}$, this stochastic choice behavior violates SST.[34] □

These examples illustrate that, when the DM can self-regulate to influence the probability that her alter ego realizes, then her ex-post choices are consistent with well-documented violations of consistency principles, such as REG and SST. Moreover, self-regulation can also lead to systematic violations of other well-known stochastic choice properties, such as the *linearity* property (LIN) in Gul and Pesendorfer [2006]. This property asserts that the probability of choosing p from $\{p, r\}$ should be the same as the probability of choosing $\alpha p + (1 - \alpha)q$ from $\alpha\{p, r\} + (1 - \alpha)\{q\}$ for all $q \in P$ and $\alpha \in (0, 1)$. LIN is satisfied by any random Strotz model (e.g., the dual-self model). However, in Examples 3 and 4, p is chosen in menu $\{p, r\}$ with probability $\frac{1}{2}$. On the other hand, for any $\alpha < \frac{\bar{c}}{\frac{1}{6}(u(p)-u(r))}$, lottery $\alpha p + (1 - \alpha)q$ is chosen in menu $\alpha\{p, r\} + (1 - \alpha)\{q\}$ with probability $\frac{1}{3}$, violating LIN. As such, similar to the violations of IND for deterministic choice (e.g., the common ratio effect), changes in the incentives naturally lead to violations of LIN for stochastic choice.

[33] The strong stochastic transitivity principle asserts that if the probability of selecting q is higher than 0.5 in $\{p, q\}$ and the probability of selecting r is higher than 0.5 in $\{q, r\}$, then the probability of selecting r in $\{p, r\}$ should be higher than both probabilities (see, e.g., Rieskamp, Busemeyer, and Mellers [2006]).

[34] In particular, these specific choice patterns fit the experimental findings for stochastic choice among gambles in Mellers and Biagini [1994].

8 Conclusion

In this paper, we study the behavior of a DM who can exercise costly effort to regulate herself, thereby reducing internal conflicts between her normative preferences and mood-driven choices. We provide an axiomatic characterization of self-regulation preferences and show how the costs of self-regulation can be elicited and compared across individuals. We also show that the self-regulation model generalize many well-known models of internal conflict and argue that self-regulation can also be relevant in a variety of other settings, including choice with addiction, inattention, cognitive load, status quo bias, and selfishness.

We show that self-regulation (i) induces an increase in desire for commitment, leading to systematic violations of Set Independence, which is satisfied by many prominent models of internal conflicts in the literature, (ii) provides a novel source for well documented violations of consistency principles for ex-post deterministic choice (e.g., the attraction effect and the common ratio effect) and stochastic choice (e.g., the asymmetrical dominance effect and the comparability effect), and (iii) generates unintended income effects in a consumption-saving problem. The self-regulation model is therefore sufficiently general to rationalize a wide variety of mood-driven choices and internal conflicts, while having enough structure to identify meaningful behavioral parameters from choice data.

A Appendix

A.1 Preliminaries

Let Σ denote the Borel sigma-algebra over $\mathcal{V}$, and let $B(\Sigma)$ be the set of bounded Σ-measurable functions mapping $\mathcal{V}$ to $\mathbb{R}$. When endowed with the sup-norm metric, $B(\Sigma)$ 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 $k \in \Phi_c$,[35] *(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 $\alpha\varphi, \alpha\varphi + (1-\alpha)k \in \Phi$, *(iv) vertically invariant* if $I(\varphi + k) = I(\varphi) + k$ for all $\varphi \in \Phi$ and $k \in \Phi_c$ such that $\varphi + k \in \Phi$, and a *(v) niveloid* if $I(\varphi) - I(\psi) \leq \sup\limits_{v \in \mathcal{V}}(\varphi(v) - \psi(v))$ for all $\varphi, \psi \in \Phi$.[36]

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 $\varphi_A^u(v) = \max_{p \in M_v(A)} u(p)$ for all $v \in \mathcal{V}$. 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 $u_* = \min\limits_{p \in P} u(p)$ and $u^* = \max\limits_{p \in P} u(p)$. 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 used to establish the results in the text. The first lemma characterizes the dominance relation. The second lemma provides a representation for a binary relation satisfying Axioms 1–6. The final lemma establishes that there is a common solution to the self-regulation

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

[36] Clearly, a niveloid is Lipschitz continuous. Moreover, Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini [2014] show that a niveloid is a monotone vertically invariant functional, while the converse is true whenever its domain is a tube.

choice problem for a collection of menus if and only if the DM is neutral with respect to the mixture of these menus.

A characterization of dominance: In the following, fix some $u \in \mathcal{V}$. Let $A_{p,q} = \{r \in co(A) : u(p) \geq u(r) \geq u(q)\}$ for $A \in \mathcal{A}$ and $p, q \in P$ such that $u(p) \geq u(q)$, and let $A_p = \{r \in co(A) : u(r) = u(p)\}$ if $u(p) = u(q)$. Now define a partial order $\trianglerighteq_u$ on $\mathcal{A}$ by $A \trianglerighteq_u B$ if for all $p, q \in P$, $u(p) > u(q)$ implies $A_p[\frac{1}{2}]B_q \supset B_p[\frac{1}{2}]A_q$.

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

Proof. **[Necessity]:** 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 $A_p[\frac{1}{2}]B_q \supset B_p[\frac{1}{2}]A_q$, $\frac{1}{2}v(r) + \frac{1}{2}v(s') \geq \frac{1}{2}v(s) + \frac{1}{2}v(r')$, 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 $A_s[\frac{1}{2}]B_r \supset B_s[\frac{1}{2}]A_r$. Since $r \in A_r$ and $s \in B_s$, both of these sets are non-empty. Since $A_s[\frac{1}{2}]B_r \supset B_s[\frac{1}{2}]A_r$, 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, $\varphi_A \geq \varphi_B$.

[Sufficiency]: 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 $A_p[\frac{1}{2}]B_q \supset B_p[\frac{1}{2}]A_q$.

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. Thus, 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. Let $r_{w,u} = (u \cdot r) \cdot u + (w \cdot r) \cdot w \in \mathbb{R}^n$ denote the projection of any $r \in P$ 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 $\bar{v}(p^A_{w,u}) = \bar{v}(q^A_{w,u})$. Without loss of generality, let $\bar{v}(p^B_{w,u}) > \bar{v}(p^A_{w,u}) = \bar{v}(q^A_{w,u}) > \bar{v}(q^B_{w,u})$.

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 $v_a^B = \alpha^B(a)u + \left(\sqrt{1-(\alpha^B(a))^2}\right) w \in \mathcal{V}$ satisfies $v_a^B(r[a^*]) = v_a^B(r[a]_{w,u}^B)$ 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, let $\alpha_B^* < \alpha_A^*$. 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 $v = v_{u,w}^\alpha = \alpha u + \left(\sqrt{1-\alpha^2}\right) w$ for some $w \in \mathcal{W}$ and some $\alpha \in (-1, 1)$. Moreover, for any $E \in \mathcal{A}$, and for any $v_{u,w}^\alpha, v_{u,w}^\beta \in \mathcal{V}$ with $w \in W$ and $\alpha, \beta \in (-1, 1)$, we have $M_{v_{u,w}^\alpha}(E_s \cup E_t) \cap M_w(E_t) = \emptyset$ implies $M_{v_{u,w}^\beta}(E_s \cup E_t) \cap M_w(E_t) = \emptyset$ 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 $\max_{e \in C_q} v \cdot e > v \cdot s$ and let $c \in C_q$ such that $v \cdot c \geq v \cdot e$ for all $e \in C_q$. Now let $\alpha = \frac{u \cdot s - u \cdot c}{(v \cdot c - v \cdot s) + (u \cdot s - u \cdot c)} \in (0, 1)$ and $w = \alpha v + (1-\alpha)u$. Then $w \cdot s = w \cdot c \geq w \cdot e$ 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 $A_p[\frac{1}{2}]B_q \supset B_p[\frac{1}{2}]A_q$. □

Implications of Axioms 1–6: The following lemma obtains several results which we use in proving Theorems 1 and 2 in the text:

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 commitment equivalent $p_A \in P$.

(iii) The function c^* defined on $\Delta(\mathcal{V})$ by $c^*(\pi) = \sup_{A \in \mathcal{A}} (\langle \varphi_A^u, \pi \rangle - u(p_A))$ for all $\pi \in \Delta(\mathcal{V})$ is non-negative lower-semicontinuous and proper.

(iv) The functional $U : \mathcal{A} \to \mathbb{R}$, defined by $U(A) = \max_{\pi \in \Delta(\mathcal{V})} (\langle \varphi_A^u, \pi \rangle - c^*(\pi))$ for all $A \in \mathcal{A}$ 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 6, we have $\{p\} \succsim \{q\}[\frac{1}{2}]\{p\}$. This implies $\{q\}[\frac{1}{2}]\{p\} \succsim \{q\}$ by Axiom 3, and so we must have $\{p\} \sim \{q\}[\frac{1}{2}]\{p\}$. By Axiom 3 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 $b_A, w_A \in A$ such that $\{b_A\} \succsim \{p\} \succsim \{w_A\}$ for all $p \in A$. Clearly $\{b_A\} \succsim A \succsim \{w_A\}$ by Axiom 5, and so by Axiom 2, the following (non-empty) sets, whose union is equal to $[0, 1]$, must be closed: $\{\alpha \in [0, 1] : \{b_A\}[\alpha]\{w_A\} \succsim A\}$ and $\{\alpha \in [0, 1] : A \succsim \{b_A\}[\alpha]\{w_A\}\}$. Since $[0, 1]$ is a connected set, these two sets must intersect; that is, there must exist some $\alpha \in [0, 1]$ such that $A \sim \{b_A\}[\alpha]\{w_A\}$. Let $p_A \in P$ be equal to $\alpha b_A + (1 - \alpha) w_A$. Finally, note that if $A \in \mathcal{A}^o$, then $b_A, w_A \in P^o$, 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 $A_p[\frac{1}{2}]\{w_A\}_q \supset \{w_A\}_p[\frac{1}{2}]A_q$ for all $p, q \in P$ such that $\{p\} \succ \{q\}$. To see why, note that either $\{w_A\}_p = \emptyset$ or $A_q = \emptyset$. Hence, by Axiom

5, $A \succsim \{w_A\}$ 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 $A \in \mathcal{A}$. 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 $\mathcal{A}$.

Let $I^o : \Phi^o \to \mathbb{R}$ be a functional defined by $I^o(\varphi_A) = u(p_A)$ for all $A \in \mathcal{A}$, where p_A denote a commitment equivalent of A. For any two menus $A, B \in \mathcal{A}^o$ with commitment equivalents p_A and p_B, $A \succsim B$ if and only if $\{p_A\} \succsim \{p_B\}$, and so $I^o(\varphi_A) \geq I^o(\varphi_B)$ if and only if $A \succsim B$. Moreover, the following argument shows that I^o is monotonic, and so it is well-defined.

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 5, $co(A) \succsim co(B)$. As a result, Axiom 4 implies $I^o(\varphi_A) \geq I^o(\varphi_B)$.

For the rest of the proof, we proceed in steps to establish that I^o is a normalized convex 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 6, 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

5 $\{b\}[\alpha]\{\bar{p}\} \succsim A[\alpha]\{\bar{p}\} \succsim \{w\}[\alpha]\{\bar{p}\}$. 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 3, 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.

Step 5 (I^o is a niveloid): Since I^o is vertically invariant, 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, Lemma 22] establishes that I^* is also convex. We now show that I^* must also be monotone. By the first paragraph in the proof of Maccheroni et al. [2004, Lemma 24], 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 Maccheroni et al. [2004, Lemma 23]. 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 $I(\varphi_A) = \lim_{m\to\infty} I^o(\varphi_A^m)$ for all $A \in \mathcal{A}$. 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

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

showing that I is convex. As a result, I is a normalized convex niveloid which assumes values in $K = [u_*, u^*]$.

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 Maccheroni et al. [2004, Lemma 27] establishes that $I(\varphi) = \max_{\pi\in\Delta(\mathcal{V})}\left(\langle\varphi,\pi\rangle - c^*(\pi)\right)$ for all $\varphi \in \Phi$, where $c^* : \Delta(\mathcal{V}) \to [0,\infty]$ is defined as in Part (iii).

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 $I^o(\varphi^m_A) \geq u(p) > u(q) \geq I^o(\varphi^m_B)$. As such, we obtain $I(\varphi_A) \geq u(p) > u(q) \geq I(\varphi_B)$ since weak inequalities are preserved in the limit, and so $I(\varphi_A) > I(\varphi_B)$.

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

Since I is continuous, $I^o(\varphi^m_A) \geq u(p) > u(q) \geq I^o(\varphi^m_B)$ for all $m \geq M$ for some$M \in \mathbb{N}$ 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 $\succsim$. $\square$

Neutral desire for commitment: Our final lemma is used in the proofs of Proposition 2 and Proposition 1. It characterizes when there is a common solution to the self-regulation choice problem for a finite collection of menus.

Let $U : \mathcal{A} \to \mathbb{R}$ and $\mathcal{D} : \mathcal{A} \twoheadrightarrow \Delta(\mathcal{V})$ represent, respectively, the value function and policy correspondence of the self-regulation choice problem with parameters (u, c). That is, for any $A \in \mathcal{A}$ let $U(A) = \max_{\pi\in\Delta(\mathcal{V})} (\langle \varphi_A, \pi\rangle - c(\pi))$ and $\mathcal{D}(A) = \arg\max_{\pi\in\Delta(\mathcal{V})} (\langle \varphi_A, \pi\rangle - c(\pi))$.

Let $\sum_{i=1}^N \alpha_i A_i = \{\sum_{i=1}^N \alpha_i p_i \ : \ p_i \in A_i, \ \forall i = 1, ..., N\}$ for $A_1, ..., A_N \in \mathcal{A}$ and $\alpha_1, ..., \alpha_N \in (0, 1)$ such that $\sum_{i=1}^N \alpha_i = 1$. We observe 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, $U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i)$ if and only if $\bigcap_i \mathcal{D}(A_i) \neq \emptyset$.*

Proof. **[Necessity]:** Let $A_1, ..., A_N \in \mathcal{A}$ and $\alpha_1, ..., \alpha_N \in (0, 1)$ with $\sum_{i=1}^N \alpha_i = 1$ and $U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i)$. We proceed by induction on N. If $N = 1$, then the result trivially holds. Now suppose that $N > 1$, and the implication holds 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, we have $U(B) \leq \sum_{i=2}^N \left(\frac{\alpha_i}{1-\alpha_1}\right) U(A_i)$ 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, $\sum_{i=2}^N \alpha_i U(A_i) = (1 - \alpha_1)U(B)$ 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

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

Multiplying both sides of the inequality by $\alpha_1/(1-\alpha_1)$ and adding $(1-\alpha_1)\langle\varphi_B,\pi\rangle$, we get $\langle\varphi_B,\pi\rangle - U(B) \geq \langle\alpha_1\varphi_{A_1} + (1-\alpha_1)\varphi_B,\pi\rangle - U(A_1[\alpha_1]B)$ which implies that $\langle\varphi_B,\pi\rangle - U(B) \geq c(\pi)$, and so $\pi \in \mathcal{D}(B)$.

By an analogous argument, $\pi \in \mathcal{D}(A_1)$ and thus, $\mathcal{D}(A_1[\alpha_1]B) \subset \mathcal{D}(A_1)\cap\mathcal{D}(B)$. 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 $\mathcal{D}(\sum_i \alpha_i A_i) \subset \mathcal{D}(A_i)$ for all $i = 1, ..., N$. Since $\mathcal{D}(\sum_i \alpha_i A_i) \neq \emptyset$, we have $\bigcap_i \mathcal{D}(A_i) \neq \emptyset$.

[**Sufficiency**]: Let $\pi \in \bigcap_i \mathcal{D}(A_i)$. Then $\sum_i \alpha_i U(A_i) = \langle\sum_i \alpha_i\varphi_{A_i},\pi\rangle - c(\pi)$ implying $\sum_i \alpha_i U(A_i) \leq U(\sum_i \alpha_i A_i)$. On the other hand, the convexity property of U implies $\sum_i \alpha_i U(A_i) \geq U(\sum_i \alpha_i A_i)$, and so $U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i)$. □

A.3 Proofs for the results in the text

Proof of Theorem 1: It is straightforward to show that a self-regulation preference satisfies Axioms 1–6 (in particular, Axiom 5 follows from the necessity 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-regulation preference. □

Proof of Proposition 1: Let $\succsim$ be a self-regulation preference with a representation (u, c). It is straightforward to show that if there is a common maximizer $\pi \in \Delta(\mathcal{V})$ which solves the self-regulation problem of $\succsim$ for each menu $A \in \mathcal{A}$, then $\succsim$ satisfies the NDC axiom. For the converse, define the value function $U : \mathcal{A} \to \mathbb{R}$ as in Section A.2.

Let $A, B \in \mathcal{A}$ and $\alpha \in (0,1)$, and let p_A and p_B be commitment equivalents of A and B, respectively. By NDC, $\{p_A\}[\alpha]\{p_B\} \sim \{p_A\}[\alpha]B \sim A[\alpha]B$. 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 $\sum_i \alpha_i = 1$, $U(\sum_i \alpha_i A_i) = \sum_i \alpha_i U(A_i)$.

Let $\mathcal{D} : \mathcal{A} \twoheadrightarrow \Delta(\mathcal{V})$ be the policy correspondence defined as in Section A.2. By Lemma 3, it follows that $\bigcap_i \mathcal{D}(A_i) \neq \emptyset$. Hence, the collection of closed sets $\{\mathcal{D}(A) : A \in \mathcal{A}\}$ has the finite intersection property. Since $\Delta(\mathcal{V})$ is compact, it follows that there exists some $\pi \in \bigcap_{A\in\mathcal{A}} \mathcal{D}(A)$, and so $U(A) = \langle\varphi_A,\pi\rangle - c(\pi)$ for all menus $A \in \mathcal{A}$. Thus, $A \succsim B$ if and only if $\langle\varphi_A,\pi\rangle \geq \langle\varphi_A,\pi\rangle$ for all menus $A, B \in \mathcal{A}$. Moreover, by Theorem 3 in Dekel and Lipman [2012], π is unique. □

Proof of Proposition 2: It is straightforward to prove that a preference $\succsim$ defined in Proposition 2 is a self-regulation preference which satisfies the WNDC axiom.

For the converse, suppose $\succsim$ is a self-regulation preference which satisfies the WNDC axiom. Let (u, c) represent $\succsim$, and assume without loss of generality that c is the minimal cost function. Let $U : \mathcal{A} \to \mathbb{R}$ be given by $U(A) = \max_{\pi \in \Delta(\mathcal{V})} \langle \varphi_A, \pi \rangle - c(\pi)$.

Let $A \in \mathcal{A}$ and $p \in P$ be such that $\{p\} \sim A$. By WNDC, $A \sim \alpha A + (1-\alpha)\{p\}$ for any $\alpha \in (0, 1)$, and so $U(\alpha A + (1-\alpha)\{p\}) = \alpha U(A) + (1-\alpha)U(\{p\})$. As such, by Lemma 3, there exists some $\pi_A \in \mathcal{D}(A) \cap \mathcal{D}(\{p\})$, where $\mathcal{D} : \mathcal{A} \twoheadrightarrow \Delta(\mathcal{V})$ is the policy correspondence defined as Section A.2. Since c is the minimal cost function, it is grounded and so $c(\pi_A) = 0$. Hence, $U(A) = \langle \varphi_A, \pi_A \rangle - c(\pi_A) = \max_{\pi \in \Pi} \langle \varphi_A, \pi \rangle$ where $\Pi = \{\pi \in \Delta(\mathcal{V}) : c(\pi) = 0\}$. □

Proof of Theorem 2: Let $\succsim$ be a self-regulation preference represented by (u, c). By Lemma 2, (u, c^*) also represents $\succsim$. It therefore remains to show that $c \geq c^*$ (establishing c^* as the minimal 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 (u, c^*) be a representation for a self-regulation preference where c^* is the minimal cost function. 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 monotonicity, let $\pi, \rho \in \Delta(\mathcal{V})$ with $\pi \trianglerighteq^*_u \rho$. Then by Theorem 4 and Lemma 6 in Dekel and Lipman [2012], $\langle \varphi_A, \pi \rangle \geq \langle \varphi_A, \rho \rangle$ for all $A \in \mathcal{A}$. 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-regulation 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 $\pi \in \Delta(\mathcal{V})$,

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

[(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 $\{p\} \succsim_2 A$ follows since,

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

□

A.4 Monotonicity and WSB

Similar to Axiom 5, Weak Set Betweenness can be thought of as a monotonicity condition. However, the intuition behind Weak Set Betweenness (WSB) uses only the idea that the DM anticipates her ex-post choice behavior to satisfy WARP, and not necessarily IND.[37] The stronger requirement of IND allows our monotonicity axiom (Axiom 5) to make additional comparisons of menus where WSB cannot make any comparison. The following figures illustrate two examples of such menu comparisons, where we focus on a setting with three prizes, so each lottery can be

[37]To illustrate, suppose that ex-post choices satisfy WARP and consider two menus A and B. WARP implies all lotteries chosen in menu $A \cup B$ are either all chosen in menu A or all chosen in menu B. As a result, when $\{p\} \succsim \{q\}$ for all $p \in A$ and $q \in B$, then for any lottery chosen in B, there is a better lottery chosen in $A \cup B$, and for any lottery chosen in $A \cup B$, there is a better lottery chosen in A. As such, A should be preferred over $A \cup B$, which should be preferred over B. Hence, IND is not needed by ex-post choices to derive WSB for menu choices.

represented by a point in the Marschak-Machina triangle.

Figure 2(a) depicts two menus $A = \{p, q\}$ and $B = \{p, q, p', q'\}$, where the dashed lines labeled $u(p)$ and $u(q)$ represent two indifference curves of the normative utility u and the arrow indicates the direction of improvement. Menu B is a union of A and its translation $A' = \{p', q'\}$ parallel to the indifference curves; that is, $p' - p = q' - q$ with $u(p') = u(p)$ and $u(q') = u(q)$. WSB makes no prediction about how the DM ranks menu A versus menu B, allowing for the possibility that the DM could strictly prefer either one of these menus. However, since B is a union of A and its translation A', a choice correspondence that satisfies IND leads to the choice of q (respectively, p) in menu A if and only if it leads to a choice of q or q' (respectively, p or p') in menu B. Intuitively, a DM who anticipates that her ex-post choices will be consistent with IND should therefore be indifferent between menus A and B, and this indifference is implied by Axiom 5.

(a) Menus A and B　　(b) Menus A" and B"

Figure 2: Monotonicity and WSB

Figure 2(b) depicts two other menus $A'' = \{r, p, q\}$ and $B'' = \{r, p, q, p', q'\}$, which are obtained by adding an alternative r to the menus A and B, respectively. Moreover, r satisfies $p = \alpha r + (1 - \alpha)q$ for some $\alpha \in (0, 1)$; hence, the DM strictly prefers lottery r to lottery p. Similar to the choice situation in Figure 2(a), a choice correspondence that satisfies IND leads to the choice of q in menu A'' if and only if it leads to the choice of q or q' in menu B''. On the other hand, if the choice correspondence leads to the choice of p, p' or r in menu B'', then it leads to the choice of only r in menu A''. As a result, a DM who anticipates that ex-post choices will be consistent with IND should not prefer menu B'' over menu A''. While WSB makes no prediction about how the DM ranks A'' versus B'', Axiom 5 implies that the DM weakly prefers A'' to B''.

A.5 DDC and ex-post choice

There are several models of internal conflicts in the literature that, in contrast to the implications of the self-regulation model, can imply a decreasing desire for commitment (DDC). For instance the convex self-control model in Noor and Takeoka [2010], and the willpower model in Masatlioglu, Nakajima, and Ozdenoren [2017] are two such models. Below we briefly discuss how these two models can also generate different ex-post choice patterns than the self-regulation model. In particular, we show that these models are incompatible with the choice patterns given in Section 7.

Convex self-control model

According to the convex self-control model in Noor and Takeoka [2010], the DM evaluates menu A in terms of $\max_{p\in A}\left(u(p)-\varphi\left(\max_{q\in A} v(q)-v(p)\right)\right)$, for $u,v\in\mathcal{V}$ and a convex function $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ with $\varphi(0)=0$. In particular, the convexity of φ leads to menu preferences that exhibit DDC. Now, consider Example 1, where there is a DM with normative utility that satisfies $u(p)>u(q)>u(r)$. We have argued that the DM chooses q from $\{p,q\}$, r from $\{q,r\}$, and yet chooses p from $\{p,q,r\}$ (violating WARP). These ex-post choice patterns are not consistent with the convex self-control model. To see this, note that $u(p)>u(q)$ and q chosen from $\{p,q\}$ imply that $v(q)>v(p)$ and $u(p)-u(q)<\varphi\left(v(q)-v(p)\right)$. Likewise, $u(q)>u(r)$ and r chosen from $\{q,r\}$ implies that $v(r)>v(q)$ and $u(q)-u(r)<\varphi\left(v(r)-v(q)\right)$. As a result, $u(p)-u(r)<\varphi\left(v(r)-v(q)\right)+\varphi\left(v(q)-v(p)\right)$. Since φ is convex and $\varphi(0)=0$, φ is superadditive and so $\varphi\left(v(r)-v(q)\right)+\varphi\left(v(q)-v(p)\right)\leq\varphi\left(v(r)-v(p)\right)$. Thus, $u(p)-u(r)<\varphi\left(v(r)-v(p)\right)$, and therefore the DM does not choose p from the menu $\{p,q,r\}$.

Willpower model

According to the willpower model in Masatlioglu, Nakajima, and Ozdenoren [2017], the DM chooses a lottery p from menu A which solves $\max_{p \in A} u(p)$ subject to $\max_{q \in A} v(q) - v(p) \leq w(p)$, for $u, v \in \mathcal{V}$ and $w : P \to \mathbb{R}_+$. This model can imply preferences over menus that satisfy DDC, for example, when w is constant. Now, consider Example 2, where there is a DM with normative utility that satisfies $u(p) > u(r)$. We have argued that the DM chooses p from $\{p, r\}$, and yet chooses $\frac{1}{4}r + \frac{3}{4}p$ from $\frac{1}{4}\{p, r\} + \frac{3}{4}\{p\}$ (violating IND). These ex-post choice patterns are not consistent with the willpower model. To see this, note that p chosen from $\{p, r\}$ implies $v(r) - v(p) \leq w(p)$. It follows that $\frac{1}{4}\left(v(r) - v(p)\right) \leq w(p)$, and therefore $v\left(\frac{1}{4}v(r) + \frac{3}{4}v(p)\right) - v(p) \leq w(p)$. As a result, the DM has enough willpower to choose p within menu $\frac{1}{4}\{p, r\} + \frac{3}{4}\{p\}$, and therefore does not choose the normatively worse alternative $\frac{1}{4}r + \frac{3}{4}p$.

References

Ahn, D., R. Iijima, Y. Le Yaouang, and T. Sarver (2017): "Behavioral Characterizations of Naivete for Time Inconsistent Preferences," Working Paper No.238, Economic Research Initiatives at Duke (ERID).

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*, 503–546.

Arrow, K. J. and L. Hurwicz (1972): *Uncertainty and Expectations in Economics: Essays in Honour of G.L.S. Shackle*, Oxford: Basil Blackwell, chap. An Optimality Criterion for Decision-Making under Ignorance, 1–11.

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.

Bernheim, B. D. and A. Rangel (2004): "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, 94, 1558–1590.

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

Buturak, G. and O. Evren (2017): "Choice Overload and Asymmetric Regret," *Theoretical Economics*, 12, 1029–1056.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini (2014): "Niveloids and Their Extensions: Risk Measures on Small Domains," *Journal of Mathematical Analysis and Applications*, 413, 343–360.

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 (2017): "Rationally Inattentive Preferences and Hidden Information Costs," *Theoretical Economics*, 12, 621–654.

Dean, M., O. Kibris, and Y. Masatlioglu (2017): "Limited Attention and Status Quo Bias," *Journal of Economic Theory*, 169, 93 – 127.

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

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

Dillenberger, D. and P. Sadowski (2012): "Ashamed to be Selfish," *Theoretical Economics*, 7, 99 – 124.

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

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

Epstein, L. G. and I. Kopylov (2007): "Cold Feet," *Theoretical Economics*, 2, 231–259.

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

——— (2015): "Hidden Actions and Preferences for Timing of Resolution of Uncertainty," *Theoretical Economics*, 10, 489–541.

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

Grant, S., S.-L. Hsieh, and M.-Y. Liang (2017): "Costly Self-Control and Limited Willpower," Working Paper.

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

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

——— (2007): "Harmful Addiction," *The Review of Economic Studies*, 74 (1), 147–172.

Herstein, I. N. and J. Milnor (1953): "An Axiomatic Approach to Measurable Utility," *Econometrica*, 291–297.

Huber, J., J. W. Payne, and C. Puto (1982): "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, 9, 90–98.

Huber, J. and C. Puto (1983): "Market Boundaries and Product Choice: Illustrating Attraction and Substitution Efffects," *Journal of Consumer Research*, 10, 31–44.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 263–291.

Koida, N. (2017): "Anticipated Stochastic Choice," *Economic Theory*, 1–30.

Kopylov, I. (2012): "Perfectionism and Choice," *Econometrica*, 80–5, 1819–1843.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 112(2), 443–477.

Lipman, B. and W. Pesendorfer (2013): *Advances in Economics and Econometrics: Tenth World Congress*, Cambridge University Press, vol. 1, chap. Temptation, 243–304.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): "When more is less: Limited consideration," *Journal of Economic Theory*, 170, 70 – 85.

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.

Manzini, P. and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82(3), 1153–1176.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed Attention," *American Economic Review*, 102(5), 2183–2205.

Masatlioglu, Y., D. Nakajima, and E. Ozdenoren (2017): "Willpower and Compromise Effect," Discussion Paper DP12354, Centre for Economic Policy Research.

MASATLIOGLU, Y. AND E. A. OK (2005): "Rational Choice with Status Quo Bias," *Journal of Economic Theory*, 121, 1–29.

MELLERS, B. A. AND K. BIAGINI (1994): "Similarity and Choice," *Psychological Review*, 101, 505–518.

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

NOOR, J. (2007): "Commitment and Self-Control," *Journal of Economic Theory*, 135, 1–34.

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

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

——— (2018): "Impatience as Selfishness," Working Paper.

O'DONOUGHE, T. AND M. RABIN (2001): "Choice and Procrastination," *The Quarterly Journal of Economics*, 116 (1), 121–160.

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

PHELPS, E. AND R. POLLAK (1968): "On Second-Best National Saving and Game-Equilibrium Growth," *The Review of Economic Studies*, 35(2), 185–199.

PRELEC, D. AND G. LOWENSTEIN (1991): "Decision Making over Time and under Uncertainty: A Common Approach," *Management Science*, 37, 770–786.

RIESKAMP, J., J. R. BUSEMEYER, AND B. A. MELLERS (2006): "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice," *Journal of Economic Literature*, 44, 631–661.

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.