Commitments and Weak Resolve[*†]

Igor Kopylov *Jawwad Noor*

March 8, 2016

Abstract

To avoid harmful temptations, people may voluntarily constrain their feasible choices. Such *commitments* require *resolve*, that is, a clear view of one's long-term normative goals and future temptations. Resolve can be weakened by temporary doubts about normative objectives and self-control costs. Such doubts can make people refuse commitments and later succumb to feasible temptations. We model such behavioral patterns and several types of resolve in a three-period extension of Gul and Pesendorfer's (2001) menu framework. Our main utility representation portrays an agent who is ex ante confident about her ex post normative objectives, but becomes less resolved at the interim stage when she is tempted to doubt her normative objectives. All components of this representation are derived in an essentially unique way from axioms imposed on choice behavior. We show by example that concerns about weak resolve can motivate the use of slack commitments and cold-turkey abstention methods.

[*]Early drafts of this paper were titled "Self-Deception and Choice".

[†]Kopylov is at the Dept of Economics, University of California, Irvine; E-mail: ikopylov@uci.edu. Noor is at the Dept of Economics, Boston University; Email: jnoor@bu.edu.
The authors would like to thank Larry Epstein, Bart Lipman, Stergios Skaperdas, Faruk Gul, and two anonymous referees as well as seminar participants at Harvard-MIT, NYU, Michigan, Zurich, UC San Diego, USC for useful comments. Noor would like to thank the economics dept at MIT for their hospitality where part of this research was completed. The usual disclaimer applies.

1 Introduction

People use various *commitments* (Schelling [23], Bryan, Karlan, and Nelson [7]) that avoid harmful temptations. For example, they may choose to keep only healthy foods and drinks at home, request self-exclusions from casino gambling, use the StayFocused application to limit Internet surfing, etc. Such commitments can be modeled via choices among *menus*—sets of consumption alternatives that become feasible later. In the menu framework, a ranking

$$\{x\} \succ \{x, y\} \tag{1}$$

describes an agent who commits ex ante to choose x ex post by making the possible temptation y unavailable.

Gul and Pesendorfer [12] (henceforth GP) model a preference for commitment like (1) for an agent who finds x to be normatively superior to y ex ante, but needs a costly *self-control* to choose x over the imminent temptation y ex post. In GP's two-period setting, commitments can be made ex ante without any costs.[1] Moreover, agents are assumed to have a clear view of their normative objectives, future temptations, and self-control costs. Such clarity of mind is called *resolve*. Behaviorally, it can be expressed by GP's Set-Betweenness axiom.

Despite its importance in self-regulation, resolve can be a difficult cognitive task. Weakness of resolve and the associated commitment problems are common among recovering addicts. The psychological literature describes various strategies people adopt in order to bolster their resolve and ignore their urges (Trope and Fishback [27] , Baumeister et al [5]).[2]

In this paper, we model agents who lose their initial resolve and start to doubt their normative objectives and temptations. We call such ambivalence *weak resolve*. For a behavioral expression of weak resolve, consider an agent who exhibits the ranking

$$a \cup \{y\} \succ_1 \{x\} \succ_1 \{y\} \tag{2}$$

[1] In multi-period extensions (e.g. Noor [20]), commitments can also require costly self-control to resist future temptations.

[2] For instance, an agent may self-impose rewards for good behavior and penalties for bad (Ainslie [1], Becker [6]). They may link the attainment of a goal with their sense of self-worth (Mischel [19]). They may manage their attention by distracting themselves (Mischel [18]) or by elaborating on the reasons why achieving the goal is important (Kuhl [16]).

even if she routinely succumbs to the temptation y in the menu $a \cup \{y\}$. For example, a compulsive gambler may normatively prefer to watch a movie (x) rather than excessively gamble (y). Yet she can doubt her normative objectives when she contemplates between going to the theatre $\{x\}$ versus going to the casino $a \cup \{y\}$. To justify her casino trip in her mind, the agent may engage in arguments like "maybe, it is okay for me to do some recreational gambling" or "I can easily resist the temptation y in favor of better normative alternatives in a". When in the casino, the agent may routinely succumb to y rather than exert self-control. In this example, we think of the existence of doubts about normative preference as an example of weak resolve.

Here the commitment ranking $\{x\} \succ_1 \{y\}$ reveals that the agent does have some persistence in her normative views that is sufficiently strong to choose a commitment to x rather than y. But the preference $a \cup \{y\} \succ_1 \{x\}$ reflects the influence of her self-talk that causes her to lose her clarity of mind and induce a belief that she can benefit from the flexibility of the menu $a \cup \{y\}$. Not withstanding this belief, the temptation y in the menu $\{x, y\}$ may be irresistible ex post when consumption is imminent.

1.1 Framework and Utility Representations

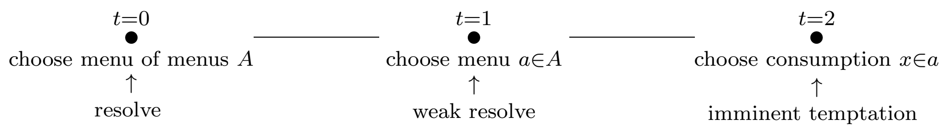
The combination of the ranking (2) and the subsequent choice of the temptation y in $a \cup \{y\}$ is inconsistent with all standard models of commitment and flexibility, including GP's costly self-control and the Strotz-like dual-self representations, the models of random temptations as in Dekel, Lipman, Rustichini [8] and multi-period dynamic temptations as in Kopylov [14] and Noor [20]. All of these models imply the ranking $\{y\} \succeq_1 a \cup \{y\}$ if y is selected in the menu $a \cup \{y\}$ with certainty. Moreover, if $a = \{x\}$, then the ranking $\{x, y\} \succ_1 \{x\} \succ_1 \{y\}$ violates GP's Set-Betweenness directly.[3]

By contrast, we assume that the choice of commitment $\{x\}$ over the menu $a \cup \{y\}$ can require a cognitive effort beyond resisting the temptating appeal of y. Hence, the ranking (2) can hold even if y is chosen in $a \cup \{y\}$ ex post.

We add a third time period to accommodate this departure from standard models. The time line is as follows. Temptation is experienced in the ex post stage at the time of choice from a menu, and potentially also in the interim

[3] In this case, the model of perfectionism in Kopylov [15] does not accommodate (2) either because x is the normatively best element in $\{x, y\}$.

stage where the agent picks a menu. In the ex ante stage she is in a cold resolved state and accurately anticipates future choices and self-control costs. This perspective dictates her preference over the *menus of menus* she will face in the interim period. The corresponding ex ante preferences $\succeq$ over menus of menus is our primitive. The interim choices $a \in A$ and ex post choices $x \in a$ are described by observable choice sets $C(A)$ and $C(a)$.



For example, the binary relation $\succeq_1$ in (2) represents interim choices in menus A, and the ex post choice of y in $a \cup \{y\}$ is written as $C(a \cup \{y\}) = \{y\}$.

Formally, we extend GP's model of temptation and self-control to three time periods (Theorems 1 and 2 below). The utility representation for ex ante preferences over menus of menus A is

$$U_0(A) = \max_{a \in A} \left[U(a) - \max_{b \in A}(V(b) - V(a)) \right],$$

where U, V are linear utility over menus and

$$U(a) = \max_{x \in a}[u(x) - \max_{y \in a}(v(y) - v(x))],$$

for some expected utility indices u and v over consumption lotteries $x, y, \ldots$. Analogously to GP's setting, the agent's choices $a \in A$ and $x \in a$ at the interim and ex post stages maximize the corresponding normative utilities U and u net of self-control costs arising from temptations V and v respectively. GP's axioms imply resolve at the ex ante stage, but impose no structure on the temptation indices V and v beyond continuity and linearity.

The choices at the interim and ex post stages are described by

$$C(A) = \arg\max_{a \in A}[U + V](a) = \arg\max_{a \in A} W(a)$$
$$C(a) = \arg\max_{x \in a}[u + v](a) = \arg\max_{x \in a} w(x).$$

Theorem 2 characterizes these representations. The freedom in the specification of V allows interim choices to maximize an arbitrary linear and continuous function W.

Our main results (Theorem 3 and 4) focus on the structure of the interim functions V and W that can be interpreted in terms of weak resolve. We characterize three nested functional forms:

$$V(a) = \kappa U(a) + \sum_{i=1}^{m} \max_{x \in a} u_i(x), \qquad \text{for } \kappa \geq 0, \tag{3}$$

$$V(a) = \kappa U(a) + \lambda \max_{y \in a} v(y), \qquad \text{for } \kappa \geq 0, \lambda > 0, \tag{4}$$

$$V(a) = \kappa U(a) + \lambda \max_{y \in a} v(y), \qquad \text{for } \kappa \geq 0 \text{ and } \lambda \in (0, \kappa + 1], \tag{5}$$

where $u_1, \ldots, u_m$ are linear continuous functions. The temptation utility V is therefore determined by a combination of normative perspective U and other factors u_i. This combination captures a tension between the agent's resolve to maximize U and various doubts that she may have about her normative objectives and future self-control costs. Intuitively, while her doubts urge her to behave in ways that deviate from U, her resolve curbs these urges s captured by V. Behaviorally, these doubts can make flexibility tempting even without adding greater ex post temptations in representation (3). Special cases (4) and (5) make the agent's interim doubts more specific and aligned with future temptations. Representations (3)–(5) are essentially unique under suitable regularity conditions on the functions u, v, and $u_1, \ldots, u_m$.

The axioms that characterize the nested functional forms (3)–(5) are imposed for all interim menus a and b:

- Monotone Temptations (MT): $\{a \cup b, b\} \succeq \{a \cup b\}$,
- Persistent Temptations (PT): $\{a\} \succ \{a, b\} \Rightarrow \{a\} \succ \{a \cup b\}$,
- Preference for Earlier Decisions (PED): $\{a, b\} \succeq \{a \cup b\}$.

Each of these conditions is stronger than the previous one and captures a more specific form of resolve. Moreover, these axioms, especially MT and PED, appear sufficiently tractable for empirical tests.

1.2 Temptations vs Weak Resolve

In our model, despite wavering in the interim period, the agent maintains some of her ex ante normative perspective. We have interpreted this as an expression of resolve, albeit weakened by doubts. There is also a second expression in our model. All our representations exhibit the feature that some future temptations are ignored by the agent, as opposed to resisted via costly self-control. This admits an interpretation in terms of resolve as well. For instance, by not taking antabuse tablets, an alcoholic has an opportunity to indulge later. An alcoholic without any resolve may be tempted to not comply with her treatment: she may be tempted to forgo the antabuse tablets since she may lose sight of her normative goals. An alcoholic with strong resolve, on the other hand, may be sufficiently focused on her normative goals to the extent that she may not even be tempted by noncompliance - her resolve may manifest in *urge-control*.

Behaviorally, this is expressed by a violation of the so-called Temptation Stationarity axiom used in Kopylov [14] and Noor [20], which asserts the equivalence

$$\{a\} \succ \{a, b\} \quad \Leftrightarrow \quad \{a\} \succ \{a \cup b\},$$

that is, menu b tempts if and only if it contains tempting alternatives. The rankings

$$\{a\} \succ \{a \cup b\} \text{ and } \{a\} \not\succ \{a, b\}$$

are interpreted as a possible behavioral expression of strong resolve. This interpretation is most relevant when the agent exhibits some such instances, while also exhibiting other instances where there are menus a', b' such that $\{a'\} \succ \{a' \cup b'\}$ and $\{a'\} \succ \{a', b'\}$. In the representations (3)–(5), the urge-control created by strength of resolve is expressed in the fact that normative utility U of a menu tempers the value of its temptation utility V.

The remainder of this paper proceeds as follows. Section 2 introduces the primitives of the model and presents a benchmark three-period extension of GP's model. Section 3 presents our main results. Sections 4 contain applications and discusses related literature. Proofs are relegated to appendices.

2 Temptations and Self-Control in a Three-Period Framework

Adapt GP's menu framework to three time periods—ex ante, interim, and ex post. Let $X = \{x, y, z, \dots\}$ be the set of all Borel probability measures on a compact set Z of deterministic consumptions. Endow X with the weak convergence topology and the Prohorov metric.

Let $\mathcal{M}_1 = \{a, b, c, \dots\}$ be the set of all non-empty compact subsets $a \subset X$. Endow $\mathcal{M}_1$ with the Hausdorff metric. Let $\mathcal{M}_0 = \{A, B, C, \dots\}$ be the set of all non-empty compact subsets $A \subset \mathcal{M}_1$. Elements $A \in \mathcal{M}_0$ and $a \in \mathcal{M}_1$ are both called *menus* and distinguished by the upper and lower cases in the notation. Endow $\mathcal{M}_0$ with the Hausdorff metric. Define mixtures

$$\alpha a + (1-\alpha) b = \{\alpha x + (1-\alpha) y \ : \ x \in a, \ y \in b\}$$
$$\alpha A + (1-\alpha) B = \{\alpha a + (1-\alpha) b \ : \ a \in A, \ b \in B\}$$

for all $\alpha \in [0,1]$ and menus $a, b \in \mathcal{M}_1$ and $A, B \in \mathcal{M}_0$. Then both $\mathcal{M}_0$ and $\mathcal{M}_1$ are compact (see Theorem 3.71 in Aliprantis and Border [2]) and the mixture operations in these spaces are continuous.

Assume that consumption lotteries $x \in X$ are resolved after the ex post stage. Interpret any menu $a \in \mathcal{M}_1$ as an interim action that restricts the ex post choice to the set $a \subset X$. Similarly, interpret any menu $A \in \mathcal{M}_0$ as an ex ante action that restricts the interim choice to the set $A \subset \mathcal{M}_1$.

Let a binary relation $\succeq$ on $\mathcal{M}_0$ be the agent's ex ante weak preference over menus $A \in \mathcal{M}_0$. Write the symmetric and asymmetric parts of this relation as $\sim$ and $\succ$ respectively.

Adapt GP's list of axioms for the preference $\succeq$.

Axiom 1 (Order)**.** $\succeq$ *is complete and transitive.*

Axiom 2 (Continuity)**.** *For all menus $A \in \mathcal{M}_0$, the sets $\{B \in \mathcal{M}_0 : B \succeq A\}$ and $\{B \in \mathcal{M}_0 : B \preceq A\}$ are closed.*

Axiom 3 (Independence)**.** *For all $\alpha \in [0,1]$ and menus $A, B, C \in \mathcal{M}_0$,*

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

Axiom 4 (Set-Betweenness)**.** *For all menus $a, b \in \mathcal{M}_1$ and $A, B \in \mathcal{M}_0$,*

$$\{a\} \succeq \{b\} \quad \Rightarrow \quad \{a\} \succeq \{a \cup b\} \succeq \{b\},$$
$$A \succeq B \quad \Rightarrow \quad A \succeq A \cup B \succeq B.$$

Order and Continuity are standard. To motivate Independence, assume indifference between the menu $\alpha A + (1-\alpha) C$ and a hypothetical lottery $\alpha \circ A + (1-\alpha) \circ C$ that is resolved immediately after the ex ante stage and yields the menus A or C with probabilities α and $1-\alpha$ respectively.[4] Then Independence can be explained by the standard separability argument.

Set-Betweenness is imposed separately over menus $A \in \mathcal{M}_0$ and over singleton menus $\{a\}$ that commit the interim choice to be $a \in \mathcal{M}_1$. Assume that the agent evaluates menus A and $\{a\}$ via her anticipated

- interim choice $c \in A$ and strongest temptation $b \in A$,
- ex post choice $x \in a$ and strongest temptation $y \in a$

respectively. Then both parts of Set-Betweenness in our three-period framework can be motivated exactly as in GP's two-period setting.

To establish uniqueness in the representation results below, assume

Axiom 5 (Generic Temptations (GT))**.** *There exist $x, y, x', y' \in X$ and $a, b, a', b' \in \mathcal{M}_1$ such that*

$$\{\{x\}\} \sim \{\{x,y\}\} \succ \{\{y\}\} \quad \textit{and} \quad \{\{x'\}\} \succ \{\{x',y'\}\}$$
$$\{a\} \sim \{a,b\} \succ \{b\} \quad \textit{and} \quad \{a'\} \succ \{a',b'\}.$$

The ranking $\{\{x\}\} \sim \{\{x,y\}\} \succ \{\{y\}\}$ asserts that normatively inferior alternatives y need not necessarily tempt x ex post. For example, if y combines x with a monetary penalty, then it is compelling that y should be both normatively inferior and less tempting than x. The ranking $\{a\} \sim \{a,b\} \succ \{b\}$ has a similar motivation in terms of interim choices.

The ranking $\{\{x'\}\} \succ \{\{x',y'\}\}$ excludes the case when the agent is never tempted ex post. In this case the function U has the form $U(A) = \max_{x\in a} u(x)$. Similarly, the ranking $\{a'\} \succ \{a',b'\}$ excludes the case when the agent has no interim temptations, In this case, $U_0(A) = \max_{a\in A} U(a)$.

Say that a function $u : X \to \mathbb{R}$ is *linear* if for all $\alpha \in [0,1]$ and $x, y \in X$,

$$u(\alpha x + (1-\alpha)y) = \alpha u(x) + (1-\alpha)u(y).$$

Let $\mathcal{U}$ be the set of all continuous linear functions $u : X \to \mathbb{R}$. Similarly, define the set $\mathcal{U}_1$ of all continuous linear functions $V : \mathcal{M}_1 \to \mathbb{R}$.

[4]GP assume the same indifference in the two-period setting. In our framework, the agent's interim choice $\alpha a + (1-\alpha)c$ in $\alpha A + (1-\alpha)C$ and her ex post choice $\alpha x + (1-\alpha)y$ in $\alpha a + (1-\alpha)c$ determine her consumptions $x \in a \in A$ and $y \in c \in C$ contingent on the resolution of the lottery between the menus A and C after the ex post stage. Indifference between the early and late resolution of this lottery implies indifference between $\alpha A + (1-\alpha)C$ and $\alpha \circ A + (1-\alpha) \circ C$.

Theorem 1. *$\succeq$ satisfies Axioms 1–4 if and only if $\succeq$ is represented by a utility function U_0 such that for all $A \in \mathcal{M}_0$ and $a \in \mathcal{M}_1$,*

$$\begin{aligned} U_0(A) &= \max_{a \in A} \left[U(a) - \max_{b \in A} (V(b) - V(a)) \right] \\ U(a) &= \max_{x \in a} [u(x) - \max_{y \in a} (v(y) - v(x))], \end{aligned} \tag{6}$$

where $u, v \in \mathcal{U}$ and $V \in \mathcal{U}_1$.

Moreover, if $\succeq$ satisfies GT, then representation (6) is unique up to a positive linear transformation of the tuple (u, v, V): any other representation (6) with tuple (u', v', V') can be written as $u' = \alpha u + \beta_u$, $v' = \alpha v + \beta_v$, and $V' = \alpha V + \beta_V$ for some $\alpha > 0$ and $\beta_u, \beta_v, \beta_V \in \mathbb{R}$.

This result is a direct consequence of Kopylov [14, Theorem 1]. GP's representations (6) are derived jointly for two utility functions U_0 on $\mathcal{M}_0$ and U on $\mathcal{M}_1$. The *commitment utility* indices U and u represent the agent's ex ante normative perspective for the interim and ex post choices. The components $\max_{b \in A}(V(b) - V(a))$ and $\max_{y \in a}(v(y) - v(x))$ can be interpreted as non-negative self-control costs that the agent first incurs to choose $a \in A$ rather than $b \in A$, and then $x \in a$ rather than $y \in a$.

Note that Theorem 1 does not relate the interim temptation index V with ex post indices u and v. For example, if $v = 0$, then the agent expects to maximize u ex post and obeys *strategic rationality*

$$\{a\} \succeq \{b\} \quad \Rightarrow \quad \{a\} \sim \{a \cup b\}$$

for all $a, b \in \mathcal{M}_1$. However, she may still be tempted at the interim stage and exhibit a preference for commitment $A \succ A \cup B$ for some menus $A, B \in \mathcal{M}_0$.

2.1 Interim and Ex Post Choices

The above interpretations for representations (6) suggest that interim and ex post choices should strike an optimal compromise between long-term normative objectives and costs of resisting temptations. Therefore, agents should maximize the functions

$$\begin{aligned} W(a) &= U(a) + V(a) \\ w(x) &= u(x) + v(x). \end{aligned}$$

respectively.

To model interim and ex post choices formally, consider an additional primitive. For any menus $A \in \mathcal{M}_0$ and $a \in \mathcal{M}_1$, let $C(A) \subset A$ and $C(a) \subset a$ be non-empty sets of all alternatives that the decision maker can accept in menus A and a at the interim and ex post stages respectively. For notational convenience, we use the same symbol $C(\cdot)$ to describe the observable choices for both time periods.

Impose two standard conditions on the choice rule $C(\cdot)$.

Axiom 6 (Weak Axiom of Revealed Preference (WARP))**.** *For all $A, B \in \mathcal{M}_1$, $a, b \in \mathcal{M}_0$, and $x, y \in X$,*

$$\begin{array}{lcl} x \in C(a),\ y \in a,\ y \in C(b),\ x \in b & \Rightarrow & x \in C(b) \\ a \in C(A),\ b \in A,\ b \in C(B),\ a \in B & \Rightarrow & a \in C(B). \end{array}$$

Axiom 7 (Closed Graph)**.** *The set*

$$\{(A, a, x) : A \in \mathcal{M}_0, a \in C(A), x \in C(a)\}$$

is closed in $\mathcal{M}_0 \times \mathcal{M}_1 \times X$.

Arrow [4] shows that WARP is necessary and sufficient for the choice rule $C(\cdot)$ at the interim and ex post stages to be rationalized by some complete and transitive preferences $\succeq_1$ and $\succeq_2$ respectively. Arrow's result applies here because menus include arbitrary finite sets. Closed Graph implies that $\succeq_1$ and $\succeq_2$ are continuous.

Axiom 8 (Consistency)**.** *For all $A \in \mathcal{M}_0$, $a \in \mathcal{M}_1$, and $x \in X$,*

$$\begin{array}{lcl} A \succ A \setminus \{a\} & \Rightarrow & C(A) = \{a\} \\ \{a\} \succ \{a \setminus \{x\}\} & \Rightarrow & C(a) = \{x\}. \end{array} \tag{7}$$

This condition requires that the agent may strictly benefit ex ante from a feasible alternative $a \in A$ or $x \in a$ only if she actually chooses a in A or x in a respectively at the interim stage or ex post stages respectively. This condition is similar to Axiom 4 (Approximate Improvements are Chosen) in Dekel, Lipman and Rustichini [8].

Theorem 2. *Suppose that $\succeq$ satisfies axioms 1–5 and is represented by* (6) *with a tuple (u, v, V). Then a choice rule $C(\cdot)$ satisfies WARP, Closed Graph, and Consistency if and only if for all $A \in \mathcal{M}_0$ and $a \in \mathcal{M}_1$,*

$$\begin{aligned} C(A) &= \arg\max_{a\in A}[U+V](a) = \arg\max_{a\in A} W(a) \\ C(a) &= \arg\max_{x\in a}[u+v](a) = \arg\max_{x\in a} w(x). \end{aligned} \tag{8}$$

Together with Theorem 1, this result characterizes consistent representations for choice behavior in all three time periods. Note that the temptation function V in Theorem 2 is still an arbitrary linear, continuous, non-constant function that does not have any relation with the ex post indices u and v. This freedom in the specification of V implies that the function W maximized at the interim stage need not have any additional structure beyond linearity and continuity. Thus the interim choice behavior can be consistent with strategic rationality, GP's model, and various extensions of this model to random and multiple temptations (e.g. Stovall [24]).

Our next goal is to obtain a more specific structure for V that relates it with normative and temptation indices.

3 Main Representation Results

To refine the three-period model (6), consider first a simple monotonicity condition on interim temptations.

Axiom 9 (Monotone Temptations (MT))**.** *For all $a, b \in \mathcal{M}_1$,*

$$\{a \cup b, b\} \succeq \{a \cup b\}.$$

The exclusion of the ranking $\{a \cup b\} \succ \{a \cup b, b\}$ means that the smaller menu b cannot tempt the more flexible one $a \cup b$. This constraint is plausible because all consumptions in b that the decision maker may possibly consider at the interim stage are also feasible in $a \cup b$. Implicit in this interpretation is that an agent's interim choice of menu is based on the use of rationales that translate into maximizing a ranking of alternatives. Note that the agent has only one rationale ex ante (the normative preference) but multiple ones arise in the interim period. The intepretation is that she begins to doubt her normative objectives because her resolve weakens as the agent gets closer to the moment of consumption (the ex post period).

Next we adapt the Finiteness axioms of Kopylov [13] for the three-period framework.

Axiom 10 (Finiteness)**.** *There is n such that for any sequence of n menus $a_1, \ldots, a_n \in \mathcal{M}_1$, there is $j \in \{1, \ldots, n\}$ such that for all $A \in \mathcal{M}_0$,*

$$\left\{ \bigcup_{i \in \{1,\ldots,n\}} a_i \right\} \cup A \sim \left\{ \bigcup_{i \in \{1,\ldots,n\} \setminus \{j\}} a_i \right\} \cup A.$$

This axiom must hold if the agent's interim evaluation of any menu a is determined by finitely many factors $\max_{x \in a} u_k(x)$.

Say that a list of functions $u_1, \ldots, u_n \in \mathcal{U}$ is *redundant* if $u_i = \alpha u_j + \beta$ for some $j \neq i$, $\alpha \geq 0$ and $\beta \in \mathbb{R}$. By convention, any sum over an empty index set is zero.

Theorem 3. *$\succeq$ satisfies Axioms 1–5, Finiteness and MT if and only if $\succeq$ is represented by a utility function U_0 such that for all $A \in \mathcal{M}_0$ and $a \in \mathcal{M}_1$,*

$$U_0(A) = \max_{a \in A} \left[U(a) - \max_{b \in A}(V(b) - V(a)) \right] \tag{9}$$

$$U(a) = \max_{x \in a}[u(x) - \max_{y \in a}(v(y) - v(x))], \tag{10}$$

$$V(a) = \kappa U(a) + \sum_{i=1}^{m} \max_{x \in a} u_i(x), \tag{11}$$

where $\kappa \geq 0$, the lists of functions $u + v, u_1, \ldots, u_m \in \mathcal{U}$ and $u + v, v, u$ are not redundant, and $V \neq \kappa U$.

Moreover, $\succeq$ has another representation (9)–(11) with parameters $\kappa' \geq 0$ and functions $u', v', u'_1, \ldots, u'_{m'} \in \mathcal{U}$ if and only if $\kappa' = \kappa$, $m' = m$, $u' = \alpha u + \beta_u$, $v' = \alpha v + \beta_v$, and $u'_1, \ldots, u'_m$ can be permuted by some $\pi : \{1, \ldots, m\} \to \{1, \ldots, m\}$ so that $u'_{\pi(i)} = \alpha u_i + \beta_i$ for some $\alpha > 0$ and $\beta_u, \beta_v, \beta_i \in \mathbb{R}$.

This result derives the added structure (11) for the temptation utility V from the MT axiom. This structure suggests that the interim temptation appeal $V(a)$ of a menu a combines the normative utility $U(a)$ and the maxima of several other factors $u_1, \ldots, u_m$ in a.

The interpretation is as in the Introduction. Note that the functions $u_1, \ldots, u_m$ may include positive multiples μu and λv of the normative and temptation functions u and v respectively with parameters $\mu > 0$ and $\lambda > 0$.

The corresponding representation for interim choices is

$$W(a) = (\kappa + 1)U(a) + \lambda \max_{y \in a} v(y) + \mu \max_{z \in a} u(z) + \sum_{k=1}^{m} \max_{x \in a} u_k(x).$$

Here the decision maker's interim choices can be directly swayed by future temptations that she may fail to recognize as such. Moreover, she may be misled by "good intentions": the desire to maximize the normative function u without taking future self-control costs into consideration. Finally, her choices can be also guided by other factors $u_1, \ldots, u_m$ that differ from $u + v$, u, and v.

3.1 Specializations

Consider now a special case where the agent's only possible urge at the interim stage is to maximize her future temptation utility. In other words, the agent may doubt the distinction between normative and temptation preferences at the interim stage. In the absence of other doubts, she should comply with

Axiom 11 (Persistent Temptations (PT))**.** *For all $a, b \in \mathcal{M}_1$,*

$$\{a\} \succ \{a, b\} \quad \Rightarrow \quad \{a\} \succ \{a \cup b\}.$$

This condition requires that if a menu b is more tempting than another menu a at the interim stage, then it should also provide a greater temptation ex post. Unlike MT, interim temptations must be coherent with the ex post temptation index v. Note that PT implies MT because the ranking $\{a \cup b\} \succ \{a \cup b, b\}$ implies $\{a \cup b\} \succ \{a \cup b\}$, which is a contradiction. On the other hand, PT is weaker than the Temptation Stationarity in Kopylov [14] and Noor [20], which asserts the equivalence

$$\{a\} \succ \{a, b\} \quad \Leftrightarrow \quad \{a\} \succ \{a \cup b\}.$$

The content of PT is that it does not require that ex post temptations must necessarily be tempting at the interim stage. In fact, the rankings $\{a\} \succ \{a \cup b\}$ and $\{a\} \not\succ \{a, b\}$ may reflect *urge-control* at the interim stage that arises from strength of resolve. For instance, a smoker may successfully ignore his urge to view smoking as normatively acceptable and wear a nicotine patch at the interim stage, but fail to resist smoking if he does not wear the patch ex post.

Another way to relate interim doubts and ex post temptations is

Axiom 12 (Preference for Earlier Decisions (PED))**.** *For all* $a, b \in \mathcal{M}_1$*,*

$$\{a, b\} \succeq \{a \cup b\}$$

This axiom assumes that the agent should weakly prefer ex ante to make the choice between the menus a and b at the interim stage rather than ex post. Note that any consumption x in $a \cup b$ can be made in two cognitive steps by selecting either a or b first, and then picking x in the selected submenu. The interpretation is that doubts are easier to resist than direct temptations: resisting the doubts about the normative harm of smoking is easier than resisting the urge to smoke when cigarettes are immediately available. Alternatively, one can interpret PED in terms of temporal distance to final temptations. The notion that an agent can make better choices from a distance is common in the literature (GP [12], Noor [21]).

Extending the idea of PED, one can assume that for all $n = 1, 2, \ldots$ and menus $a_1, \ldots, a_n \in \mathcal{M}_1$,

$$\{a_1, \ldots, a_n\} \succeq \left\{ \bigcup_{i=1 \ldots n} a_i \right\}.$$

The stronger axiom turns out to be equivalent to PED in our model because it is satisfied by the utility representation that is characterized by PED in Theorem 3 below.[5]

As $\{a\} \succ \{a, b\}$ implies $\{a\} \succ \{a, b\} \succeq \{a \cup b\}$, then PED implies PT, and the three axioms are nested:

$$\text{PED} \quad \Rightarrow \quad \text{PT} \quad \Rightarrow \quad \text{MT}.$$

Theorem 4. $\succeq$ *satisfies Axioms 1–5 and PT if and only if* $\succeq$ *is represented by* (9) *where for all* $a \in \mathcal{M}_1$*,*

$$U(a) = \max_{x \in a}[u(x) - \max_{y \in a}(v(y) - v(x))], \tag{12}$$

$$V(a) = \kappa U(a) + \lambda \max_{y \in a} v(y), \tag{13}$$

[5] PED is also implied by the combination of Stovall's [25] axioms Monotonicity and Option to Commit in his model of uncertain temptations.

where $\kappa \geq 0$, $\lambda > 0$, and the functions $u+v, u, v \in \mathcal{U}$ are not redundant.

If PED holds as well, then representation (13) *must hold with $\lambda \leq \kappa + 1$.*

These representations are unique up to a positive linear transformation of the functions u and v.

This result restricts the specification of V to just two components: the normative component $U(a)$ and the appeal of future temptations $\max_{x \in a} v(x)$. Note that Finiteness is not assumed, but is implied by the list of axioms in Theorem 4.

The interim utility W takes the form

$$W(a) = U(a) + V(a) = (\kappa + 1) \max_{x \in a} w(x) + (\lambda - \kappa - 1) \max_{y \in a} v(y).$$

If $\lambda > \kappa + 1$, then this representation predicts a strict preference for flexibility $C(\{\{x\}, \{y\}, \{x, y\}\}) = \{\{x, y\}\}$ for x and y such that $w(x) > w(y)$ and $v(y) > v(x)$.

If $1 + \kappa > \lambda$, then

$$W(a) = (1 + \kappa - \lambda) \max_{x \in a} \left[u^*(x) - \max_{y \in a}(v(y) - v(x)) \right].$$

where $u^* = \frac{1+\kappa}{1+\kappa-\lambda} u + \frac{\lambda}{1+\kappa-\lambda} v$. In this case, interim choices represented by W conform to GP's model where the ex ante normative perspective u is distorted to u^* in the direction of the temptation utility v. The agent as portrayed by this representation expects her interim self to relax her normative standards to u^*, but maintain her ex ante evaluation of future self-control costs.

The limiting case when $1 + \kappa = \lambda$ and

$$W(a) = \max_{x \in a} [u(a) + v(a)]$$

portrays an agent who relaxes her normative standards to $u^* = u + v$ and ignores any possibility of future temptation. It is as if by adjusting her normative perspective she resolves all internal conflicts.

To get another interpretation for the parameters κ and λ in terms of choice behavior, consider a pair of preferences $\succeq$ and $\succeq^*$ over $\mathcal{M}_0$. Call this pair *comparable* if both $\succeq$ and $\succeq^*$ satisfy Axioms 1–5 and PT, and the two rankings agree on the domain of singleton menus so that

$$\{a\} \succeq \{b\} \quad \Leftrightarrow \quad \{a\} \succeq^* \{b\}$$

for all $a, b \in \mathcal{M}_1$.

By Theorem 3, any comparable pair of preferences $\succeq$ and $\succeq^*$ can be represented by (12)-(13) with components (u, v, κ, λ) and $(u^*, v^*, \kappa^*, \lambda^*)$ respectively. Moreover, the functions U and U^* represent the same preference on $\mathcal{M}_1$ and hence, by GP's Theorem, one can take $u = u^*$ and $v = v^*$.

Say that $\succeq^*$ is *less resolved* than $\succeq$ if for all menus $a, b \in \mathcal{M}_1$,

$$\{a\} \succ \{a, b\} \quad \Rightarrow \quad \{a\} \succ^* \{a, b\}. \tag{14}$$

In other words, if interim resolve reveals doubts that a is normatively better than b for the preference $\succeq$, then it does so for $\succeq^*$ as well. In that sense $\succeq$ is more successful at urge-control.

Theorem 5. *Let $\succeq$ and $\succeq^*$ be a comparable pair of preferences. Then $\succeq^*$ is less resolved than $\succeq$ if and only if the two preferences have representations (12)-(13) such that $\frac{\kappa^*}{\lambda^*} \leq \frac{\kappa}{\lambda}$.*

This result suggests that the ratio $\frac{\kappa}{\lambda}$ is positively related with the effectiveness of resolve in controlling urges and can be interpreted as a degree of resolve.

4 Discussion

4.1 Slack Commitments and Abstention

While *binding* commitments make it physically impossible to succumb to temptations in a given time period, many other commitments, which we call *slack*, keep temptations feasible within the relevant time period. Instead of physical barriers, slack commitments impose emotional or monetary penalties for succumbing to temptations. For example, people often make promises and vows to themselves or to other individuals. Such resolutions can be emotionally costly to break, but put no physical constraints on behavior. There are financial commitments (such as illiquid real-estate or retirement investments in Laibson [17]) that make overspending more costly, albeit still possible. By definition, *deadlines* are slack commitments as well: it is always feasible to ignore them, but doing so incurs various penalties. Yet people impose deadlines on themselves (see Ariely and Wertenbroch [3]). In the long run, almost all commitments must be slack because most barriers to temptations are transient in space or time.

In principle, slack commitments can be modeled via choices among two-period menus. For instance, a ranking

$$\{x, y \ominus c\} \succ \{x, y\} \tag{15}$$

describes an agent who commits ex ante to pay a penalty c if she chooses y over x ex post. Here $y \ominus c$ denotes the combination of the consumption y and the penalty c.

Similarly to Strotz [26], one can explain slack commitments by *non-stationary preferences*. For example, ranking (15) is plausible for agents who favor x over y ex ante, expect to choose y over x ex post, but find the penalty c sufficient to make $y \ominus c$ less desirable than x ex post. GP's model can accommodate (15) as well when the imposed penalty c can alleviate the self-control that the agent needs ex post to resist the temptation y in the menu $\{x, y\}$ in favor of the normatively superior x.

Yet, slack commitments derived from these standard arguments have some problematic features. First, such commitments must be always kept.[6] But personal resolutions and deadlines are routinely broken: Bryan, Karlan, and Nelson [7] quote a survey where only 12% of subjects managed to keep any of their New Year resolutions. Gine, Karlan, and Zinman [11] report that more than 80% of their subjects restarted smoking after making financial commitments not to do so.

Second, the standard models predict that if $q^* > 0$ is a normatively optimal level of some tempting consumption, then agents should never opt for complete abstinence via a *cold-turkey* slack commitment that imposes a penalty c on any positive consumption $q > 0$. In both models of Strotz and GP, it should be strictly better to use a more permissive commitment that imposes the same penalty c but only if the tempting consumption exceeds q^*. However, cold-turkey commitments are popular, presumably due to the sentiment echoed in St. Augustine's quote that "complete abstinence is easier than perfect moderation". Indeed, gamblers may swear off gambling altogether even if they find recreational value in small-scale wagers, and dieters often prohibit even a little consumption of unhealthy foods.

Our *three-period* extension of GP's model predicts that it may be optimal for agents to use slack commitments that

[6] More precisely, an agent who plans to succumb to the temptation $y \ominus c$ in the menu $\{x, y \ominus c\}$ ex post must have a ranking $\{x, y\} \sim \{y\} \succ \{y \ominus c\} \sim \{x, y \ominus c\}$ ex ante in both models.

- get broken in the presence of ex post temptations,
- require full abstention rather than perfect moderation.

For example, an agent can go on a diet that completely excludes unhealthy foods in order to motivate herself not to buy such foods in grocery stores, but she may still expect to break her commitment when immediate temptations are available in a restaurant or at a party.

The idea is that the mild penalties may suffice to stave off urges in the interim period via resolve, even if they would be insufficient to battle temptation ex post. Full abstention may also be more effective than perfect moderation in affecting interim urges in the interim period. In both cases, the dependence of temptation utility on normative value is exploited by hurting the normative value of a menu appropriately in order to manage urges.

We illustrate the above arguments with two examples.

Example 1. Consider an agent who can avoid some ex post temptations (like staying on the couch) through interim commitments (say by going to the gym). Let the agent's ex post menu be $a = \{x, y\}$ if she keeps the temptation y feasible, and $b = \{x, z\}$ if she goes to the gym. Here x denotes a little exercise and z denotes a full-blown workout with the gym equipment. Let

	$x \ominus c$	$y \ominus c$	z
u	$20 - c$	$-c$	30
v	50	100	0

Here the temptation function v is unaffected by the commitment penalties, and the lack of resolve manifests itself through factors $u_1 = 2u$ and $u_2 = v$. Let $\kappa = 0$. Let $a_c = \{x \ominus c, y \ominus c\}$ and $b_c = \{x \ominus c, z\}$. Then

$$U_0(\{a_c, b_c\}) = \begin{cases} W(a_c) - V(a_c) = U(a_c) = u(y \ominus c) = -c & \text{if } c < 5 \\ W(b_c) - V(a_c) = -10 + c & \text{if } c \in [5, 15] \\ W(b_c) - V(b_c) = U(b_c) = 20 - c & \text{if } c \geq 15. \end{cases}$$

This utility function is maximized by taking $c = 15$. In this case b_c is chosen at the interim stage, and $x \ominus c$ is picked ex post. Thus the slack commitment is used to police the interim choice, and is broken ex post. If the penalty is imposed only on the temptation y, then the self-control costs cannot be decreased at the interim stage by the penalty c.

Example 2. Suppose that the agent has the interim choice between commitment to full abstention $a = \{x\}$ and access to tempting consumption $b = \{x, z, y\}$, where z represents perfect moderation and y is a harmful level of consumption. For example, b may describe keeping a pack of cigarettes at home, and z having one cigarette with the morning coffee, and y smoking the whole pack in a day. Let

	x	$y \ominus c$	$z \ominus c$
u	20	$-c$	$30 - c$
v	0	100	50

Let $u_1 = 2u$ and $u_2 = v$. Let $\kappa = 0$. Let $b_c = \{x, y \ominus c, z \ominus c\}$. Then

$$U_0(\{a, b_c\}) = \begin{cases} W(b_c) - V(b_c) = U(b_c) = u(y \ominus c) = -c & \text{if } c \leq \frac{100}{3} \\ W(a) - V(b_c) = -100 + 2c & \text{if } c \in [\frac{100}{3}, 60] \\ W(a) - V(a) = U(a) = 20 & \text{if } c \geq 60. \end{cases}$$

This utility function is maximized by taking $c \geq 60$. In this case, the commitment to full abstention is slack at the ex ante stage and becomes binding at the interim stage when a is chosen. Note that if c is imposed only on y, then U_0 does not increase.

4.2 Related Literature

The idea that temptation may influence the choice of menu is present in Noor [20, 21] and Noor and Ren [22]. Our most restrictive representation is related to the model in Noor [21]. There are two important differences, however. First, the model appears in [21] mainly as a means to axiomatically unify other temptation models in the literature. This paper provides a novel perspective on that model. Second, our choice domain differs substantially from [21] and, in particular, a counterpart of our axioms do not appear in [21].

Kopylov [14, Corollary 4] uses a stronger version of Monotonic Temptations to model tempting flexibility. Yet his model assumes that there is no preference for commitment at the interim stage. This assumption alleviates the technical analysis, but fails to accommodate any effects of weak resolve on the use of interim commitments.

We noted in footnote 5 the relationship of PED with the axioms in the independent work of Stovall [25]. The objective of that work is to identify uncertainty about normative preferences that is resolved by the interim stage when commitments can be still available, but temptations are not yet present. Stovall rules out the possibility of interim temptations via the Monotonicity axiom, thereby diverging from our paper.

Finally, in terms of stories for how an agent may be tempted to change her ex ante perspective, Epstein [9] and Epstein et al [10] model an agent who is tempted to retroactively change her beliefs over a state space. In our model the agent is tempted to doubt and possibly adjust her normative goals.

A Appendix: Proofs of Theorems

The proofs rely on the results of Kopylov [14, 13], but require substantial additional effort to pin down the components of the temptation value V in Theorems 3 and 4. The fact that V can be represented by finitely many "subjective states" is shown via techniques of Kopylov [13]. The identity of "subjective states" and the inequalities $\kappa \geq 0$, $\lambda > 0$, $\kappa + 1 \geq \lambda$ are established through model-specific arguments.

In the proofs, we adopt the following notation. For any function $u \in \mathcal{U}$ and any menu $a \in \mathcal{M}_1$, write

$$u(a) = \max_{x \in a} u(x),$$

and let

$$\mathcal{T}(u) = \{\alpha u + \beta : \alpha \geq 0, \beta \in \mathbb{R}\}$$

be the set of all non-negative transformations of the function u.

Lemma A.1 in Kopylov [15] asserts that for any $u_1, \ldots, u_S \in \mathcal{U}$, there are elements $x_1, \ldots, x_S \in X$ such that for all $i, j \in \{1, \ldots, S\}$,

$$u_i \notin \mathcal{T}(u_j) \quad \Leftrightarrow \quad u_i(x_i) > u_i(x_j) \tag{16}$$

$$u_i \in \mathcal{T}(u_j) \quad \Leftrightarrow \quad u_i(x_i) = u_i(x_j). \tag{17}$$

Obviously, it follows that $u_i(x_i) \geq u_i(x_j)$ for all i, j.

It follows that if u, v and $u+v$ are not redundant, then for all $\alpha, \beta, \gamma \in \mathbb{R}$,

$$\alpha u + \beta v + \gamma = 0 \quad \Rightarrow \quad \alpha = \beta = \gamma = 0. \tag{18}$$

Indeed, suppose that $\alpha u(x) + \beta v(x) + \gamma = 0$ for all $x \in X$. If $\alpha = 0$ and $\beta \neq 0$, then v is a constant function. If $\alpha \neq 0$ and $\beta = 0$, then u is a constant function. If $\alpha \neq 0$ and $\beta \neq 0$, then either $u + v \in \mathcal{T}(u)$ or $u + v \in \mathcal{T}(v)$. In all of these cases, the list of functions u, v, $u + v$ is not redundant. Thus $\alpha = \beta = \gamma = 0$ must hold in (18).

A.1 Proofs of Theorems 1 and Theorem 2

The first part of Theorem 1 follows immediately from Kopylov [14, Theorem 1]. Suppose that $\succeq$ is represented by a triple (u, v, V). Let $w = u + v$ and $W = U + V$. Then U_0 and U can be written as

$$U_0(A) = \max_{a \in A} W(a) - \max_{b \in A} V(b) \tag{19}$$

$$U(a) = w(a) - v(a). \tag{20}$$

for all $A \in \mathcal{M}_0$ and $a \in \mathcal{M}_1$.

Generic Temptations imply that

- u, v and $u + v$ are not redundant.
- U, V and $U + V$ are not redundant either.[7]

Indeed, if $v = \alpha u + \beta$ for $\alpha \geq 0$, then $\{\{x\}\} \succ \{\{x, y\}\}$ is impossible. If $v = \alpha u + \beta$ for $\alpha < 0$, then $\{\{x\}\} \sim \{\{x, y\}\} \succ \{\{y\}\}$ is impossible. Other situations when u, v, and $u + v$ are redundant are reduced to the above two cases, and hence are also impossible.

The above non-redundancy implies that there exist $x^*, y^* \in X$ and $a^*, b^* \in \mathcal{M}_1$ such that

$$\{\{x^*\}\} \succ \{\{x^*, y^*\}\} \succ \{\{y^*\}\} \tag{21}$$

$$\{a^*\} \succ \{a^*, b^*\} \succ \{b^*\}. \tag{22}$$

Turn to Theorem 2. Suppose that $C(\cdot)$ is represented by (8). WARP and Closed Graph are standard implications. Consistency must hold because

$$A \succ A \setminus \{a'\} \;\Rightarrow\; \max_{a \in A} W(a) - \max_{b \in A} V(b) > \max_{a \in A \setminus \{a'\}} W(a) - \max_{b \in A \setminus \{a'\}} V(b) \;\Rightarrow\; \max_{a \in A} W(a) > \max_{a \in A \setminus \{a'\}} W(a) \;\Rightarrow\; C(A) = \{a\}.$$

[7] Redundancy for functions in $\mathcal{U}_1$ is defined similarly to redundancy for functions in $\mathcal{U}_0$.

The second part is analogous.

Conversely, suppose that $C(\cdot)$ satisfies WARP, Closed Graph, and Consistency. Let $u_0 = u$, $u_1 = u+v$, and $u_2 = v$. Take $x_0, x_1, x_2 \in X$ that satisfy condition (16).

Take any menu $a \in \mathcal{M}_1$ and $x_* \in a$ such that $u_1(x_*) \geq u_1(y)$ for all $y \in a$. Fix any $\alpha \in (0,1)$. Let $z = \alpha x_1 + (1-\alpha)x_*$ and

$$b = \{z\} \cup (\alpha\{x_0, x_2\} + (1-\alpha)a).$$

Then $u_1(z) > u_1(y)$ for all $y \in b$ such that $y \neq z$. Yet $u_0(z) < u_0(\alpha x_0 + (1-\alpha)x_*)$ and $u_2(z) < u_2(\alpha x_2 + (1-\alpha)x_*)$. Then $b \succ b \setminus \{z\}$. By Consistency, $C(b) = \{z\}$. As α can be arbitrarily small, then by Closed Graph, $x_* \in C(a)$.

Take any $y \in a$ such that $u_1(y) < u_1(x_*)$. Let $Y = \{x \in X : u_1(y) < u_1(x) < u_1(x_*)\}$. Then Y is a mixture space. When restricted to Y, the linear functions u_0, u_1, u_2 are not redundant. Therefore, x_0, x_1, x_2 that satisfy (16) can be found in Y. Let $b = \{x_0, x_1, x_2, y\}$. Then $b \succ b \setminus \{x_1\}$. Yet x_1 is not the perfect element in b. By Consistency, $C(b) = \{x_1\}$.

Let $a' = a \cup \{x_1\}$. Note that $x_* \in \arg\max_{z \in a'} u_1(z)$. Therefore $x_* \in C(a')$. Suppose that $y \in C(a)$. By WARP, $y \in C(a')$ because $x_* \in C(a') \cap a$, and $y \in C(b)$ because $x_1 \in C(b) \cap a'$. Yet $C(b) = \{x_1\}$. This contradiction implies that $y \notin C(a)$. Thus $x \in C(a)$ if and only if $u_1(x) \geq u_1(y)$ for all $y \in a$.

The second part or (8) is analogous.

A.2 Proofs of Theorem 3 and 4

Suppose that $\succeq$ satisfies Axioms 1–5, Monotonic Temptations, and Finiteness. Then $\succeq$ can be represented by (6) for some $u, v \in \mathcal{U}$ and $V \in \mathcal{U}_1$.

Consider the temptation preference $\succeq_1$ that is represented by V on $\mathcal{M}_1$. This preference satisfies all conditions—Order, Continuity, Independence, and Finiteness—in Kopylov [13, Theorem 2.1]. In particular, his Finiteness for $\succeq_1$ follows from our Finiteness for $\succeq$ because there exists n such that for any sequence of n menus $a_1, \dots, a_n \in \mathcal{M}_1$, there is $j \in \{1, \dots, n\}$ such that

$$V\left(\bigcup_{i \in \{1,\dots,n\}} a_i\right) = V\left(\bigcup_{i \in \{1,\dots,n\} \setminus \{j\}} a_i\right).$$

Thus, $\succeq_1$ over $a \in \mathcal{M}_1$ can be represented by

$$V_1(a) = \sum_{i=1}^{m} u_i(a) - \sum_{j=1}^{k} v_j(a) \tag{23}$$

where $m, n \geq 0$ and the functions $u_1, \ldots, u_m, v_1, \ldots, v_k \in \mathcal{U}_0$ are not redundant.

As both $V_1, V \in \mathcal{U}_1$ are linear, continuous, and represent the same preference relation $\succeq_1$ one can take $V_1 = V$ without loss in generality. Take

$$x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+k}$$

that satisfy (16) for the functions $u_1, \ldots, u_m, v_1, \ldots, v_k$.

We claim that if the number of negative components $k > 0$, then $k = 1$ and $v_1 \in \mathcal{T}(v)$. Indeed, suppose that there is $v_i \notin \mathcal{T}(v)$. Let $a = \{x_1, \ldots, x_{m+k}\}$ and $b = a \setminus \{x_i\}$. If $v_i \in \mathcal{T}(w)$, then $U(a) > U(b)$ because $w(a) > w(b)$ and $v(a) = v(b)$, but $V(b) > V(a)$ because $v_i(b) < v_i(a)$.. The ranking $\{a\} = \{a \cup b\} \succ \{a \cup b, b\}$ contradicts MT. Rewrite V as

$$V(a) = \kappa w(a) + \sum_{i=2}^{m} u_i(A) - \sum_{j=1}^{k} v_j(A)$$

where $\kappa \geq 0$ and the functions $w, u_2, \ldots, u_m, v_1, \ldots, v_k$ are not redundant (if w is redundant with some u_j, just exclude u_j from the list of $u_1, \ldots, u_m$.) Take $u_1 = w$. Let $a = \{x_1, \ldots, x_{m+k}\}$ and $b = a \setminus \{x_i, x_1\}$. Then $U(a) > U(b)$ because $w(a) > w(b)$ and $v(a) = v(b)$, but $V(b) > V(a)$ because $v_i(b) < v_i(a)$. Thus we arrive at the contradiction with MT again: $\{a\} = \{a \cup b\} \succ \{a \cup b, b\}$.

Let $v_1 = \mu v$ for $\mu \geq 0$. Rewrite V further as

$$V(a) = \kappa U(a) + (\kappa - \mu) v(a) + \sum_{i=2}^{m} u_i(A).$$

No other negative components other than v have been shown to exist. Moreover, $\kappa - \mu \geq 0$ because otherwise, one can take a and b such that $v(b) < v(a)$, $w(b) < w(a)$, $u_i(a) = u_i(b)$, and $U(a) > U(b)$ with a very small difference $U(a) - U(b)$. Then $V(b) > V(a)$ and $\{a \cup b\} \succ \{a \cup b, b\}$. Thus V can be written in the required way (11) where $u_1 = (\mu - \kappa) w$.

Uniqueness follows from the uniqueness in Kopylov [13, Theorem 2.1].

Turn to Theorem 4. Suppose that $\succeq$ satisfies PT. (Finiteness and MT need not be assumed any more, but they will follow from the obtained representations.) Without loss in generality, assume that

$$u(x^*) = v(x^*) = w(x^*) = V(\{x^*\}) = 0. \tag{24}$$

Lemma 6. *There are unique $\kappa, \rho \in \mathbb{R}$ such that for all $a \in \mathcal{M}_1$,*

$$V(a) = \kappa w(a) + \rho v(a). \tag{25}$$

Proof. We claim first that for all $a, b \in \mathcal{M}_1$,

$$w(a) = w(b),\ v(a) = v(b) \quad \Rightarrow \quad V(a) = V(b). \tag{26}$$

Show this claim by contradiction. Take any $a, b \in \mathcal{M}_1$ such that $w(a) = w(b)$ and $v(a) = v(b)$. Then $U(a) = w(a) - v(a) = w(b) - v(b) = U(b)$.

Suppose that $V(a) \neq V(b)$. Wlog, let $V(b) > V(a)$. Modify a, b by small increments so that $w(a) > w(b)$, $v(a) = v(b)$, but $V(b) > V(a)$ by continuity. Then $\{a\} \sim \{a \cup b\} \succ \{b\}$, but $\{a\} \succ \{a, b\}$. This contradicts PT.

Take any three menus $a_1, a_2, a_3 \in \mathcal{M}_1$. Let $a = \cup_{i=1}^3 a_i$. Wlog $w(a) = w(a_1) = w(a_1 \cup a_2)$ and $v(a) = v(a_1 \cup a_2)$. By (26), $V(a) = V(a_1 \cup a_2)$. Let $\succeq_V$ be the ranking that V represents on $\mathcal{M}_1$. Theorem 2.1 in Kopylov [13] implies that this ranking can be represented also by

$$V'(a) = \gamma_1 u_1(a) + \gamma_2 u_2(a) \tag{27}$$

where $\gamma_1, \gamma_2 \in \{-1, 1\}$ and $u_1, u_2 \in \mathcal{U}$. Moreover, if $u_1 \in \mathcal{T}(u_2)$ or $u_2 \in \mathcal{T}(u_1)$, then representation (27) can be rewritten with just one non-zero component. In this case, wlog let $u_2 = 0$.

As $V', V \in \mathcal{U}_1$ are both linear and continuous, then wlog, $V' = V$. Let $u_3 = w$, $u_4 = v$. Take $x_1, x_2, x_3, x_4 \in X$ that satisfy (16) and (17). Let $a = \{x_1, x_3, x_3, x_4\}$ and $b = \{x_2, x_3, x_4\}$. Then $w(a) = w(b) = w(x_3)$ and $v(a) = v(b) = v(x_4)$. By (26), $V(a) = V(b)$. The equalities $u_2(a) = u_2(b) = u_2(x_2)$ and $V'(a) = V'(b)$ imply that $u_1(a) = u_1(b)$. By (17), $u_1 \in \mathcal{T}(u_i)$ for some $i > 1$. By convention, if $u_1 \in \mathcal{T}(u_2)$, then $u_2 = 0$ and hence, u_1 is constant. It contradicts the fact that V is not constant. Therefore, $u_1 \in \mathcal{T}(w)$ or $u_1 \in \mathcal{T}(v)$.

Analogously, by taking $b = \{x_1, x_3, x_4\}$ one can show that $u_2 \in \mathcal{T}(u_j)$ for some $j \neq 2$. If $u_2 = 0$, then trivially $u_2 \in \mathcal{T}(w) \cup \mathcal{T}(v)$. If $u_2 \neq 0$, then by convention $u_2 \notin \mathcal{T}(u_1)$. Thus $u_2 \in \mathcal{T}(w)$ or $u_2 \in \mathcal{T}(v)$. As both u_1 and u_2 belong to $\mathcal{T}(w) \cup \mathcal{T}(v)$, then for all $a \in \mathcal{M}_1$,

$$V'(a) = \gamma_1 u_1(a) + \gamma_2 u_2(a) = \kappa w(a) + \rho v(a) + \beta$$

for some $\kappa, \rho, \beta \in \mathbb{R}$. The normalization (24) implies that $\beta = 0$.

To show the uniqueness of κ and ρ, suppose that $V(a) = \kappa' w(a) + \rho' v(a)$ for some $\kappa', \rho' \in \mathbb{R}$. Then $\rho = \rho'$ because

$$V(\{x^*, y^*\}) = \rho v(y^*) = \rho' v(y^*)$$

and $v(y^*) > v(x^*) = 0$. Thus $\kappa w(y^*) = \kappa' w(y^*) = V(\{y^*\}) - \rho v(y^*)$. As $w(y^*) < w(x^*) = 0$, then $\kappa = \kappa'$. $\square$

The previous lemma implies that for all menus $a \in \mathcal{M}_1$,

$$V(a) = \kappa w(a) + \rho v(a) = \kappa U(a) + \lambda v(a) \tag{28}$$

where $\lambda = \kappa + \rho$. It remains to show that $\kappa \geq 0$, $\lambda > 0$, and $\lambda - \kappa = \rho \leq 1$ for the case when PED holds.

Show that $\kappa \geq 0$. If $\kappa < 0$, then

$$\{\{x^*, y^*\}\} \succ \{\{x^*, y^*\}, \{y^*\}\}$$

which contradicts MT and a fortiori, PT.

Show that $\lambda > 0$. Otherwise, one can take a and b such that $v(b) < v(a)$, $w(b) < w(a)$, and $U(a) > U(b)$ with a very small difference $U(a) - U(b)$. Then $V(b) > V(a)$ and $\{a \cup b\} \succ \{a \cup b, b\}$. Thus $\lambda > 0$.

Assume that $\succeq$ satisfies PED and show that $\rho \leq 1$. Suppose that $\rho > 1$. Take $\alpha, \beta \in (0, 1)$ such that

$$\alpha(\rho - 1)(v(y^*) - v(x^*)) > (1 - \beta)(\kappa + 1)(w(x^*) - w(y^*)).$$

Let $a = \{x^*, \alpha x^* + (1 - \alpha)y^*\}$, $b = \{\beta x^* + (1 - \beta)y^*, y^*\}$, and $c = \{x^*, y^*\}$. Then

$$\begin{aligned} U(a) = w(x^*) - v(y^*) + \alpha(v(y^*) - v(x^*)) > U(c) = w(x^*) - v(y^*) > \\ w(x^*) - v(y^*) - (1 - \beta)(w(x^*) - x(y^*)) = U(b). \end{aligned}$$

$$\begin{aligned} W(a) = (1 + \kappa)w(x^*) + (\rho - 1)v(\alpha x^* + (1 - \alpha)y^*) = \\ W(c) - \alpha(\rho - 1)(v(y^*) - v(x^*)) < W(c) - (1 - \beta)(\kappa + 1)(w(x^*) - w(y^*)) = \\ (1 + \kappa)w(\beta x^* + (1 - \beta)y^*)) + (\rho - 1)v(y^*) = W(b). \end{aligned}$$

As $U(a) > U(b)$ and $W(a) < W(b)$, then $V(b) > V(a)$. Thus $\{a, b\} \sim \{b\}$. However,

$$U(a \cup b) = w(a \cup b) - v(a \cup b) = w(x^*) - v(y^*) = U(c) > U(b).$$

Thus $\{a \cup b\} \succ \{b\}$, which contradicts PED.

Thus the temptation function V has the required form (28).

Suppose conversely that $\succeq$ has the required utility representation (6), where $u, v \in \mathcal{U}$ are independent functions, and the function V has the form (28) for some parameters $\kappa \geq 0$, $0 < \lambda \leq \kappa + 1$. Axioms 1–4 must hold by Theorem 1.

Show that $\succeq$ satisfied PED. Take any menus $a, b \in \mathcal{M}_1$. Wlog $\{a\} \succeq \{b\}$. Consider three cases.

Case 1. $V(a) \geq V(b)$. By (19), $\{a\} \sim \{a, b\}$. By Set-Betweenness, $\{a, b\} \succeq \{a \cup b\}$.

Case 2. $w(b) \geq w(a)$. As $w(a) - v(a) \geq w(b) - v(b)$, then $v(b) \geq v(a)$. Thus $\{a \cup b\} \sim \{b\}$ and by Set-Betweenness, $\{a, b\} \succeq \{a \cup b\}$.

Case 3. $V(b) > V(a)$ and $w(a) > w(b)$. As $\kappa U(b) + \lambda v(b) > \kappa U(a) + \lambda v(a)$, then $v(b) > v(a)$. As $\lambda - \kappa \leq 1$, then

$$U_0(\{a, b\}) \geq W(a) - V(b) = U(a) + (V(a) - V(b)) =$$
$$(w(a) - v(a)) + \kappa(w(a) - w(b)) - (\lambda - \kappa)(v(b) - v(a)) \geq$$
$$(w(a) - v(a)) - (v(b) - v(a)) = w(a) - v(b) = U_0(\{a \cup b\}).$$

Thus $\{a, b\} \succeq \{a \cup b\}$, and PED holds.

A.3 Proof of Theorem 5.

Suppose that $\succeq$ and $\succeq^*$ have representations (12)-(13) with components (u, v, κ, λ) and $(u, v, \kappa^*, \lambda^*)$. Let $w = u + v$. Note that $U(a) = U^*(a) = w(a) - v(a)$ for all $a \in \mathcal{M}_1$. Consider two cases.

Case 1. $\frac{\kappa^*}{\lambda^*} \leq \frac{\kappa}{\lambda}$. Take any $a, b \in \mathcal{M}_1$ such that $\{a\} \succ \{a, b\}$. Then $U(a) > U(b)$ and $V(b) > V(a)$. Therefore, $U^*(a) > U^*(b)$ and

$$V^*(a) - V^*(b) = \kappa^*[U(a) - U(b)] + \lambda^*(v(a) - v(b)) =$$
$$= \lambda^* \left(\tfrac{\kappa^*}{\lambda^*}[U(a) - U(b)] + (v(a) - v(b))\right) \leq$$
$$\lambda^* \left(\tfrac{\kappa}{\lambda}[U(a) - U(b)] + (v(a) - v(b))\right) = \tfrac{\lambda^*}{\lambda}[V(a) - V(b)] < 0.$$

Thus, $\{a\} \succ^* \{a, b\}$, and $\succeq^*$ is less resolved than $\succeq$.

Case 2. $\frac{\kappa^*}{\lambda^*} > \frac{\kappa}{\lambda}$. As u, v, and $u + v$ are not redundant, then the functions $\kappa u + \lambda v$ and $\kappa^* u + \lambda^* v$ are not redundant. Indeed, by (18),

$$(\kappa^* u + \lambda^* v) = \alpha(\kappa u + \lambda v) + \gamma$$

implies that $\gamma = 0$, $\alpha\lambda = \lambda^*$ $\alpha\kappa = \kappa^*$, and hence, $\frac{\kappa^*}{\lambda^*} = \frac{\kappa}{\lambda}$.

By (16), there are $x, y \in X$ such that $V^*(\{x\}) > V^*(\{y\})$ and $V(\{y\}) > V(\{x\})$. Then $U(\{x\}) = u(x) > u(y) = U(\{y\})$ because

$$\frac{\kappa^*}{\lambda^*}[u(x) - u(y)] > v(y) - v(x) > \frac{\kappa}{\lambda}[u(x) - u(y)].$$

Then $\{\{x\}\} \succ \{\{x\}, \{y\}\} \succeq \{\{y\}\}$, but $\{\{x\}\} \sim^* \{\{x\}, \{y\}\} \succ^* \{\{y\}\}$. Thus $\succeq^*$ is not less resolved than $\succeq$.

References

[1] G. Ainslie. Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, 82:463–496, 1975.

[2] C. Aliprantis and K. Border. *Infinite Dimensional Analysis*. Springer, 1999.

[3] D. Ariely and K. Wertenbroch. Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, 13:219–224, 2002.

[4] K. Arrow. Rational choice functions and orderings. *Econometrica*, 26:121–127, 1959.

[5] R. Baumeister, T. Heatherton, and D. Tice. *Losing Control: How and Why People Fail at Self-Regulation*. Academic Press, San Diego, CA, 1994.

[6] H. Becker. Notes on the concept of commitment. *American Journal of Sociology*, 66:32–40, 1960.

[7] G. Bryan, D. Karlan, and S. Nelson. Commitment devices. *Annual Review of Economics*, 2:671–698, 2010.

[8] E. Dekel, B. L. Lipman, and A. Rustichini. Temptation-driven preferences. *Review of Economic Studies*, 76:937–971, 2009.

[9] L. Epstein. An axiomatic model of non-Bayesian updating. *Review of Economic Studies*, 73:413–436, 2006.

[10] L. Epstein, J. Noor, and A. Sandroni. Non-bayesian updating: A theoretical framework. *Theoretical Economics*, 3:193–229, 2008.

[11] X. Gine, D. Karlan, and J. Zinman. Put your money where your butt is: A commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2:213–35, 2010.

[12] F. Gul and W. Pesendorfer. Temptation and self-control. *Econometrica*, 69:1403–1435, 2001.

[13] I. Kopylov. Finite additive utility representations for preferences over menus. *Journal of Economic Theory*, 144:354–374, 2009.

[14] I. Kopylov. Temptations in general settings. *The B.E. Journal of Theoretical Economics*, 9(31), 2009.

[15] I. Kopylov. Perfectionism and choice. *Econometrica*, 80:1819–1843, 2012.

[16] J. Kuhl. *Volitional Aspects of Achievement Motivation and Learned Helplessness: Toward a Comprehensive Theory of Action Control*. Academic Press, New York, 1984. In B. A. Maher (Ed.), Progress in experimental personality research Vol 13, pp 99-171.

[17] D. Laibson. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112:443–477, 1997.

[18] W. Mischel. *Processes in Delay of Gratification*. Academic Press, New York, 1974. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology Vol 7,.

[19] W. Mischel. Convergences and challenges in the search for consistency. *American Psychologist*, 39:351–364, 1984.

[20] J. Noor. Commitment and self-control. *Journal of Economic Theory*, 135:1–34, 2007.

[21] J. Noor. Temptation, welfare, and revealed preference. Mimeo, Boston University, 2009.

[22] J. Noor and L. Ren. Guilt and choice. Mimeo, Boston University, 2009.

[23] T. Schelling. *Strategies of Commitment and Other Essays*. Harvard University Press, Cambridge, 2006.

[24] J. Stovall. Multiple temptations. *Econometrica*, 78:349–376, 2010.

[25] J. Stovall. Temptation with uncertain normative preferences. Mimeo, University of Warick, 2013.

[26] R. Strotz. Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23:165–180, 1956.

[27] Y. Trope and A. Fishbach. Counteractive self-control in overcoming temptation. *Journal of Personality and Social Psychology*, 29:493–506, 2000.