

Temptation–Driven Preferences¹

Eddie Dekel²

Barton L. Lipman³

Aldo Rustichini⁴

July 2004

Preliminary and Incomplete Draft

¹This draft is very preliminary. Dekel and Rustichini have not yet “approved” this draft, so any mistakes are Lipman’s fault. Thanks to Massimo Marinacci, Ben Polak, and Phil Reny for helpful discussions.

²Economics Dept., Northwestern University, and School of Economics, Tel Aviv University
E–mail: dekel@nwu.edu.

³Boston University. E–mail: blipman@bu.edu. This work was begun while this author was at the University of Wisconsin.

⁴University of Minnesota. E–mail: arust@econ.umn.edu.

Abstract

“My own behavior baffles me. For I find myself not doing what I really want to do but doing what I really loathe.” Saint Paul

Gul–Pesendorfer [2001] give a decision–theoretic model of temptation building on a framework of Dekel, Lipman, and Rustichini [2001]. We argue that the axiom they use to identify temptation, set betweenness, rules out plausible and interesting forms of temptation. We propose a simple alternative axiom we call DFC, *desire for commitment*. This axiom characterizes temptation as situations where given any set of alternatives, the agent prefers committing herself to some particular item from the set rather than leaving herself the flexibility of choosing later. We show that this axiom (together with a somewhat more technical condition) characterizes a natural generalization of the Gul–Pesendorfer representation.

1 Introduction

Gul–Pesendorfer [2001] (henceforth GP) give a decision–theoretic model of temptation building on a framework of Dekel, Lipman, and Rustichini [2001] (DLR). The set of preferences considered by DLR included preferences “driven by” a desire for flexibility, preferences “driven by” temptation, and those which combined both considerations. GP focus on temptation alone by considering a particular subset of these preferences which are only affected by temptation. We argue that the axiom they use to identify temptation, set betweenness, rules out plausible and interesting forms of temptation. In this sense, they identify a subset of the preferences which are driven by temptation.

We propose a simple alternative axiom we call DFC, *desire for commitment*, and show that this characterizes a natural generalization of the GP representation. DFC simply says that given any set of alternatives, the agent at least weakly prefers to commit herself to some option from this set rather than make a choice from the set. In this sense, DFC is exactly the statement that there is no value to flexibility but the agent may fear being tempted to choose “inappropriately.” Given the natural interpretation of the axiom and the intuitive nature of the representation it generates, we conclude that DFC is a natural way to identify from the large set considered by DLR those preferences which are “temptation–driven.”

We also give some special cases of the main representation and the additional axioms which correspond to these.

In the next section, we present the basic model and state our research goals more precisely. In the process, we sketch the relevant results in DLR and GP. In Section 3, we give examples to motivate the issues and illustrate the kinds of representations we are interested in. In Section 4, we give representation results. First, we characterize the set of preferences originally considered by DLR which also satisfy DFC. These preferences generate only a weak form of the representation we are interested in. We then show that an additional axiom gives a more interesting strengthened representation. Section 5 contains characterizations of some special cases. In Section 6, we briefly discuss directions for further research.

2 The Model

Let B be a finite set of *prizes* and let $\Delta(B)$ denote the set of probability distributions on B . A typical subset of $\Delta(B)$ will be referred to as a *menu* and denoted x (or \tilde{x} , x' , \bar{x} , y , etc.), while a typical element of $\Delta(B)$, a *lottery*, will be denoted by β . The agent has

a preference relation \succ on the set of closed nonempty subsets of $\Delta(B)$ which is denoted X . Given menus x and y and a number $\lambda \in [0, 1]$, let

$$\lambda x + (1 - \lambda)y = \{\beta \in \Delta(B) \mid \beta = \lambda\beta' + (1 - \lambda)\beta'', \text{ for some } \beta' \in x, \beta'' \in y\}$$

where, as usual, $\lambda\beta' + (1 - \lambda)\beta''$ is the probability distribution over B giving b probability $\lambda\beta'(b) + (1 - \lambda)\beta''(b)$.

The relevant axioms used in DLR [2001] are:

Axiom 1 (Weak Order) \succ is asymmetric and negatively transitive.

Axiom 2 (Continuity) The strict upper and lower contour sets, $\{x' \subseteq \Delta(B) \mid x' \succ x\}$ and $\{x' \subseteq \Delta(B) \mid x \succ x'\}$, are open (in the Hausdorff topology).

Axiom 3 (Independence) If $x \succ x'$, then for all $\lambda \in (0, 1]$ and all \bar{x} ,

$$\lambda x + (1 - \lambda)\bar{x} \succ \lambda x' + (1 - \lambda)\bar{x}.$$

While DLR discuss other representations, the relevant one for our purposes is

Definition 1 An **additive EU representation** is a set S , a state-dependent utility function¹ $U : \Delta(B) \times S \rightarrow \mathbf{R}$, and a finitely additive measure μ with full support on S such that (i) $V(x)$ defined by

$$V(x) = \int_S \max_{\beta \in x} U(\beta, s) \mu(ds)$$

is continuous and represents \succ and (ii) each $U(\cdot, s)$ is an expected-utility function in the sense that

$$U(\beta, s) = \sum_{b \in B} \beta(b)U(b, s).$$

The representation theorem² in DLR [2001] is:

¹To be more precise, S is required to be a measure space and U measurable with respect to this space.

²This result differs slightly from that in DLR [2001] in two respects. First, DLR included a requirement that S be nonempty as part of the definition of an additive EU representation and correspondingly included a nontriviality axiom. Second, DLR required that no state be redundant implying, in particular, that there is no s such that $U(\cdot, s)$ is a constant function. Since GP do not rule out such representations, we omit these issues to avoid irrelevant details in comparing our results to GP.

Theorem 1 *The preference \succ has an additive EU representation if and only if it satisfies weak order, continuity, and independence.*

The additive EU representation is easiest to understand in the case where S is finite, a case we focus on for the rest of this paper. In this case, we have

$$V(x) = \sum_{s \in S} \mu(s) \max_{\beta \in x} U(\beta, s),$$

where $\mu(s) \neq 0$ but can be positive or negative. It is convenient to rewrite this as follows. Let $\{s_1, \dots, s_I\}$ denote the set of *positive states* — those with $\mu(s) > 0$ — and let $\{s_{I+1}, \dots, s_{I+J}\}$ denote the set of *negative states* — those with $\mu(s) < 0$. For $i = 1, \dots, I$, let $w_i(\beta) = U(\beta, s_i)\mu(s_i)$. For $j = 1, \dots, J$, let $v_j(\beta) = U(\beta, s_{I+j})|\mu(s_{I+j})|$. Then we can write

$$V(x) = \sum_{i=1}^I \max_{\beta \in x} w_i(\beta) - \sum_{j=1}^J \max_{\beta \in x} v_j(\beta).$$

If there were no negative states, this representation would have an obvious interpretation. Think of the w_i 's as different utility functions the agent might have at some later date when she will choose from the menu she picks today. At the point when she will make this choice, she will know which of the w_i 's is her utility function and, naturally, will choose the item from the menu which maximizes this utility. Her *ex ante* evaluation of the menu, then, is simply the expected value of the maximum. If each of the w_i 's is equally likely, we obtain the value above. This is exactly the interpretation originally offered by Kreps [1979, 1992] who first considered preferences over sets as a way of modeling preferences for flexibility. Obviously, though, the presence of the negative states makes this interpretation awkward.

One way to reach a clearer understanding of this representation, then, is to rule out the negative states. DLR show that the axiom which does this is one of Kreps' main axioms, namely monotonicity.

Axiom 4 (Monotonicity) *If $x \subset x'$, then $x' \succeq x$.*

DLR show

Theorem 2 *The preference \succ has an additive EU representation with a positive measure μ if and only if it satisfies weak order, continuity, independence, and monotonicity.*

Rephrased, if \succ satisfies monotonicity, then the additive EU representation contains no negative states. This result is the generalization of Kreps' [1979] representation theorem.

Intuitively, monotonicity is the statement that we are considered agents who always at least weakly value flexibility. Such agents either are not concerned about temptation or, at least, value flexibility so highly as to outweigh such considerations. In this case, the additive EU representation is easy to understand as describing a forward-looking agent with some beliefs about what her possible future needs are.

GP take a different approach. They first recognized that temptation and self-control could be studied using this sets of lotteries framework if one does not impose monotonicity. If the agent might be tempted in the future to consume something she currently doesn't want herself to consume, this is revealed by a preference for commitment, not flexibility. GP's [2001] representation theorem differs from Theorem 1 in two respects. First, their result is more general in that they allow B to be compact rather than assuming it to be finite, an issue we ignore henceforth. Second, they add an axiom which they call *set betweenness*:

Axiom 5 (Set Betweenness) *If $x \succeq y$, then $x \succeq x \cup y \succeq y$.*

To understand this axiom, suppose the agent is deciding where to eat lunch and wishes to consume a healthy meal. Think of x , y , and $x \cup y$ as the menus available at the three possible restaurants. Suppose x consists only of a single healthy food item, say broccoli, while y consists only of some fattening food item, say french fries. Then the fact that the agent wants to consume a healthy meal suggests $x \succ y$. How should the agent rank the menu $x \cup y$ relative to the other two? A natural hypothesis is that the third restaurant would fall in between the other two in the agent's ranking. It would be better than the menu with only french fries since the agent might choose broccoli given the option. On the other hand, the third menu would be worse than the menu with only broccoli since the agent might succumb to temptation or, even if she didn't succumb, might suffer from the costs of maintaining self-control in the face of the temptation. Hence $x \succ x \cup y \succ y$, in line with what set betweenness requires.

The representation GP consider³

Definition 2 *A self control representation is a pair of functions (u, v) , $u : \Delta(B) \rightarrow \mathbf{R}$, $v : \Delta(B) \rightarrow \mathbf{R}$, such that each is an expected utility function and the function V_{GP} defined*

³They also consider another representation, the overwhelming temptation representation, which is not relevant for our purposes.

by

$$V_{GP}(x) = \max_{\beta \in x} [u(\beta) + v(\beta)] - \max_{\beta \in x} v(\beta)$$

represents \succ .

GP show

Theorem 3 *A self control representation exists if and only if the preference satisfies weak order, continuity, independence, and set betweenness.*

To interpret this representation, first note that we can think of u as the “commitment preference” — that is, what the agent would choose if she could commit herself *ex ante*. Specifically, $V(\{\beta\}) = u(\beta)$ for any β . For any menu x and any $\beta \in x$, let

$$c(\beta, x) = \left[\max_{\beta' \in x} v(\beta') \right] - v(\beta).$$

Intuitively, c is the foregone utility according to v from choosing β from x instead of the β' which v finds optimal. It is easy to see that

$$V_{GP}(x) = \max_{\beta \in x} [u(\beta) - c(\beta, x)].$$

In this form, it is natural to interpret c as the cost of the self-control needed to choose β from x . Given this, v is naturally interpreted as the temptation utility since it is what determines the self-control cost.

Another way to understand the GP representation is to relate it to DLR. It is easy to see that the self-control representation is exactly an additive EU representation with one positive state and one negative state where $w_1 = u + v$ and $v_1 = v$. In this sense, the only difference between GP’s representation and the DLR representation with one positive and one negative state is a change of variables.

One way to think about these results is to begin by considering the set of preferences satisfying weak order, continuity, and independence. For brevity, we will refer to these as *DLR preferences*. Intuitively, if we consider the subset of these DLR preferences which are monotonic, we are restricting attention to agents who value flexibility but are not affected by temptation. It seems very natural to call such preferences *flexibility-driven*, both because the axiom and the representation it generates seem to describe such an agent.

Analogously, we will refer to preferences which exhibit a concern about temptation but no value to flexibility *per se* as *temptation-driven*. It seems natural to say that

the subset of DLR preferences that satisfy set betweenness are temptation–driven preferences. However, set betweenness does not appear to be as obvious a statement of “temptation–driven preferences” as monotonicity does for “flexibility–driven.” In fact, it is not hard to give examples of behavior which appears to be temptation–driven but where set betweenness is violated. This suggests that set betweenness is stronger than just a restriction to temptation–driven preferences. Our goal in this paper is to identify and give a representation theorem for the full class of temptation–driven DLR preferences.

An axiom which seems to be a more natural way to define temptation–driven is

Axiom 6 (DFC: Desire for Commitment) *A preference \succ satisfies DFC if for every x , there is some $\alpha \in x$ such that $\{\alpha\} \succeq x$.*

Intuitively, this axiom seems to be a necessary condition to say that a preference is temptation–driven. The axiom says that there is no value to flexibility associated with x , only potential costs due to temptation leading the agent to choose some point other than α .

3 Motivating Examples and Some Alternative Representations

In this section, we give two examples to illustrate our argument that set betweenness is stronger than a restriction to temptation–driven preferences. We also use these examples to suggest some representations that may be of interest.

Example 1.

Consider an agent who is trying to diet and so would like to commit herself to eating only broccoli. There are two kinds of snacks available: chocolate cake and high fat potato chips. Let b denote the broccoli, c the chocolate cake, and p the potato chips. The following ranking seems quite natural:

$$\{b\} \succ \{b, c\}, \{b, p\} \succ \{b, c, p\}.$$

That is, the agent would like to commit herself to eating only broccoli, so $\{b\}$ is the best of these four menus. If she has both broccoli and a fattening snack available, the temptation of the snack will lower her utility. If she has broccoli and *both* fattening snacks available, she is still worse off since two snacks are harder to resist than one.

Two snacks could be worse than one for at least two reasons. First, it could be that the agent is unsure what kind of temptation will strike. If the agent would be in a mood for a salty snack, then she may be able to control herself easily if only the chocolate cake is available as an alternative to broccoli. Similarly, if she is in the mood for a sweet snack, she may be able to control herself if only the potato chips are available. But if she has both available, she is more likely to be hit by a temptation she cannot avoid. Hence the effect on choice is likely to be stronger. Second, even if she resists temptation, the psychological cost of self-control seems likely to be higher in the presence of two snacks than in the presence of one.⁴

This preference violates set betweenness. Note that $\{b, c, p\}$ is strictly worse than $\{b, c\}$ and $\{b, p\}$ even though it is the union of these two sets. Hence set betweenness implies that two temptations can *never* be worse than each of the temptations separately. In GP, temptation is one dimensional in the sense that any menu has a most tempting option and only this temptation is relevant to the self-control costs. That is, there is no interaction between temptations in determining the self-control costs.

It is not hard to give additive EU representations that can model either of the two reasons stated above for two snacks to be worse than one. To see this, define utility functions u , v_1 , and v_2 by

	u	v_1	v_2
b	3	2	2
c	0	0	6
p	0	6	0

Define V_1 by the following natural generalization of GP:

$$V_1(x) = \frac{1}{2} \sum_{i=1}^2 \left[\max_{\beta \in x} [u(\beta) + v_i(\beta)] - \max_{\beta \in x} v_i(\beta) \right].$$

Equivalently, let

$$c_i(\beta, x) = \left[\max_{\beta' \in x} v_i(\beta') \right] - v_i(\beta).$$

Then

$$V_1(x) = \frac{1}{2} \sum_{i=1}^2 \max_{\beta \in x} [u(\beta) - c_i(\beta, x)].$$

Intuitively, the agent doesn't know whether the temptation that will strike is the one described by v_1 and cost function c_1 (where she is most tempted by the potato chips) or v_2 and cost function c_2 (where she is most tempted by the chocolate cake) and gives probability 1/2 to each possibility. It is easy to verify that this gives $V_1(\{b\}) = 3$,

⁴GP [2001, 1408–1409] mention this kind of intuition as one reason why set betweenness may be violated.

$V_1(\{b, c\}) = V_1(\{b, p\}) = 3/2$, and $V_1(\{b, c, p\}) = 0$, yielding the ordering suggested above.

Alternatively, define V_2 by a different generalization of GP:

$$V_2(x) = \max_{\beta \in x} [u(\beta) + v_1(\beta) + v_2(\beta)] - \max_{\beta \in x} v_1(\beta) - \max_{\beta \in x} v_2(\beta).$$

Here we can think of cost of choosing β from menu x as

$$c(\beta, x) = \left[\max_{\beta \in x} v_1(\beta) + \max_{\beta \in x} v_2(\beta) \right] - v_1(\beta) - v_2(\beta).$$

It is not hard to see that this cost function has the property that resisting two temptations is harder than resisting one. More specifically, it is easy to verify that $V_2(\{b\}) = 3$, $V_2(\{b, c\}) = V_2(\{b, p\}) = -1$, and $V_2(\{b, c, p\}) = -5$, again yielding the ordering suggested above.

Example 2.

Consider again the dieting agent facing multiple temptations, but now suppose the two snacks available are high fat chocolate ice cream (c) and low fat chocolate frozen yogurt (y). In this case, it seems natural that the agent might have the following rankings:

$$\{b, y\} \succ \{y\} \quad \text{and} \quad \{b, c, y\} \succ \{b, c\}.$$

In other words, the agent would rather have a chance of sticking to her diet rather than committing herself to violating it so $\{b, y\} \succ \{y\}$. Also, if the temptation of the ice cream is unavoidable, it's better to also have the low fat frozen yogurt around. If so, then when temptation strikes, the agent may be able to resolve her hunger for chocolate in a less fattening way.

Again, GP cannot have this. This is not a violation of set betweenness but instead a violation of the combination of set betweenness and independence. To see why this cannot occur in their model, note that

$$V(\{b, y\}) = \max\{u(b) + v(b), u(y) + v(y)\} - \max\{v(b), v(y)\}$$

while $V(\{y\}) = u(y) = u(y) + v(y) - v(y)$. Obviously, $\max\{v(b), v(y)\} \geq v(y)$. So $V(\{b, y\}) > V(\{y\})$ requires $\max\{u(b) + v(b), u(y) + v(y)\} > u(y) + v(y)$ or $u(b) + v(b) > u(y) + v(y)$. Given this,

$$\max\{u(b) + v(b), u(c) + v(c), u(y) + v(y)\} = \max\{u(b) + v(b), u(c) + v(c)\}.$$

Since

$$\max\{v(b), v(c), v(y)\} \geq \max\{v(b), v(c)\},$$

we get $V(\{b, c, y\}) \leq V(\{b, c\})$. That is, we must have $\{b, c\} \succeq \{b, c, y\}$.⁵

Intuitively, in GP, $\{b, y\} \succ \{y\}$ implies that the agent will never choose frozen yogurt when broccoli is available. Hence the only effect frozen yogurt can have when broccoli is available is to increase self-control costs. The possibility that y could be a compromise against some worse temptation is not allowed.

For a simple additive EU representation which allows the intuitive preference suggested above, define

	u	v
b	6	0
c	0	8
y	4	6

and let

$$V_3(x) = \frac{1}{2} \max_{\beta \in x} u(\beta) + \frac{1}{2} \left\{ \max_{\beta \in x} [u(\beta) + v(\beta)] - \max_{\beta \in x} v(\beta) \right\}.$$

Intuitively, there is a probability of 1/2 that the agent avoids temptation and chooses according to the commitment preference u . With probability 1/2, the agent is tempted, however, and has a preference of the form characterized by GP. This gives $V_3(\{b, y\}) = 5 > 4 = V_3(\{y\})$ and $V_3(\{b, c, y\}) = 5 > 3 = V_3(\{b, c\})$, in line with the intuitive story.

The three representations used in these examples share certain features in common. First, all are additive EU representations. That is, all the preferences involved satisfy weak order, continuity, and independence. While we do not wish to argue that these axioms are innocuous, it is not clear why temptation should require some violation of these properties. Second, in all cases, the representation is written in terms of the utility functions for the negative states and u , the commitment utility. Equivalently, we can write the representation in terms of the commitment utility and various possible cost functions where these costs are generated from different possible temptations. In this sense, different positive states correspond to different degrees of or different types of temptation, but share a common view of what is “truly best” as embodied in u . Put differently, there is no uncertainty about “true preferences” and hence no “true” value to flexibility, only uncertainty about temptation.

Restricting attention to additive EU representations with finitely many states, a general kind of representation which fits with these criteria is

⁵We cannot show this directly from the axioms. We do know, however, that it cannot be demonstrated from set betweenness alone — independence is essential to this conclusion. More specifically, this preference is consistent with set betweenness if independence is violated or independence if set betweenness is violated.

Definition 3 A temptation representation is a function V_T representing \succ such that

$$V_T(x) = \sum_{i=1}^I q_i \max_{\beta \in x} [u(\beta) - c_i(\beta, x)]$$

where $q_i > 0$ for all i , $\sum_i q_i = 1$, and

$$c_i(\beta, x) = \left[\sum_{j \in J_i} \max_{\beta' \in x} v_j(\beta') \right] - \sum_{j \in J_i} v_j(\beta)$$

where u and each v_j is an expected-utility function.

Note that $\sum_i q_i = 1$ implies that $V_T(\{\beta\}) = u(\beta)$, so u is the commitment utility.

Intuitively, we can think of the q_i 's as the probabilities over the I different ways temptation may affect the agent where the way in which temptation affects the agent is measured by control cost function c_i .

We can think of this as generalizing GP in two directions. First, more than one temptation can affect the agent at a time. That is, the cost of self-control may depend on more than one temptation utility. Second, the agent is uncertain which temptation or temptations will affect her. That is, we further generalize by assuming the agent doesn't know which of many generalized GP representations will describe her.

A similar idea is

Definition 4 A weak temptation representation is a function V_w representing \succ such that

$$V_w(x) = \sum_{i=1}^{I'} q_i \max_{\beta \in x} [u(\beta) - c_i(\beta, x)] + \sum_{i=I'+1}^I \max_{\beta \in x} [-c_i(\beta, x)]$$

where $q_i > 0$ for all i , $\sum_i q_i = 1$, and

$$c_i(\beta, x) = \left[\sum_{j \in J_i} \max_{\beta' \in x} v_j(\beta') \right] - \sum_{j \in J_i} v_j(\beta)$$

where u and each v_j is an expected-utility function.

Obviously, a temptation representation is a special case of a weak temptation representation where $I' = I$.

A temptation representation seems more natural than a weak temptation representation. As we will see, DFC only generates a weak temptation representation — an additional axiom is needed to obtain a temptation representation.

4 Results

We first show that DFC is the additional axiom needed to go from a finite state additive EU representation to a weak temptation representation. After this, we introduce the additional axiom needed to generate a temptation representation and demonstrate that it does so.

Notation. In what follows, we write u , v_j , etc., to denote the vector giving the payoffs to the pure outcomes associated with utility function u , v_j , etc. We will always write these as column vectors. Because there are n pure outcomes, then, these are n by 1. We will write lotteries as 1 by n row vectors, so $\beta \cdot u = u(\beta)$, etc. Also, $\mathbf{1}$ denotes the n by 1 vector of 1's.

Theorem 4 *Suppose \succ has an additive EU representation of the form*

$$V(x) = \sum_{i=1}^I \max_{\beta \in x} w_i(\beta) - \sum_{j=1}^J \max_{\beta \in x} v_j(\beta).$$

Define u by $u(\beta) = V(\{\beta\})$, so $u = \sum_i w_i - \sum_j v_j$. Suppose \succ satisfies DFC. Then there are positive scalars a_i , $i = 1, \dots, I$, and b_{ij} , $i = 1, \dots, I$, $j = 1, \dots, J$ and scalars c_i , $i = 1, \dots, I$ such that $\sum_i a_i = \sum_i b_{ij} = 1$ for all j and

$$w_i = a_i u + \sum_j b_{ij} v_j + c_i \mathbf{1}$$

for all i .

Proof. Suppose not. Let Z denote the set of nI by 1 vectors $(z'_1, \dots, z'_I)'$ such that

$$z_i = a_i u + \sum_j b_{ij} v_j + c_i \mathbf{1}, \quad \forall i$$

for scalars a_i , b_{ij} , and c_i satisfying the conditions of the theorem. So if the theorem does not hold, the vector $(w'_1, \dots, w'_I)' \notin Z$. Since Z is obviously closed and convex, the separating hyperplane theorem implies that there is a vector p such that

$$p \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_I \end{pmatrix} > p \cdot \begin{pmatrix} z_1 \\ \vdots \\ z_I \end{pmatrix}, \quad \forall \begin{pmatrix} z_1 \\ \vdots \\ z_I \end{pmatrix} \in Z.$$

Write $p = (p_1, \dots, p_I)$ where each p_i is a 1 by n vector. So

$$\sum_i p_i \cdot w_i > \sum_i p_i \cdot z_i, \quad \forall \begin{pmatrix} z_1 \\ \vdots \\ z_I \end{pmatrix} \in Z.$$

Equivalently,

$$\sum_i p_i \cdot w_i > \sum_i a_i p_i \cdot u + \sum_j \sum_i b_{ij} p_i \cdot v_j + \sum_i c_i p_i \cdot \mathbf{1}$$

for any a_i, b_{ij} , and c_i such that $a_i \geq 0$ for all i , $b_{ij} \geq 0$ for all i and j , and $\sum_i a_i = \sum_i b_{ij} = 1$ for all j . Since c_i is arbitrary in both sign and magnitude, we must have $p_i \cdot \mathbf{1} = 0$ for all i . If not, we could find a c_i which would violate the inequality above.

Also, for every choice of $a_i \geq 0$ such that $\sum_i a_i = 1$,

$$\max_i p_i \cdot u \geq \sum_i a_i p_i \cdot u$$

with equality for an appropriately chosen (a_1, \dots, a_I) . Similarly, for any nonnegative b_{ij} 's with $\sum_i b_{ij} = 1$,

$$\max_i p_i \cdot v_j \geq \sum_i b_{ij} p_i \cdot v_j$$

with equality for an appropriately chosen (b_{1j}, \dots, b_{Ij}) . Hence the inequality above implies

$$\sum_i p_i \cdot w_i > \max_i p_i \cdot u + \sum_j \max_i p_i \cdot v_j.$$

Write p_i as (p_{1i}, \dots, p_{ni}) . Without loss of generality, we can assume that $|p_{ki}| \leq 1/n$ for all k and i . (Otherwise we could divide both sides of the inequality above by $n \max_{k,i} |p_{ki}|$ and redefine p_i to have this property.) Let β denote the probability distribution $(1/n, \dots, 1/n)$. For each i , let $\alpha_i = p_i + \beta$. Note that $\alpha_{ki} = p_{ki} + 1/n$ and so $\alpha_{ki} \geq 0$ for all k, i . Also, $\alpha_i \cdot \mathbf{1} = p_i \cdot \mathbf{1} + \beta \cdot \mathbf{1} = 1$. Hence each α_i is a probability distribution. Substituting $\alpha_i - \beta$ for p_i ,

$$\sum_i \alpha_i \cdot w_i - \sum_i \beta \cdot w_i > \max_i \alpha_i \cdot u - \beta \cdot u + \sum_j \max_i \alpha_i \cdot v_j - \sum_j \beta \cdot v_j.$$

By definition of u , $\sum_i w_i = u + \sum_j v_j$. Hence this is

$$\sum_i \alpha_i \cdot w_i - \sum_j \max_i \alpha_i \cdot v_j > \max_i \alpha_i \cdot u.$$

Let $x = \{\alpha_1, \dots, \alpha_I\}$. Then

$$V(x) \geq \sum_i \alpha_i \cdot w_i - \sum_j \max_i \alpha_i \cdot v_j > \max_i \alpha_i \cdot u = \max_{\alpha \in x} u(\alpha).$$

But this contradicts DFC. ■

Corollary 1 *has a weak temptation representation if and only if it has a finite state additive EU representation and satisfies DFC.*

Proof. The necessity of \succ having a finite state additive EU representation is obvious. For necessity of DFC, suppose \succ has a weak temptation representation. For any menu x and any i , let α_i denote a maximizer of $a_i u(\beta) + \sum_{j \in J_i} v_j(\beta)$ over $\beta \in x$. Then

$$\begin{aligned} V_w(x) &= \sum_i [a_i u(\alpha_i) + \sum_{j \in J_i} v_j(\alpha_i)] - \sum_i \sum_{j \in J_i} \max_{\beta \in x} v_j(\beta) \\ &\leq \sum_i [a_i u(\alpha_i) + \sum_{j \in J_i} v_j(\alpha_i)] - \sum_i \sum_{j \in J_i} v_j(\alpha_i) \\ &= \sum_i a_i u(\alpha_i) \\ &\leq \max_{\beta \in x} u(\beta). \end{aligned}$$

Hence DFC must hold.

For sufficiency, note that the preceding theorem implies

$$\begin{aligned} V(x) &= \sum_i \max_{\beta \in x} [a_i u(\beta) + \sum_j b_{ij} v_j(\beta) + c_i] - \sum_j \max_{\beta \in x} v_j(\beta) \\ &= \sum_i \max_{\beta \in x} [a_i u(\beta) + \sum_j b_{ij} v_j(\beta)] - \sum_j \max_{\beta \in x} v_j(\beta) + \sum_i c_i. \end{aligned}$$

But

$$u + \sum_j v_j = \sum_i w_i = \sum_i a_i u + \sum_i \sum_j b_{ij} v_j + \sum_i c_i \mathbf{1}.$$

Since $\sum_i a_i = \sum_i b_{ij} = 1$ for all j , this says

$$u + \sum_j v_j = u + \sum_j v_j + \sum_i c_i \mathbf{1},$$

so $\sum_i c_i = 0$.

Let I_+ denote the set of i such that $a_i > 0$. For each $i \in I_+$, let $q_i = a_i$. Let K denote the number of (i, j) pairs for which $b_{ij} > 0$. For each such (i, j) , let $k(i, j)$ denote a distinct element of $\{1, \dots, K\}$. For each $i \in I_+$ and each j such that $b_{ij} > 0$, define a utility function $\hat{v}_{k(i,j)} = [b_{ij}/a_i] v_j$ and let $k(i, j) \in J_i$. For each $i \notin I_+$ and each j with $b_{ij} > 0$, define a utility function $\hat{v}_{k(i,j)} = b_{ij} v_j$ and let $k(i, j) \in J_i$. So for $i \in I_+$,

$$u_i = a_i u + \sum_j b_{ij} v_j = q_i [u + \sum_{j \in J_i} \hat{v}_j].$$

For $i \notin I_+$,

$$u_i = \sum_j b_{ij} v_j = \sum_{j \in J_i} \hat{v}_j.$$

Also,

$$\begin{aligned} \sum_j \max_{\beta \in x} v_j(\beta) &= \sum_j \sum_i b_{ij} \max_{\beta \in x} v_j(\beta) \\ &= \sum_{i \in I_+} \sum_{j \in J_i} q_i \max_{\beta \in x} \hat{v}_j(\beta) + \sum_{i \notin I_+} \sum_{j \in J_i} \max_{\beta \in x} \hat{v}_j(\beta). \end{aligned}$$

Hence

$$V(x) = \sum_{i \in I_+} q_i \max_{\beta \in x} [u(\beta) - c_i(\beta, x)] + \sum_{i \notin I_+} \max_{\beta \in x} [-c_i(\beta, x)]$$

where

$$c_i(\beta, x) = \left[\sum_{j \in J_i} \max_{\beta' \in x} \hat{v}_j(\beta') \right] - \sum_{j \in J_i} \hat{v}_j(\beta).$$

Hence V is a weak temptation representation. ■

Clearly, the key step needed to go from a weak temptation representation to a temptation representation is to extend Theorem 4 to a result which ensures that the a_i 's are strictly positive. With this done, the proof of Corollary 1 directly yields existence of a temptation representation.

To state the axiom needed for this purpose requires additional notation and terminology. Given a menu x , let $B(x)$ denote the set of $\alpha \in x$ such that $\{\alpha\} \succeq \{\alpha'\}$ for all $\alpha' \in x$. That is, $B(x)$ is the set of best commitments in x .

Definition 5 *A menu x is temptation-free if there is an $\alpha \in B(x)$ such that $\{\alpha\} \sim x$.*

To see the reason for the name, note that if x is temptation-free, then an agent facing x is indifferent between committing herself to some choice from x and not doing so. In this sense, the options in x which are not in $B(x)$ do not tempt the agent away from her optimal commitment.

For any lottery β and $\varepsilon > 0$, let $N_\varepsilon(\beta)$ denote the ε neighborhood of β . The key axiom is

Axiom 7 (Domination) *If $x \cup \{\beta\}$ is a temptation-free menu with $\beta \notin B(x \cup \{\beta\})$, then there is an $\varepsilon > 0$ such that for all $\hat{\beta} \in N_\varepsilon(\beta)$ and all x' ,*

$$x' \cup x \succeq x' \cup x \cup \{\hat{\beta}\}.$$

To see the intuition for this axiom, suppose $\{\alpha, \beta\}$ is temptation-free and that $\{\alpha\} \succ \{\beta\}$. To say that $\{\alpha, \beta\}$ is temptation-free, then, implies that $\{\alpha\} \sim \{\alpha, \beta\} \succ \{\beta\}$. Intuitively, this says that β never tempts the agent away from α — regardless of what temptation might strike, the agent chooses α from $\{\alpha, \beta\}$ without incurring any cost of self-control. That is, α dominates β in the sense that it is strictly better according to the commitment preference and at least as good according to any of the possible temptations. In light of this dominance, then, β has no effect on any menu containing α . That is, $x' \cup \{\alpha\} \sim x' \cup \{\alpha, \beta\}$.

However, suppose we change β a small amount to $\hat{\beta}$. By continuity, if this change is sufficiently small, it cannot reverse the strict commitment preference for α . However,

if under some temptation, the agent is indifferent between α and β , a small change in β could break this tie in favor of β . If this occurs, choosing α from $\{\alpha, \hat{\beta}\}$ will require some self-control by the agent. Hence a small change in β to $\hat{\beta}$ could only cause $\hat{\beta}$ to be a bad option to have when α is available. That is, we would necessarily have $x' \cup \{\alpha\} \succeq x' \cup \{\alpha, \hat{\beta}\}$.

Lemma 1 *A preference \succ has a temptation representation only if it satisfies dominance.*

Proof. Assume \succ has a temptation representation denoted V . Fix any temptation-free menu $y = x \cup \{\beta\}$ with $\beta \notin B(y)$. By definition of a temptation representation, we have

$$V(y) = \sum_i q_i \max_{\alpha \in y} [u(\alpha) + \sum_{j \in J_i} v_j(\alpha)] - \sum_i q_i \sum_{j \in J_i} \max_{\alpha \in y} v_j(\alpha).$$

For each i , let α_i denote a maximizer of $u + \sum_{j \in J_i} v_j$ over $\alpha \in y$. Let α^* denote any element of $B(y)$. Then

$$V(y) - u(\alpha^*) = \left[\sum_i q_i u(\alpha_i) - u(\alpha^*) \right] + \sum_i q_i \sum_{j \in J_i} \left[v_j(\alpha_i) - \max_{\alpha \in y} v_j(\alpha) \right].$$

Because $q_i > 0$ for all i and $\sum_i q_i = 1$, $\alpha^* \in B(y)$ implies $\sum_i q_i u(\alpha_i) \leq u(\alpha^*)$, so the first term in brackets is weakly negative. Obviously, for all i and j , the second term in brackets is weakly negative as well. By hypothesis, y is temptation-free, so $\{\alpha^*\} \sim y$. Hence $u(\alpha^*) = V(y)$. Hence both terms must be zero.

Note that

$$\left[\sum_i q_i u(\alpha_i) - u(\alpha^*) \right] = 0$$

if and only if $\alpha_i \in B(y)$ for all i . Also,

$$\sum_i q_i \sum_{j \in J_i} \left[v_j(\alpha_i) - \max_{\alpha \in y} v_j(\alpha) \right] = 0$$

if and only if for all i and all $j \in J_i$, α_i maximizes v_j over $\alpha \in y$. From the above, every $\alpha_i \in B(y)$. Since $\beta \notin B(y)$, we cannot have $\beta = \alpha_i$ for any i . Hence

$$\max_{\alpha \in x \cup \{\beta\}} v_j(\alpha) = \max_{\alpha \in x} v_j(\alpha) \geq v_j(\beta), \quad \forall j.$$

Also, the fact that $\beta \notin B(x \cup \{\beta\})$ implies

$$\max_{\alpha \in x \cup \{\beta\}} u(\alpha) = \max_{\alpha \in x} u(\alpha) > u(\beta).$$

Hence for all i , we have

$$\max_{\alpha \in x} \left[u(\alpha) + \sum_{j \in J_i} v_j(\alpha) \right] > u(\beta) + \sum_{j \in J_i} v_j(\beta).$$

By continuity, then, for any $\hat{\beta}$ sufficiently close to β ,

$$\max_{\alpha \in x} \left[u(\alpha) + \sum_{j \in J_i} v_j(\alpha) \right] > u(\hat{\beta}) + \sum_{j \in J_i} v_j(\hat{\beta}).$$

Hence for any x' , we can write $V(x' \cup x \cup \{\hat{\beta}\})$ as

$$\begin{aligned} & \sum_i q_i \max_{\alpha \in x' \cup x \cup \{\hat{\beta}\}} \left[u(\alpha) + \sum_{j \in J_i} v_j(\alpha) \right] - \sum_i q_i \sum_{j \in J_i} \max_{\alpha \in x' \cup x \cup \{\hat{\beta}\}} v_j(\alpha) \\ &= \sum_i q_i \max_{\beta \in x' \cup x} \left[u(\alpha) + \sum_{j \in J_i} v_j(\alpha) \right] - \sum_i q_i \sum_{j \in J_i} \max_{\alpha \in x' \cup x \cup \{\hat{\beta}\}} v_j(\alpha) \\ &\geq \sum_i q_i \max_{\beta \in x' \cup x} \left[u(\alpha) + \sum_{j \in J_i} v_j(\alpha) \right] - \sum_i q_i \sum_{j \in J_i} \max_{\alpha \in x' \cup x} v_j(\alpha) \\ &= V(x' \cup x). \end{aligned}$$

So $x' \cup x \succeq x' \cup x \cup \{\hat{\beta}\}$. Hence \succ satisfies dominance. ■

Theorem 5 \succ has a temptation representation if and only if it has a finite state additive EU representation, satisfies DFC, and satisfies dominance.

Proof. In light of Lemma 1, necessity is obvious. Henceforth let \succ denote a preference with a finite state additive EU representation V which satisfies DFC and dominance. Before moving to the main part of the proof of sufficiency, we handle a couple of simple special cases to simplify the main analysis. First, it is easy to see that if \succ has a finite state additive EU representation, then it has such a representation which is nonredundant in the sense that no u_i or v_j is a constant function and no two states correspond to the same preference. On the other hand, this nonredundant representation could have $I = 0$, $J = 0$, or both. We first handle these cases, then subsequently focus on the case where $I \geq 1$, $J \geq 1$, no state is a constant preference, and no two states have the same preference.

If $I = J = 0$, the preference is trivial in the sense that $x \sim x'$ for all x and x' . In this case, the preference is obviously represented by the temptation representation

$$V(x) = \max_{\beta \in x} [u(\beta) + v(\beta)] - \max_{\beta \in x} v(\beta)$$

where v and u are constant functions. If $I = 0$ but $J \geq 1$, then we have

$$V(x) = K - \sum_j \max_{\beta \in x} v_j(\beta)$$

for an arbitrary constant K . Let w_1 denote a constant function equal to K and define $u = w_1 - \sum_j v_j$. Then

$$V(x) = \max_{\beta \in x} [u(\beta) + \sum_j v_j(\beta)] - \sum_j \max_{\beta \in x} v_j(\beta),$$

giving a temptation representation. Finally, suppose $J = 0$. To satisfy DFC, we must then have $I = 1$, so $V(x) = \max_{\beta \in x} w_1(\beta) + K$ for an arbitrary constant K . Let v_1 be a constant function equal to K and define $u = w_1 - v_1$. Then obviously

$$V(x) = \max_{\beta \in x} [u(\beta) + v_1(\beta)] - \max_{\beta \in x} v_1(\beta),$$

giving a temptation representation.

The remainder of the proof shows the result for the case where the finite additive EU representation has $I \geq 1$ positive states and $J \geq 1$ negative states, none of which are constant and no two of which correspond to the same preference over menus. Following GP, we refer to this as a *regular* representation.

Lemma 2 *Suppose \succ satisfies dominance and has a regular, finite state additive EU representation given by*

$$V(x) = \sum_i \max_{\beta \in x} w_i(\beta) - \sum_j \max_{\beta \in x} v_j(\beta).$$

Fix any interior β and any x such that $x \cup \{\beta\}$ is temptation-free and $\beta \notin B(x \cup \{\beta\})$. Then there is no i with

$$w_i(\beta) = \max_{\alpha \in x \cup \{\beta\}} w_i(\alpha).$$

Proof. Suppose not. Suppose there is an interior β , an x such that $x \cup \{\beta\}$ is temptation-free and $\beta \notin B(x \cup \{\beta\})$, and an i with

$$w_i(\beta) = \max_{\alpha \in x \cup \{\beta\}} w_i(\alpha).$$

Because $\beta \notin B(x \cup \{\beta\})$, we know that $u(\beta) < \max_{\alpha \in x} u(\alpha)$. By hypothesis, the additive EU representation is regular so w_i is not constant. Because w_i is not constant and β is interior, for any $\varepsilon > 0$, we can find a $\hat{\beta}$ within an ε neighborhood of β such that $u_i(\hat{\beta}) > u_i(\beta)$. Hence $u_i(\hat{\beta}) > \max_{\alpha \in x} u_i(\alpha)$. Obviously, if ε is sufficiently small, we will have $u(\hat{\beta})$ close to $u(\beta)$ and hence $u(\hat{\beta}) < \max_{\alpha \in x} u(\alpha)$.

Let \hat{J} denote the set of j such that

$$v_j(\hat{\beta}) > \max_{\alpha \in x} v_j(\alpha).$$

For each $j \in \hat{J}$, we can find a γ_j such that $v_j(\hat{\beta}) = v_j(\gamma_j)$ and $w_i(\hat{\beta}) > w_i(\gamma_j)$. To see that this must be possible, note that the selection of j implies that w_i and $-v_j$ do not represent the same preference. By hypothesis, the additive EU representation is regular so w_i and v_j do not represent the same preference and neither is constant. Hence the v_j indifference curve through $\hat{\beta}$ must have a nontrivial intersection with the u_i indifference curve through $\hat{\beta}$. Hence such a γ_j must exist.

Let x' denote the collection of these γ_j 's. (If $\hat{J} = \emptyset$, then $x' = \emptyset$.) By dominance,

$$x' \cup x \cup \{\hat{\beta}\} \preceq x' \cup x.$$

Recall that w_i ranks $\hat{\beta}$ above any $\alpha \in x$ and above any of the γ_j 's. Also, by construction of the γ_j 's, each v_j ranks some point in $x' \cup x$ (at least weakly) above $\hat{\beta}$. Hence

$$V(x' \cup x \cup \{\hat{\beta}\}) = w_i(\hat{\beta}) + \sum_{k \neq i} \max_{\alpha \in x' \cup x \cup \{\hat{\beta}\}} w_k(\alpha) - \sum_j \max_{\alpha \in x' \cup x} v_j(\alpha).$$

Using the fact that the w_i comparison of $\hat{\beta}$ to any $\alpha \in x$ or any γ_j is strict, this expression is

$$> \max_{\alpha \in x' \cup x} w_i(\alpha) + \sum_{k \neq i} \max_{\alpha \in x' \cup x \cup \{\hat{\beta}\}} w_k(\alpha) - \sum_j \max_{\alpha \in x' \cup x} v_j(\alpha).$$

Obviously, this is

$$\geq \sum_k \max_{\alpha \in x' \cup x} w_k(\alpha) - \sum_j \max_{\alpha \in x' \cup x} v_j(\alpha) = V(x' \cup x).$$

Hence $x' \cup x \cup \{\hat{\beta}\} \succ x' \cup x$, contradicting dominance. ■

To complete the proof of Theorem 5, we use the following result from Rockafellar [1970] (Theorem 22.2, pages 198–199):

Lemma 3 *Let $z_i \in \mathbf{R}^N$ and $Z_i \in \mathbf{R}$ for $i = 1, \dots, m$ and let ℓ be an integer, $1 \leq \ell \leq m$. Assume that the system $z_i \cdot y \leq Z_i$, $i = \ell + 1, \dots, m$ is consistent. Then one and only one of the following alternatives holds:*

(a) *There exists a vector y such that*

$$\begin{aligned} z_i \cdot y &< Z_i, \quad i = 1, \dots, \ell \\ z_i \cdot y &\leq Z_i, \quad i = \ell + 1, \dots, m \end{aligned}$$

(b) *There exist non-negative real numbers $\lambda_1, \dots, \lambda_m$ such that at least one of the numbers $\lambda_1, \dots, \lambda_\ell$ is not zero, and*

$$\sum_{i=1}^m \lambda_i z_i = 0$$

$$\sum_{i=1}^m \lambda_i Z_i \leq 0.$$

It is easy to use this result to show that if we have some equality constraints, we simply drop the requirement that the corresponding λ 's are non-negative.

Fix \succ with a regular finite state additive EU representation which satisfies DFC and dominance. We use Lemma 3 to show that there exists $a_1, \dots, a_I, b_{11}, \dots, b_{IJ}$, and c_1, \dots, c_I such that

$$\begin{aligned} a_i u + \sum_j b_{ij} v_j + c_i \mathbf{1} &= w_i \\ \sum_i a_i &= 1 \\ \sum_i b_{ij} &= 1, \quad \forall j \\ -b_{ij} &\leq 0 \\ -a_i &< 0. \end{aligned}$$

Because DFC implies that a weak temptation representation exists, the part of the system with only weak inequality constraints is obviously consistent. To state the alternatives implied by the lemma in the most straightforward way possible, let λ_{ik} denote the real number corresponding to the equation

$$a_i u(k) + \sum_j b_{ij} v_j(k) + c_i = w_i(k)$$

where k denotes the k th pure outcome. We use $\bar{\mu}$ to correspond to the equation $\sum_i a_i = 1$, μ_j for the equation $\sum_i b_{ij} = 1$, φ_{ij} for $-b_{ij} \leq 0$, and ψ_i for $-a_i < 0$. Hence Lemma 3 implies that either a strict representation exists or there exists $\lambda_{ik}, \bar{\mu}, \mu_j, \varphi_{ij}$, and ψ_i such that

$$\begin{aligned} \varphi_{ij} &\geq 0, \quad \forall i, j \\ \psi_i &\geq 0, \quad \forall i, \text{ strictly for some } i \\ \sum_k \lambda_{ik} u(k) + \bar{\mu} - \psi_i &= 0, \quad i = 1, \dots, I \\ \sum_k \lambda_{ik} v_j(k) + \mu_j - \varphi_{ij} &= 0, \quad i = 1, \dots, I; j = 1, \dots, J \\ \sum_k \lambda_{ik} &= 0, \quad i = 1, \dots, I \\ \sum_i \sum_k \lambda_{ik} w_i(k) + \bar{\mu} + \sum_j \mu_j &\leq 0 \end{aligned}$$

Assume, contrary to the claim above, that no a_i 's, b_{ij} 's, and c_i 's exist satisfying the conditions postulated. Then by Lemma 3, there must be a solution to this system of equations. Note that we cannot have a solution to these equations with $\lambda_{ik} = 0$ for all i and k . To see this, note that we would then have $\bar{\mu} = \psi_i$ for all i and hence $\bar{\mu} > 0$. Also, we would have $\mu_j = \varphi_{ij}$ and hence $\mu_j \geq 0$ for all j . But then the last equation gives $\bar{\mu} + \sum_j \mu_j \leq 0$, a contradiction. Since $\sum_k \lambda_{ik} = 0$, this implies $\max_{i,k} \lambda_{ik} > 0$. Without loss of generality, then, we can assume that $\lambda_{ik} < 1/n$ for all i and k . (Recall that there are n pure outcomes.) Otherwise, we can divide through all equations by $2n \max_{i,k} \lambda_{ik}$ and redefine all variables appropriately.

Rearranging the equations gives

$$\begin{aligned} \sum_k \lambda_{ik} u(k) + \bar{\mu} &= \psi_i \geq 0, \quad \forall i \text{ with strict inequality for some } i \\ \sum_k \lambda_{ik} v_j(k) + \mu_j &= \varphi_{ij} \geq 0, \quad \forall i, j \\ \sum_i \sum_k \lambda_{ik} w_i(k) + \bar{\mu} + \sum_j \mu_j &\leq 0 \end{aligned}$$

For each i , define an interior probability distribution α_i by $\alpha_i(k) = (1/n) - \lambda_{ik}$. Because $\lambda_{ik} < 1/n$ for all i and k , we have $\alpha_i(k) > 0$ for all i and k . Also, $\sum_k \alpha_i(k) = 1 - \sum_k \lambda_{ik} = 1$. Letting β denote the probability distribution $(1/n, \dots, 1/n)$, we can rewrite the above as

$$\begin{aligned} u(\beta) + \bar{\mu} &\geq u(\alpha_i), \quad \forall i \text{ with strict inequality for some } i \\ v_j(\beta) + \mu_j &\geq v_j(\alpha_i), \quad \forall i \\ \sum_i w_i(\beta) + \bar{\mu} + \sum_j \mu_j &\leq \sum_i w_i(\alpha_i). \end{aligned}$$

The first inequality implies

$$u(\beta) + \bar{\mu} \geq \max_i u(\alpha_i) \tag{1}$$

with a strict inequality for some i . The second inequality implies

$$\sum_j v_j(\beta) + \sum_j \mu_j \geq \sum_j \max_i v_j(\alpha_i). \tag{2}$$

Turning to the third inequality, recall that $\sum_i w_i = u + \sum_j v_j$. Hence the third inequality is equivalent to

$$u(\beta) + \sum_j v_j(\beta) + \bar{\mu} + \sum_j \mu_j \leq \sum_i w_i(\alpha_i).$$

Summing equations (1) and (2) yields

$$u(\beta) + \sum_j v_j(\beta) + \bar{\mu} + \sum_j \mu_j \geq \max_i u(\alpha_i) + \sum_j \max_i v_j(\alpha_i)$$

so

$$\sum_i w_i(\alpha_i) - \sum_j \max_i v_j(\alpha_i) \geq u(\beta) + \sum_j v_j(\beta) + \bar{\mu} + \sum_j \mu_j - \sum_j \max_i v_j(\alpha_i) \geq \max_i u(\alpha_i). \quad (3)$$

Let $x = \{\alpha_1, \dots, \alpha_I\}$. Then

$$V(x) \geq \sum_i w_i(\alpha_i) - \sum_j \max_i v_j(\alpha_i) \geq \max_i u(\alpha_i).$$

By DFC, $\max_i u(\alpha_i) \geq V(x)$. Hence

$$V(x) = \sum_i w_i(\alpha_i) - \sum_j \max_i v_j(\alpha_i) = \max_i u(\alpha_i).$$

Hence x is a temptation-free menu. Note that the first equality above implies that α_i maximizes w_i for all i . Also, the second equality together with equation (3) implies that the weak inequalities in equations (1) and (2) must be equalities. In particular, then,

$$u(\beta) + \bar{\mu} = \max_i u(\alpha_i).$$

However, recall that

$$u(\beta) + \bar{\mu} \geq u(\alpha_i), \quad \forall i \text{ with strict inequality for some } i$$

That is, there must be some k for which $u(\alpha_k) < \max_i u(\alpha_i)$. Hence $x \neq B(x)$. But α_i maximizes w_i for every i , contradicting Lemma 2.

Hence there must exist such a_i , b_{ij} , and c_i . It is easy to use the proof of Corollary 1 to complete the construction of a temptation representation. ■

5 Special Cases

We can completely characterize the preferences corresponding to two special cases of temptation representations, specifically two of the three representations used in the examples. First, consider a representation of the form

$$V_{1P}(x) = \max_{\beta \in x} [u(\beta) + \sum_{j=1}^J v_j(\beta)] - \sum_{j=1}^J \max_{\beta \in x} v_j(\beta)$$

which we call a *one positive state representation*. Equivalently,

$$V_{1P}(x) = \max_{\beta \in x} [u(\beta) - c(\beta, x)]$$

where

$$c(\beta, x) = \left[\sum_{j=1}^J \max_{\beta' \in x} v_j(\beta') \right] - \sum_{j=1}^J v_j(\beta).$$

We call this representation one–positive state because if we have a finite state additive EU representation with one positive state, it can always be written this way by a generalization of the change of variables discussed in Section 2. Specifically, suppose we have a representation of the form

$$V(x) = \max_{\beta \in x} w_1(\beta) - \sum_{j=1}^J \max_{\beta \in x} v_j(\beta).$$

The commitment utility u is defined by $u(\beta) = V(\{\beta\}) = w_1(\beta) - \sum_j v_j(\beta)$. Hence we can change variables to rewrite V in the form of V_{1P} .

The one positive state representation turns out to correspond to a particular half of set betweenness. Specifically,

Axiom 8 (Positive Set Betweenness) \succ *satisfies positive set betweenness if whenever $x \succeq y$, we have $x \succeq x \cup y$.*

For future use, we define the other half similarly:

Axiom 9 (Negative Set Betweenness) \succ *satisfies negative set betweenness if whenever $x \succeq y$, we have $x \cup y \succeq y$.*

To see the intuition, suppose \succ satisfies positive set betweenness and suppose $x \succeq y$. Then $x \cup y$ is bounded “on the positive side” in the sense that $x \succeq x \cup y$. Hence the flexibility of being able to choose between x and y has only negative consequences. That is, the flexibility to choose between x and y cannot be better than x , though it can, conceivably, be worse than y . Hence the uncertainty the agent faces regarding her tastes is entirely on the negative side. this implies that there may be multiple negative states but can only be one positive one.

Positive set betweenness can be seen as a strengthening of DFC in the sense that

Lemma 4 *Suppose \succ is a weak order satisfying positive set betweenness. Then for any finite menu x , there is some $\alpha \in x$ with $\{\alpha\} \succeq x$.*

Proof. Consider a menu $x = \{\alpha, \beta\}$ where, without loss of generality, $\{\alpha\} \succeq \{\beta\}$. By positive set betweenness, $\{\alpha\} \succeq \{\beta\}$ implies $\{\alpha\} \succeq \{\alpha, \beta\}$, giving the desired conclusion for this menu. Now suppose we have shown that for every menu with cardinality less than or equal to k , the conclusion of the lemma holds. Consider any menu x with cardinality $k+1$. Let α satisfy $\{\alpha\} \succeq \{\beta\}$ for all $\beta \in x$. Let α' satisfy $\{\alpha'\} \succeq \{\beta\}$ for all $\beta \in x \setminus \{\alpha\}$. By the induction hypothesis, $\{\alpha'\} \succeq x \setminus \alpha$. By definition, $\{\alpha\} \succeq \{\alpha'\}$, so since \succ is a weak order, we have $\{\alpha\} \succeq x \setminus \{\alpha\}$. Hence by positive set betweenness, we have $\{\alpha\} \succeq x$, the desired conclusion. ■

We have

Theorem 6 \succ has a one positive state representation if and only if it has a finite state additive EU representation and satisfies positive set betweenness.

Proof. (Necessity.) The necessity of \succ having a finite state additive EU representation is obvious. So let us show that if \succ has a finite state additive EU representation with only one positive state and $x \succeq y$, then $x \succeq x \cup y$. It is not hard to see that

$$V(x \cup y) = \sum_i \max \left\{ \max_{\beta \in x} w_i(\beta), \max_{\beta \in y} w_i(\beta) \right\} - \sum_j \max \left\{ \max_{\beta \in x} v_j(\beta), \max_{\beta \in y} v_j(\beta) \right\}.$$

When there is only one positive state, $I = 1$, so we can rewrite this as

$$\begin{aligned} V(x \cup y) &= \max \left\{ \max_{\beta \in x} w_1(\beta), \max_{\beta \in y} w_1(\beta) \right\} \\ &\quad - \sum_j \max \left\{ \max_{\beta \in x} v_j(\beta), \max_{\beta \in y} v_j(\beta) \right\}. \end{aligned}$$

Hence

$$\begin{aligned} V(x \cup y) &\leq \max \left\{ \max_{\beta \in x} w_1(\beta), \max_{\beta \in y} w_1(\beta) \right\} \\ &\quad - \max \left\{ \sum_j \max_{\beta \in x} v_j(\beta), \sum_j \max_{\beta \in y} v_j(\beta) \right\} \\ &\leq \max \left\{ \max_{\beta \in x} w_1(\beta) - \sum_j \max_{\beta \in x} v_j(\beta), \right. \\ &\quad \left. \max_{\beta \in y} w_1(\beta) - \sum_j \max_{\beta \in y} v_j(\beta) \right\} \\ &= \max \{V(x), V(y)\} = V(x). \end{aligned}$$

Hence $x \succeq x \cup y$.

(Sufficiency.) Suppose \succ has a finite state additive EU representation and satisfies positive set betweenness. Assume, contrary to our claim, that this representation has more than one positive state. (It is sufficient to show that there is only one positive state since, as shown above, the change of variables then yields the form V_{1P} .) So \succ has a representation of the form

$$V(x) = \sum_{i=1}^I \max_{\beta \in x} w_i(\beta) - \sum_{j=1}^J \max_{\beta \in x} v_j(\beta)$$

where $I \geq 2$. Without loss of generality, we can assume that w_1 and w_2 represent different preferences over $\Delta(B)$ — otherwise, we can rewrite the representation to combine these two states into one. Let \hat{x} denote a sphere in the interior of $\Delta(B)$. Let

$$x = \left[\bigcap_{i=1}^I \{\beta \in \Delta(B) \mid w_i(\beta) \leq \max_{\beta' \in \hat{x}} w_i(\beta')\} \right] \cap \left[\bigcap_{j=1}^J \{\beta \in \Delta(B) \mid v_j(\beta) \leq \max_{\beta' \in \hat{x}} v_j(\beta')\} \right].$$

Because \hat{x} is a sphere and because I and J are finite, there must be a w_i indifference curve which makes up part of the boundary of x for $i = 1, 2$. Fix a small $\varepsilon > 0$. For $i = 1, 2$ and $k = 1, \dots, I$, let $\varepsilon_k^i = 0$ for $k \neq i$ and $\varepsilon_k^i = \varepsilon$. Finally, for $i = 1, 2$, let y_i equal

$$\left[\bigcap_{k=1}^I \{\beta \in \Delta(B) \mid w_k(\beta) \leq \max_{\beta' \in \hat{x}} w_k(\beta') - \varepsilon_k^i\} \right] \cap \left[\bigcap_{j=1}^J \{\beta \in \Delta(B) \mid v_j(\beta) \leq \max_{\beta' \in \hat{x}} v_j(\beta')\} \right].$$

Because I and J are finite, if ε is sufficiently small,

$$\max_{\beta \in y_i} w_k(\beta) = \max_{\beta \in x} w_k(\beta), \quad \forall k \neq i$$

and

$$\max_{\beta \in y_i} v_j(\beta) = \max_{\beta \in x} v_j(\beta), \quad \forall j.$$

Hence $x \sim y_1 \cup y_2$. Also,

$$\max_{\beta \in y_i} w_i(\beta) < \max_{\beta \in x} w_i(\beta).$$

Hence $x \succ y_i$, $i = 1, 2$. Hence $y_1 \cup y_2 \succ y_i$, $i = 1, 2$, contradicting positive set betweenness. ■

One can modify the proof of Theorem 6 in obvious ways to show

Theorem 7 \succ has a finite state additive EU representation with one negative state if and only if it has a finite state additive EU representation and satisfies negative set betweenness.

Theorem 3 is obviously a corollary to Theorems 6 and 7.

A second special case takes Theorem 7 as its starting point. This representation has one negative state but many positive states which differ only in the strength of temptation in that state. Specifically, we define an *uncertain strength of temptation representation* to be one which takes the form

$$V_{US}(x) = \sum_{i=1}^I q_i \left[\max_{\beta \in x} [u(\beta) + \gamma_i v(\beta)] - \gamma_i \max_{\beta \in x} v(\beta) \right]$$

where $q_i > 0$ for all i and $\sum_i q_i = 1$. Equivalently,

$$V_{US}(x) = \sum_i q_i \max_{\beta \in x} [u(\beta) - \gamma_i c(\beta, x)]$$

where

$$c(\beta, x) = [\max_{\beta' \in x} v(\beta')] - v(\beta).$$

In this representation, the temptation is always v , but the strength of the temptation (as measured by γ_i) is random. The probability that the strength of the temptation is γ_i is given by q_i .

We have

Theorem 8 \succ has an uncertain strength of temptation representation if and only if it has a finite state additive EU representation and satisfies DFC and negative set betweenness.

Proof. Necessity is obvious. For sufficiency, assume \succ has a finite state additive EU representation and satisfies DFC and negative set betweenness. We know from Theorem 7 that it has only one negative state. Using this and Theorem 4, we see that \succ can be represented by a function V of the form

$$V(x) = \sum_{i=1}^I \max_{\beta \in x} [a_i u(\beta) + b_i v(\beta)] - \max_{\beta \in x} v(\beta)$$

where $a_i \geq 0$ and $b_i \geq 0$ for all i and $\sum_i a_i = \sum_i b_i = 1$. (The argument in the proof of Corollary 1 showing that $\sum_i c_i = 0$ applies here as well.)

We can assume without loss of generality that $a_i > 0$ for all i . To see this, suppose $a_1 = 0$. Then we can write

$$V(x) = \sum_{i=2}^I \max_{\beta \in x} [a_i u(\beta) + b_i v(\beta)] - \max_{\beta \in x} (1 - b_1) v(\beta).$$

If $b_1 = 1$, then $b_i = 0$ for all $i \neq 1$. Because $a_1 = 0$ and $\sum_i a_i = 1$, we then have $V(x) = \max_{\beta \in x} u(\beta)$. This is a V_{US} representation with $I = 1$ and $\gamma_1 = 0$. So suppose $b_1 < 1$. Let $\hat{v} = (1 - b_1)v$ and for $i = 2, \dots, I$, let $\hat{b}_i = b_i / (1 - b_1)$. Note that $\sum_{i=2}^I \hat{b}_i = 1$. Hence we can rewrite V as

$$V(x) = \sum_{i=2}^I \max_{\beta \in x} [a_i u(\beta) + \hat{b}_i \hat{v}(\beta)] - \max_{\beta \in x} \hat{v}(\beta).$$

Continuing as needed, we can eliminate any i with $a_i = 0$.

Given that $a_i > 0$ for all i , let $q_i = a_i$ and let $\gamma_i = b_i / a_i$. With this change of notation, V can be rewritten in the form of V_{US} . ■

6 Speculations

There are several interesting issues left to explore. In the previous section, we characterized the additional axioms needed to yield two of the three models we discussed in the examples in Section 3. We are currently working on determining the axioms which characterize the third model.

Another direction of interest is the extent to which a strict or weak representation is identified. It is not hard to show that the coefficients are unique if the vectors $u, v_1, \dots, v_J, \mathbf{1}$ are linearly independent. It is also not hard to show by example that the coefficients need not be unique otherwise. The sufficiency of this condition for uniqueness is related to the fact that in the Harsanyi aggregation theorem (Harsanyi [1955]), the coefficients of the aggregation are unique when the analog of this condition holds. In that case, the independence condition is both necessary and sufficient, not just sufficient, for uniqueness. Here, however, we have restrictions on the coefficients across i , a source of restrictions with no analog in the Harsanyi aggregation literature.⁶ Hence the independence condition may not be required for uniqueness in our case. At this point, we do not know a necessary and sufficient condition for uniqueness. We also do not know to what extent the coefficients are identified when uniqueness fails.

⁶See Weymark [1991] for a detailed discussion of when the coefficients in the Harsanyi aggregation are unique.

References

- [1] Dekel, E., B. Lipman, and A. Rustichini, “Representing Preferences with a Unique Subjective State Space,” *Econometrica*, **69**, July 2001, 891–934.
- [2] Gul, F., and W. Pesendorfer, “Temptation and Self–Control,” *Econometrica*, **69**, November 2001, 1403–1435.
- [3] Gul, F., and W. Pesendorfer, “Self–Control and the Theory of Consumption,” *Econometrica*, **72**, January 2004a, 119–158.
- [4] Gul, F., and W. Pesendorfer, “A Theory of Addiction,” Princeton University working paper, 2004b.
- [5] Harsanyi, J., “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility,” *Journal of Political Economy*, **63**, 1955, 309–321.
- [6] Kreps, D., “A Representation Theorem for ‘Preference for Flexibility’,” *Econometrica*, **47**, May 1979, 565–576.
- [7] Kreps, D., “Static Choice and Unforeseen Contingencies” in *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, P. Dasgupta, D. Gale, O. Hart, and E. Maskin, eds., Cambridge, MA: MIT Press, 1992, 259–281.
- [8] Rockafellar, R. T., *Convex Analysis*, Princeton, NJ: Princeton University Press, 1970.
- [9] Weymark, J., “A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism,” in J. Elster and J. Roemer, eds., *Interpersonal Comparisons of Well–Being*, Cambridge: Cambridge University Press, 1991, 255–320.