Temptation–Driven Preferences[1]

Eddie Dekel[2] Barton L. Lipman[3] Aldo Rustichini[4]

December 2003
Preliminary and Incomplete Draft

[1]This draft is very preliminary. Dekel and Rustichini have not yet "approved" this draft, so any mistakes are Lipman's fault. Thanks to Phil Reny for helpful discussion.

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

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

[4]University 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

1 Introduction

In a recent series of papers, Gul and Pesendorfer (GP) [2001a, 2001b, 2002] extend a framework originally due to Dekel, Lipman, and Rustichini (DLR) [2001] to the study of temptation and self–control, raising some very intriguing questions. The key idea behind these papers is originally due to Kreps [1979, 1992]. Kreps' insight was that an agent's preferences regarding flexibility reveals the utility possibilities she sees. Kreps considered preferences over *menus* — that is, sets of consumption bundles — with the interpretation that the choice of a menu corresponds to the commitment to choose one of the items on the menu at a later date. To see the kind of information revealed by the agent's preferences over menus, suppose the menu "chicken or fish" is strictly preferred to the menu "chicken" and to the menu "fish." This is naturally interpreted as saying that the agent considers it possible that she will prefer chicken to fish and possible that she will prefer fish to chicken.

DLR showed that significantly tighter results could be obtained if one considered menus of lotteries rather than menus of deterministic options. Letting B denote a finite set of consumption bundles and $\Delta(B)$ the set of probability distributions over B, DLR considered a preference relation $\succ$ over the set of closed nonempty subsets of $\Delta(B)$, denoted X. Among other things, they characterized the set of preferences over X which could be represented by a utility function V taking the form

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

where each $U(\cdot, s)$ is an expected–utility function and μ is a measure, not necessarily positive. DLR call this an *additive EU representation.* To see the idea, think of s as the state of the world which the agent does not know at the time she chooses a menu. She will learn s later and then choose from the menu. Learning s reveals her utility function over menu items, $U(\cdot, s)$. Naturally, then, she will choose the best item on the menu x according to these preferences. The value of the menu, then, is the expected utility associated with this process (though this interpretation is strained when μ is negative).

GP [2001a] recognized that temptation and self–control could also be studied using this sets of lotteries framework. 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. The key axiom GP introduce, *set betweenness*, says that for any menus x and y,

$$x \succeq y \ \text{ implies } \ x \succeq x \cup y \succeq y.$$

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.

GP show that such a preference can be represented by a utility function of the form

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

where u and v are expected utility functions. 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. In line with this, we can think of the agent's utility if she chooses β from menu x as $u(\beta) - c(\beta, x)$ where c is a self–control cost. The utility of menu x then is the maximum of this expression over $\beta \in x$. A natural way to think about this self–control cost is to view it as the foregone "temptation utility" associated with β — that is, to let $c(\beta, x) = [\max_{\beta' \in x} v(\beta')] - v(\beta)$ where we think of v as the utility of succumbing to temptation. Substituting this for c yields the GP representation.

Another way to understand the GP representation is to relate it to DLR. It is easy to show[1] that an additive EU representation for a preference satisfying set betweenness must take the form

$$\max_{\beta \in x} U_1(\beta) - \max_{\beta \in x} U_2(\beta).$$

If we change variables by letting $v = U_2$ and $u = U_1 - U_2$, we obtain the GP form of the representation.

There are many aspects of temptation which are potentially relevant to explaining consumption behavior. We see the GP representation and axioms as a very nice way of capturing one of these ideas. As shown in Gul–Pesendorfer [2002], their model of temptation has significant and often surprising implications for consumption behavior. On the other hand, there seem to be many aspects of temptation which this omits, the inclusion of which could potentially change the model's predictions. While such omission is the natural way to begin research, we believe it is time to begin introducing some of the missing elements.

[1] This is an immediate implication of GP's representation theorem. One can also prove it directly as we do in Theorems 4 and 5 below.

More specifically, the set of preferences DLR considered includes preferences driven by flexibility, preferences driven by temptation, and preferences affected by elements of both concerns. To study temptation, it is natural to focus on preferences which are only driven by temptation and not a desire for flexibility. As we argue in more detail below, GP have identified a subset of such "temptation–driven" preferences. In this paper, we seek to characterize the full subset of the preferences considered by DLR which are temptation–driven. As we will argue by means of examples, this broader class of preferences is needed if we wish to allow some realistic forms of temptation.

In the next section, we present the basic model and state our research goals more precisely. In Section 3, we give examples to motivate the issues and illustrate the kinds of representations we are interested in. As of this writing, we do not have the full characterization we seek. In Section 4, we give the results we currently have. Section 5 contains some discussion of anticipated/desired further results and concludes.

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 function*[2] $U : \Delta(B) \times S \to \mathbf{R}$*, 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 theorem[3] in DLR [2001] is:

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_+ denote the set of *positive states* — those with $\mu(s) > 0$ — and let S_- denote the set of *negative states* — those with $\mu(s) < 0$. Let $p(s) = |\mu(s)|$. Then we can write

$$V(x) = \sum_{s \in S_+} p(s) \max_{\beta \in x} U(\beta, s) - \sum_{s \in S_-} p(s) \max_{\beta \in x} U(\beta, s).$$

Another axiom DLR consider is

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

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

[3]This 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.

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.*

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 his possible future needs are.

GP's [2001a] 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, as discussed in the introduction, they add an axiom which they call *set betweenness*:

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

Their representation:

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

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

represents $\succ$.

They show that

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

Obviously, then, a self control representation is a special case of an additive EU representation. As discussed in the introduction, it is easy to rewrite an additive EU representation with one positive state and one negative state in the form of GP's self control representation.

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. We will call such preferences *flexibility–driven*. Analogously, we will refer to preferences which exhibit a concern about temptation but no value to flexibility *per se* as *temptation–driven*. Clearly, the subset of DLR preferences that satisfy set betweenness are temptation–driven preferences.

Intuitively, monotonicity is a clear statement of the idea of flexibility–driven preferences. However, set betweenness does not appear to be an obvious analog for temptation. 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.

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:

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

Intuitively, the agent doesn't know whether the temptation that will strike is the one described by v_1 (where she is most tempted by the potato chips) or v_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 $V_1(\{b\}) = 3$, $V_1(\{b, c\}) = V_1(\{b, p\}) = 3/2$, 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

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

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

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 $V(\{b, c, p\}) = -5$, 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:

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

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.

[5] We 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.

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. In particular, instead of having an arbitrary utility function for a given positive state, it is some linear combination of the commitment utility u and the negative state utility functions. 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

$$V_{GR}(x) = \sum_{i=1}^{I} q_i \left[\max_{\beta \in x} [u(\beta) + \sum_{j=1}^{J} \gamma_{ij} v_j(\beta)] - \sum_{j=1}^{J} \gamma_{ij} \max_{\beta \in x} v_j(\beta) \right]$$

where $q_i > 0$ for all i, $\sum_i q_i = 1$, and $\gamma_{ij} \geq 0$ for all i and j. The q_i's give the probabilities over the I different ways temptation may affect the agent. There are J different kinds of temptations, where γ_{ij} gives the strength of temptation j in situation i. Note that $V_{GR}(\{\beta\}) = u(\beta)$, so u is again the commitment utility. We can rewrite this representation as

$$V_{GR} = \sum_i q_i \max_{\beta \in x} [u(\beta) - c_i(\beta, x)]$$

where

$$c_i(\beta, x) = \sum_{j=1}^{J} \gamma_{ij} \max_{\beta \in x} v_j(\beta) - \sum_{j=1}^{J} \gamma_{ij} v_j(\beta).$$

This version of the representation is naturally interpreted as having uncertain self–control costs.

Our goal, restated, is to characterize the subset of the DLR preferences with a representation of form of V_{GR}.

Additionally, we wish to characterize the extent to which the representation is identified. In our discussion of Example 1, we noted two reasons why two snacks could be worse than one, each of which corresponded to its own representation. As we will see, a preference which has one of these kinds of representations cannot have the other. In this sense, the two reasons why two snacks can be worse are behaviorally distinguishable. The general question is whether behavior (preferences) can identify the set of temptations an individual is subject to, the probabilities over these various temptations, and the "strength" of these temptations.

4 Results

We can completely characterize the preferences corresponding to two special cases of the general representation, 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.* We call it this 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 the introduction. Specifically, suppose we have a representation of the form

$$\max_{\beta \in x} U(\beta, s_0)p(s_0) - \sum_{s \in S_-} \max_{\beta \in x} U(\beta, s)p(s)$$

where S_- is finite. Write $S_- = \{s_1, \ldots, s_J\}$. Define $v_j(\beta) = U(\beta, s_j)p(s_j)$ and let $u(\beta) = U(\beta, s_0)p(s_0) - \sum_{j=1}^{J} v_j(\beta)$, so that u is the commitment utility. This change of variables enables us to rewrite the representation in the form of V_{1P}.

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

Axiom 6 (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 7 (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 his tastes is entirely on the negative side. This implies that there may be multiple negative states but can only be one positive one.

We have

Theorem 4 $\succ$ *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 $x \succeq x \cup y$. By definition,

$$\begin{aligned} V(x \cup y) &= \textstyle\sum_s \max_{\beta \in x \cup y} U(\beta, s)\mu(s) \\ &= \textstyle\sum_s \max\left\{\max_{\beta \in x} U(\beta, s), \max_{\beta \in y} U(\beta, s)\right\}\mu(s). \end{aligned}$$

Let $p(s) = |\mu(s)|$. When there is only one positive state, say s^*, we can rewrite this as

$$\begin{aligned} V(x \cup y) &= \max\left\{\max_{\beta \in x} U(\beta, s^*), \max_{\beta \in y} U(\beta, s^*)\right\} p(s^*) \\ &\quad - \textstyle\sum_{s \in S_-} \max\left\{\max_{\beta \in x} U(\beta, s), \max_{\beta \in y} U(\beta, s)\right\} p(s) \end{aligned}$$

where $S_- = S \setminus \{s^*\}$. Hence

$$\begin{aligned} V(x \cup y) &\leq \max\Big\{\max_{\beta \in x} U(\beta, s^*), \max_{\beta \in y} U(\beta, s^*)\Big\} p(s^*) \\ &\quad - \max\Big\{\textstyle\sum_{s \in S_-} \max_{\beta \in x} U(\beta, s)p(s), \sum_{s \in S_-} \max_{\beta \in y} U(\beta, s)p(s)\Big\} \\ &\leq \max\Big\{\max_{\beta \in x} U(\beta, s^*)p(s^*) - \textstyle\sum_{s \in S_-} \max_{\beta \in x} U(\beta, s)p(s), \\ &\qquad\qquad \max_{\beta \in y} U(\beta, s^*)p(s^*) - \textstyle\sum_{s \in S_-} \max_{\beta \in y} U(\beta, s)p(s)\Big\} \\ &= \max\{V(x), V(y)\} = V(x). \end{aligned}$$

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}.) Let s_1 and s_2 be distinct positive states in the representation. Without loss of generality, we can assume that $U(\cdot, s_1)$ and $U(\cdot, s_2)$ 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

$$x = \bigcap_{s \in S} \{\beta \in \Delta(B) \mid u(\beta, s) \leq \max_{\beta' \in \hat{x}} u(\beta', s)\}.$$

Because $\hat{x}$ is a sphere and because S is finite, there must be a state s_i indifference curve which makes up part of the boundary of x for $i = 1, 2$. Fix a small $\varepsilon > 0$. For $i = 1, 2$, let $\varepsilon_i(s) = 0$ for $s \neq s_i$ and $\varepsilon_i(s_i) = \varepsilon$. Finally, for $i = 1, 2$, define

$$y_i = \bigcap_{s \in S} \{\beta \in \Delta(B) \mid u(\beta, s) \leq \max_{\beta' \in \hat{x}} u(\beta', s) - \varepsilon_i(s)\}.$$

Because the state space is finite, if ε is sufficiently small,

$$\max_{\beta \in y_i} u(\beta, s) = \max_{\beta \in x} u(\beta, s), \quad \forall s \neq s_i.$$

Hence $x \sim y_1 \cup y_2$. Also, because both states are positive, $x \succ y_i$, $i = 1, 2$. Hence $y_1 \cup y_2 \succ y_i$, $i = 1, 2$, contradicting positive set betweenness. ▌

One can modify the proof of Theorem 4 in obvious ways to show

Theorem 5 *$\succ$ has a finite state additive EU representation with a singleton set of negative states 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 4 and 5.

A second representation we can characterize takes Theorem 5 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 $\sum_i q_i = 1$. 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.

A necessary condition for such a representation is

Axiom 8 (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.

This axiom can be seen as a weakening of positive set betweenness in the sense that

Lemma 1 *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. ▌

This axiom is actually necessary for the existence of a V_{GR} representation and hence necessary for the special case of a V_{US} representation. To see this, suppose $\succ$ has a V_{GR} representation. For any menu x and any i, let α_i denote a maximizer of $u(\beta) + \sum_{j=1}^{J} \gamma_{ij} v_j(\beta)$ over $\beta \in x$. Then

$$\begin{aligned} V_{GR}(x) &= \textstyle\sum_{i=1}^{I} q_i \left[[u(\alpha_i) + \sum_{j=1}^{J} \gamma_{ij} v_j(\alpha_i)] - \sum_{j=1}^{J} \gamma_{ij} \max_{\beta \in x} v_j(\beta) \right] \\ &\le \textstyle\sum_{i=1}^{I} q_i \left[[u(\alpha_i) + \sum_{j=1}^{J} \gamma_{ij} v_j(\alpha_i)] - \sum_{j=1}^{J} \gamma_{ij} v_j(\alpha_i) \right] \\ &= \textstyle\sum_{i=1}^{I} q_i u(\alpha_i) \\ &\le \max_{\beta \in x} u(\beta). \end{aligned}$$

Hence DFC must hold.

We have

Theorem 6 *Suppose $\succ$ has an additive EU representation with finitely many states. Then it has a V_{US} representation if and only if it satisfies DFC and negative set betweenness.*

The proof is below to facilitate comparison to the very similar proof of Theorem 1.

Conjecture 1 *The q_i's are unique.*

We can almost show existence of the V_{GR} representation for one somewhat special case. To explain the condition involved, recall that B is a finite set. Let n denote its cardinality. Then we can treat an expected utility preference as an $1 \times n$ vector w where we write lotteries as $n \times 1$ vectors β so expected utility is $w \cdot \beta$. Let $\mathbf{1}$ denote the $1 \times n$ vector of 1's. We list this as a "near–theorem" as there is one leap of faith involved, as noted in the proof.

Near–Theorem 1 *Suppose $\succ$ has a finite state additive EU representation such that $u, v_1, \ldots, v_J$ and $\mathbf{1}$ are linearly independent. Then it has a V_{GR} representation if and only if it satisfies DFC.*

The proofs of this result and of Theorem 6 both use a result originally due to Harsanyi [1955].[6] We include a simple proof here for completeness. The result is that an expected utility preference w is a positive linear combination of other expected utility preferences $w_1, \ldots, w_M$ plus a constant if and only if w agrees with the Pareto ranking defined by $w_1, \ldots, w_M$.

Lemma 2 *Suppose w and $w_1, \ldots, w_M$ are expected utility preferences. Then there exist scalars $c_1, \ldots, c_M$ with $c_i \geq 0$ for $i = 1, \ldots, M$ and a scalar d such that*

$$w = \sum_{i=1}^{M} c_i w_i + d\mathbf{1}$$

if and only if

$$\forall \alpha, \beta \in \Delta : \quad w_i \cdot \alpha \geq w_i \cdot \beta, \ \forall i \quad \Longrightarrow \quad w \cdot \alpha \geq w \cdot \beta. \tag{1}$$

Proof. The "only if" part is obvious and so omitted. For the "if," suppose w satisfies equation (1) but cannot be written as $\sum_i c_i w_i + d\mathbf{1}$ for some $c_i \geq 0$ and some d. Let

$$W = \{w' \mid w' = \sum_i c_i w_i + d\mathbf{1}, \text{ for some } c_i \geq 0, \text{ some } d\}.$$

[6]For more on Harsanyi's theorem, including further references to the related literature, see Hammond [1992].

Obviously, W is closed and convex. Since $w \notin W$ by hypothesis, there is a separating hyperplane. So there exists a vector $p \neq 0$, $n \times 1$, such that $w \cdot p < w' \cdot p$ for all $w' \in W$. That is,

$$w \cdot p < \sum_i c_i w_i \cdot p + d\mathbf{1} \cdot p$$

for all $c_i \geq 0$ and all d. Since the sign of d is arbitrary, this implies that $\sum_k p_k = 0$. Otherwise, we can take $d \to -\infty$ or $d \to \infty$ to make $d\mathbf{1} \cdot p$ arbitrarily negative and force a contradiction. Similarly, $w_i \cdot p \geq 0$ for all i. To see this, suppose $w_i \cdot p < 0$ for some i. Then we can make c_i arbitrarily large to generate a contradiction. Finally, we must have $w \cdot p < 0$. Otherwise, $c_i = d = 0$ for all i yields a contradiction.

Hence there exists p, $n \times 1$ such that $\sum_k p_k = 0$, $w_i \cdot p \geq 0$ for all i, and $w \cdot p < 0$. I now show that equation (1) implies a contradiction. To see this, first consider any p such that $p = \alpha - \beta$ for some $\alpha, \beta \in \Delta$. (Obviously, such a p will satisfy $\mathbf{1} \cdot p = 0$.) By (1), for any such p, $w_i \cdot p \geq 0$ for all i implies $w \cdot p \geq 0$, a contradiction. So p cannot be written as $\alpha - \beta$ for $\alpha, \beta \in \Delta$.

Clearly, p does not equal $\alpha - \beta$ for any $\alpha, \beta \in \Delta$ iff $p + \beta \notin \Delta$ for any $\beta \in \Delta$. However, $\mathbf{1} \cdot (p + \beta) = 1$. Hence this holds iff for all $\beta \in \Delta$, there exists k such that $p_k + \beta_k < 0$ where $p = (p_1, \ldots, p_n)$ and analogously for β. Let $g(p) = \max_{k=1,\ldots,n} |p_k|$. We know that $p \neq 0$ so $g(p) > 0$. Let $h(p) = [1/ng(p)]p$. Clearly, $\mathbf{1} \cdot h(p) = (\mathbf{1} \cdot p)/ng(p) = 0$. Also, every component of $h(p)$ is less than or equal to $1/n$ in absolute value. Letting β^* denote the lottery with probability $1/n$ on every outcome, then, we see that $h(p) + \beta^* \in \Delta$. Hence $h(p)$ can be written as $\alpha - \beta$ for some $\alpha, \beta \in \Delta$.

Note that

$$w_i \cdot p \geq 0 \iff \frac{w_i \cdot p}{g(p)n} \geq 0 \iff w_i \cdot h(p) \geq 0.$$

Hence by equation (1), $w \cdot h(p) \geq 0$, implying $w \cdot p \geq 0$, a contradiction. ▮

Given a finite state additive EU representation, number the states as $s_1, \ldots, s_I, s_{I+1}, \ldots, s_{I+J}$ where the first I states are positive and the remaining J are negative. We can then define $u_i(\beta) = U(\beta, s_i)\mu(s_i)$ for $i = 1, \ldots, I$ and $v_j(\beta) = U(\beta, s_{I+j}\mu(s_j)$ for $j = 1, \ldots, J$. This rewrites the representation in the more useful form

$$V(x) = \sum_{i=1}^{I} \max_{\beta \in x} u_i(\beta) - \sum_{j=1}^{J} \max_{\beta \in x} v_j(\beta). \tag{2}$$

Lemma 3 *Suppose $\succ$ has a finite state additive EU representation of the form*

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

If $\succ$ *satisfies DFC, then there exist constants* a_i*,* b_{ij}*, and* c_i*,* $i = 1, \dots, I$ *and* $j = 1, \dots, J$ *such that* $a_i \geq 0$*,* $b_{ij} \geq 0$*, and*

$$u_i(\beta) = a_i u(\beta) + b_{ij} v_j(\beta) + c_i$$

for all β *and all* i *where* u *is defined by*

$$u(\beta) = V(\{\beta\}) = \sum_i u_i(\beta) - \sum_j v_j(\beta).$$

Proof. The claim follows from Lemma 2 if we can show that u_i agrees with the Pareto ranking of u and $v_1, \dots, v_J$. In other words, we need to show that if $\succ$ satisfies DFC, then for all α and β,

$$u(\alpha) \geq u(\beta), \quad v_j(\alpha) \geq v_j(\beta), \ \forall j$$

implies

$$u_i(\alpha) \geq u_i(\beta), \quad \forall i.$$

To see this, suppose not. Then we have some α and β such that

$$u(\alpha) \geq u(\beta), \quad v_j(\alpha) \geq v_j(\beta), \ \forall j$$

but $u_i(\beta) > u_i(\alpha)$ for some i. Hence

$$\begin{aligned} V(\{\alpha, \beta\}) &= \textstyle\sum_i \max\{u_i(\alpha), u_i(\beta)\} - \sum_j v_j(\alpha) \\ &> \textstyle\sum_i u_i(\alpha) - \sum_j v_j(\alpha) \\ &= u(\alpha) \geq u(\beta). \end{aligned}$$

Since $u(\beta) = V(\{\beta\})$ for all β, this implies $\{\alpha, \beta\} \succ \{\alpha\} \succeq \{\beta\}$, contradicting DFC. ▌

Proof. Necessity of the two axioms is obvious. For sufficiency, first observe that negative set betweenness implies one negative state. So we have an additive EU of the form

$$V(x) = \sum_{i=1}^{I} \max_{\beta \in x} u_i(\beta) - \max_{\beta \in x} v(\beta).$$

By Lemma 3, DFC implies that there are nonnegative coefficients a_i and b_i and coefficients c_i such that $u_i(\beta) = a_i u(\beta) + b_i v(\beta) + c_i$ for all β and all i.

First, suppose u is a constant function. In this case, $V(x)$ must be constant in x. To see this, note that if u is constant, we can write

$$V(x) = \left[\sum_i b_i - 1\right] \max_{\beta \in x} v(\beta) + K$$

where K is a constant. If $\sum_i b_i - 1 = 0$, then $V(x)$ is a constant as asserted. If $\sum_i b_i > 1$, then V ranks singletons the same way v does. By definition, u is the way V ranks singletons and u is constant. Hence v must be constant, so V is constant as asserted. Finally, if $\sum_i b_i < 1$, V ranks constants the same way $-v$ does. Hence u and $-v$ must represent the same preference. Hence since u is constant, v must be, again implying V is constant. When V is constant, we can define any constant u and v and trivially write the representation in the form of V_{US}. (Alternatively, we can add a nontriviality axiom and eliminate this case.) So we can assume without loss of generality that u is not a constant function.

Next, suppose v is a constant function. In this case, we can write

$$V(x) = \left[\sum_i a_i\right] \max_{\beta \in x} u(\beta) + K$$

where K is a constant. By definition, $V(\{\beta\}) = u(\beta)$, so either $u(\beta) = K/[1 - \sum_i a_i]$ for all β and hence is constant (ruled out above) or $\sum_i a_i = 1$ and $K = 0$. Hence we have $V(x) = \max_{\beta \in x} u(\beta)$. This is a V_{US} representation where $I = 1$ and $\gamma_1 = 0$. So we can assume without loss of generality that v is not a constant function.

Note that $u_i(\beta) = a_i u(\beta) + b_i v(\beta) + c_i$ implies

$$\sum_{i=1}^{I} u_i(\beta) = u(\beta) \sum_i a_i + v(\beta) \sum_i b_i + \sum_i c_i. \tag{3}$$

But by definition, $u(\beta) = \sum_i u_i(\beta) - v(\beta)$, so this implies

$$u(\beta)\left[1 - \sum_i a_i\right] = v(\beta)\left[\sum_i b_i - 1\right] + \sum_i c_i$$

or, more simply,

$$u(\beta)[1 - A] = v(\beta)[B - 1] + C. \tag{4}$$

From the above, neither u nor v is a constant function. Hence $A = 1$ if and only if $B = 1$.

First, suppose $B \neq 1$ and $k_0 = (1 - A)/(B - 1) > 0$. Hence we can rewrite equation (4) as $v(\beta) = k_0 u(\beta) + k_1$ for all β where $k_1 = -C/(B - 1)$. Hence $u_i(\beta) = (a_i + b_i k_0) u(\beta) + c_i + b_i k_1$ where $a_i + b_i k_0 \geq 0$ for all i. So

$$V(x) = K_0 \max_{\beta \in x} u(x) + K_1$$

for some constants K_0 and K_1. But $V(\{\beta\}) = u(\beta)$ for all β. Since u is not constant, we must have $K_0 = 1$ and $K_1 = 0$. Thus we have a V_{US} representation where $I = 1$ and $\gamma_1 = 0$.

Next, suppose $B \neq 1$ and $k_0 = (1 - A)/(B - 1) < 0$. We can rewrite equation (4) as $u(\beta) = (1/k_0)v(\beta) - (k_1/k_0)$, so

$$u_i(\beta) = \left[\frac{a_i}{k_0} + b_i\right] v(\beta) + c_i - a_i \frac{k_1}{k_0} = (a_i + b_i k_0) u(\beta) + c_i + b_i k_1.$$

Note that $(a_i/k_0) + b_i > 0$ iff $a_i + b_i k_0 < 0$. In other words, for any i, we can either write u_i as a positive affine transformation of u or as a positive affine transformation of v. Hence

$$V(x) = \left[\sum_{i|a_i+b_ik_0\geq 0}(a_i + b_ik_0)\right] \max_{\beta\in x} u(\beta) - \left[1 - \sum_{i|a_i+b_ik_0<0}\left(\frac{a_i}{k_0} + b_i\right)\right] \max_{\beta\in x} v(\beta) + K$$

for some constant K.

Letting K_u denote the term multiplying $\max u$ and K_v the term multiplying $\max v$, we have

$$V(x) = K_u \max_{\beta\in x} u(\beta) - K_v \max_{\beta\in x} v(\beta) + K.$$

By definition, $V(\{\beta\}) = u(\beta)$, so $(1-K_u)u(\beta) = -K_v v(\beta)+K = -K_v k_0 u(\beta)-K_v k_1+K$. Since u is not constant, we have $1 - K_u = -K_v k_0$ and $-K_v k_1 + K = 0$.

Case 1. $K_u = 1$, so $K_v = 0$. In this case, we have $V(x) = \max_{\beta\in x} u(\beta) + K$. Since $V(\{\beta\}) = u(\beta)$, we must have $K = 0$. This corresponds to V_{US} with $I = 1$ and $\gamma_1 = 0$.

Case 2. $K_u < 1$, so $-K_v k_0 > 0$, implying $K_v > 0$. Let $\hat{v} = K_v v$. Then we have a V_{US} representation with $\hat{v}$ as the negative state utility, $I = 1$, and $\gamma_1 = 1$. To see this, simply note that

$$u(\beta)+\hat{v}(\beta) = u(\beta)+K_v v(\beta) = u(\beta)+K_v k_0 u(\beta)+K_v k_1 = (1+K_v k_0)u(\beta)+K = K_u u(\beta)+K.$$

Hence

$$\max_{\beta\in x}[u(\beta) + \hat{v}(\beta)] - \max_{\beta\in x} \hat{v}(\beta) = K_u \max_{\beta\in x} u(\beta) - K_v \max_{\beta\in x} v(\beta) + K = V(x).$$

Case 3. $K_u > 1$, so $-K_v k_0 < 0$ implying $K_v < 0$. Fix any α and β such that $u(\alpha) > u(\beta)$. (Since u is not constant, this must be possible.) Then

$$\begin{aligned} V(\{\alpha, \beta\}) &= K_u u(\alpha) - K_v v(\beta) + K \\ &= K_u u(\alpha) - K_v k_0 u(\beta) - K_v k_1 + K \\ &= K_u u(\alpha) + (1 - K_u)u(\beta). \end{aligned}$$

Hence $V(\{\alpha, \beta\}) > V(\{\alpha\}) = u(\alpha)$ if $(K_u - 1)u(\alpha) > (K_u - 1)u(\beta)$. Since $K_u > 1$ and $u(\alpha) > u(\beta)$, this holds. But then we have $\{\alpha, \beta\} \succ \{\alpha\} \succ \{\beta\}$, contradicting DFC.

The only possibility left for equation (4), then, is that $A = B = 1$ and $C = 0$. Without loss of generality, we can assume that $a_i > 0$ for all i. To see that this is without loss of generality, suppose $a_i = 0$ for some i, say i^*. Then $u_{i^*}(\beta) = b_{i^*} v(\beta) + c_{i^*}$. In this case, we have

$$V(x) = \sum_{i\neq i^*} \max_{\beta\in x} u_i(\beta) - (1 - b_{i^*}) \max_{\beta\in x} v(\beta) + c_{i^*}.$$

Because $\sum_i b_i = 1$, we know that $b_{i^*} \leq 1$. If it equals 1, then $b_i = 0$ for all $i \neq i^*$. In this case, we have

$$V(x) = \left[\sum_{i \neq i^*} a_i\right] \max_{\beta \in x} u(\beta) + C = \max_{\beta \in x} u(\beta)$$

where the second equality follows from $\sum_i a_i = 1$, $a_{i^*} = 0$, and $C = 0$. But then we have a V_{US} representation with $I = 1$ and $\gamma_1 = 0$. So suppose $b_{i^*} < 1$. In this case, we can define $\hat{v} = (1 - b_{i^*})v - c_{i^*}$, write

$$V(x) = \sum_{i \neq i^*} \max_{\beta \in x} u_i(\beta) - \max_{\beta \in x} \hat{v}(\beta)$$

and repeat the analysis with the smaller set of positive states and redefined utility function for the negative state.

So assume that $a_i > 0$ for all i. Let $q_i = a_i$ and let $\gamma_i = b_i/a_i$. Since $a_i > 0$ and $b_i \geq 0$, this ensures that $q_i > 0$ and $\gamma_i \geq 0$ as required. Also, $\sum_i a_i = A = 1$, so $\sum_i q_i = 1$ as required.

Note that

$$\begin{aligned} V(x) &= \textstyle\sum_i \max_{\beta \in x} u_i(\beta) - \max_{\beta \in x} v(\beta) \\ &= \textstyle\sum_i \max_{\beta \in x} [a_i u(\beta) + b_i v(\beta) + c_i] - \max_{\beta \in x} v(\beta) \\ &= \textstyle\sum_i \max_{\beta \in x} q_i [u(\beta) + \gamma_i v(\beta)] - \max_{\beta \in x} v(\beta) + C \\ &= \textstyle\sum_i q_i \left\{\max_{\beta \in x} [u(\beta) + \gamma_i v(\beta)] - \gamma_i \max_{\beta \in x} v(\beta)\right\} \end{aligned}$$

where the last equality follows from $\sum_i q_i \gamma_i = \sum_i b_i = B = 1$ and $C = 0$. Hence $\succ$ has a V_{US} representation. ▌

Near–Proof of Near–Theorem 1. This is very similar to the previous proof. Again, the necessity of DFC is obvious. Analogously to the previous proof, Lemma 3 implies that there are nonnegative numbers a_i and b_{ij} and numbers c_i such that $u_i = a_i u + \sum_i b_{ij} v_j + c_i \mathbf{1}$. Now we add

Leap of Faith: $a_i > 0$ for all i.

Again, we get

$$\sum_i u_i = uA + \sum_j B_j v_j + C\mathbf{1}.$$

Again, $\sum_i u_i = u + \sum_j v_j$, so this implies

$$u\,[1 - A] = \sum_j [B - 1]\, v_j + C\mathbf{1}.$$

By hypothesis, the vectors $u, v_1, \ldots, v_J$ and $\mathbf{1}$ are linearly independent. Hence this implies $A = 1$, $B_j = 1$ for all j, and $C = 0$. So let $q_i = a_i$ and $\gamma_{ij} = b_{ij}/q_i$. This specification ensures that $q_i > 0$ for all i, $\sum_i q_i = 1$, and $\gamma_{ij} \geq 0$ for all i and j as required.

Note that

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

where the last equality uses $\sum_i c_i = 0$ and $\sum_i q_i \gamma_{ij} = \sum_i d_{ij} = 1$ for all j. Hence we have a V_{GR} representation. ▌

5 Speculations

There are several interesting issues left to explore. An obvious open question is the characterization of the set of DLR preferences with a V_{GR} representation. At one point, we had conjectured that a necessary and sufficient condition for a preference with a finite state additive EU representation to have a V_{GR} representation was DFC. It is easy to see that this is necessary and Lemma 3 suggests that it is almost sufficient. We now have examples which show that DFC is not, however, sufficient in general.

Another direction of interest is the extent to which such a representation is identified. It is not hard to see that the γ_{ij}'s cannot be identified in general. To see why, suppose we have a representation with negative state utilities $v_1, \ldots, v_J$ and coefficients q_i, $i = 1, \ldots, I$, and γ_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$. Then we also have a representation where we replace v_j by, say, $2v_j$ and replace γ_{ij} by $(1/2)\gamma_{ij}$ for each i. In other words, we can rescale the v_j's and correspondingly rescale the γ_{ij}'s in a way which leaves the representation essentially unchanged.

On the other hand, this is not true of the q_i's in general. This leaves open the possibility that these coefficients are identified. We conjecture that they are identified when the vectors u and $v_1, \ldots, v_J$ are linearly independent. One can show that they typically are not identified otherwise.

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 2001a, 1403–1435.

[3] Gul, F., and W. Pesendorfer, "A Theory of Addiction," Princeton University working paper, 2001b.

[4] Gul, F., and W. Pesendorfer, "Self–Control and the Theory of Consumption," Princeton University working paper, 2002.

[5] Hammond, P., "Harsanyi's Utilitarian Theorem: A Simpler Proof and Some Ethical Connotations," in *Rational Interaction: Essays in Honor of John Harsanyi*, R. Selten, ed., Berlin: Springer-Verlag, 1992, 305–319.

[6] Harsanyi, J., "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, **63**, 1955, 309–321.

[7] Kreps, D., "A Representation Theorem for 'Preference for Flexibility'," *Econometrica,* **47**, May 1979, 565–576.

[8] 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.

[9] Ziegler, G., *Lectures on Polytopes*, New York: Springer–Verlag, 1995.