Costly Self-Control and Limited Willpower

Simon Grant
Research School of Economics,
Australian National University
and
School of Economics,
University of Queensland.

Sung-Lin Hsieh
Department of Economics
National Taiwan University

Meng-Yu Liang
Institute of Economics,
Academia Sinica.

September 26, 2015

Abstract

We construct a representation theorem for individual choice among sets of lotteries, from which the individual will later choose a single lottery. In particular, our axioms building on those in Gul and Pesendorfer (2001) (GP01), allow for a preference for commitment and self-control subject to sufficient willpower. Four of the five axioms of our characterization are as in GP01 (Theorem 3) except that the independence axiom is restricted to singleton menus and those two-element menus in which any failure of self-control in the second period arises from the individual being *unwilling* to incur the cost of exercising such self-control and not from being *unable* to exert self-control because of limited willpower. We add one new axiom to regulate willpower as a limited cognitive resource in which the available 'stock' does not vary across menus. In our characterization, agents with insufficient willpower to resist temptations are bound to choose an option with lower 'compromise utility' while the behaviors of agents who resist temptations remain unchanged.

Address for Correspondence Meng-Yu Liang, Institute of Economics, Academia Sinica 128 Sec. 2, Academia Rd. Taipei, Taiwan 115.

e-mail myliang@econ.sinica.edu.tw, tel. +888-2-2782-2791#506 .

Keywords dynamic choice, self-control, will-power.

JEL Classification: D81

1 Introduction

Consider an individual who is contemplating what activity she plans to do this evening after returning home from work and before going out to dinner.[1] She may choose to work out in her local gym, to read a (literary prize-winning) novel, or to watch (trashy) TV. In the morning when she is full of vigor and good intentions, she prefers the work-out in the gym to reading the novel which in turn she prefers to watching TV. But in the evening after having spent a hard day at the office, slumping in front of the TV is more tempting than reading the novel which in turn is more tempting than the work-out in the gym. To avoid such temptations, she may try to restrict the options available to her in the evening. For example, she might solicit a ride from a colleague at work who is also going to the gym, or she might tell her roommate to hide the TV remote.[2] In lieu of taking any preventive measures such as these, she may exercise, at some cost, self-control in the manner described by **?** (hereafter, GP01) and resist the tempting option.

Unlike GP01's model, however, suppose that even if our individual is willing (and able) to exert self control to resist selecting the more tempting option in a choice between working out in the gym and reading the novel and in a choice between reading the novel and watching TV, she finds herself *unable* to resist temptation when choosing between working out in the gym and watching TV. That is, when the temptation "distance" between a pair of alternatives is too great, suppose she lacks the willpower to resist the tempting option. Notice that, as our example illustrates, this possibility results in second-period choices that violate the weak axiom of revealed preferences. In this regard, our approach is similar to the "revealed willpower" model of **?** (hereafter, MNO14) which views willpower as a cognitive resource that can explain the behavior of individuals who have imperfect control over their immediate "urges". In contrast to MNO14, however, in our model as in GP01, the exercise of self-control in resisting temptation is always costly.

To illustrate these ideas more formally, suppose there are two periods: morning and evening. The activity is undertaken in the evening when our individual chooses it from a set of available alternatives. In the morning, she chooses from among sets of available alternatives or "menus". Let $\succsim$ denote her preference relation over menus (with $\succ$ denoting strict preference and $\sim$ denoting indifference). Singleton sets describe situations in

[1] This example is based on one from Masatlioglu et al (2014, p9).

[2] Of course in the case of the latter, she still faces the temptation of reading the novel unless she asks her roommate to hide it as well, and even the ride to the gym may still leave her with the option of reading the novel if she is in the habit of carrying it around with her in her pocketbook.

which the individual has commited herself in the morning to a particular activity in the evening. Corresponding to the story above, we have $\{Gym\} \succ \{Novel\} \succ \{TV\}$. A situation where she chooses in the evening between a pair of alternatives corresponds to one of the three two-element menus, $\{Gym, Novel\}$, $\{Novel, TV\}$ and $\{Gym, TV\}$. The situation where she has not restricted her options in the evening corresponds to the menu $\{Gym, Novel, TV\}$.

As is the case in GP01, the fact that activities $Novel$ and TV are tempting for Gym and the activity TV is tempting for $Novel$ is reflected by the strict preferences $\{Gym\} \succ \{Gym, Novel\}$, $\{Gym\} \succ \{Gym, TV\}$ and $\{Novel\} \succ \{Novel, TV\}$ which means that the availability of the tempting alternative makes the individual worse off than she would be if she were able to commit to the less tempting alternative.[3] Self-control (respectively, succumbing to temptation) is captured by a strict preference (respectively, indifference) between a menu comprising that pair of alternatives and the singleton menu comprising the tempting alternative. Corresponding to the story above, we have $\{Gym, Novel\} \succ \{Novel\}$, $\{Gym, TV\} \sim \{TV\}$. and $\{Novel, TV\} \succ \{TV\}$. Finally, it will turn out that any menu is indifferent to a two-element subset that comprises the most tempting alternative and the best compromise alternative for which the individual has sufficient willpower to choose in the second period. For the story above we have $\{Gym, Novel, TV\} \sim \{Novel, TV\}$, since the individual has insufficient willpower to choose Gym in presence of the (very) tempting outcome TV.

The decision-maker is assumed to have preferences over sets (or "menus") of lotteries. Our axioms, building on those in GP01, allow for a preference for commitment and self-control *subject to sufficient willpower.* We refer to the resulting preferences as *temptation preferences with costly self-control and limited willpower.* In our main representation result (Theorem 3) we show they admit a representation of the form:

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

$$\text{s.t.} \max_{y \in A} v(y) - v(x) \leq w.$$

This is the representation obtained by GP01 *with the addition of a willpower constraint.* We interpret it as saying that the individual anticipates that in period 2 she will choose from the menu the alternative that maximizes the 'compromise utility' (the sum of the commitment utility u and the temptation urge v), subject to the difference between the most tempting urge and that of the selected alternative being no more than her

[3] In MNO14's model such strict preferences would only hold if the individual was going to succumb to temptation in the second period. For example, $\{Gym\} \succ \{Gym, Novel\}$ would entail $\{Gym, Novel\} \sim \{Novel\}$ in their model.

willpower w. Denoting this element of A by x^*, the 'utility' of the menu which guides her period 1 choice over menus is then given by the commitment utility $u\left(x^*\right)$ less the amount $\max_{y\in A} v\left(y\right) - v\left(x^*\right)$, which following GP01, we interpret as the (utility) cost of self-control. Thus the willpower w represents the *upper bound* on the self-control cost the individual is able to bear.

Notice that from the two strict preferences $\{Gym, Novel\} \succ \{Novel\}$ and $\{Novel, TV\} \succ \{TV\}$ we can infer from the representation that

$$u\left(Gym\right) + v\left(Gym\right) > u\left(Novel\right) + v\left(Novel\right) > u\left(TV\right) + v\left(TV\right)$$
$$\Rightarrow u\left(Gym\right) + v\left(Gym\right) - v\left(TV\right) > u\left(TV\right).$$

So in the absence of any constraint on the individual's willpower it would necessarily follow that $\{Gym, TV\} \succ \{TV\}$, as is indeed would be the case in GP01's model. The fact that we actually have indifference, "reveals" that $v\left(TV\right) - v\left(Gym\right) > w$, that is, she has insufficient willpower to resist the tempting option.

Four of the five axioms of Theorem 3 are as in GP01 (Theorem 3) except that the independence axiom is restricted to singleton menus and those two-element menus in which any failure of self-control in the second period arises from the individual being *unwilling* to incur the cost of exercising such self-control and not from being *unable* to exert self-control because of limited willpower. The reason is that independence may fail when 'mixing' with menus for which a failure to exercise self-control arises from insufficient willpower. To see why, recall we have noted above that the preference pattern

$$\{Gym\} \succ \{Gym, Novel\} \succ \{Novel\} \succ \{Novel, TV\} \succ \{Gym, TV\} \sim \{TV\}$$

may be interpreted as revealing that although Gym is better than $Novel$ both in terms of commitment utility and 'compromise utility' the individual's willpower is insufficient to resist choosing from the menu $\{Gym, TV\}$ the more tempting option TV. Now, if we let $[z]$ denote the degenerate lottery that yields the consequence z with probability 1, for a weight $\alpha \in (0, 1)$ let us consider the menus,

$$\alpha\{[Gym], [TV]\} + (1-\alpha)\{[TV]\} = \{\alpha[Gym] + (1-\alpha)[TV], [TV]\}$$
$$\text{and } \alpha\{[Novel], [TV]\} + (1-\alpha)\{[TV]\} = \{\alpha[Novel] + (1-\alpha)[TV], [TV]\}$$

Intuitively, if in both menus the cost of exercising self-control to resist the tempting option TV is decreasing in α and, moreover, tends to zero as α tends to zero, then for a sufficiently small α, the cost of self-control will not exhaust the individual's willpower in either menu. Thus, for a sufficiently small α, we should expect the individual to express the strict preference $\{\alpha[Gym] + (1-\alpha)[TV], [TV]\} \succ \{\alpha[Novel] + (1-\alpha)[TV], [TV]\}$. But this constitutes a violation of independence and hence motivates our restriction of its domain.

We first show (Theorem 2) that for a preference relation in which no failure of self-control may be attributed to a lack of willpower (the case studied by GP01) we obtain GP01's Theorem 3 even though we have restricted the domain in which independence applies. That is, such a family of preferences either admits GP01's costly self-control representation or their overwhelming temptation representation where the representation takes the form

$$U(A) = \max_{x \in A} u(x), \text{ s.t. } v(y) \leq v(x) \text{ for all } y \in A.$$

Preferences admitting the overwhelming temptation representation may be viewed as a special case of MNO14 with a zero willpower constraint.

Our main contribution, however, is to allow for preferences in which at least some failures of self-contriol may be attributed to a lack of willpower. To characterize these preferences, we need to be able to ascertain when one can infer that the exercise of self-control exhausts the DM's ("stock" of) willpower. To see how such an inference might be drawn, suppose, in the context of our running example above, that for some $\bar{\alpha} \in (0,1)$, it is the case that for any $\alpha \leq \bar{\alpha}$, our DM expresses the strict preference

$$\{[Gym], (1-\alpha)[Gym] + \alpha[TV]\} \succ \{(1-\alpha)[Gym] + \alpha[TV]\},$$

but for any $\alpha > \bar{\alpha}$, she expresses the *indifference*

$$\{[Gym], (1-\alpha)[Gym] + \alpha[TV]\} \sim \{(1-\alpha)[Gym] + \alpha[TV]\}.$$

If, as seems natural, we think that the cost of exerting self-control by choosing *Gym* in the presence a binary gamble with support $\{Gym, TV\}$ is (weakly) increasing in the weight the binary gamble assigns to the tempting alternative TV, then the fact that $(1-\bar{\alpha})[Gym] + \bar{\alpha}[TV]$ is the *most* tempting such binary gamble she can resist suggests that the cost of exerting such self-control does indeed exhaust her stock of willpower.

As we are viewing willpower as a limited cognitive resource that enables the DM to resist temptation, we contend that such a limit should not depend on the menu in question. Thus we should be able to conclude, from the fact that the cost of exerting self-control in a choice between $[Gym]$ and $(1-\bar{\alpha})[Gym] + \bar{\alpha}[TV]$ reaches the upper-bound on how much self-control cost the DM can bear, that this cost must be at least as great as the cost of exerting self-control in any other situation. For example, it should not exceed the cost of exerting self-control in a choice between $[Novel]$ and $[TV]$.

Now consider the pair of menus,

$$\left\{\frac{1}{2}[Gym]+\frac{1}{2}\left[Novel\right],\frac{1}{2}[Gym]+\frac{1}{2}\left[TV\right]\right\}$$
$$\text{and }\left\{\frac{1}{2}[Novel]+\frac{1}{2}\left[Gym\right],\frac{1}{2}[Novel]+\frac{1}{2}\left(1-\bar{\alpha}\right)\left[Gym\right]+\frac{1}{2}\bar{\alpha}[TV]\right\}.$$

Notice that the former corresponds to a half-half set mixture of $\{[Gym]\}$ and $\{[Novel],[TV]\}$ while the latter corresponds to a half-half set mixture of $\{[Novel]\}$ and $\{[Gym],(1-\bar{\alpha})\,[Gym]+\bar{\alpha}[TV]\}$. Independence type reasoning[4] suggests that the lottery $\frac{1}{2}[Gym]+\frac{1}{2}\,[Novel]$ will be chosen from each of these menus. Independence type reasoning also suggests the cost of self-control required to select $\frac{1}{2}[Gym]+\frac{1}{2}\,[Novel]$ from the menu $\left\{\frac{1}{2}[Gym]+\frac{1}{2}\,[Novel]\,,\frac{1}{2}[Gym]+\frac{1}{2}\,[TV]\right\}$ is half of the self-control cost required to select $[Novel]$ from the menu $\{[Novel],[TV]\}$. Similarly, the cost of self-control required to select $\frac{1}{2}[Gym]+\frac{1}{2}\,[Novel]$ from the menu $\left\{\frac{1}{2}[Novel]+\frac{1}{2}\,[Gym]\,,\frac{1}{2}[Novel]+\frac{1}{2}\,(1-\bar{\alpha})\,[Gym]+\frac{1}{2}\bar{\alpha}[TV]\right\}$ is half of the self-control cost required to select $[Gym]$ from the menu $\{[Gym],(1-\bar{\alpha})\,[Gym]+\bar{\alpha}[TV]\}$.

But for the willpower limit to exercising self-control *not* to be menu-dependent then requires that

$$\left\{\frac{1}{2}[Gym]+\frac{1}{2}\left[Novel\right],\frac{1}{2}[Gym]+\frac{1}{2}\left[TV\right]\right\}\succsim\left\{\frac{1}{2}[Novel]+\frac{1}{2}\left[Gym\right],\frac{1}{2}[Novel]+\frac{1}{2}\left(1-\bar{\alpha}\right)\left[Gym\right]+\frac{1}{2}\bar{\alpha}[TV]\right\}.$$

And this is precisely what our additional axiom (Axiom 5) ensures.

Moreover, we show (Theorem 3) our new axiom in conjunction with the other four axioms provides a characterization of temptation preferences with costly self-control and limited (but strictly positive) willpower.

2 Framework and Definitions

We consider a two-period decision problem similar to the setting in GP01. There is a finite set Z of (final) prizes (or consequences), with generic element z. Let $\Delta\left(Z\right)$ denote the set of lotteries defined on Z, with generic elements x, y, a, b, et cetera. That is, $\Delta\left(Z\right)$ may be taken to be the set of functions $x:Z\rightarrow\mathbb{R}_{+}$, for which $\sum_{z\in Z}x\left(z\right)=1$. We endow it with the topology generated by the uniform metric $d\left(x,y\right)=\max_{z\in Z}\left|x\left(z\right)-y\left(z\right)\right|$. As is standard, for any pair of lotteries x, y in $\Delta\left(Z\right)$ and any α in $[0,1]$, let $\alpha x+\left(1-\alpha\right)y$

[4] Such independence reasoning is valid here as it will turn out that these menus all lie in the domain for which we assume our restricted independence axiom applies.

denote the lottery that assigns to each prize $z \in Z$ the probability $\alpha x(z) + (1-\alpha) y(z)$. In addition, for any $z \in Z$, let $[z]$ denote the degenerate lottery which assigns probability 1 to z, that is, $[z](z) = 1$ and thus, $[z](z') = 0$ for all $z' \neq z$.

Let $\mathscr{A}$ denote the set of menus which we take to be the set of all compact subsets of $\Delta(Z)$ with generic elements A, B. We endow $\mathscr{A}$ with the (Hausdorff) topology generated by the metric

$$d_h(A,B) = \max\left\{\max_{x\in A}\min_{y\in B} d(x,y), \max_{x\in B}\min_{y\in A} d(x,y)\right\}.$$

For any pair of menus A, B in $\mathscr{A}$ and any α in $[0,1]$, let $\alpha A + (1-\alpha) B$ denote the menu in $\mathscr{A}$ given by $\{\alpha x + (1-\alpha) y : x \in A,\, y \in B\}$.

The (first-period) preferences $\succsim$ of the DM are defined on $\mathscr{A}$. As is standard, $\succ$ (respectively, $\sim$) denotes the asymmetric (respectively, symmetric) parts of $\succsim$. We consider the restriction of $\succsim$ to singleton lotteries as the DM's *commitment preferences* defined over the set of lotteries $\Delta(Z)$. That is, the DM is deemed to weakly prefer lottery x to lottery y (in terms of her commitment preferences) if $\{x\} \succsim \{y\}$.

In order to formalize the main differences between our model and that of GP01 it is convenient to introduce the following concepts and attendant notation regarding lotteries for which the individual can and cannot exert self-control with respect to her commitment preferences.

For the lottery $x \in \Delta(Z)$, we take the set of tempting alternatives for which the DM fails to exert self-control as ones for which the DM is unable or unwilling to exert self-control in a menu just comprising x and that alternative. That is, an alternative y is deemed *a tempting alternative to x for which the DM fails to exert self-control*, if, despite strictly preferring, according to her commitment preferences, lottery x to lottery y (that is, $\{x\} \succ \{y\}$), she is indifferent between $\{y\}$ and the menu $\{x,y\}$. We interpret the latter indifference as reflecting her (rational) anticipation that if she faces a choice in the second period from the menu $\{x,y\}$ she will be unable or unwilling to exert (sufficient) self-control to choose x. Formally,

$$T(x) := \{y \in \Delta(Z) : \{x\} \succ \{x,y\} \sim \{y\}\}.$$

Correspondingly, we define the set of *tempting alternatives to x for which the DM can exert costly self-control* as being,

$$S(x) := \{y \in \Delta(Z) : \{x\} \succ \{x,y\} \succ \{y\}\}.$$

Now for any lottery y in $T(x)$, there are two reasons for why the DM might anticipate she would not choose

x from the menu $\{x, y\}$: (i) the alternative y is at least as good a "compromise" alternative as x, or (ii) despite x being the better compromise candidate, the individual has insufficient willpower to resist choosing y. But in the latter case, by considering another alternative formed by taking a convex combination of x and y that is sufficiently close to x, the DM will be able (and willing) to exert self-control whenever facing a choice in the second period between x and that convex combination of x and y. Hence we divide $T(x)$ into two sets, $L(x)$ and $T(x) \backslash L(x)$, where

$$L(x) := \{y \in T(x) : (1-\alpha)x + \alpha y \in S(x), \text{ for some } \alpha \in (0,1)\}.$$

Notice that if $L(x) = \emptyset$ for all $x \in \Delta(Z)$, then this is the same as the case of unlimited willpower considered by GP01. Consider now a situation in which for some pair of lotteries x and y in $\Delta(Z)$ we have $y \in L(x)$. To aid our intuition, referring to the example discussed in the introduction, suppose that x is the degenerate lottery in which the activity working out in the gym is undertaken for sure, and y is the (also) degenerate lottery in which the activity watching TV is undertaken for sure. We interpret $y \in L(x)$ as saying that although the DM cannot resist in the second period the temptation of watching TV when facing a choice between either of them for sure, she can resist the more tempting option when the choice is between working out in the gym for sure and a binary gamble in which the probability she works out in the gym is sufficiently high thus entailing only a small complementary probability assigned to her ending up watching TV. That is, the DM has sufficient willpower to resist binary gambles that involve only a small chance of the tempting prize obtaining.

We consider the collection of all singleton menus as well as those two-element menus in which either there is self-control or any anticipated failure of self-control does *not* arise as a result of lack of willpower. That is, any failure of self-control in the second period arises solely from costly self-control. Formally, set

$$\mathscr{L}(\succsim) := \{\{x\} : x \in \Delta(Z)\} \cup \{\{x, y\} : \{x\} \succ \{y\} \text{ and } y \notin L(x)\}.$$

In addition, it will be convenient in the sequel to consider the collection of all singleton menus as well as those two-element menus in which there is a tempting alternative for which the DM can exert costly self-control. Formally, set

$$\mathscr{M}(\succsim) := \{\{x\} : x \in \Delta(Z)\} \cup \{\{x, y\} : \text{with } y \in S(x)\}.$$

3 Toward a Representation.

We impose the following axioms as in GP01 (Theorem 3) except that, for reasons we shall explain below, the independence axiom is restricted to $\mathscr{L}(\succsim)$.

Axiom 1 (Ordering). $\succsim$ *is a complete and transitive binary relation.*

Axiom 2a (Upper Semi-Continuity). *The sets* $\{B \in \mathscr{A} : B \succsim A\}$ *are closed.*

Axiom 2b (Lower von Neumann-Morgenstern Continuity). $A \succ B \succ C$ *implies* $\alpha A + (1-\alpha) C \succ B$ *for some* $\alpha \in (0,1)$.

Axiom 3 (Restricted Independence). *For any* $A, B, C \in \mathscr{L}(\succsim)$, $A \succ B$ *and* $\alpha \in (0,1)$ *implies* $\alpha A + (1-\alpha) C \succ \alpha B + (1-\alpha) C$.

Axiom 4 (Set Betweennness). $A \succsim B$ *implies* $A \succsim A \cup B \succsim B$.

When a sequence of menus passes through the willpower constraint, we cannot ensure that there will not be a corresponding discontinuity in the preferences over menus. Hence, we adopt the relaxed pair of continuity axioms used by GP01 for their theorem 3 which allowed for individuals who would choose at time 2 solely according to their temptation preferences.[5] However, unlike GP01, we do not require independence to hold on the entire relation. Instead we restrict its application to $\mathscr{L}(\succsim)$. As we argued by means of the example discussed in the introduction, independence may fail when 'mixing' with menus for which any failure to exercise self-control arises from insufficient willpower. For example, if $A \succ B$ arises because all the better choices in B are not available owing to a lack of willpower, then conceivably there might exist a weight α and some other menu C such that for the set $\alpha B + (1-\alpha) C$ the mixtures it contains with those better choices in B are now available. This in turn might result in $\alpha B + (1-\alpha) C \succsim \alpha A + (1-\alpha) C$, that is, a violation of independence.[6]

We begin the derivation of the costly self-control with limited willpower representation by noting it follows from standard arguments, the above axioms ensure the preferences admit a representation.

Lemma 3.1. *If Axioms 1, 2a, 2b hold, then there exists a function* $U : \mathscr{A} \to \mathbb{R}$ *that represents* $\succsim$.

[5] We do not impose GP's third continuity assumption (Axiom 2c, GP01, p1412), since, as they note, it is implied by Axiom 2b if the outcome set is finite as is the case in our setting.

[6] To relate to the example from the introduction, take $A = \{[Novel], [TV]\}$, $B = \{[Gym], [TV]\}$ and $C = \{[TV]\}$. Although $A \succ B$, in the introduction we argued that for sufficiently small α, $\alpha B + (1-\alpha) C \succsim \alpha A + (1-\alpha) C$.

GP01 established that given the preference relation admits a functional representation, adding set betweenness implies that each finite menu can be shown to be indifferent to an appropriately selected two-element menu.

Lemma 3.2 (GP01, Lemma 2, p1422)**.** *Let U be a function that represents some $\succsim$ satisfying Axiom 4. If $A \in \mathscr{A}$ is a finite set, then*

$$U(A) = \max_{x\in A}\min_{y\in A} U(\{x,y\}) = \min_{y\in A}\max_{x\in A} U(\{x,y\}).$$

Moreover, there is an x^*, y^* such that (x^*, y^*) solves the maxmin problem and (y^*, x^*) solves the minmax problem.

We use Lemma 3.2 to prove a result analogous to Lemma 3 in GP01 (p1422). But unlike GP01 we establish this result without assuming that the function representing $\succsim$ is linear. Instead the proof invokes Axiom 3 (restricted independence) directly.

Lemma 3.3. *Let U be a function that represents some $\succsim$ satisfying Axioms 1, 2a, 2b, 3 and 4 and $A = \alpha\{x,y\} + (1-\alpha)\{a,b\}$.*

$$\{x,y\} \succ \{y\} \text{ and } \{a,b\} \succ \{b\} \text{ implies } U(A) = \min_{y'\in A} U(\{\alpha x + (1-\alpha)a, y'\}),$$

and

$$\{x\} \succ \{x,y\},\ \{a\} \succ \{a,b\},\ y \notin L(x) \text{ and } b \notin L(a)$$
$$\text{implies } U(A) = \max_{x'\in A} U(\{x', \alpha y + (1-\alpha)b\}).$$

Lemma 3.3 enables us to define a mixture operation for the space $\mathscr{M}(\succsim)$, which we recall is comprised of the set of singleton menus and two-element menus in which there is a tempting alternative for which the DM can exert costly self-control. Since any A in $\mathscr{M}(\succsim)$ has at most two-elements, it follows that for any pair of menus A and B in $\mathscr{M}(\succsim)$ and any α in $(0,1)$, the menu $\alpha A + (1-\alpha)B$ has either one, two or four elements. So consider the following (set-)mixture operator which we denote by $h_\alpha(\cdot,\cdot)$. If $A = \{a,b\}$ and $B = \{x,y\}$ with $\{a\} \succ \{a,b\} \succ \{b\}$ and $\{x\} \succ \{x,y\} \succ \{y\}$, then the $(\alpha, 1-\alpha)$-(set-)mixing of A and B consists of taking the $(\alpha, 1-\alpha)$-convex combination of the two better alternatives from each set and the $(\alpha, 1-\alpha)$-convex combination of the two worse alternatives from each set. Thus the resulting 'mixture' set still contains only two elements. For all other possible configurations the standard operation leads to at most two elements anyway, so no modification is required in these cases. More formally, we have for any A and B in $\mathscr{M}(\succsim)$ and any α in $(0,1)$,

$$h_{\alpha}(A,B) := \begin{cases} \{\alpha a + (1-\alpha)x, \alpha b + (1-\alpha)y\} & \text{if } A = \{a,b\},\ b \in S(a),\ B = \{x,y\},\ y \in S(x), \\ \alpha A + (1-\alpha)B & \text{otherwise.} \end{cases}$$

Lemma 3.4. *If a preference $\succsim$ satisfies Axioms 3 and 4, then $\left(\mathscr{M}(\succsim), \{h_{\alpha}\}_{\alpha\in[0,1]}\right)$ is a mixture space as defined in **?**, p52.* [7]

Since it follows from Lemma 3.3 that for any A, B in $\mathscr{M}(\succsim)$, $h_{\alpha}(A,B) \sim \alpha A + (1-\alpha)B$, as a consequence of Lemma 3.4 we can apply the mixture space theorem (**?**, Theorem 8, p297) to obtain the following representation of $\succsim$ restricted to $\mathscr{M}(\succsim)$.

Theorem 1. *A preference relation satisfies Axioms 1, 2a, 2b, 3 and 4 if and only if there exists a linear function $U : \mathscr{M}(\succsim) \to \mathbb{R}$, such that for any A, B in $\mathscr{M}(\succsim)$, $U(A) \geq U(B) \Leftrightarrow A \succsim B$. Moreover, U in the representation is unique up to a positive affine transformation and its restriction to singleton sets is continuous.*

Now to extend the representation obtained in Theorem 1, notice first it follows from Axiom 4 (set-betweenness) that for *any* two-element menu either the menu is indifferent to a singleton menu that consists of just one element from that menu or that menu lies in preference terms strictly between the two singleton menus formed from its two elements. That is, $\{a\} \sim \{a,b\}$ or $\{a,b\} \sim \{b\}$ or $\{a\} \succ \{a,b\} \succ \{b\}$. For the third case, since $\{a,b\}$ is in $\mathscr{M}(\succsim)$, $U(\{a,b\})$ is already defined. For the other two cases we can simply set $U(\{a,b\})$ either to $U(\{a\})$ or to $U(\{b\})$. This provides the unique extension of the function $U(\cdot)$ from Theorem 1 to extend the representation of $\succsim$ to all two-element sets.

It remains to extend the representation to all menus. Our first step in this task is to define, as do GP01, the linear (commitment utility) function $u : \Delta(Z) \to \mathbb{R}$, by setting $u(x) := U(\{x\})$. Next, for any two lotteries a, b and any $\gamma \in (0,1)$, such that $\{a,b\} \in \mathscr{M}(\succsim)$ and $\{a, (1-\gamma)b + \gamma x\} \in \mathscr{M}(\succsim)$ for all $x \in \Delta(Z)$, we define the (temptation utility) function $v : \Delta(Z) \to \mathbb{R}$, as follows:

$$v(x; a, b, \gamma) := \frac{U(\{a,b\}) - U(\{a, (1-\gamma)b + \gamma x\})}{\gamma}.$$

An analogous result to GP01's Lemma 4 (p1423) holds, although we note that the domain of the $U(\cdot)$ in the

[7] In particular,we have for any $\alpha,\beta \in (0,1)$, and any $A,B \in \mathscr{M}(\succsim)$, $h_{\alpha}\left(h_{\beta}(A,B), B\right) = h_{\alpha\beta}(A,B)$.

statement of the next lemma is $\mathscr{M}(\succsim)$ rather than the unrestricted domain $\mathscr{A}$. However, the proof in GP01 is still valid in our setting since all two-element sets used in their proof are in $\mathscr{M}(\succsim)$.

Lemma 3.5. *Let U be a linear function that represents the restriction of some $\succsim$ to $\mathscr{M}(\succsim)$. Suppose that $(1-\gamma)b+\gamma x \in S(a)$ for all $x \in \Delta(Z)$. Then:*

(i) $\forall$ *x such that $x \in S(a)$, $v(x;a,b,\gamma) = U(\{a,b\}) - U(\{a,x\})$.*

(ii) $v(a;a,b,\gamma) = U(\{a,b\}) - U(\{a\})$.

(iii) $v(\alpha x+(1-\alpha)x';a,b,\gamma) = \alpha v(x;a,b,\gamma) + (1-\alpha)v(x';a,b,\gamma)$.

(iv) *$v(x;a,b,\gamma') = v(x;a,b,\gamma)$, for all $\gamma' \in (0,\gamma)$.*

(v) *Suppose that $(1-\gamma)b'+\gamma x \in S(a')$, for all $x \in \Delta(Z)$. Then $v(x;a,b,\gamma) = v(x;a',b',\gamma) + v(b';a,b,\gamma)$.*

Although U is linear on $\mathscr{M}(\succsim)$, we have not established that it is linear on $\mathscr{L}(\succsim)$. However, using an argument similar to the proof of Lemma 5.6 in Kreps (1988), we obtain the following weaker version of linearity.

Lemma 3.6. *Let U be a function that restricted to $\mathscr{M}(\succsim)$ is linear and represents some $\succsim$ satisfying Axioms 1, 2a, 2b, 3 and 4. If $\{x,y\} \in \mathscr{L}(\succsim)$, then for any $A \in \mathscr{M}(\succsim)$, and any $\alpha \in (0,1)$,*

$$U(\alpha\{x,y\}+(1-\alpha)A) = \alpha U(\{x,y\}) + (1-\alpha)U(A).$$

Next we adapt Lemma 5 of GP01(p1424) to our framework and establish a costly self-control representation over two-element menus where for the more preferred element (in terms of the DM's commitment preferences) the willpower constraint never binds.

Lemma 3.7. *Let U be a function that restricted to $\mathscr{M}(\succsim)$ is linear and represents some $\succsim$ satisfying Axioms 1, 2a, 2b, 3 and 4. Consider a pair of lotteries $a, y \in \Delta(Z)$ such that $U(\{a\}) \geq U(\{a,y\}) \geq U(\{y\})$ and $L(a) = \emptyset$. Suppose lottery $b \in \Delta(Z)$ and $\gamma \in (0,1)$, satisfy $(1-\gamma)b+\gamma x \in S(a)$, for all $x \in \Delta(Z)$. Then*

$$U(\{a,y\}) = \max_{x\in\{a,y\}}\{u(x)+v(x;a,b,\gamma)\} - \max_{x'\in\{a,y\}} v(x';a,b,\gamma).$$

Theorem 2. *Supppose $L(a) = \phi$ for all $a \in \Delta(Z)$. A preference $\succeq$ satisfies Axiom 1, 2a, 2b, 3, 4 if and only if there are continuous linear function u, v and a constant w such that the function U defined as*

$$U(A) = \max_{x \in A}\{u(x) + v(x)\} - \max_{y \in A} v(y)$$

$$s.t.\ \max_{y \in A} v(y) - v(x) \leq w$$

for all $A \in \mathscr{A}$ and U present $\succeq$, where w is either suficiently large so the constraint is not relevent or $w = 0$.

4 The Representation for Costly Self-Control with Limited Willpower.

In the previous section, the focus was on menus in which the willpower constraint was not binding and so we obtained results similar to GP01. To obtain a representation that allows for menus in which a failure of self-control may be due to a lack of willpower, we propose one new axiom that allows us to interpret willpower as a limited cognitive resource in which the available 'stock' does not vary across menus.

For any subset $D \subset \Delta(Z)$, let $\bar{D}$ denote its closure.

Axiom 5. *For any lotteries $a, b, x, y \in \Delta(Z)$, if $b \in \overline{L(a)} \cap S(a)$ and $y \in S(x)$, then $\{\frac{1}{2}a + \frac{1}{2}x, \frac{1}{2}a + \frac{1}{2}y\} \succsim \{\frac{1}{2}x + \frac{1}{2}a, \frac{1}{2}x + \frac{1}{2}b\}$.*

To understand this axiom, notice that since $b \in S(a)$ and $y \in S(x)$, for both menus $\{a, b\}$ and $\{x, y\}$, the individual has sufficient willpower to exert costly self-control. However, since $b \in \overline{L(a)}$, exerting such self-control exhausts her entire stock of willpower. The implication we wish to draw is that this means the self-control cost in resisting temptation in the menu $\{a, b\}$ can be no less than it is in resisting temptation in the menu $\{x, y\}$. Since by Lemma 3.4 it follows that $\frac{1}{2}a + \frac{1}{2}y \in S\left(\frac{1}{2}a + \frac{1}{2}x\right)$ and $\frac{1}{2}x + \frac{1}{2}b \in S\left(\frac{1}{2}x + \frac{1}{2}a\right)$ this means that for both menus $\{\frac{1}{2}a + \frac{1}{2}x, \frac{1}{2}a + \frac{1}{2}y\}$ and $\{\frac{1}{2}x + \frac{1}{2}a, \frac{1}{2}x + \frac{1}{2}b\}$, the individual will be choosing the lottery $\frac{1}{2}a + \frac{1}{2}x$ in period 2. However, the cost of self-control should be less in the former than it is in the latter, so the axiom requires

$$\{\frac{1}{2}a + \frac{1}{2}x, \frac{1}{2}a + \frac{1}{2}y\} \succsim \{\frac{1}{2}x + \frac{1}{2}a, \frac{1}{2}x + \frac{1}{2}b\}.$$

We require one final lemma before the main theorem.

Lemma 4.1. *Let U be a function that restricted to $\mathscr{M}(\succsim)$ is linear and represents some $\succsim$ satisfying Axioms 1, 2a, 2b, and 3-5. Suppose a lottery a satisfies $L(a) \neq \emptyset$ and suppose lottery $b \in \Delta(Z)$ and $\gamma \in (0, 1)$, satisfy*

$(1-\gamma) b + \gamma x \in S(a)$, for all $x \in \Delta(Z)$. Then

$$U(\{a,y\}) = \max_{x \in \{a,y\}} \{u(x) + v(x;a,b,\gamma)\} - \max_{x' \in \{a,y\}} v(x';a,b,\gamma),$$

$$s.t. \ \max_{x' \in \{a,y\}} v(x';a,b,\gamma) - v(x;a,b,\gamma) \leq w(a)$$

where $w(a) = \max_{x' \in S(a)} v(x';a,b,\gamma) - v(a;a,b,\gamma)$.

We are now ready to prove our main theorem.

Theorem 3. *Suppose $L(a) \neq \emptyset$ for some $a \in \Delta(Z)$. A preference relation $\succsim$ satisfies Axioms 1, 2a, 2b, 3-5 if and only if there are continuous linear function u, v and a constant $w > 0$ such that the function U defined as:*

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

$$s.t. \ \max_{y \in A} v(y) - v(x) \leq w,$$

where neither u nor v is constant and v is not an affine transformation of u except $v(\cdot) = -\alpha u(\cdot) + \beta$ for some $\alpha \in (0,1)$ and $\beta \in R$. Moreover, (u,v,w) in the representation is unique in the sense that $u' = \alpha u + \beta, v' = \alpha v + \beta', w' = \alpha w$ represent the same preferences as u, v, w.

5 Comparative measure for commitment and self-control

In this section, we define comparative measures of preference for commitment and of costly-self control. We will adapt the result of Theorem 3 in the proof. Hence, we call DM has a regular preference when there exists some $a \in \Delta(Z)$, such that $L(a) \neq \emptyset$.

Definition 5.1. *The preference $\succeq$ has a preference for commitment at A if there exists $B \subset A$ such that $B \succ A$. The preference $\succeq$ has a preference for commitment if $\succeq$ has costly self-control at some $A \in \mathscr{A}$.*

Definition 5.2. *The preferece $\succeq_1$ has greater preference for commitment than $\succeq_2$ if, for all $A \in \mathscr{A}$, $\succeq_2$ has preference for commitment at A implies $\succeq_1$ has preference for commitment at A.*

Since a preference for commitment considers the existence of lotteries $a, b \in \Delta(Z)$ such that $\{a\} \succ \{a,b\}$, whether $L(a) = \phi$ or not is irrelevent. Hence, Theorem 8 in GP01 remains valid.

Theorem 4. *Let $\succeq_1, \succeq_2$ be two regular temptation preferences. Then, $\succeq_1$has greater preference for commitment than $\succeq_2$if and only if there exists u_2, v_2 such that $(u_2,\gamma v_2)$ represents $\succeq_2$ and*

$$u_2 = \alpha u_1 + (1-\alpha)\, v_1,$$

$$v_2 = \beta u_1 + (1-\beta)\, v_1,$$

for some $\alpha, \beta \in [0,1]$ and some $\gamma > 0$.

The following discussion is related to Theorem 9 of GP01.

Definition 5.3. *The preference $\succeq$ has more costly self-control at A if there exissts B, C such that $A = B \cup C$ and $B \succ A \succ C$. The preference $\succeq$ has constly self-control if $\succeq$ has costly self-control at some $A \in \mathscr{A}$.*

Definition 5.4. *The preference $\succeq_1$has more costly self-control than $\succeq_2$if, for all $A \in \mathscr{A}$, $\succeq_2$has costly self-control at A implies $\succeq_1$has self-control at A.*

Theorem 9 of GP01 considers a situation that $\{a\} \succ_2 \{a,b\} \succ_2 \{b\}$ implies $\{a\} \succ_1 \{a,b\} \succ_1 \{b\}$. Hence, when $L_1(a) \neq \phi$, the condition that $u_2 + v_2$ and v_2 are convex combination of $u_1 + v_1$ and v_1 is no longer a sufficient condition for $\succeq_1$being more costly self-control than $\succeq_2$. To see this, first from Theorem 3 we always can sellect u_1, v_1 such that there is a lottery a in the interior of $\Delta(Z)$ such that $u_1(a) = 0$ and $v_1(a) = 0$. Given u_1, v_1, let $\mathscr{U}_1 = \{u_1(x), v_1(x) : x \in \Delta(Z)\} \subset \mathbb{R}^2$ be the domain of (u_1, v_1) on $\Delta(Z)$. For u_2, v_2 satisfying

$$\begin{pmatrix} u_2 + v_2 \\ v_2 \end{pmatrix} = \begin{pmatrix} \alpha & 1-\alpha \\ \gamma\beta & \gamma(1-\beta) \end{pmatrix} \begin{pmatrix} u_1 + v_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} \alpha u_1 + v_1 \\ \gamma\beta u_1 + \gamma v_1 \end{pmatrix} \tag{1}$$

for some $\alpha, \beta \in [0,1]$ and some $\gamma > 0$, we cannot obtain $S_2(a) \subset S_1(a)$ without further restrictions. Please see Figure 1. .

Note that (1) implies that

$$\begin{aligned} u_2 &= (\alpha - \gamma\beta)u_1 + (1-\gamma)v_1 &(2)\\ &= (\alpha - \beta)u_1 + \frac{1-\gamma}{\gamma}v_2 &(3) \end{aligned}$$

Hence, if we restrict our comparison of two preferences in cases where $\alpha - \gamma\beta > 0$ and $\alpha - \beta > 0$, then we require $\succ_1, \succ_2$satisfy the following conditions.

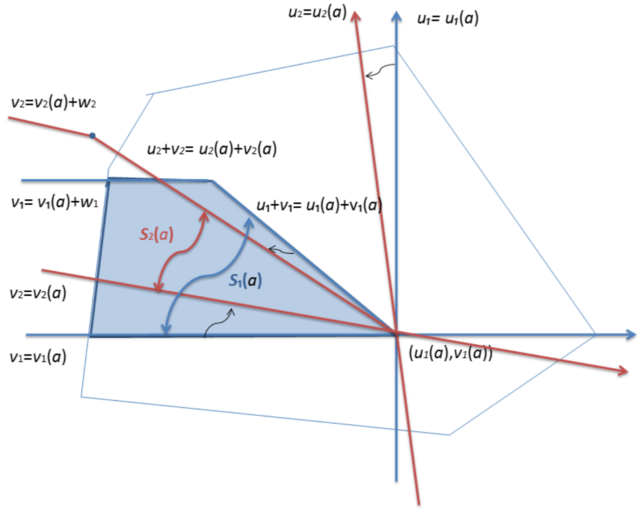

Figure 1: $S_2(a) \not\subset S_1(a)$

1. if $v_1(a) = v_1(b)$ then $\{a\} \succ_1 \{b\}$ if and only if $\{a\} \succ_2 \{b\}$, and

2. if $v_2(a) = v_2(b)$ then $\{a\} \succ_2 \{b\}$ iff $\{a\} \succ_1 \{b\}$.

The first condition is from equation 2 and the second condition is from 3. These two conditions on preferences $\succeq_1, \succeq_2$are to require DM 1 and DM2 have the same ranking among lotteries which are equally tempting for at least one of them. With this restriction, we rule out cases that $b \in S_1(a)$ but $\{b\} \succ_2 \{a\}$. Hence, we have the following theorem.

Theorem 5. *Let $\succeq_1, \succeq_2$be two regular self-control preferences satisfy conditions 1 and 2. Let (u_1, v_1) be a representation of $\succeq_1$. Then, $\succeq_1$has more constly self-control than $\succeq_2$if and only if there exists u_2, v_2 such that (u_2, v_2) represents $\succeq_2$ and satisfy the followings:*

1.
$$\begin{pmatrix} u_2 + v_2 \\ v_2 \end{pmatrix} = \begin{pmatrix} \alpha & 1-\alpha \\ \gamma\beta & \gamma(1-\beta) \end{pmatrix} \begin{pmatrix} u_1 + v_1 \\ v_1 \end{pmatrix},$$
for some $\alpha, \beta \in [0,1], \gamma > 0$, $\alpha - \gamma\beta > 0$ and $\alpha - \beta > 0$.

2. If $L_1(a) \neq \phi$, then if $w_2 \geq \gamma\frac{\alpha-\beta}{\alpha}w_1$, we have $(u_1(a) - \frac{w_1}{\alpha}, v_1(a) + w_1) \notin \mathscr{U}_1$, and if $w_2 < \gamma\frac{\alpha-\beta}{\alpha}w_1$, we have $(u_1(a) - \frac{\gamma w_1 - w_2}{\gamma\beta}, v_1(a) + w_1) \notin \mathscr{U}_1$.

Proof. If $L_1(a) \neq \phi$, from conditions 1 and 2, we only need to get conditions such that $S_1(a) \subset S_2(a)$. Let b' be the intersection point of $u_2(b) + v_2(b) = u_2(a) + v_2(a)$ and $v_2(b) - v_2(a) = w_2$. Hence, by (1)we have $v_1(b') = v_1(a) + \frac{\alpha}{\gamma(\alpha-\beta)} w_2$. If $v_1(b') \geq v_1(a) + w_1$, i.e. $w_2 \geq \gamma \frac{\alpha-\beta}{\alpha} w_1$, then the critical point b_1 would be the intersection of $u_2(b) + v_2(b) = u_2(a) + v_2(a)$ and $v_1(b) - v_1(a) = w_1$. Hence, $S_2(a) \subset S_1(a)$ if and only if $(u_1(b_1), v_1(b_1)) \notin \mathscr{U}_1$. If $v_1(b') < v_1(a) + w_1$, then the critical point b_2 would be the intersection of $v_2(b) - v_2(a) = w_2$ and $v_1(b) - v_1(a) = w_1$. Hence, $S_2(a) \subset S_1(a)$ if and only if $(u_1(b_2), v_1(b_2)) \notin \mathscr{U}_1$. □

6 Appendix

Proof of Lemma 3.3. By Lemma 3.2, there exists (x^*, y^*) such that $A \sim \{x^*, y^*\}$ and (x^*, y^*) solves the maxmin problem. First we show that $\{x, y\} \succ \{y\}$ and $\{a, b\} \succ \{b\}$ implies $x^* = \alpha x + (1-\alpha)a$. By Axiom 3, we have

$$A \succ \alpha\{y\} + (1-\alpha)\{a, b\},$$

$$A \succ \alpha\{x, y\} + (1-\alpha)\{b\}.$$

Suppose $x^* = \alpha x + (1-\alpha)b$. Then, since $A \sim \{x^*, y^*\}$ and it solves the maxmin problem, we have

$$A \preceq \{\alpha x + (1-\alpha)b, \alpha y + (1-\alpha)b\} = \alpha\{x, y\} + (1-\alpha)\{b\} \prec A,$$

which yields a contradiction. Similarly, if $x^* = \alpha y + (1-\alpha)a$, then

$$A \preceq \{\alpha y + (1-\alpha)b, \alpha y + (1-\alpha)a\} = \alpha\{y\} + (1-\alpha)\{a, b\} \prec A.$$

If $x^* = \alpha y + (1-\alpha)b$, then

$$A \preceq \{\alpha y + (1-\alpha)b, \alpha y + (1-\alpha)b\} = \{\alpha y + (1-\alpha)b\} \prec \alpha\{y\} + (1-\alpha)\{a, b\} \prec A.$$

Hence, $x^* = \alpha x + (1-\alpha)a$. Suppose that we have $\{x\} \succ \{x, y\}$ and $\{a\} \succ \{a, b\}$ with $y \notin L(x)$ and $b \notin L(a)$. Then we can apply Axiom 3 and obtain

$$\alpha\{x\} + (1-\alpha)\{a, b\} \succ A,$$

$$\alpha\{x, y\} + (1-\alpha)\{a\} \succ A.$$

Then since $A \sim \{y^*, x^*\}$ and it solves the minmax problem, we can use a similar argument as above to show $y^* = \alpha y + (1-\alpha)b$. □

Proof of Lemma 3.4. First we will show that $h_\alpha(A, B) \in \mathscr{M}(\succeq)$ for any $A, B \in \mathscr{M}(\succeq)$. From Lemma 3.3, it is known that h_α is either a singleton set or a two-element set. If h_α is a singleton set, then obviously it is in $\mathscr{M}(\succeq)$. If h_α has two elements, then it only takes one of the two possible forms, either $h_\alpha(\{a, b\}, \{x\})$, or $h_\alpha(\{a, b\}, \{x, y\})$ with $b \in S(a)$ and $y \in S(x)$. By Axiom 3, we have

$$\begin{aligned}\{\alpha a + (1-\alpha)x\} = \alpha\{a\} + (1-\alpha)\{x\} &\succ \{\alpha a + (1-\alpha)x, \alpha b + (1-\alpha)x)\} \\ &\succ \alpha\{b\} + (1-\alpha)\{x\} = \{\alpha b + (1-\alpha)x\}\end{aligned}$$

Hence, $h_\alpha(\{a, b\}, \{x\}) \in \mathscr{M}(\succeq)$. By Axiom 3, we also have

$$\begin{aligned}\{\alpha a + (1-\alpha)x\} \succ \alpha\{a, b\} + (1-\alpha)\{x\} &\succ \alpha\{a, b\} + (1-\alpha)\{x, y\} \\ &\succ \alpha\{a, b\} + (1-\alpha)\{y\} \succ \alpha\{b\} + (1-\alpha)\{y\} = \{\alpha b + (1-\alpha)y\}\end{aligned}$$

Hence, $h_\alpha(\{a, b\}, \{x, y\}) \in \mathscr{M}(\succeq)$ as well.

Next we will show that $h_\alpha((h_\beta(A, B), B) = h_{\alpha\beta}(A, B)$ for any $A, B \in \mathscr{M}(\succeq)$. We only deal with the case when $A = \{x, y\}$ and $B = \{a, b\}$ with $y \in S(x)$ and $b \in S(a) \succ \{b\}$ because for the rest cases, the argument is similar but easier.

$$\begin{aligned}&h_\alpha(h_\beta(\{x, y\}, \{a, b\}), \{a, b\}) \\ &= h_\alpha(\{\beta x + (1-\beta)a, \quad \beta y + (1-\beta)b\}, \{a, b\}) \\ &= \{\alpha(\beta x + (1-\beta)a) + (1-\alpha)a, (1-\alpha)(\beta y + (1-\beta)b) + (1-\alpha)b\} \\ &= h_{\alpha\beta}(\{x, y\}, \{a, b\}).\end{aligned}$$

□

Proof of Lemma 3.6. If $\{x, y\}$ satisfies $\{x\} \succ \{x, y\} \succ \{y\}$, the linearity is already proven in lemma 4. We have to deal with $\{x, y\}$ with $\{x\} \sim \{x, y\} \succ \{y\}$ or $\{x\} \succ \{x, y\} \sim \{y\}$. If $\{x, y\}$ satisfies $\{x\} \succ \{x, y\} \sim \{y\}$, we claim that $\alpha\{x, y\} + (1-\alpha)A \sim \alpha\{y\} + (1-\alpha)A$ for all $A \in \mathscr{M}(\succeq)$. By Axiom 3, $\{x\} \succ \{y\}$ implies $\alpha\{x\} + (1-\alpha)A \succ \alpha\{y\} + (1-\alpha)A$, and Axiom 4 further implies $\alpha\{x\} + (1-\alpha)A \succsim \alpha\{x, y\} + (1-\alpha)A \succsim \alpha\{y\} + (1-\alpha)A$. In this case, we only have to show that $\alpha\{x, y\} + (1-\alpha)A \succ \alpha\{y\} + (1-\alpha)A$ will lead to a contradiction. Let us take $A = \{a, b\}$ with $\{a\} \succ \{a, b\} \succ \{b\}$. Suppose that $\alpha\{x, y\} + (1 - \alpha)\, A \succ \alpha\{y\} + (1 - \alpha)\, A$. In this case, we have $\{x\} \succ \{x, y\} \sim \{y\}$. Since $\{x\} \succ \{y\}$, letting $\alpha, \beta \in (0, 1)$ and applying Axiom 3, we obtain

$$\beta\{x\} + (1 - \beta)\{y\} \succ \beta\{y\} + (1 - \beta)\{y\} = \{y\} \sim \{x, y\},$$

and

$$\alpha\{x'\} + (1 - \alpha)\{a, b\} \succ \alpha\{x, y\} + (1 - \alpha)\{a, b\},$$

where $\{x'\} = \beta\{x\} + (1 - \beta)\{y\}$.

Since $\alpha\{x'\} + (1 - \alpha)\{a, b\} \succ \alpha\{x, y\} + (1 - \alpha)\{a, b\} \succ \alpha\{y\} + (1 - \alpha)\{a, b\}$, von Neumann-Morgenstern continuity implies there exists some $\gamma \in (0, 1)$ such that

$$\begin{aligned} &\alpha\{x\} + (1 - \alpha)\{a, b\} \\ &\succ \gamma\left(\alpha\{x'\} + (1 - \alpha)\{a, b\}\right) + (1 - \gamma)\left(\alpha\{x'\} + (1 - \alpha)\{a, b\}\right). \end{aligned}$$

We only deal with the case when $A = \{a, b\}$ with $\{a\} \succ \{a, b\} \succ \{b\}$ because when A is a singleton set, the argument is similar but easier. Since $\alpha\{x'\} + (1 - \alpha)\{a, b\}$ and $\alpha\{y\} + (1 - \alpha)\{a, b\}$ are both in $\mathscr{M}(\succeq)$, we use Lemma 3.3 to obtain the first "$\sim$" below

$$\begin{aligned} \alpha\{x, y\} + (1 - \alpha)\{a, b\} &\succ \gamma\left(\alpha\{x'\} + (1 - \alpha)\{a, b\}\right) + (1 - \gamma)\left(\alpha\{y\} + (1 - \alpha)\{a, b\}\right) \\ &\sim h_\gamma\left(\alpha\{x'\} + (1 - \alpha)\{a, b\}, \alpha\{y\} + (1 - \alpha)\{a, b\}\right) \\ &\sim \alpha\left(\gamma\{x'\} + (1 - \gamma)\{y\}\right) + (1 - \alpha)\{a, b\} \\ &\succ \alpha\{x, y\} + (1 - \alpha)\{a, b\}, \end{aligned}$$

which yields a contradiction. The last "$\succ$" uses the fact that $\{x'\} \succ \{y\}$, $\gamma\{x'\} + (1 - \alpha)\{y\} \succ \{y\} \sim \{x, y\}$ and Axiom 4. Since $\alpha\{x, y\} + (1 - \alpha)\, A \sim \alpha\{y\} + (1 - \alpha)\, A$, we have $U(\alpha\{x, y\} + (1 - \alpha)\, A) = U(\alpha\{y\} + (1 - \alpha)\, A)$. Using Theorem 1, we have $U(\alpha\{y\} + (1 - \alpha)\, A) = \alpha U(\{y\}) + (1 - \alpha)\, U(A) = \alpha U(\{x, y\}) + (1 - \alpha)\, U(A)$ for all $A \in \mathscr{M}(\succeq)$ For $\{x, y\}$ with $\{x\} \sim \{x, y\} \succ \{y\}$, Axioms 3 and 4 imply $\alpha\{x\} + (1 - \alpha)\{a, b\} \succsim \alpha\{x, y\} + (1 - \alpha)\{a, b\}$ in the previous discussion. By using the fact that $\{x\} \succ \{y\}$, and applying a similar argument, we can rule out the possibility that $\alpha\{x\} + (1 - \alpha)\{a, b\} \succ \alpha\{x, y\} + (1 - \alpha)\{a, b\}$ and further obtain $U(\alpha\{x, y\} + (1 - \alpha)\, A) = U(\alpha\{x\} + (1 - \alpha)\, A) = \alpha U(\{x\}) + (1 - \alpha)\, U(A) = \alpha U(\{x, y\}) + (1 - \alpha)\, U(A)$ for all $A \in \mathscr{M}(\succeq)$ □

Proof of Lemma 3.7. For the case where $U(\{a\}) > U(\{a,y\}) > U(\{y\})$, by GP(2001) we know $v(y;a,b,\gamma) \geq v(a;a,b,\gamma)$ and $u(a) + v(a;a,b,\gamma) - v(y;a,b,\gamma) > u(y) + v(y;a,b,\gamma) - v(y;a,b,\gamma)$. Let $A = (1-\gamma)\{a,b\}+\gamma\{a,y\}$. Since $\{a,b\} \in \mathscr{M}(\succeq)$, by Lemma 3.6 we have $U(A) = (1-\gamma)U(\{a,b\})+\gamma U(\{a,y\})$ for $\{a,y\} \in \mathscr{L}(\succeq)$. For the case where $U(\{a\}) = U(\{a,y\}) > U(\{y\})$, the first part of Lemma 3.3 establishes that $U(A) = \min_{z\in A} U(\{a,z\})$. Hence, we have $v(a;a,b,\gamma) \geq v(y;a,b,\gamma)$ for the by the same argument in GP(2001). For the case where $U(\{a\}) > U(\{a,y\}) = U(\{y\})$ and $y \in T(a) \setminus L(a)$, we will show $v(y;a,b,\gamma) \geq v(a;a,b,\gamma) + u(a) - u(y)$. From GP(2001) this is equivalent to show that

$$U(\{a,(1-\gamma)b+\gamma y\}) \leq (1-\gamma)U(\{a,b\}) + \gamma U(\{a,y\}) = U(A)$$

The above inequality holds because of the second part of Lemma 3.3 $U(A) = \max_{w\in A}(\{w,(1-\gamma)b+\gamma y\})$. □

Proof of Lemma 4.1. For the case where $y \in S(a)$, we have $v(y;a,b,\gamma) \leq \max_{z\in S(a)} v(z;a,b,\gamma)$. Hence, by Lemma 3.7, we have the desired result for all $\{a,y\} \in \mathscr{L}(\succeq)$. The remaining part of the proof is to show that if $y \in L(a)$, then we must have $v(y;a,b,\gamma) - v(a;a,b,\gamma) \geq w(a)$, which is equivalent to show that $v(y;a,b,\gamma) > \max_{z\in S(a)} v(z;a,b,\gamma)$. Since $y \in L(a)$, $\exists\bar{\gamma} \in (0,1)$ such that $\gamma' y + (1-\gamma')a \in S(a)$ if $\gamma' \leq \bar{\gamma}$ and $\gamma' y + (1-\gamma')a \in L(a)$ if $\gamma' > \bar{\gamma}$. Let $\bar{y} = \bar{\gamma} y + (1-\bar{\gamma})a$. Since $\bar{y} \in S(a)$, by Lemma 3.5 (iii), we have $\bar{\gamma} v(y;a,b,\gamma) + (1-\bar{\gamma})v(a;a,b,\gamma) = v(\bar{y};a,b,\gamma) \geq v(a;a,b,\gamma)$. Hence, $v(y;a,b,\gamma) \geq v(a;a,b,\gamma)$. Moreover, $u(a) + v(a;a,b,\gamma) > u(\bar{y}) + v(\bar{y};a,b,\gamma) = \bar{\gamma}(u(y) + v(y;a,b,\gamma)) + (1-\bar{\gamma})(u(a) + v(a;a,b,\gamma))$. Hence, $u(a) + v(a;a,b,\gamma) > u(y) + v(y;a,b,\gamma)$. We claim that $\max_{z\in S(a)} v(z;a,b,\gamma) = v(\bar{y};a,b,\gamma)$. For any $b' \in S(a)$, from Axiom 5, we have $\frac{1}{2}\{a\} + \frac{1}{2}\{a,b'\} \succeq \frac{1}{2}\{a\} + \frac{1}{2}\{a,\bar{y}\}$. Since $\{a,b'\}$ and $\{a,\bar{y}\}$ are in $\mathscr{M}(\succeq)$, we have $\frac{1}{2}u(a) + \frac{1}{2}U(\{a,b'\}) \geq \frac{1}{2}u(a) + \frac{1}{2}U(\{a,\bar{y}\})$, $U(\{a,b'\}) = u(a) + v(a;a,b,\gamma) - v(b';a,b,\gamma)$ and $U(\{a,\bar{y}\}) = u(a) + v(a;a,b,\gamma) - v(\bar{y};a,b,\gamma)$. Hence, $v(b';a,b,\gamma) \leq v(\bar{y};a,b,\gamma)$. □

Note that if $y \in \overline{L(x)} \cap S(x)$, then we can find a sequence of y_i converges to y such that $y_i \in L(x)$. Hence, when $L(x) \notin \emptyset$, the claim 1 in GP's proof of Theorem 3 does not hold. However, we have the folowings:

Lemma 6.1. *If there exists some pair $x, y' \in \Delta(Z)$ such that $y' \in S(x)$, then there is a $\gamma > 0$ such that $(1-\gamma)y + \gamma a \in S(x)$ for all $a \in \Delta(Z)$ and $y = \frac{x+y'}{2}$.*

Proof. From Lemma 3.4 and $\{x,y\} = \frac{1}{2}\{x\} + \frac{1}{2}\{x,y'\}$, we know $y \in S(x)$. Suppose there is a $\gamma_z > 0$ such that $(1-\gamma_z)y + \gamma_z[z] \in S(x)$ for all $z \in Z$. Since Z is finite, letting $\gamma = \min_{z\in Z}\{\gamma_z\}$ and applying Lemma 3.4, we can obain the desired result. Hence, for any $z \in Z$, let $y_i := (1-\gamma_i)y + \gamma_i[z]$, and $x_i := (1-\gamma_i)x + \gamma_i[z]$

and $\gamma_i \to 0$. We will show that $y_i \in S(x)$ for sufficinetly large i. Suppose to the contrary that we can find a subsequence $y_{i'}$ from y_i, such that $y_{i'} \notin S(x)$. We show all the possible alteratives will lead to a contradiction.
,

Case 1. Suppose we have $\{x, y_{i'}\}\tilde{}\{x\}$ for all i'. By Axiom 2a, we have $\{x, y\} \succeq \{x\}$, which contradicting $y \in S(x)$.

Case 2. Suppose we have $y_{i'} \in L(x)$ for all i'. By the definition of $L(x)$, there exists $\alpha \in (0,1)$ such that $(1-\alpha)x + \alpha y_{i'} \in S(x)$. Let $a = \frac{1-\alpha}{2-\alpha}y' + (1 - \frac{1-\alpha}{2-\alpha})((1-\alpha)x + \alpha y_{i'})$. By Lemma 3.4, we have $a \in S(x)$. However, we can rewrite $a = \left(1 - \frac{\alpha\gamma_i}{2-\alpha}\right) y + \frac{\alpha\gamma_i}{2-\alpha}[z]$. Hence, for all $\gamma_j \leq \frac{\alpha\gamma_i}{2-\alpha}$, we have $y_j \in S(x)$, which yields a contradiction.

Case 3. Suppose we have $y_{i'} \in T(x) \backslash L(x)$ and $y \in T(x_{i'})$ for all i'. By Axiom 4, we have

$$U(\{x, y_{i'}, x_{i'}, y\}) \leq \max\{U(\{x, y_{i'}\}), U(\{x_{i'}, y)\} = \max\{U(\{y_{i'}\}), U(\{y\})\}$$

and

$$U(\{x, y_{i'}, x_{i'}, y\}) \geq \min\{U(\{x, y\}), U(\{x_{i'}, y_{i'}\})\}.$$

Note that $\{x_{i'}, y_{i'}\} = (1-\gamma_{i'})\{x, y\} + \gamma_{i'}\{[z]\}$. By Lemma 3.6, we have $U(\{x_{i'}, y_{i'}\}) = (1-\gamma_{i'})U(\{x, y\}) + \gamma_{i'}U(\{[z]\})$. Since $U(\{x, y\}) > U(\{y\})$, we obtain a contradiction.

Case 4. Suppose we have $y_{i'} \in T(x) \backslash L(x)$ and $y \in S(x_{i'})$ for all i'. We apply Lemma 3.6 to obtain

$$\frac{1}{2}U(\{x, y_{i'}\}) + \frac{1}{2}U(\{x_{i'}, y\}) = U\left(\frac{1}{2}\{x, y_{i'}\} + \frac{1}{2}\{x_{i'}, y\}\right),$$

therefore the same argument in GP01 (Claim 1, page 1426) follows. □

Now we are ready to prove our main theorem.

Proof of Theorem 3. We first show that the "only if" part of Theorem 3 holds. Since there exists a in $\Delta(Z)$ such that $L(a) \neq \emptyset$, we have $S(a) \neq \emptyset$. By Lemma 6.1, there exists $b \in S(a)$, $\gamma \in (0,1)$ satisfy $(1-\gamma)b + \gamma a' \in S(a)$ for all $a' \in \Delta(Z)$. By Lemma 4.1, we can let $u(a') := U(\{a'\})$, $v(a') := v(a'; a, b, \gamma)$ for all $a' \in \Delta(Z)$ and $w = v(\bar{a}; a, b, \gamma) - v(a; a, b, \gamma)$, where $\bar{a} \in \overline{L(a)} \cap S(a)$. By the first part of the proof of Lemma ***3.7***, we know $u(a) + v(a) > u(b) + v(b)$ and $v(b) > v(a)$. Hence, neither u nor v is constant and $v(\cdot) \neq -\alpha u(\cdot) + \beta$ for some $\alpha \in (-\infty, 0] \cup [1, \infty)$ and $\beta \in R$.

Now consider the set $A = \{x, y'\}$, where x and y' are in the relative interior of $\Delta(Z)$. Assume without loss of generality, that $u(x) \geq u(y')$. Since x is in the interior of $\Delta(Z)$ and $b \in S(a)$, we can select $\alpha \in (0,1)$ and

$x' \in \Delta(Z)$ such that $\{x,y\} = \alpha\{x'\} + (1-\alpha)\{a,b\} \in \mathscr{M}(\succeq)$. Hence, by Lemma 6.1, there exists $\gamma' \in (0,1)$ such that $(1-\gamma')\,y + \gamma' a' \in S(x)$ for all $a' \in \Delta(Z)$. If $L(x) = \emptyset$, then $\{x, y'\} \in \mathscr{L}(\succeq)$. Hence, we can apply Lemma 3.7 and obtain

$$U(\{x,y'\}) = \max_{x'' \in \{x,y'\}} \{u(x'') + v(x''; x, y, \gamma')\} - \max_{x'' \in \{x,y'\}} v(x''; x, y, \gamma').$$

Note that the willpower constraint is relevant only when $y' \in S(x)$. Hence, we need to show that $v(y'; x, y, \gamma') - v(x; x, y, \gamma') \leq w$ if $y' \in S(x)$. Since $y' \in S(x)$, by Axiom 5, we have $\frac{1}{2}\{a\} + \frac{1}{2}\{x, y'\} \succeq \frac{1}{2}\{x\} + \frac{1}{2}\{a, a^*\}$. By Lemma 3.6, we then have $\frac{1}{2}u(a) + \frac{1}{2}U(\{x,y'\}) \geq \frac{1}{2}u(x) + \frac{1}{2}U(\{a,\bar{a}\})$, where $U(\{x,y'\}) = u(x) + v(x; x, y, \gamma') - v(y'; x, y, \gamma')$ and $U(\{a,\bar{a}\}) = u(a) - w$. Hence, we have $v(y'; x, y, \gamma') - v(x; x, y, \gamma') \leq w$. Let $\gamma^* = \min\{\gamma, \gamma'\}$. By Lemma 3.5 (iv), $v(\cdot; a, b, \gamma^*) = v(\cdot; a, b, \gamma)$ and $v(\cdot; x, y, \gamma^*) = v(\cdot; x, y, \gamma')$. By Lemma 3.5 (v), for an appropriate constant k, $v(\cdot; a, b, \gamma^*) = v(\cdot; x, y, \gamma^*) + k$ and hence it follows that

$$\begin{gathered} U(\{x,y'\}) = \max_{x'' \in \{x,y'\}} \{u(x'') + v(x'')\} - \max_{y'' \in \{x,y'\}} v(y'') \\ s.t. \max_{y'' \in \{x,y'\}} v(y'') - v(x'') \leq w \end{gathered}$$

If $L(x) \neq \emptyset$, we can apply Lemma 4.1,

$$\begin{gathered} U(\{x,y'\}) = \max_{x'' \in \{x,y'\}} \{u(x'') + v(x''; x, y, \gamma')\} - \max_{y'' \in \{x,y'\}} v(y''; x, y, \gamma'), \\ \text{s.t.} \max_{y'' \in \{x,y'\}} v(y''; x, y, \gamma') - v(x''; x, y, \gamma') \leq w(x), \end{gathered}$$

where $w(x) = \max_{y'' \in S(x)} v(y''; x, y, \gamma') - v(x; x, y, \gamma')$. Take $\bar{x} \in \overline{L(x)} \cap S(x)$. By Axiom 5, we have $\frac{1}{2}\{a\} + \frac{1}{2}\{x, \bar{x}\} \sim \frac{1}{2}\{x\} + \frac{1}{2}\{a, \bar{a}\}$. By Lemma 3.6, we then have $\frac{1}{2}u(a) + \frac{1}{2}U(\{x, \bar{x}\}) = \frac{1}{2}u(x) + \frac{1}{2}U(\{a, \bar{a}\})$, where $U(\{x, \bar{x}\}) = u(x) + v(x; x, y, \gamma') - v(\bar{x}; x, y, \gamma')$ and $U(\{a, \bar{a}\}) = u(a) - w$. Hence, we have $w(x) = v(\bar{x}; x, y, \gamma') - v(x; x, y, \gamma') = w$. Follow the same argument as above, we have

$$\begin{gathered} U(\{x,y'\}) = \max_{x'' \in \{x,y'\}} \{u(x'') + v(x'')\} - \max_{y'' \in \{x,y'\}} v(y'') \\ s.t. \max_{y'' \in \{x,y'\}} v(y'') - v(x'') \leq w. \end{gathered}$$

Now consider an arbitrary finite set A. We know that

$$\begin{aligned} U(A) &= \max_{x\in A}\min_{y\in A} U(\{x,y\}) \\ &= \max_{x\in A}\min_{y\in A}\left\{\begin{array}{c} \max_{x'\in\{x,y\}}\{u(x')+v(x')\}-\max_{y'\in\{x,y\}} v(y') \\ s.t.\, \max_{y'\in\{x,y\}} v(y')-v(x')\leq w \end{array}\right\} \\ &= \max_{x\in A}\min_{y\in A}\left\{\begin{array}{c} \max_{x'\in\{x,y\}}\{u(x')+v(x')\} \\ s.t.\, \max_{y'\in\{x,y\}} v(y')-v(x')\leq w \end{array}\right\}+\min_{y\in A}\{-v(y)\} \end{aligned}$$

Let $y^*\in\arg\max_{y\in A} v(y)$. If ,$x\in A$ such that $v(y^*)-v(x)>w$, then x does not solve the constraint maxminmax problem because for the pair $\{x,y^*\}$ we would choose y^* instead of x. Hence, x will not survive after the second requirement, i.e., $\min_{y\in A}$ when we take $y=y^*$. Now consider any $x\in A$ such that $v(y^*)-v(x)\leq w$, then for any pair $\{x,x'\}$ where $x'\in A$, we have $v(x')-v(\ x^*)\leq w$. Hence, if $v(y^*)-v(x')\leq w$, then we choose x over x' only when $u(x)+v(x)\geq u(x')+v(x')$. Hence, we have

$$\begin{aligned} U(A) = \max_{x\in A}\{u(x)+v(x)\}-\max_{y\in A}\{v(y)\} \\ s.t. \max_{y\in A} v(y)-v(x)\leq w \end{aligned}$$

□

References

Gul, F., and W. Pesendorfer (2001): "Temptation and Self-Control,"*Econometrica*, 69(6), 1403-1435.

Herstein, I. N., and J. Milnor (1953): "An Axiomatic Approach to Measurable Utility," *Econometrica*, 21(2), 291-297.

Kreps, D. (1988): *Notes on the theory of choice.* Westview Press.

Masatlioglu, Y., D. Nakajima, and E. Ozdenoren (2014): "Revealed Willpower," Working Paper.