Menu Choice, Environmental Cues and Temptation: A "Dual Self" Approach to Self-control*

Kalyan Chatterjee† R. Vijay Krishna‡

November, 2005

Abstract

In this paper we consider the following two-period problem of self-control. In the first period, an individual has to decide on the set of feasible choices from which she will select one in the second period. In the second period, the individual *might* choose an alternative that she would find inferior in the first period. This eventuality need not occur with certainty but might be triggered by the nature of the set chosen in the first period. We propose a model for this problem and axioms for first-period preferences, in which the second period choice could be interpreted as being made by an "alter-ego" who appears with some probability. We provide a discussion of the behavioural implications of our model as compared with existing theories.

*We wish to thank Sophie Bade, Siddhartha Bandyopadhyay, Susanna Esteban, Kfir Eliaz, Drew Fudenberg, Ed Green, Faruk Gul, John Hardman Moore, Jawwad Noor, József Sákovics, Rani Spiegler and Jonathan Thomas for helpful comments and Avantika Chowdhury for research assistance. We are especially grateful to Bart Lipman for insightful and thought-provoking discussions.

†Department of Economics, Pennsylvania State University. *Email*: kchatterjee@psu.edu, *URL:* http://www.econ.psu.edu/people/biographies/chatterjee_bio.htm

‡Department of Economics, University of Edinburgh. *Email*: vijay.krishna@ed.ac.uk, *URL:* http://homepages.ed.ac.uk/vkrishna

1 Introduction

As first pointed out by Schelling (1978, 1984), the self-control problems we face in our everyday lives pose a dilemma for the usual model of rational choice. Schelling's 1984 Ely lecture to the American Economic Association starts off with the woman who, before delivery, asks her obstetrician to withhold anaesthesia during delivery but, while in the throes of delivering her child, experiences extreme pain and changes her mind. This example is characteristic of a class of phenomena characterised by individuals "attempting to overrule one's own preferences," as Schelling puts it. There are several ways of understanding these phenomena and many of them are mentioned in this article of Schelling's, all using the language of multiple selves. For example, one could regard one type of behaviour as a mistake; constraining oneself enables us to "separate the anomalous behaviour from the rational; we take sides with whichever consumer self appeals to us as the authentic representation of values." One can also, as Schelling says, treat these selves more symmetrically and ask "To which patient is a physician obligated ... the one asking for anaesthesia or the one who asked that it be withheld?" And later in the article, "... without necessarily taking sides..., we can say that it looks as if different selves took turns, each self wanting its own values to govern what the other selves will do... ."

Schelling also recognises the difficulties with speaking of "multiple selves"; he writes that he is only secure using this terminology among economists because questions might arise, in law, for example, as to which self was party to a contract or violated the law.

An important feature of the examples in Schelling's work is that each self obtains utility even when it is absent (when another self is in control), as in the woman who after the delivery would have preferred not to have asked for anaesthesia during the delivery. David Gauthier (1987) has an interesting example of the person who preferred rock music at 20 and prefers classical music at 40, but the 40-year self would have preferred to have preferred classical music when he was 20. (Likewise the 20-year old would no doubt want not to like classical music more than rock at 40, even though "he" would have been gone for 20 years by the time his classical-music-loving self is in control.)

As noted above, Schelling introduces three distinct ways of thinking about people who "overrule their own preferences." The first is that they make mistakes, the second is that they have preferences that change over time and the third is that decision makers have multiple selves, one of whom makes the actual choice. Our model, closest to the third approach, seeks to explain the behaviour described by Schelling as a temporary loss of control, with choices being *made as if* by a virtual alter-ego with different preferences but being *evaluated* in accordance with the decision-maker's own long-term preferences. Moreover, this loss of control is not certain to happen (or certain not to happen); there is some uncertainty as to whether an individual will be tempted or not.

It is best to think of our approach in an intertemporal context, though explicit consideration of time is deferred to another paper. An individual has to make a choice from a set of lotteries at some point in the future. At an earlier point in time, she gets to choose a *menu* of lotteries from which to make his future choice. We are interested in the following behaviour, which the individual considers a possibility. The menu chosen might trigger temptation in the next period (modelled here as an alternative self, with different preferences from the initial one, assuming control of making a choice from the menu) or it might not (so that the same self remains in charge of the choice). The probability of the alternate self taking over is menu-dependent. Whichever self is in charge of making a choice from the menu in the second period makes its own most preferred choice. Each self does not particularly care about the *utilities* of the other self, so this is not an interdependent utilities model, but does care about the *choice* made (purchases of classical music CDs vs. other kinds in the second period). Let x be a typical menu and β a typical lottery in the menu. This gives the decision-maker a utility function of the following form:

$$U(x) = (1-\rho_x)\max_{\beta\in x} u(\beta) + \rho_x \max_{\beta\in B_v(x)} u(\beta)$$

where U is the individual's utility from a menu, u her utility from a lottery, v is the alter-ego's utility from a lottery and $B_v(x)$ is the set of v-maximisers in x. We shall take ρ_x[1] to be the probability that the individual gets tempted (when faced with the

[1]Needless to say, ρ_x must satisfy certain regularity properties. For more on this, see §3.

choices in the menu x), which results in the alter-ego making a choice. We shall say that a utility function over menus which takes such a form admits a *dual self* representation. We provide axioms for first period preferences over menus so that the decision-maker's utility from a menu is given by the equation above. Thus, a decision-maker who satisfies our axioms behaves *as if* there is a probability of his getting tempted, when a choice has to be made from a menu, which is represented as the choice being made by an alter-ego. It should be emphasised that the alter-ego (and his utility function v) is subjective, as is the probability, ρ_x, of getting tempted. The only observables are first period choices over menus. We introduce a slight asymmetry between the selves in that the alter-ego, if he has to make a choice, will choose, among his most preferred alternatives, that which is most preferred by the decision-maker. Since we are characterising the decision-maker's utility, how she breaks ties does not really matter to us.

Several recent papers have focused on the problems raised by Schelling. The paper closest in spirit to ours is the innovative paper by Bernheim and Rangel (2004), who specifically deal with addiction and are clear that, in their view, the individual who takes drugs is making a mistake caused by overestimating the amount of pleasure consumption would involve relative to the long-term costs of such consumption. The selves are not treated symmetrically; drug consumption is anomalous and abstaining from it rational. Their model also explicitly takes into account the effect of environmental cues in triggering the change of the controlling self, from *cold* to *hot*. Here the cold self is supposed to be the preference that usually represents the agent, while the hot self is the one who makes the anomalous choices. Fudenberg and Levine (2005) adopt an explicitly dual self model for these dynamic choice problems and focus on the game between the selves rather than on the axiomatisation of a virtual dual self model, as we do here. Eliaz and Spiegler (2004) study contracting issues with several preference representations, including dual selves.[2]

The interpretation of anomalous choices as mistakes might be problematic, because the mistakes appear to be systematically in one direction (no one makes a mistake by

[2]In this context, also see Esteban and Miyagawa (2005a,b) and Esteban, Miyagawa and Shum (2003) who also have an example where a car could be tempting, but a different one from our paper.

consuming too little chocolate cake). The behaviour might instead indicate either cognitive limitations on the effects of the cake or a temporary loss of self-control. Nor is it wholly satisfactory to think of the problem of self-control as one of changing preferences, because the individual concerned might continue to prefer something, say good health to bad, while engaging in behaviour, say smoking, which seems antithetical to such preferences.

Also breaking the link between choice and preference has problems; for instance, how are we to talk about welfare if we cannot infer preferences from the choices that are made? Our main contribution in this paper is to point out that if the domain of choice is appropriately defined (as choice of feasible sets), then we can talk about decision makers who make choices over decision problems who behave *as if* the actual choice (from the feasible set) *may* be made by the alter-ego. Since the alter-ego supposedly only shows up in the presence of certain cues, we consider the decision-maker's preferences over decision problems where the alter-ego does not affect her (i.e. first-period menu choice problems), thus maintaining the link between choice and welfare.

In terms of formalism, our paper is closest to Dekel, Lipman and Rustichini (2001, 2005) (henceforth DLR and DLR05 respectively) and Gul and Pesendorfer (2001) (henceforth GP). We discuss these papers in more detail in §5. Samuelson and Swinkels (2004) explore the evolutionary foundations of temptation. They develop a model where endowing humans with utilities of menus that depend on unchosen alternatives is an optimal choice for nature from an evolutionary perspective.

Gul and Pesendorfer also emphasise the choice and welfare issue raised earlier in this section and the difficulties an explicitly multiple selves model might cause with respect to this issue. As an illustration of the importance of this point, we note that Bernheim and Rangel refer to addicts' description of past use of addictive substances as a mistake. However, verbal communication (from the addict) could be quite unreliable. What does suggest that substance (ab)use may indeed be considered a mistake by the abuser is reflected in the observation that agents notice their susceptibility to certain cues and anticipate making ex-ante inferior choices and act so as to *manage* their addiction in a sophisticated way. Thus, agents reveal that their choices (from a menu) may not always reflect their true preferences and act in order to constrain themselves suitably.[3]

The remainder of the paper is structured as follows. In §2, we illustrate the workings of our model by looking more closely at Schelling's patient (§2.1). We also discuss a simple bargaining model (§2.2) where the importance of the "timing of temptation" (i.e. the instant at which the agent feels tempted) is illustrated. In §3 we introduce our model, in §4 the axioms and our representation theorem and sketch the proof of the representation theorem in §4.2. We compare our axioms with those of Gul and Pesendorfer (2001) in §5.1 and with those of Dekel, Lipman and Rustichini (2001, 2005) in §5.2 and explore the workings of our model extended to the case of multiple exogenous states of the world in §5.3. §6 concludes and proofs are in the Appendix.

2 Examples

In this section, we consider some examples that illustrate the behavioural contrasts between our proposal and those advanced in the literature, notably the influential paper of GP.

2.1 Schelling's Patient

To recall, the example is of a woman who asks her doctor to refuse any pleas she might make for an anaesthetic to relieve her pain during delivery. The patient might not be aware of the extent of the possible pain, so any choice here is a lottery. Suppose that the possible choices are $\beta =$ *anaesthetic* and $\alpha =$ *no anaesthetic* and that the woman can also constrain her future choices by choosing to have the baby in a remote cottage without

[3]Bernheim and Rangel also have a discussion of American tourists in Britain who persist in looking left first when crossing; clearly by so doing they are not signalling their preference for a collision with a vehicle to no collision. But most tourists, realising this will happen, probably constrain themselves to wait for a "Walk" signal or use a zebra crossing in circumstances where they would cross the road without waiting in their home country. Thus, choices over these decision problems do, in fact, reflect their preferences.

access to painkillers and physicians. Then, assuming that other risks are kept constant between the choice of the hospital and the cottage, the hospital involves a menu $x_1 = \{\alpha, \beta\}$ and the cottage $x_2 = \{\alpha\}$. One would expect the patient to prefer x_2 if she feels she might be tempted by the availability of the painkiller in the hospital.[4]

We rationalise the preferences above by saying that the woman behaves *as if* she has an alter-ego who, she believes, will take over the decision-making in the hospital with some probability. We assume too that the untempted self (with utility function $u(\cdot)$) would prefer α and the alter-ego, i.e. the tempted self (with utility $v(\cdot)$) would prefer β, each "self" in the absence of the other.

In our specification, the untempted type, given she goes to the hospital, would always forgo the anaesthetic before she goes into delivery but with positive probability ask for it above some pain threshold during the actual procedure. [5]

Such first-period preferences are also consistent with the axioms of Gul and Pesendorfer (2001). We discuss their paper in more detail later on but, in brief, GP's patient would choose $\widehat{\beta} \in x$ to maximise $u(\widehat{\beta}) - c(\widehat{\beta}, x)$ but would incur self-control costs of $c(\widehat{\beta}, x) := \max_{\beta' \in x} v(\beta') - v(\widehat{\beta})$. (Recall that $\max_{\beta' \in x} v(\beta') = v(\beta)$.) Thus, if she goes to the hospital, her utility would be $\max_{\widehat{\beta} \in x_1} \left\{ u(\widehat{\beta}) + v(\widehat{\beta}) \right\} - v(\beta)$. If the first term were maximised by α, the patient would ask the doctor to forgo an anaesthetic and *knows that she will not change her mind during childbirth*, but would get a lower utility from the menu x_1 than from x_2 where there are no self-control costs. If the first term were maximised by β, then the menu x_1 would yield the utility $u(\beta)$ and x_2 would yield $u(\alpha) + v(\alpha) - v(\alpha) = u(\alpha)$. In other words, the patient has a very high cost of self-control

[4]Our colleague Sophie Bade has pointed out that there is an echo of Bernheim-Rangel here, in that someone addicted to painkillers might, for long-run health reasons, want to avoid being tempted by them.

[5]The patient's instructions to the doctor could be interpreted as an expression of the preferences given by the $u(\cdot)$ utility function and asking for the anaesthetic later the similar expression of the $v(\cdot)$ utility function. We could also think of the instructions prior to entering the hospital as creating another menu item β', where now there is some probability that the doctor would refuse the anaesthetic even if it were to be asked for. Such a menu item would replace β and therefore make the hospital more preferred for the long-run type than the presence of β would entail.

and gives in to her temptation.

In GP, either could happen but only one will and the decision-maker knows which outcome will occur. If both x_1 and x_2 are available both GP and our model would predict the choice of x_2.

The only difference between our paper and GP, in this example, is therefore in implied second-period behaviour. If the woman chooses the hospital, she will be tempted *with some probability* in our framework.

2.2 A Bargaining Application

In this application, we shall demonstrate the kinds of behaviour that are likely to emerge if the agent has self-control problems. From a technical perspective, we shall also show how our model can be adapted to a situation where there might be exogenous states, an issue addressed in §5.3.

Suppose the decision-maker wants to buy a car. There is a dealer who stocks two kinds of cars, the sedate type B and the souped-up coupé type C. Let us describe the environment in detail.

Seller. The seller chooses the types of cars available on display. His actions lead to a probability distribution over the menus $\{B, C\}$, $\{B\}$ and $\{C\}$. Let the actions be so that the probability of $\{B\}$ be $1 - q$. The cars are both worth 0 to him.

Decision-maker. He values both types of cars equally at b. He has an outside option worth 0. He also has an alter-ego who values the type B car at $\widehat{b}$ and the type C car at c. His outside option is $\widehat{c}$ is C is not present and $\overline{c}$ if C is present on the lot. (The motivation for this is that the alter-ego has different costs of walking away if car C is present on the lot.) The probability of temptation is such the alter-ego makes the decision with probability ρ° if C is present on the lot and $\rho_\circ$ otherwise. (We shall assume that $c - \overline{c} \geqslant b - \widehat{b}$.)

The Mechanism. The agent chooses whether or not to go to the dealership. After he makes his choice and *before* he reaches the dealership, the lottery over menus is resolved for the seller. Each player then makes a take-it-or-leave-it offer with probability ½. Let us assume, for simplicity, that the seller knows whether he is bargaining with the alter-ego or with the decision-makers's long-run self.[6] Notice that while the seller may offer a single car, the set of prizes (and hence the menu) is still infinite as the decision-maker also cares about the price he pays for the car.

Bargaining If C is present on the lot and the alter ego appears (an event with joint probability $q\rho^\circ$), the price will be

$$p_{c,\overline{c}} = \begin{cases} c - \overline{c} & \text{if seller makes offer;} \\ 0 & \text{otherwise.} \end{cases}$$

Here we assume that the alter ego chooses the coupe if it is present; he will compare the utilities from the two cars if both are present and the prices charged by the seller in that eventuality will make him indifferent between the two models and his outside option. The decision-maker's utility is then

$$\tfrac{1}{2}(b - (c - \overline{c})) + \tfrac{1}{2}b = b - \tfrac{1}{2}(c - \overline{c}).$$

If C is not present on the lot and the alter ego appears (an event of probability $(1-q)\rho_\circ$), a similar calculation gives the decision-maker's utility as

$$b - \tfrac{1}{2}(b - \widehat{c}).$$

Note that $\overline{c} < \widehat{c} \leqslant 0$ is a likely valuation of the outside options in that the tempted self faces both the temptation of a souped up car as well as being tempted to pay a higher price for both types of cars.

If the alter ego does not appear (an event with probability $[1-(1-q)\rho_\circ - q\rho^\circ]$), the decision-maker's utility will be $\tfrac{1}{2}b$. Thus the decision-maker's expected utility from visiting the dealership will be

$$U^* = \tfrac{1}{2}[1-(1-q)\rho_\circ - q\rho^\circ]\, b + (1-q)\rho_\circ(\tfrac{1}{2}b + \widehat{c}) + q\rho^\circ(b - \tfrac{1}{2}(c - \overline{c})).$$

[6]This may be because the alter-ego is unable to hide his enthusiasm for the coupé or displays otherwise benign signs which are noticed by the astute dealer.

The value of q will be chosen in equilibrium by the dealer in order to make it worthwhile for the decision-maker to come to the dealership, even with the prospect of being tempted by the coupé deterring this choice. In other words, the dealer will ensure that

$$U^* = q\,[\rho^\circ(b - ½(c - \overline{c})) - \rho_\circ\widehat{c}] + (½b + \rho_\circ\widehat{c}) \geqslant 0.$$

3 The Model

We have in mind a decision-maker who faces a two-period decision problem. In the first period, the agent chooses the set of alternatives from which a consumption choice will be made in the second period. Nevertheless, as in Kreps (1979), DLR and GP, we shall only look at first period choices. Let us now describe the ingredients more formally. (The basic objects of analysis are exactly the same as GP.)

The set of all prizes is Z where (Z, d) is a compact metric space. The space of probability measures on Z is denoted by Δ (with generic elements being denoted by $\alpha, \beta, \ldots$) and is endowed with the topology of weak convergence. This topology is metrisable and we let d_p be a metric which generates this topology. As in GP, the objects of analysis are subsets of Δ. Let $\mathscr{A}$ be the set of all closed subsets of Δ (with generic elements, called *menus*, denoted by $x, y, \ldots$) endowed with the Hausdorff metric

$$d_h(x,y) := \max\left\{\max_x \min_y d_p(\alpha,\beta), \max_y \min_x d_p(\alpha,\beta)\right\}.$$

Convex combinations of elements $x, y \in \mathscr{A}$ is defined as follows. We let $\lambda x + (1-\lambda)y := \{\gamma = \lambda\alpha + (1-\lambda)\beta : \alpha \in x, \beta \in y\}$ where $\lambda \in [0,1]$. (This is the so-called Minkowski sum of sets.) We are interested in binary relations $\succcurlyeq$ which are subsets of $\mathscr{A} \times \mathscr{A}$. (In the sequel, read $A \to B$ as A implies B unless the arrow is a limit. In either case, it should be clear from the context what the intended arrow denotes and no confusion should arise.)

Before we impose axioms on $\succcurlyeq$, it may be worthwhile to dwell on the implications of the model. The use of subsets of lotteries over Z as the domain for preferences instead of subsets of Z itself was first initiated by DLR in this context and is reminiscent of the approach pioneered by Anscombe and Aumann (1963). From a normative point of view, this approach should not be troublesome as long as our decision-makers are able to conceive of the lotteries they consume and agree with the axioms we impose on them. But from a revealed preference perspective, are decision-makers faced with menus of lotteries? As noted by Kreps (1988, pp. 101) (in the context of the Anscombe-Aumann theory), if decision-makers are not faced with choices of lotteries, our assumption that they are can be quite burdensome, especially from a descriptive point of view.

Nevertheless, it could be argued that such menus of lotteries are, in fact, objects of choice. A patient who chooses to go to a hospital (in Schelling's example) is, arguably, choosing a menu of lotteries with the level of pain being an uncontrolled random event. Similarly, a seafood fancier who goes to a restaurant not knowing the quality of the shrimp he is about to get, is doing the same. It is also possible that the menu of lotteries could arise from a non-degenerate mixed strategy played by an opponent, for instance in determining the set of objects available for sale by, say, a car dealer. There is, of course, the analytical benefit of our approach, which is the use of the additional structure a linear space provides. (Prime examples of this are Anscombe-Aumann and DLR.)

4 Axioms and Representations

We impose the following axioms on our preferences.

Axiom 1 (Preferences) $\succcurlyeq$ *is a complete and transitive binary relation.*

Axiom 2 (Continuity) *The sets* $\{y : y \succcurlyeq x\}$ *and* $\{y : x \succcurlyeq y\}$ *are closed.*

Axiom 3 (Independence) $x \succ y$ *and* $\lambda \in (0,1]$ *implies* $\lambda x + (1-\lambda)z \succ \lambda y + (1-\lambda)z$.

The first axiom is standard. Axiom 2 is a continuity requirement in the Hausdorff topology. The motivation for *Independence* is the familiar one and some normative arguments in its favour are given in DLR and GP. It basically says that our decision-maker does not distinguish between simple and compound lotteries and all that matters to her are the prizes. Nevertheless, as noted by Fudenberg and Levine (2005), this may not be an innocuous assumption.

Let us define

$$B(x) := \bigcap_{\beta \in x} \{\alpha \in x : \{\alpha\} \succcurlyeq \{\beta\}\}$$

and

$$W(x) := \bigcap_{\beta \in x} \{\alpha \in x : \{\beta\} \succcurlyeq \{\alpha\}\}$$

to be the sets of best and worst elements respectively in x. In light of *Continuity* and the compactness of x, it follows that both $B(x)$ and $W(x)$ are well defined. Our next axiom captures the essence of temptation.

Axiom 4 (Temptation) *For all* x*,* $B(x) \succcurlyeq x \succcurlyeq W(x)$*.*

Temptation says that insofar as the presence of alternatives different from the best alternative in the menu affects the decision-maker, it does not make the decision-maker worse off than his worst choice in the menu. Note that this implicitly rules out any role for flexibility. (If it is the case that for some $\alpha, \beta \in \Delta$, $\{\alpha, \beta\} \succ \{\alpha\}, \{\beta\}$, the decision-maker can be said to have a preference for flexibility. *Temptation* rules this out.) But it also says that the cost of temptation (i.e. the cost of not being able to choose the best alternative) is bounded. In particular, it rules out situations like Sen's rational donkey, which starves because it is unable to make a choice between two equally acceptable alternatives. In other words, it is never the case that "analysis is paralysis."

A decision-maker who faces no temptation would simply pick the best lottery in any menu. Our decision-makers however do not always do so. Consider three prizes, broccoli (b), rich chocolate cake (c) and deep-fried Mars bars (m). Let us suppose the decision-maker has the following preferences over the prizes in the morning: $\{b\} \succ \{c\} \succ \{m\}$. A "standard" decision-maker would always pick her most preferred alternative in any menu she encounters in the afternoon. But we are interested in decision-makers who are not immune to temptations. Suppose the decision-maker has preferences over the following menus: $\{b\} \succ \{b,c\} \succcurlyeq \{c\}$. We can then conclude that the presence of c in the menu, which makes the decision-maker worse off, is the source of temptation. (Formally, let $\{\beta\}$ be any lottery. Say that $\{\alpha\}$ *tempts* $\{\beta\}$ if $\{\beta\}$ is superior to $\{\alpha\}$ and the addition of $\{\alpha\}$ to $\{\beta\}$ makes the agent strictly worse off, i.e. $\{\beta\} \succ \{\alpha,\beta\} \succcurlyeq \{\alpha\}$.)

Now, also suppose $\{b\} \succ \{b,m\} \succcurlyeq \{m\}$. It is then reasonable to expect that a menu which consists of b and a lottery over c and m also makes the decision-maker worse off as compared to the menu which consists only of b. In other words, it is reasonable to expect that for all $\lambda \in [0,1]$, $\{b\} \succ \{b, \lambda c + (1-\lambda)m\}$. This is reflected in our next axiom.

Axiom 5 (Regularity) *$\{\beta\} \succ \{\beta,\alpha_1\}$ and $\{\beta\} \succ \{\beta,\alpha_2\}$ implies $\{\beta\} \succ \{\beta, \lambda\alpha_1 + (1-\lambda)\alpha_2\}$ for all $\lambda \in [0,1]$.*

Our next axiom is an excision axiom in that it allows us to *excise* elements from a menu without affecting the value of the menu to the decision-maker. Let us say that $\beta \in x$ is *untempted in* x if there exists $\alpha' \in x$ such that $\{\beta\} \succ \{\alpha'\}$ and for all $\alpha \in x$, $\{\beta\} \not\succ \{\beta,\alpha\}$. Consider once again the three prizes, broccoli (b), rich chocolate cake (c) and deep-fried Mars bars (m). As before, our decision-maker has the following preferences over the prizes in the morning: $\{b\} \succ \{c\} \succ \{m\}$ and both m and c tempt b, i.e. $\{b\} \succ \{b,c\}, \{b,m\}$. Thus, the presence of c and m make the decision-maker strictly worse off. Now, suppose that adding m to the menu $\{c\}$ does not affect the decision-maker, i.e. $\{c\} \sim \{c,m\}$. This implies that the "real" temptation comes from the rich chocolate cake and the addition of m to the menu $\{b,c\}$ should leave the agent indifferent, (i.e. $\{b,c\} \sim \{b,c,m\}$). Our next axiom formalises this idea.

Axiom 6 (AoM: Additivity of Menus) *For x,y finite, if $\beta \in x \cup y$ is untempted in $x \cup y$, then*

$$\{\beta\} \sim \{\beta\} \cup y \longleftrightarrow \{\beta\} \cup x \cup y \sim \{\beta\} \cup x.$$

We now define the linear functionals relevant to a dual self representation. As is standard, we shall say that $U : \mathscr{A} \to \mathbb{R}$ is *linear* if $U(\lambda x + (1-\lambda)y) = \lambda U(x) + (1-\lambda)U(y)$ for all $x, y \in \mathscr{A}$ and $\lambda \in (0,1)$ and that it *represents* $\succcurlyeq$ if it is the case that $U(x) \geqslant U(y)$ if and only if $x \succcurlyeq y$. The functions $u, v : \Delta \to \mathbb{R}$ are linear if similar conditions hold. Let $B_v(x) = \arg\max_{\beta \in x} v(\beta)$ be the set of v-maximisers in x (with a similar definition for B_u). Let $\beta^*_x \in B_u(x)$ and let $\widehat{\beta}_x \in B_u(B_v(x))$.

For any menu x, we shall say that the decision-maker, when confronted with a choice from the menu x, gets "tempted" with probability ρ_x. If ρ is to be consistent with linearity of U, then it must be the case that for all $\lambda \in (0,1)$,

$$\rho_{\lambda x + (1-\lambda) y} = \frac{\lambda \rho_x \delta_x + (1-\lambda)\rho_y \delta_y}{\lambda \delta_x + (1-\lambda)\delta_y} \tag{$\clubsuit$}$$

where $\delta_x := u(\beta^*_x) - u(\widehat{\beta}_x)$ and $\delta_y := u(\beta^*_y) - u(\widehat{\beta}_y)$. (The expression follows from the linearity of u and v and the observation that if β^*_x (resp. $\widehat{\beta}_x$) maximises u (resp. v) over x and if β^*_y (resp. $\widehat{\beta}_y$) maximises u (resp. v) over y, then $\lambda\beta^*_x + (1-\lambda)\beta^*_y$ (resp. $\lambda\widehat{\beta}_x + (1-\lambda)\widehat{\beta}_y$) maximises u (resp. v) over $\lambda x + (1-\lambda)y$. Let us first formalise this suggestive terminology.

Definition 4.0.1. *A linear functional $U : \mathscr{A} \to \mathbb{R}$ admits a* ***dual self*** *representation if there exist continuous linear functionals $u, v : \Delta \to \mathbb{R}$ unique up to affine transformations with $u := U|_\Delta$, a correspondence $\varphi : \mathscr{A} \twoheadrightarrow [0,1]$ that admits a selection ρ so that*

$$U(x) = (1-\rho_x)\max_{\beta \in x} u(\beta) + \rho_x \max_{\beta \in B_v(x)} v(\beta).$$

Moreover, (i) $\varphi(x) \subset (0,1]$ is a singleton if $B_u(x) \cap B_v(x) = \varnothing$, (ii) $\rho_x = \rho_{x+c}$ for all signed measures c such that $c(\Delta) = 0$ and $x + c \in \mathscr{A}$ and x such that $\varphi(x)$ is a singleton, (iii) ρ is consistent with the linearity of U and (iv) $\rho_x = \rho_{x \cap \left\{\alpha : \{\alpha\} \succcurlyeq \{\widehat{\beta}\}\right\}}$ where $\widehat{\beta} \in B_u(B_v(x))$.

The dual self theorem below says that when faced with choices of menus, the decision-maker who satisfies Axioms 1–6 behaves *as if* he has an alter-ego who has a utility function over lotteries given by v. Moreover, this alter-ego chooses the lottery in his most-preferred set (in x) which maximises the decision-maker's utility. Also, the decision-maker behaves *as if* he will be tempted (i.e. the probability that the choice will be made by the alter-ego) with a probability of ρ_x when faced with the menu x.

It should be emphasised that ρ_x and v are subjective and hence unobservable. The only observables here are first period behaviour. Furthermore, the decision-maker behaves as if second period choice from a convex menu is from an extreme face of the menu. (That the decision-maker is indifferent between any menu and its convex hull is proved in Lemma A.0.2 in the Appendix.) Notice also that if ρ is to be linear, it must be constant. The reason for this is easy to see. Since both U and u are linear (a fact which follows from *Independence*) and U depends on u and ρ multiplicatively, it must be the case that either ρ is non-linear or it is constant. Moreover, it is demonstrated in proposition 4.0.5 below that if ρ is to be a continuous selection, it must be constant.

The definition also says that while the probability of temptation depends on the menu, this dependence is relative in the sense that the probability of getting tempted is translation invariant (i.e. $\rho_x = \rho_{x+c}$ for all signed measures c such that $x + c \in \mathscr{A}$ and $c(\Delta) = 0$). This is a due to the more general fact that *Independence* and *Continuity* ensures that the preferences themselves are translation invariant. This is made precise below.

Definition 4.0.2. *A binary relation $\succcurlyeq$ is* ***translation invariant*** *if $x \succcurlyeq y$ implies $x + c \succcurlyeq y + c$ for all signed measures c such that $c(\Delta) = 0$ and $x + c, y + c \in \mathscr{A}$.*

Lemma 4.0.3 (Translation Invariance). *Let $\succcurlyeq$ satisfy Axioms 1, 2 and 3. Then $\succcurlyeq$ is translation invariant.*

Proof. See appendix. □

But notice that the translation invariance of ρ only applies to menus where temptation is meaningful, i.e. menus where the decision-maker and his alter-ego do not share a common maximiser. In addition, if there are lotteries α and β such that α tempts β, then the alter-ego makes the choice with positive probability (for any menu where temptation is meaningful in the sense defined above) if his best choice differs from the best choice of the cold-self. Finally, as a consequence of *AoM*, we find that removing lotteries that give the alter-ego less utility than her best choice and are worse (for the decision-maker) than the alter-ego's choice leaves the value of the menu (to the decision-maker) unchanged. We now state the dual self theorem formally.

Theorem 4.0.4 (Dual Selves). *A binary relation $\succcurlyeq$ satisfies Axioms 1–6 if and only if there exists a continuous linear functional $U : \mathscr{A} \to \mathbb{R}$, unique up to affine transformation, that represents $\succcurlyeq$ and U admits a dual self representation.*

Proof. See appendix. □

We should point out that in the dual self representation above, ρ cannot be constant as this would violate the continuity of $\succcurlyeq$. We now show that if $\succcurlyeq$ is continuous, then there exists no continuous selection ρ.

Proposition 4.0.5. *Let ρ be a continuous selection. Then ρ is constant.*

Proof. Suppose ρ is a continuous selection and suppose it is not constant. Then, there exist menus x, y such that $\rho_x \neq \rho_y$. Then, for any singleton $\{\beta\}$, $\rho_{\lambda\{\beta\}+(1-\lambda)x} = \rho_x$ and $\rho_{\lambda\{\beta\}+(1-\lambda)y} = \rho_y$. (This follows from equation (♣) above.) But for any $\varepsilon > 0$, there exists λ large enough such that

$$d_h(\lambda\{\beta\} + (1-\lambda)x, \lambda\{\beta\} + (1-\lambda)y) < \varepsilon$$

but ρ_x and ρ_y remain just as far apart, contradicting the continuity of ρ. □

Thus, no matter what additional axioms we choose, we cannot have continuous ρ except if ρ is constant. Furthermore, if ρ is constant, then it must be the case that $\succcurlyeq$ is discontinuous. We examine the case of constant ρ in the next section.

4.1 Other Representations

The dual self representation allows many possibilities. Indeed, it encompasses the GP representation (see §5.1 below) and allows ρ to depend on arbitrarily many items in a menu. We can, nevertheless, say that there can be *no* continuous selection ρ. This was demonstrated in proposition 4.0.5. Moreover, the dual self theorem rules out situations where the decision maker has no self control, i.e. decision makers who have a dual self representation with $\rho_x = 1$ for all x. This is because such a decision maker's preferences are typically only upper-semicontinuous (and not continuous). In this section, we introduce another excision axiom which will give us this possibility and also the case where $\rho_x \in [0,1]$ is a constant for all x. Let us first weaken *Continuity* appropriately.

Axiom 2a: (Upper Semicontinuity) The sets $\{y \in \mathscr{A} : y \succcurlyeq x\}$ are closed.

Axiom 2b: (Lower von Neumann-Morgenstern Continuity) $x \succ y \succ z$ implies $\lambda x + (1-\lambda)z$ for some $\lambda \in (0,1)$.

Axiom 2c: (Lower Singleton Continuity) The sets $\{\alpha : \{\beta\} \succcurlyeq \{\alpha\}\}$ are closed.

Axioms 2a–c are identical to Axioms 2a–c in §3 of GP. They weaken *Continuity* just enough to enable us to have a linear utility representation that admits a dual self representation. Notice also that Axioms 2a–c are strictly weaker than Axiom 2.

We have already seen one kind of excision in *AoM*. Another kind of excision is the notion that the only items in a menu that matter to a decision-maker are the alternative he would have chosen were he not tempted and the item in the menu that causes him maximal temptation. For instance, suppose the decision-maker's preferences are as follows: $\{b\} \succ \{b,c\} \succ \{c\} \succ \{c,m\} \succ \{m\}$. Thus, although c tempts b, c itself is tempted by m. Then, whenever both are present, we will require that the decision-maker is unaffected by the presence of c. In other words, $\{b,c,m\} \sim \{b,m\}$. We shall formalise this below. Let us say that $\beta \in x$ is *tempted* if there exists $y \subset x$ such that $\{\beta\} \succ \{\beta\} \cup y$.

Axiom 7 (SoM: Separability of Menus) If x is finite and $\beta \notin B(x \cup \{\beta\})$ is tempted,

$$x \cup \{\beta\} \sim x.$$

SoM says that the only alternative that matters in a menu (other than the decision-maker's best alternative in the menu) is the object that is maximally tempting. We want to express the idea that *if* the agent succumbs to temptation, he will fall all the way and choose the most tempting alternative. This is a strong assumption, but it has the advantage of providing a lot of structure to the dual self representation.

Theorem 4.1.1. *A binary relation $\succcurlyeq$ satisfies Axioms 1, 2a–c, 3–7 if and only if there exists an upper semicontinuous linear functional $U : \mathscr{A} \to \mathbb{R}$, unique up to affine transformation, that represents $\succcurlyeq$ and admits a dual self representation wherein*

$$U(x) = (1-\rho)\max_{\beta\in x} u(\beta) + \rho \max_{\beta\in B_v(x)} u(\beta).$$

Proof. See Appendix B. □

Notice that as opposed to theorem 4.0.4, we have weakened here the axiom *Continuity* but also imposed an additional excision axiom *SoM*. A reasonable question to ask is, What can we say about the case where we only weaken *Continuity* but do not require *SoM*? In such a case, we can prove a weaker version of theorem 4.0.4 in that the dual self representation only holds for either finite menus or menus that are the convex hulls of finitely many points. Indeed, the proof of theorem 4.1.1 proceeds by first constructing such a representation. *SoM* then allows us to identify, for each menu, a two-element subset that is equivalent to the original menu. Using this identification enables us to establish the theorem for finite menus. *Upper Semicontinuity* then lets us extend the result to arbitrary menus.

4.2 Proof-sketch of Theorem 4.0.4

The "only if" part of the proof is straightforward and is omitted. Here, we only sketch the "if" part. The proof proceeds through a number of simple of arguments which we describe below.

1. *Representing* $\succcurlyeq$. An application of the mixture space theorem (lemma A.0.3) shows that *Preferences, Continuity* and *Independence* guarantee the existence of a continuous linear functional U unique up to affine transformation which represents $\succcurlyeq$. Also U restricted to singletons is continuous.

2. *The alter-ego's preferences.* For lotteries α, β such that $\{\beta\} \succ \{\beta, \alpha\} \succcurlyeq \{\alpha\}$, we stipulate that this must be because the alter-ego strictly prefers α to β. From *Regularity* we see that for each $\beta \in \Delta$, the set $\beta_+ := \{\alpha : \{\beta\} \succ \{\beta, \alpha\} \succcurlyeq \{\alpha\}\}$ is convex. Repeated application of *AoM* tells us that $\beta_- := \{\alpha : \{\beta\} \sim \{\beta, \alpha\} \succ \{\alpha\}\}$ is also convex. Thus, β_+ and β_- are disjoint, convex sets and there exists a linear functional $v_\beta)$ which separates them. Furthermore, $\alpha \in \beta_+$ if and only if $v_\beta(\alpha) > v_\beta(\beta)$, i.e. adding a lottery to $\{\beta\}$ makes the decision-maker worse off if and only if the "alter-ego" strictly prefers the new lottery to β. However, this linear functional depends on β. We show next that the same linear functional performs the separation action for all lotteries in the domain.

3. *Translation Invariance.* We say that U is *translation invariant* if $U(x) \geqslant U(y)$ if and only if $U(x+c) \geqslant U(y+c)$ for all signed measures c such that $c(\Delta) = 0$ and $x+c, y+c \in \mathscr{A}$. That U is translation invariant follows from *Continuity* and *Independence*. We use this property to show that there is an essentially unique linear functional which performs the separation described in the previous step for each lottery β. Thus, there exists a continuous linear functional which represents the alter-ego.

4. *Finite menus.* For any finite menu x, let $\beta_x^* \in x$ be such that $u(\beta_x^*) = \max_{\beta \in x} u(x)$ and let $\widehat{\beta}_x \in x$ be such that $u(\widehat{\beta}_x) = \max_{B_v(x)} u(\beta)$. *Temptation* and repeated application of *AoM* implies that $u(\beta_x^*) \geqslant U(x) \geqslant u(\widehat{\beta}_x)$.

5. *Arbitrary menus.* Using *Continuity*, we show that for all x, it is the case that $u(\beta_x^*) \geqslant U(x) \geqslant u(\widehat{\beta}_x)$ where β_x^* and $\widehat{\beta}_x$ are defined as above.

6. *Representation.* A simple application of the intermediate value theorem gives us the desired representation and ρ_x for each x.

5 Other Models of Temptation

5.1 The Requirement of *Set Betweenness*

Gul and Pesendorfer (2001) introduce a condition on preferences called *Set Betweenness* (SB). This requirement says that for all menus $x, y \in \mathscr{A}$,

$$x \succcurlyeq y \longrightarrow x \succcurlyeq x \cup y \succcurlyeq y. \tag{SB}$$

Their representation theorem (as it pertains to us) says that if $\succcurlyeq$ satisfies axioms 1, 2a–c, 3 and *Set Betweenness*, then the utility of a menu is given by a function U defined either as

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

or

$$U(x) = \max_{\beta \in B_v(x)} u(\beta)$$

where $U : \mathscr{A} \to \mathbb{R}$ is linear and upper-semicontinuous, $u, v : \Delta \to \mathbb{R}$ are continuous, linear functionals unique up to the same affine transformation and $u = U|_{\Delta}$. The first representation obtains if $\succcurlyeq$ is continuous and the second obtains if $\succcurlyeq$ is upper semicontinuous but not continuous. The second kind of representation described above is to account for the possibility of "overwhelming temptation" where the agent always succumbs to temptation. In our terms, this corresponds to the case where the alter-ego always makes the choice. But to better understand the representation, let us assume for the moment, that $U(x)$ is continuous. This immediately rules out the overwhelming temptation representation. Let us write

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

and interpret $c(\beta, x)$ to be the cost imposed by the temptation (or the alter-ego) whenever the most tempting item is not chosen. We can now rewrite

$$U(x) = \max_{\beta \in x} \{u(\beta) - c(\beta, x)\}$$

which says that the utility to the decision maker of a menu is determined (additively) by the utility of the best choice in the menu and the cost it imposes through its selection.

How does this compare with our representation? In other words, suppose the binary relation $\succcurlyeq$ satisfies Axioms 1, 2, 3 and *Set Betweenness*, does it satisfy our axioms? To see that the answer is affirmative, notice that if $\succcurlyeq$ satisfies Axioms 1, 2 and 3, then (by lemma A.0.3) there exists a continuous, linear functional $U(x)$ (unique up to affine transformation) which represents $\succcurlyeq$ and $u = U|_\Delta$ is continuous and linear.

Now consider lotteries α and β so that $\{\beta\} \succ \{\beta, \alpha\} \succcurlyeq \{\alpha\}$. In GP's model, this is possible if and only if $v(\alpha) > v(\beta)$. Thus, it must be that our alter-ego can be represented by v. From GP's main representation theorem (as described above), it is now routine to verify that

$$\max_{\beta \in x} u(x) \geqslant U(x) \geqslant \max_{\beta \in B_v(x)} u(\beta)$$

so that there exists some ρ_x such that

$$U(x) = (1 - \rho_x) \max_{\beta \in x} u(\beta) + \rho_x \max_{\beta \in B_v(x)} u(\beta).$$

The case where

$$U(x) = \max_{\beta \in B_v(x)} u(\beta)$$

corresponds to the case where $\rho_x = 1$, i.e. the decision-maker gets tempted with probability 1. Moreover, ρ also satisfies the other requirements of Theorem 4.0.4 so that $U(x)$ is consistent with a dual self representation.

Thus, GP's representation implies our representation. Put another way, a decision-maker who satisfies Axioms 1–3 and *Set Betweenness* also behaves as if he has an alter-ego who makes a choice with probability ρ_x when the menu chosen is x. Thus any differences between the two models can only be detected by looking at second period choice. The examples below shows there exist preferences which satisfy our axioms but do not satisfy *Set Betweenness*.

Example 5.1.1. Let Z be any compact metric space where $|Z| \geqslant 3$. Let $\alpha_i \in \Delta$ for $i = 1, 2, 3, 4$ so that $\frac{1}{2}(\alpha_1 + \alpha_2) = \frac{1}{2}(\alpha_3 + \alpha_4)$. Assume that the decision maker gets tempted with a constant probability $\rho = ½$. Let u be such that $u(\alpha_1) > u(\alpha_3) > u(\alpha_4) > u(\alpha_2)$ and v such that $v(\alpha_4) > v(\alpha_2) > v(\alpha_1) > v(\alpha_3)$. It follows from our representation theorem that if $x := \{\alpha_1, \alpha_2\}$ and $y := \{\alpha_3, \alpha_4\}$, then $U(x) = U(y)$, i.e. $x \sim y$. But $U(x \cup y) = u\,(½(\alpha_1 + \alpha_4)) > U(x)$ so $x \cup y \succ x \sim y$. Thus, our preferences do not satisfy *Set Betweenness* when there is a constant probability (not equal to 1) of the decision-maker getting tempted. ♦

The example above relies on ρ being constant so that preferences are only upper semi-continuous and not continuous. We now show there are also continuous preferences which satisfy our axioms (so ρ is not constant) and do not satisfy GP's axioms. We first begin with some preliminaries.

An obvious consequence of GP's representation theorem is that any menu x is equivalent to a certain two-element subset. For two-element sets $\{\alpha, \beta\}$ where α tempts β, it is straightforward to note that

$$U(\{\alpha, \beta\}) = \max_{\beta' \in x} \{u(\beta') + v(\beta')\} - v(\alpha).$$

Since the preferences also satisfy our axioms, we find that

$$\rho_{\{\alpha,\beta\}} = \min\left\{ \frac{v(\alpha) - v(\beta)}{u(\beta) - u(\alpha)}, 1 \right\},$$

so that

$$U(\{\alpha, \beta\}) = (1 - \rho_x) u(\beta) + \rho_x u(\alpha).$$

Thus, for general x,

$$U(x) = (1 - \rho_x) \max_{\beta \in x} u(\beta) + \rho_x \max_{\beta \in B_v(x)} u(\beta)$$

where $\rho_x = \rho_{\{\alpha,\beta\}}$ with $\beta \in B_{u+v}(x)$ and $\alpha \in B_v(x)$. Now suppose there exists another utility function which satisfies our *Set Betweenness* given by

$$\widetilde{U}(x) = \max_{\beta \in x} \{u(\beta) + kv(\beta)\} - \max_{\beta \in x} kv(\beta)$$

where $k \in [0,\infty)$. The only difference between U and $\widetilde{U}$ is the factor k. Note that for any α,β where $u(\beta) > u(\alpha)$, $U(\{\beta\}) > U(\{\beta,\alpha\})$ if and only if $\widetilde{U}(\{\beta\}) > \widetilde{U}(\{\beta,\alpha\})$.

From the discussion above, we can write

$$U(x) = (1-\rho_x)\max_{\beta\in x} u(\beta) + \rho_x \max_{\beta\in B_v(x)} u(\beta)$$

and

$$\widetilde{U}(x) = (1-\widetilde{\rho}_x)\max_{\beta\in x} u(\beta) + \widetilde{\rho}_x \max_{\beta\in B_v(x)} u(\beta).$$

Now it is easy to see that if $\overline{U} := \lambda U + (1-\lambda)\widetilde{U}$, then

$$\overline{U}(x) = (1-\overline{\rho}_x)\max_{\beta\in x} u(\beta) + \overline{\rho}_x \max_{\beta\in B_v(x)} u(\beta)$$

where $\overline{\rho}_x := \lambda\rho_x + (1-\lambda)\widetilde{\rho}_x$. But there is no reason to expect that $\overline{U}$ will, in general, satisfy *Set Betweenness*. Indeed, the example below from Dekel, Lipman and Rustichini (2005) shows that this indeed may be the case.

Example 5.1.2. Consider a weak-willed dieter who faces choices over broccoli (b) multiple temptations in the form of rich chocolate ice cream (c) and low-fat yogurt (y). A natural set of rankings over menu is:

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

Dekel, Lipman and Rustichini (2005) show that there is no GP representation which is consistent with the above ordering. In other words, any extension of the ordering above to the space of all menus would violate one of GP's axioms.

Nevertheless, there is a dual self representation that is consistent with the above ordering. To see this most transparently, let u and v be as following:

	b	c	y
u	6	0	4
v	0	8	6

so that utility over menus is given by

$$U(x) = ½ \max_{\beta \in x} u(\beta) + ½ \left\{ \max_{\beta \in x} [u(\beta) + v(\beta)] - \max_{\beta \in x} v(\beta) \right\}.$$

In other words, with probability ½, the dieter is a standard agent and with probability ½ he takes on GP type preferences. Note that $U(\{b,y\}) = 5 > 4 = U(\{y\})$ and $U(\{b,c,y\}) = 5 > 3 = U(\{b,c\})$. Let us consider preferences which satisfy Axioms 2 and 3 and *Set Betweenness*. Such preferences have a GP representation (u,v). Such preferences also have a dual self representation (u,v,ρ) where ρ is the probability of getting tempted. Then, in the example above, his preferences can also be given by

$$U(x) = \left(1 - \frac{\rho_x}{2}\right) \max_{\beta \in x} u(\beta) + \frac{\rho_x}{2} \max_{\beta \in B_v(x)} v(\beta),$$

which is a dual self representation that does not satisfy *Set Betweenness*. ♦

5.2 Temptation and Multiple States

In another recent paper, Dekel, Lipman and Rustichini (2005) consider generalisations of GP preferences. Their starting point is the subjective state space approach pioneered in DLR. DLR (and Dekel et al., 2005) show that when Z is a finite set of prizes, for any continuous preference $\succcurlyeq$ over menus satisfying Lipschitz continuity (in the Hausdorff topology) and *Independence*, there exists an essentially unique measure space S (which is the state space), a finite measure μ, functions $U : Z \times S \to \mathbb{R}$ so that $U(\beta,\cdot)$ is measurable with respect to S and $U(\cdot,s)$ is an expected utility function, and a continuous function V where, for each menu x,

$$V(x) = \int \max_{\beta \in x} U(\beta,s)\, \mathrm{d}\mu(s)$$

so that V represents $\succcurlyeq$. Note that μ is a signed measure and is not necessarily a probability measure. Indeed, because of the state dependent utility functions, it is not possible to pin down the signed measure μ. For a finite state space, their representation then becomes

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

We shall call the above a finite additive EU representation of the preference $\succcurlyeq$. An axiomatisation of such a (finite state) representation is provided by Dekel et al. (2005) where it is shown that a *Finiteness* axiom[7] is also necessary (and sufficient) in addition to *Independence* and *Continuity*. While we do not wish to formally describe *Finiteness*, we should point out that an implication of a finite state additive EU representation (which is the class of preferences considered in DLR05) is that there exists an integer $N > 0$ so that for all menus x, there exists a sub-menu $x' \subset x$ such that $x' \sim x$ and the cardinality of x' is at most N.

To see the representation more clearly, DLR05 make the following observation. If the decision-maker were to be uncertain about his future tastes (and not face any temptation), then one would expect the measure to be positive. But in such a case, one would also expect his preferences to satisfy *Monotonicity*, i.e. $x' \subset x$ implies $x \succcurlyeq x'$. They then show the following:

Theorem 5.2.1 (DLR05). *The preference $\succcurlyeq$ has a finite additive EU representation with a positive measure μ if and only if it satisfies* Continuity, Independence, Finiteness *and* Monotonicity.

Such a result is intuitive and is the appropriate generalisation of the result in Kreps (1979). The main question in DLR05 is to find the appropriate analogue of *Monotonicity* that captures the idea of temptation. An immediate axiom that one might use is *Non-monotonicity*, namely that there exists a menu x with $x' \subset x$ such that $x' \succ x$. Thus, there has to be some instance when the decision-maker does not value flexibility and therefore the measure μ cannot be positive. But DLR05 want to look at a smaller class of preferences. They want to look at the class of preferences where the decision-maker is certain about the preferences of his untempted self. The only uncertainty is about what form the temptation takes. A further generalisation that they consider is in the form of the temptation cost. This is described below.

To model the uncertainty in temptation, DLR05 look at different linear functions

[7] For a full description of this axiom, see Dekel, Lipman and Rustichini (2005) or Dekel et al. (2005).

$v_j : Z \to \mathbb{R}$ which give rise to different cost functions

$$c_i(\beta, x) := \sum_{j \in J_i} \max_{\beta' \in x} v_j(\beta') - \sum_{j \in J_i} v_j(\beta).$$

The key to uncertainty about temptation is then that the decision-maker is uncertain about which cost function he will be facing. Thus, the generalisation of GP which obtains, namely the *temptation representation*, can be written as

$$U(x) := \sum q_i \max_{\beta \in x} \{u(\beta) - c_i(\beta, x)\}$$

where $q_i > 0$ and $\sum q_i = 1$ which means that q_i can be interpreted as the probability that the decision-maker is faced with the ith temptation in the form of the cost function c_i. The function u is such that $u(\beta) := U(\{\beta\})$ and is called the *commitment utility*. (All the functions u and v_j are expected utility functions.)

To characterise such a preference, DLR05 introduce two more axioms. The first says that if the decision-maker could commit himself to a certain item in a menu, he would. This is made precise in the following axiom.

Axiom (DFC: Desire for Commitment) There exists $\alpha \in x$ such that $\{\alpha\} \succcurlyeq x$.

It is obvious that *Temptation* implies *DFC*. However, there exist temptation representations that satisfy *Temptation* but not admit a dual self representation. This is demonstrated in example 5.2.3 below. The second axiom introduced in DLR05 is as follows.

Axiom (Domination) [8] If there exists $\alpha \in B(x \cup \{\beta\})$ such that $\{\alpha\} \sim x$ and $\beta \notin B(x \cup \{\beta\})$, then there exists $\varepsilon > 0$ such that for all $\widehat{\beta} \in N_\varepsilon(\beta)$ and all x',

$$x' \cup x \succcurlyeq x' \cup x \cup \{\widehat{\beta}\}.$$

The main representation theorem in DLR05 is then the following:

[8] This axiom is from an earlier version of their paper. In the present version, it is called *AIC: Approximate Improvements are Chosen*.

Theorem 5.2.2. *A preference relation $\succcurlyeq$ has a temptation representation if and only if it has a finite state additive EU representation and satisfies* DFC *and* Dominance.

It is straightforward to show that our representation implies *Domination*. Therefore, it would seem that any *continuous* preference (in the Hausdorff topology) which has a dual self representation strictly belongs to the class of preferences identified in DLR05. But this is not the case, primarily because we do not have a finiteness axiom. But let us first look at a a couple of examples where the decision-maker satisfies the DLR05 axioms but does not have a dual self representation.

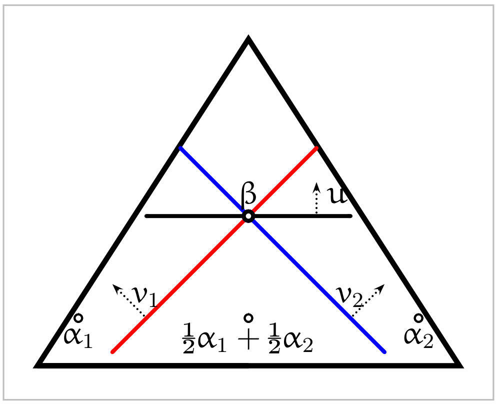
Example 5.2.3. Let Z be a finite set and let Δ be the space of lotteries over Z. Define utility function u, v_1 and v_2 so that

	u	v_1	v_2
α	0	2	2
β	0	1	6

Now let $c(\gamma, x) := \sum_j \left[\max_{\gamma' \in x} v_j(\gamma')\right] - \sum_j v_j(\gamma)$. Then $U(x) := \max_{\beta \in x}\left[u(\gamma) - c(\gamma, x)\right]$ is a temptation representation.

Let $x := \{\alpha, \beta\}$. Then $c(\alpha, x) = v_1(\alpha) + v_2(\beta) - v_1(\alpha) - v_2(\alpha) = 4$ and $c(\beta, x) = v_1(\alpha) - v_1(\beta) = 1$. This means that $U(x) = \max_{\gamma \in x}\{u(\gamma) - c(\gamma, x)\} = -1$. Thus, $\{\alpha\} \sim \{\beta\} \succ x$ which is not possible in a dual self representation. (Note that this example does not violate *Temptation* because $B(x) = x = W(x)$.) ♦

Notice that if in the temptation representation, each cost function depends on only one temptation i.e. each J_i is a singleton, such an example could not arise. In a dual self representation, the decision maker does not care about the utility level of the alter-ego. She only cares about the choice made by the alter-ego insofar as it affects her own utility level. If we were to interpret the different v_j's in a temptation representation as belonging to different selves who are the cause of the temptation, then we could say that a temptation representation is an interdependent utilities model, where the decision maker (or the "commitment self" in the terminology of DLR05) cares about the utility levels of the different selves. We shall now see another example of a temptation representation which does not admit a dual self representation. This time our axiom *Regularity* will be violated.



Example 5.2.4. Suppose the decision-maker has utility function u and is faced with two temptations denoted by expected utility functions v_1 and v_2. Also, suppose that v_j is not a convex combination of v_i and u for $i \neq j$. Then, such a configuration might look as in the figure above. But such a configuration would violate our axiom *Regularity*. The violation is because the lottery α_1 and α_2 which both tempt β, but there also exist lotteries over α_1 and α_2 which do not tempt β. ♦

It should be remarked that the construction above holds in any temptation representation where there is uncertainty about the form the temptation will take. We shall now show that there exist preferences which have dual self representations but do not satisfy finiteness. Recall that GP preferences of the following form

$$U_i(x) = \max_{\beta \in x}\{u(\beta) + kv(\beta)\} - \max_{\beta \in x} kv(\beta)$$

for $k \in [0, \infty)$, have a dual self representation with

$$\rho_x(k) = \min\left\{k\frac{v(\alpha) - v(\beta)}{u(\beta) - u(\alpha)}, 1\right\}$$

where $\beta \in B_{u+kv}(x)$ and $\alpha \in B_{kv}(x)$. Now, let (k_i) be a sequence such that $k_i \in [0, \infty)$ and the following holds: for $x := \bigcup_i \{\beta_i\} \cup \alpha$, $\beta_i \in B_{u+k_i v}(x)$, $\beta_j \in B_{u+k_i v}(x)$ implies $k_i = k_j$ and $\alpha \in B_v(x)$.

Let (λ_i) be another sequence such that $\lambda_i \in [0,1]$ for each i and $\sum_i \lambda_i = 1$. Now, it easy to see that

$$U(x) = (1-\rho_x)\max_{\beta\in x} u(\beta) + \rho_x \max_{\beta\in B_v(x)} u(\beta)$$

where $\rho_x = \sum_i \lambda_i \rho_x(k_i)$, is part of a dual self representation with ρ_x satisfying all the necessary conditions for this to be so. By choosing the k_i's and λ_i's appropriately, it is therefore possibly to construct dual self representations which do not satisfy finiteness. In other words, these utility functions are such that the utility of a menu can depend on an arbitrarily large (countable) sub-menu. This is because each $\rho(k_i)$ depends on $\{\beta_i, \alpha\}$ and we can make the set of β_i's arbitrarily large. Indeed, this is the spirit of example 5.1.2.

We can further generalise this construction (as suggested by Bart Lipman). Let μ be a Borel probability measure on $[0,1)$ and let $\rho_x = \int \rho_x(k)\, d\mu(k)$ so that we have a dual self representation which again does not satisfy finiteness.

5.3 Exogenous States of the World

Our representation admits a straightforward extension to finite exogenous states. This would be the formal equivalent of the model studied by Bernheim and Rangel (2004) limited to two periods. Formally, let S be a finite set of states with the probability that state $s \in S$ occurs being given by π_s. The state is realised after the decision-maker chooses the menu. We take this to be some set of exogenous circumstances that affect the agent only inasmuch as they affect the likelihood of his getting tempted. Note that the agent's utility function does not change across states nor does his alter-ego's. The only thing that changes is the probability of getting tempted. In particular, we are looking for a utility function (over menus) that looks like the following:

$$U(x) = \sum_{s\in S} \pi_s \left((1-\rho_x^s)\max_{\beta\in x} u(\beta) + \rho_x^s \max_{\beta\in B_v(x)} u(\beta) \right).$$

In particular, note that the probability of getting tempted can depend on *both the menu and the state.*

Example 5.3.1. Let $S := \{0, 1, \ldots, n\}$ and let ρ_i^x be the probability of getting tempted in the state of the world i. Then one specification could be the following:

$$\rho_x^i < \rho_x^{i+1}$$

for all x and $\rho_x^0 = 0$ for all x. If in addition we assumed $\rho_x^i = \rho_y^i$ for all $x, y \in \mathscr{A}$ for all $i \in S$, we would get the Bernheim-Rangel model. ♦

6 Conclusion

In this paper, we consider a decision-maker faced who has to decide on the set of feasible choices from which an actual choice will be made at a later point in time. We rule out the case where the decision-maker may prefer larger sets of feasible choices due to a preference for flexibility.

Our main contribution is to provide axioms on first period preferences that enable us to interpret this problem as a decision-maker who behaves as if he has an alter-ego (with preferences different from his own), who makes the actual choice from the menu with some probability. Doing so enables us to address problems where decision-makers demonstrate apparent dynamic inconsistency (i.e. make ex-post choices that are inferior from an ex-ante perspective) and make unambiguous welfare statements in these situations. We also relate our model to the influential papers of Gul and Pesendorfer (2001) and Dekel, Lipman and Rustichini (2005).

Possible extensions include more explicit study of dynamic applications. We intend to expand on the bargaining example discussed in brief in this paper, but this is left for a future paper.

A Proof of Theorem 4.0.4

The "only if" part of the proof is straightforward. Here we shall demonstrate the "if" part of the proof. We first begin with a crucial lemma which can also be found in DLR. The proof is presented here for expositional ease. Recall that for any set x, its convex hull is denoted by $\text{conv}(x)$.

Lemma A.0.2. *Let* $\succcurlyeq$ *satisfy* Independence. *Then for all finite* x, $x \sim \text{conv}(x)$. *Furthermore, if* $\succcurlyeq$ *satisfies* Continuity, *then* $x \sim \text{conv}(x)$ *for all* $x \in \mathscr{A}$.

Proof. We shall proceed through a series of simple arguments.

Step 0. Let us denote the standard $(n-1)$-simplex by $\Delta^{n-1} := \{p \in \mathbb{R}^n : \sum p_i = 1\}$ and its vertices by $\text{Vert}(\Delta^{n-1})$. It is straightforward to verify that for all $\lambda \in (0, 1/n)$, $\lambda \text{Vert}(\Delta^{n-1}) + (1-\lambda)\Delta^{n-1} = \Delta^{n-1}$. Now, let x be the vertices of the $(n-1)$-simplex $\text{conv}(x)$. In other words, $|x| = n$, $\dim(\text{conv}(x)) = n-1$ and $\text{conv}(x)$ is linearly isomorphic to Δ^{n-1}. Thus, for all $\lambda \in (0, 1/n)$, $\lambda x + (1-\lambda)\text{conv}(x) = \text{conv}(x)$.

Step 1. Let x now be any finite set. We know that $\text{conv}(x)$ has a simplicial decomposition, K such that $\text{Vert}(K) = x$ and each simplex has the same dimension as $\text{conv}(x)$. Now apply the result of Step 0 to each simplex to conclude that for any $\lambda \in \left(0, \frac{1}{\dim(\text{conv}(x))+1}\right)$, $\lambda x + (1-\lambda)\text{conv}(x) = \text{conv}(x)$.

Step 2. For any finite x, suppose $\text{conv}(x) \succ x$. Then, for any $\lambda \in \left(0, \frac{1}{\dim(\text{conv}(x))+1}\right)$, $\text{conv}(x) = \lambda \text{conv}(x) + (1-\lambda)\text{conv}(x) \succ \lambda x + (1-\lambda)\text{conv}(x) = \text{conv}(x)$ which is a contradiction (where the relation in the middle follows from Independence). Alternatively, suppose $x \succ \text{conv}(x)$. Then $\text{conv}(x) = \lambda x + (1-\lambda)\text{conv}(x) \succ \lambda \text{conv}(x) + (1-\lambda)\text{conv}(x) = \text{conv}(x)$ which is again a contradiction. Thus, $x \sim \text{conv}(x)$.

Step 3. Let x be an arbitrary closed set. Then (because x is compact), there exists a sequence (x_k) of finite sets such that $x_k \subset x$ and $x_k \to x$ (in the Hausdorff metric). Thus, $\text{conv}(x_k) \to \text{conv}(x)$. From the continuity of $\succcurlyeq$, we see that $x \sim \text{conv}(x)$. □

We shall now show that there exists a continuous linear functional that represents preferences. (This is Proposition 2 in DLR.) Recall that $\mathscr{A}$ is the space of all closed subsets of Δ.

Lemma A.0.3. *If $\succcurlyeq$ satisfies Axioms 1, 2 and 3, then there exists a continuous linear functional $U : \mathscr{A} \to \mathbb{R}$ that represents $\succcurlyeq$. Furthermore, U is unique up to affine transformations.*

Proof. Let $X \subset \mathscr{A}$ be the space of all closed *convex* subsets of Δ. Notice that X endowed with the Minkowski sum is a mixture space. It only remains to verify the mixture space axioms (see Kreps, 1988, page 52). By assumption, *Independence* holds. *Continuity* ensures that vN-M continuity is also satisfied. Thus, by the mixture space theorem, there exists a linear functional $V : X \to \mathbb{R}$ so that for all $x, y \in X$, $V(x) \geqslant V(y)$ if and only if $x \succcurlyeq y$.

We now extend V to all menus. Let us define $U : \mathscr{A} \to \mathbb{R}$ as follows: for all $x \in \mathscr{A}$, let $U(x) := V(\text{conv}(x))$. It is easily seen that U represents $\succcurlyeq$. All that remains to be shown is that U is linear.

From lemma A.0.2, it follows that $\lambda x + (1-\lambda)y \sim \text{conv}\,(\lambda x + (1-\lambda)y)$. Also $x \sim \text{conv}(x)$ and $y \sim \text{conv}(y)$. From *Independence* it follows that $\lambda x + (1-\lambda)y \sim \lambda\text{conv}(x) + (1-\lambda)y$ and $\lambda\text{conv}(x) + (1-\lambda)y \sim \lambda\text{conv}(x) + (1-\lambda)\text{conv}(y)$, i.e. $\lambda x + (1-\lambda)y \sim \lambda\text{conv}(x) + (1-\lambda)\text{conv}(y)$. Therefore,

$$\begin{aligned} U(\lambda x + (1-\lambda)y) &= U(\lambda\text{conv}(x) + (1-\lambda)\text{conv}(y)) \\ &= V(\lambda\text{conv}(x) + (1-\lambda)\text{conv}(y)) \\ &= \lambda V(\text{conv}(x)) + (1-\lambda)V(\text{conv}(y)) \\ &= \lambda U(x) + (1-\lambda)U(y). \end{aligned}$$

□

Note the heavy use of *Independence* in the proof. We are identifying the mixture which gives x with probability μ and y with probability $1-\mu$ with the convex combination of x and y, $\mu x + (1-\mu)y$, an identification which lies at the heart of the mixture space theorem. Let us define $u(\alpha) := U(\{\alpha\})$ and interpret it to be the decision-maker's utility from a lottery (in the untempted state). It is clear that u is a continuous, linear function. Another important property of preferences that we shall make us of is translation invariance. This is made precise below.

Definition A.0.4. *A binary relation $\succcurlyeq$ is **translation invariant** if $x \succcurlyeq y$ implies $x+c \succcurlyeq y+c$ for all signed measures c such that $c(\Delta)=0$ and $x+c, y+c \in \mathscr{A}$.*

Lemma A.0.5 (Translation Invariance). *Let $\succcurlyeq$ satisfy Axioms 1 – 3. Then $\succcurlyeq$ is translation invariant.*

Proof. Let $x \succcurlyeq y$ and c such that $c(\Delta)=0$ and $x+c, y+c \in \mathscr{A}$. Simple geometry shows that for all $\lambda \in (0,1)$,

$$\lambda y+(1-\lambda)(x+c)=\lambda\{\lambda y+(1-\lambda)(y+c)\}+(1-\lambda)\{\lambda x+(1-\lambda)(x+c)\}.$$

Then, since $\succcurlyeq$ is reflexive,

$$\lambda y+(1-\lambda)(x+c)\sim\lambda\{\lambda y+(1-\lambda)(y+c)\}+(1-\lambda)\{\lambda x+(1-\lambda)(x+c)\}.$$

From *Independence* we get

$$\lambda x+(1-\lambda)(x+c)\succcurlyeq\lambda y+(1-\lambda)(x+c).$$

Combining the relations above

$$\lambda x+(1-\lambda)(x+c)\succcurlyeq\lambda\{\lambda y+(1-\lambda)(y+c)\}+(1-\lambda)\{\lambda x+(1-\lambda)(x+c)\}.$$

From *Independence* we see that

$$\lambda x+(1-\lambda)(x+c)\succcurlyeq\lambda y+(1-\lambda)(y+c).$$

But from lemma A.0.3 above, there exists a continuous linear functional U that represents $\succcurlyeq$. Thus,

$$U(\lambda x+(1-\lambda)(x+c))\geqslant U(\lambda y+(1-\lambda)(y+c)).$$

Using the linearity of U and rearranging terms gives us for each $\lambda \in (0,1)$,

$$\lambda[U(x)-U(y)]+(1-\lambda)[U(x+c)-U(y+c)]\geqslant 0. \tag{$\bigstar$}$$

Now suppose by way of contradiction, $y+c \succ x+c$, i.e. $U(y+c) > U(x+c)$. This would mean that there exists some $\lambda \in (0,1)$ which does not satisfy ($\bigstar$) yielding the desired contradiction. □

Lemma A.0.6. *Suppose Axioms 1–6 hold. Then there exists a continuous, linear functional* $v : \Delta \to \mathbb{R}$ *such that (i)* $\{\beta\} \succ \{\alpha, \beta\} \succcurlyeq \{\alpha\}$ *if and only if* $v(\beta) < v(\alpha)$ *and (ii) for all* x*, there exists* $\widehat{\beta}_x \in x$ *such that* $U(x) \geqslant u(\widehat{\beta}_x)$ *and* $u(\widehat{\beta}_x) = \max_{\beta \in B_v(x)} u(\beta)$.

Proof. See §A.1 below. □

Now, there exists $\beta^* \in x$ so that $u(\beta^*_x) \geqslant U(x)$ from where we can determine ρ_x using the Intermediate Value Theorem, which completes the proof. The translation invariance of ρ follows from the translation invariance of U (see lemma A.0.5 above) and the other properties of ρ are also easily obtained.

A.1 The Alter-ego's Preferences

In this section, we shall construct the alter-ego's preferences via some revealed preference arguments thereby providing a proof of lemma A.0.6.

Let us define $\beta_+ := \{\alpha : \{\beta\} \succ \{\beta, \alpha\} \succcurlyeq \{\alpha\}\}$. From *Regularity*, it follows that β_+ is convex. Let us also define $\beta_- := \{\alpha : \{\beta, \alpha\} \sim \{\beta\} \succ \{\alpha\}\}$. The lemma below shows that β_- is also convex.

Lemma A.1.1. *Suppose* $\succcurlyeq$ *satisfies Axioms 1,3 and 4. Then,* β_- *is convex.*

Proof. Let $\alpha_1, \alpha_2 \in \beta_-$. By *Independence* and $\{\beta\} \sim \{\beta, \alpha_2\}$,

$$\{\beta\} \sim \lambda\{\beta\} + (1-\lambda)\{\beta, \alpha_2\}.$$

Independence and $\{\beta\} \sim \{\beta, \alpha_1\}$ also implies

$$\lambda\{\beta\} + (1-\lambda)\{\beta, \alpha_2\} \sim \lambda\{\beta, \alpha_1\} + (1-\lambda)\{\beta, \alpha_2\}.$$

Transitivity of $\succcurlyeq$ implies

$$\{\beta\} \sim \lambda\{\beta, \alpha_1\} + (1-\lambda)\{\beta, \alpha_2\}.$$

But note that

$$\lambda\{\beta, \alpha_1\} + (1-\lambda)\{\beta, \alpha_2\} = \{\beta, \lambda\alpha_1 + (1-\lambda)\beta, \lambda\beta + (1-\lambda)\alpha_2, \lambda\alpha_1 + (1-\lambda)\alpha_2\}.$$

Applying *AoM* twice, we find $\{\beta\} \sim \{\beta, \lambda\alpha_1 + (1-\lambda)\alpha_2\}$. Since $\{\beta\} \succ \{\alpha_1\}$, *Independence* gives us $\{\beta\} \succ (1-\lambda)\{\beta\} + \lambda\{\alpha_1\}$. Also, $\{\beta\} \succ \{\alpha_2\}$ and *Independence* implies $(1-\lambda)\{\beta\} + \lambda\{\alpha_1\} \succ \lambda\{\alpha_1\} + (1-\lambda)\{\alpha_2\}$. By the transitivity of $\succ$, $\{\beta\} \succ \{\lambda\alpha_1 + (1-\lambda)\alpha_2\}$. Thus, β_- is convex. □

Let us recall some definitions of objects in linear spaces. An *affine subspace* (or linear variety) of a vector space is a translation of a subspace. A *hyperplane* is a *maximal* proper affine subspace. If H is a hyperplane in a vector space $\mathbb{V}$, then there is a linear functional f on $\mathbb{V}$ and a constant c such $H = \{x : f(x) = c\}$. Moreover, if H is closed if and only if f is continuous (Luenberger, 1969, pp. 129, 130). For notational ease, we shall write H as $[f = c]$. Similarly, (two of) the negative and positive half spaces are represented as $[f \leqslant c]$ and $[f > c]$ respectively. For any subset $S \subset \mathbb{V}$, let $\text{aff}(S)$ denote the affine subspace generated by S, i.e. the smallest affine subspace that contains S. Also, let $\text{ri}\, C$ denote the relative interior of a convex set C in (a normed vector space) $\mathbb{V}$.

Lemma A.1.2. *Let $\beta^* \in \text{ri}\,\Delta$. Then there exists $v : \Delta \to \mathbb{R}$ which is continuous and linear so that $\beta^*_- \subset [v \leqslant v(\beta^*)]$ and $\beta^*_+ \subset [v > v(\beta^*)]$.*

Proof. Recall that space of all signed (countably additive) measures on (Z, d) is a Banach space $\mathbb{V}$ with the total variation norm (see, for instance, Aliprantis and Border, 1999, pp. 360). Notice that $\Delta \subset \mathbb{V}$. Indeed, the minimal subspace of $\mathbb{V}$ that contains Δ is $\mathbb{V}$. Thus, $\text{aff}(\Delta)$ is a hyperplane. Let $\text{aff}(\Delta) = x_0 + \mathbb{M}$ where $\mathbb{M}$ is a subspace of $\mathbb{V}$. (If $\beta^*_+ = \varnothing$, let $v = u = U|_\Delta$. Henceforth, we shall assume β^*_+ is not empty.)

Now, $(\beta^*_- - x_0) \cap (\beta^*_+ - x_0) = \varnothing$. Furthermore, $\beta^*_+ - x_0$ contains an interior point (in the topology relative to $\mathbb{M}$). (To see this, suppose not. Then, there exists $\alpha \in \beta_+ - x_0$ and a sequence (α_n) such that $\alpha_n \to \alpha$ and $\alpha_n \in \beta^*_- - x_0$. That is $\{\beta^*\} \sim \{\beta^*, \alpha_n\}$ for all n and $\{\beta\} \succ \{\beta, \alpha\}$. But this contradicts *Continuity*. Indeed, it contradicts the upper semicontinuity of the preferences.) It is also straightforward to verify that $\beta^*_- - x_0$ contains no interior points of $\beta^*_+ - x_0$. Thus, by the Separating Hyperplane Theorem (Luenberger, 1969, pp. 133), there exists a closed hyperplane that separates $\beta^*_- - x_0$ and $\beta^*_+ - x_0$. Moreover, $\beta^* - x_0 \in H$. Let f be a linear functional on $\mathbb{M}$ such that $H = [f = c]$

for some constant c and $\beta^*_+ - x_0 \subset [f > c]$. From *Continuity* (more specifically, from the upper semicontinuity of preferences) it follows that $(\beta^*_- - x_0) \cap H \neq \varnothing$. Now define $v(\alpha) := f(\alpha - x_0)$. Thus, v is linear on Δ and continuous. Furthermore, $\beta^*_- \subset [v \leqslant v(\beta^*)]$, $\beta^* \cap [v = v(\beta^*)] \neq \varnothing$ and $\beta^*_+ \subset [v > v(\beta^*)]$. □

We have thus far established that for some β^*, there exists a continuous linear functional v that represents the alter-ego's preferences at that point. We will now show that there is a single continuous, linear functional which represents the alter-ego's preferences over the entire domain. (We shall use *Translation Invariance* towards this end.) Note that for $\beta \in \text{ri}\,\Delta$, there exists $\varepsilon > 0$ such that $N_\varepsilon(\beta) \subset \text{ri}\,\Delta$. Also recall a fact about the Hausdorff metric, d_h. For all $\lambda \in [0,1]$,

$$d_h(\{\beta\}, \{\beta, \lambda\alpha + (1-\lambda)\beta\}) = (1-\lambda)d(\{\alpha\}, \{\beta\}).$$

Lemma A.1.3. *For all $\beta \in \Delta$, $[v = v(\beta)]$ separates β_- and β_+.*

Proof. Suppose not. Then there exists $\beta \in \Delta$ such that either

(i) $\exists\ \alpha \in \beta_-$ such that $v(\alpha) \geqslant v(\beta)$, or

(ii) $\exists\ \alpha \in \beta_+$ such that $v(\beta) \geqslant v(\alpha)$.

Let us consider the first possibility.

Let $c = \beta^* - \beta$. Since $\beta^* \in \text{ri}\,\Delta$, there exists $\varepsilon > 0$ such that $N_\varepsilon(\beta^*) \subset \text{ri}\,\Delta$. From *Independence*, we can assume $\alpha \in N_\varepsilon(\beta)$. Thus, $\alpha + c \in N_\varepsilon(\beta^*)$. Since v is continuous and linear, $v(\alpha + c) > v(\beta + c) = v(\beta^*)$. This implies that $\{\beta + c\} \succ \{\beta + c, \alpha + c\}$ which, by *Translation Invariance*[9], is equivalent to $\{\beta\} \succ \{\beta, \alpha\}$ which is a contradiction of the hypothesis that $\alpha \in \beta_-$.

The second possibility is taken care of with a similar argument, thus establishing the desired result. □

Lemma A.1.4. *For all finite x*

$$\max_{\beta \in x} u(\beta) \geqslant U(x) \geqslant \max_{\beta \in B_v(x)} u(\beta).$$

[9]Notice that we only require *Translation Invariance* to hold for two-element subsets.

Proof. Let $\beta^* \in x$ such that $u(\beta^*) = \max_{\beta \in x} u(\beta)$ and let $\widehat{\beta} \in x$ such that $u(\widehat{\beta}) = \max_{\beta \in B_v(x)} u(\beta)$. Let $x' := \{\alpha \in x : u(\alpha) \geqslant u(\widehat{\beta})\}$ and $y := \{\alpha \in x : u(\alpha) < u(\widehat{\beta})\}$. Then, $x = x' \cup y$ and by *Temptation*, $\beta^* \succcurlyeq x' \succcurlyeq \widehat{\beta}$. Let $y := \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. It then follows that for each $\alpha_i \in y$, $\{\beta\} \sim \{\beta, \alpha_i\}$. By *AoM*, it follows that $\{\beta\} \sim \{\beta, \alpha_1, \alpha_2\}$. Repeatedly applying *AoM* implies $\{\beta\} \sim \{\beta\} \cup y$. Once again applying *AoM* implies $\{\beta\} \cup x' \cup y \sim \{\beta\} \cup x' = x'$. Thus, $x \sim x'$ and $\{\beta\} \succcurlyeq x \succcurlyeq \{\widehat{\beta}\}$. □

Lemma A.1.5. *For any x,*

$$\max_{\beta \in x} u(\beta) \geqslant U(x) \geqslant \max_{\beta \in B_v(x)} u(\beta).$$

Proof. Let $\beta^* \in x$ such that $u(\beta^*) = \max_{\beta \in x} u(\beta)$ and let $\widehat{\beta} \in x$ such that $u(\widehat{\beta}) = \max_{\beta \in B_v(x)} u(\beta)$. Let (x_k) be a sequence of finite sets where $|x_k| = k$, $x_k \subset x$, $x_k \to x$ and $\beta^*, \widehat{\beta} \in x_k$ for each k.

By *Temptation*, $u(\beta^*) \geqslant U(x_k)$ for each k. Hence, by *Continuity*, $u(\beta^*) \geqslant U(x)$. Also, for each k, $U(x_k) \geqslant u(\widehat{\beta})$. Once again, *Continuity* implies that $U(x) \geqslant u(\widehat{\beta})$. This gives us the desired result. □

B Proof of Theorem 4.1.1

We shall prove the theorem for finite menus. Towards this end, we show that there exists a linear functional that represents preferences over essentially finite menus (which are defined below). Also, for finite menus x, $x \sim \text{conv}(x)$ and *Translation Invariance* holds. *SoM* is then used to derive the representation for finite menus. A straightforward continuity argument then extends the result to arbitrary menus.

Let us call a set $x \subset \Delta$ *essentially finite* if x is finite or if x is the convex hull of a finite set. Let us denote by $\mathscr{A}_0$, the space of all essentially finite menus, i.e. the space of all essentially finite subsets of Δ. Also define $X_0 \subset X$ as the space of all closed convex subsets of Δ so that for each $x \in X_0$, there exists a finite set x_0 such that $x = \text{conv}(x_0)$. In other words X_0 consists of all closed convex sets that are the convex hulls of finite sets. Before we begin, recall that lemma A.0.2 shows that for each finite x, $x \sim \text{conv}(x)$.

Lemma B.0.6. *Let $\succcurlyeq$ satisfy Axioms 1, 2a–c and 3. Then there exists an upper-semicontinuous linear functional $U : \mathscr{A}_0 \to \mathbb{R}$ such that for all $x, y \in \mathscr{A}_0$, $x \succcurlyeq y$ if and only if $U(x) \geqslant U(y)$. Also, U is unique up to affine transformation.*

Proof. The proof is the same as that of lemma A.0.3. We shall only provide a sketch here. Notice that X_0 is a mixture space. Since *Independence* and von Neumann-Morgenstern continuity also hold, there exists a $V : X_0 \to \mathbb{R}$, unique up to affine transformation, so that for all $x, y \in X_0$, $x \succcurlyeq y$ if and only if $V(x) \geqslant V(y)$. Now define $U(x) := V(\text{conv}(x))$ for all $x \in \mathscr{A}_0$. Linearity of U is demonstrated as in lemma A.0.3. Upper-semicontinuity follows from Axiom 2a. □

Another property that we shall establish for essentially finite menus is translation invariance.

Lemma B.0.7. *Let $\succcurlyeq$ satisfy Axioms 1, 2a–c and 3. Then for all $x, y \in \mathscr{A}_0$, $x \succcurlyeq y$ implies $x + c \succcurlyeq y + c$ for all signed measures c such that $c(\Delta) = 0$ and $x + c, y + c \in \mathscr{A}_0$.*

Proof. For all $x \in \mathscr{A}_0$, if c is a signed measure such that $c(\Delta) = 0$ and $x + c \subset \Delta$, then $x + c \in \mathscr{A}_0$. The proof is now similar to the proof of lemma A.0.5. Notice that the proof of lemma A.0.5 only relied on the existence of linear functional that represented preferences. From lemma B.0.6, such a functional exists which gives us the desired result. □

We shall now construct the alter-ego's preferences. The construction in §A.1 goes through without change. Recall that in the construction of v, we only used the *Upper Semicontinuity* of preferences and *Translation Invariance* for two-element subsets. Repeated application of *AoM* as in §A.1 gives us the following lemma.

Lemma B.0.8. *Suppose Axioms 1–6 hold. Then there exists a continuous, linear functional $v : \Delta \to \mathbb{R}$ such that (i) $\{\beta\} \succ \{\alpha, \beta\} \succcurlyeq \{\alpha\}$ if and only if $v(\beta) < v(\alpha)$ and (ii) for all $x \in \mathscr{A}_0$, there exists $\widehat{\beta}_x \in x$ such that $U(x) \geqslant u(\widehat{\beta}_x)$ and $u(\widehat{\beta}_x) = \max_{\beta \in B_v(x)} u(\beta)$.*

It follows from *Temptation* that for all $x \in \mathscr{A}_0$, $\max_{\beta \in x} u(\beta) \geqslant U(x) \geqslant u(\widehat{\beta}_x)$. The dual self representation follows immediately giving us ρ_x for each x. The properties of ρ which are required for it to be part of a dual self representation are easily verified.

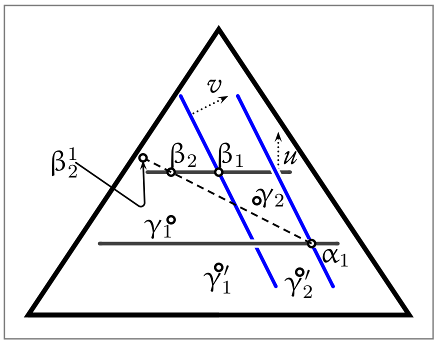
We now prove a simple lemma which shows that we can restrict attention to essentially finite menus that lie entirely in the relative interior of Δ. (In what follows, we shall denote the ε-neighbourhood of a point $\beta \in \Delta$ by $N_\varepsilon(\beta)$ and the diameter of a set x by $\operatorname{diam}(x)$. We shall also repeatedly use the fact that ρ must be consistent with the linearity of U, i.e. (♣) holds.)

Lemma B.0.9. *For all $y \in \mathscr{A}_0$ and for all $\bar{\varepsilon} > 0$, there exists $x \in \mathscr{A}_0$ so that $x \subset \operatorname{ri}\Delta$, $\rho_y = \rho_x$ and* $\operatorname{diam}(x) < \bar{\varepsilon}$.

Proof. Let y be a menu and let $\widehat{\beta} \in B_u(y)$ be an extreme point of $\operatorname{conv} y$. Also, let $\bar{\varepsilon} > 0$. Then, for all $\lambda \in (0,1)$, $\rho_{\lambda\{\widehat{\beta}\}+(1-\lambda)y} = \rho_y$. Moreover, for all $\varepsilon > 0$, there exists $\lambda_\varepsilon \in (0,1)$ such that $\lambda_\varepsilon\{\widehat{\beta}\} + (1-\lambda)y \subset N_\varepsilon(\widehat{\beta})$. Let us now take $\varepsilon \in (0, \bar{\varepsilon}/2)$ so that for some $\beta^* \in \operatorname{ri}\Delta$, $N_\varepsilon(\beta^*) \subset \operatorname{ri}\Delta$. Let $c := \beta^* - \widehat{\beta}$ be a signed measure so that $c(\Delta) = 0$. By the translation invariance property of ρ, it follows that for $x := \lambda_\varepsilon\{\widehat{\beta}\} + (1-\lambda_\varepsilon)y + c$, $\rho_x = \rho_{\lambda_\varepsilon\{\widehat{\beta}\}+(1-\lambda)y} = \rho_y$. □

Lemma B.0.10. *Let $\succcurlyeq$ have a dual self representation and satisfy* SoM. *Then, for all finite x, for any $\beta \in B_u(x)$ and for any $\alpha \in B_u(B_v(x))$, $x \sim \{\beta, \alpha\}$.*

Proof. Let $\widehat{x} := \{\beta_1, \dots, \beta_m\} \cup x' \cup \{\alpha_1, \dots, \alpha_n\} \cup y$ where $\beta_i \in B_u(\widehat{x})$ for $i = 1, \dots, m$, $\alpha_j \in B_u(B_v(\widehat{x}))$ for $j = 1, \dots, n$, $u(\beta_1) > u(\gamma) > u(\alpha_1)$ for all $\gamma \in x'$ and $u(\alpha_1) > u(\gamma')$ for all $\gamma' \in y$. (Note that by definition, $v(\alpha_1) > v(\beta_i)$ and $v(\alpha_1) \geqslant v(\gamma')$ for all $\gamma' \in y$.)



Since α_1 is untempted in $\widehat{x}$ (which means, among other things, that $\{\alpha_1\} \sim \{\alpha_1\} \cup y$), by *AoM*, $\widehat{x} \sim \widehat{x} \setminus y$. Also, by *AoM*, $\widehat{x} \sim \{\beta_1, \dots, \beta_m\} \cup x' \cup \{\alpha_1\}$. By *SoM*, $\widehat{x} \sim \{\beta_1, \dots, \beta_m\} \cup \{\alpha_1\}$. Let $x := \{\beta_1, \dots, \beta_m\} \cup \{\alpha_1\}$. From lemma B.0.9, we can assume, without loss of generality, that $x \subset \operatorname{ri}\Delta$.

Let (β_i^k) be a sequence in $\operatorname{ri}\Delta$ such that $\{\beta_i^k\} \succ \{\beta_i^{k+1}\} \succ \{\beta_i\}$ and $\beta_i^k \in \operatorname{conv}\left(\{\beta_i^1, \alpha_1\}\right)$ for each k and $\lim_{k\to\infty} \beta_i^k = \beta_i$. Since $x \subset \operatorname{ri}\Delta$, it is clear that such a sequence always exists.

Let $x_i^k = \{\beta_i^k, \alpha_1\}$. By *SoM*, it follows that $\{\beta_1, \dots, \beta_i^k, \dots, \beta_m\} \cup \{\alpha_1\} \sim x_i^k$. Furthermore, $x_i^k = \lambda_k x_i^1 + (1-\lambda_k)\{\alpha_1\}$ for some

$\lambda_k \in (0,1)$ and $\lambda_k > \lambda_{k+1}$. Therefore, $\rho_{x_i^k} = \rho_{x_i^1}$. This implies that $x_i^k \succ x_i^{k+1}$. By *Upper-Semicontinuity*, it now follows that

$$x = \lim_k \{\beta_1, \ldots, \beta_i^k, \ldots, \beta_m\} \cup \{\alpha_1\} \sim \lim_k x_i^k = x_i$$

which gives us the desired result. □

Lemma B.0.11. *Let $\succcurlyeq$ have a dual self representation and satisfy* SoM. *Then for all β, α, α' such that $\{\beta\} \succ \{\alpha\} \sim \{\alpha'\}$ and α, α' tempt β, it is the case that $\{\beta, \alpha\} \sim \{\beta, \alpha'\}$. Hence, $\rho_{\{\beta,\alpha\}} = \rho_{\{\beta,\alpha'\}}$.*

Proof. By lemma B.0.9, we can assume that $\beta \in \text{ri}\,\Delta$, so there exists $\varepsilon > 0$ such that $N_\varepsilon(\beta) \subset \text{ri}\,\Delta$. We can also assume that $\alpha, \alpha' \in N_{\varepsilon/4}(\beta)$.

Let $v(\beta) < v(\alpha) < v(\alpha')$ and let $c := \alpha' - \alpha$, $c' := \alpha - \beta$. By hypothesis, α tempts β, so that $\alpha + c = \alpha'$ tempts $\beta + c = \beta'$. By *Translation Invariance*, $\{\beta, \alpha\} \sim \{\beta', \alpha'\}$. Also, $\{\beta\} \sim \{\beta'\}$.

To see this, suppose the contrary, i.e. suppose $\{\beta\} \nsim \{\beta'\}$. By definition, $\beta' = \beta + c = \beta + (\alpha' - \alpha)$. By *Translation Invariance*, $\{\beta\} + c' \nsim \{\beta'\} + c'$, i.e. $\{\beta\} + (\alpha - \beta) \nsim \{\beta\} + (\alpha' - \alpha) + (\alpha - \beta)$ which is equivalent to $\{\alpha\} \nsim \{\alpha'\}$ which contradicts the hypothesis.

Hence β, β', α' and α form the vertices of a parallelogram. By Lemma B.0.10, $\{\beta, \alpha'\} \sim \{\beta, \beta', \alpha'\} \sim \{\beta', \alpha'\}$. This proves that $\{\beta, \alpha\} \sim \{\beta, \alpha'\}$. Since $\{\alpha\} \sim \{\alpha'\}$, it follows from the representation that $\rho_{\{\beta,\alpha\}} = \rho_{\{\beta,\alpha'\}}$. □

Proof of Theorem 4.1.1 for finite x. From lemma B.0.10, it follows that for any x, there exist elements $\beta, \alpha \in x$ such that $\{\beta, \alpha\} \sim x$. Therefore, we can restrict attention to two element subsets. Let $x = \{\beta, \alpha\}$ and $y = \{\beta', \alpha'\}$ where $x, y \in \text{ri}\,\Delta$, $\varepsilon = \text{diam}(x) \geqslant \text{diam}(y) > 0$ and $N_\varepsilon(\beta) \subset \text{ri}\,\Delta$.

Let $c := \beta - \beta'$. Then, $y + c \subset N_\varepsilon(\beta)$ and $\rho_y \sim \rho_{y+c}$. If $u(\alpha) > u(\alpha' + c)$, then there exists $\lambda \in (0,1)$ so that $u(\lambda\beta + (1-\lambda)(\alpha' + c)) = u(\alpha)$. Appealing to lemma B.0.11 now gives us the desired result. (The case where $u(\alpha) \leqslant u(\alpha' + c)$ is dealt with in a similar fashion.) □

We can now prove Theorem 4.1.1 for arbitrary menus.

Proof of Theorem 4.1.1. Let $x \in \mathscr{A}$, $\beta \in B_u(x)$ and $\alpha \in B_u(B_v(x))$. If $\beta \in B_u(B_v(x))$, then by *Temptation*, we are done. Let us assume this isn't the case.

Consider a sequence (x_k) such that for each k, $x_k \in \mathscr{A}_0$, $x_k \subset x$, $|x_k| < |x_{k+1}|$ and $\lim_k x_k = x$. Define $\alpha_k := \lambda_k \beta + (1-\lambda_k)\alpha$ for $\lambda_k \in (0,1)$. We will also require that for each k, $\beta \in x_k$ and $\alpha_k \in B_u(B_v(x_k))$. Then, $x_k \sim \{\beta, \alpha_k\}$ and $U(x_k) = \rho U(\{\beta\}) + (1-\rho)U(\{\alpha_k\})$.

Now, $\lim_k \{\beta, \alpha_k\} = \{\beta, \alpha\}$. Also, for each k, $\{\beta, \alpha_k\} \succ \{\beta, \alpha_{k+1}\}$, i.e. $x_k \succ x_{k+1}$. From *Upper Semicontinuity*, it follows that $U(x) = \lim_k U(x_k) = \lim_k U(\{\beta, \alpha_k\}) = U(\{\beta, \alpha\})$. □

References

Ahn, David (2005). "Ambiguity without a State Space," mimeo, University of California, Berkeley.

Aliprantis, Charalambos D. and Kim C. Border (1999). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 2nd ed., Springer, New York.

Anscombe, Frank J. and Robert J. Aumann (1963). "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, **34**, 199–205.

Bernheim, B. Douglas and Antonio Rangel (2004). "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, **94(5)**, 1558–1590.

Dekel, Eddie, Barton L. Lipman and Aldo Rustichini (2001). "Representing Preferences with a Unique Subjective State Space," *Econometrica*, **69(4)**, 891–934.

Dekel, Eddie, Barton L. Lipman and Aldo Rustichini (2005). "Temptation-Driven Preferences," mimeo, Boston Unversity.

Dekel, Eddie, Barton L. Lipman, Aldo Rustichini and Todd Sarver (2005). "Representing Preferences with a Unique Subjective State Space: Corrigendum," working paper.

Esteban, Susanna and Eiichi Miyagawa (2005), "Optimal Menu of Menus with Self-Control Preferences," mimeo, Pennsylvania State University.

Esteban, Susanna and Eiichi Miyagawa (2005), "Temptation, Self-Control, and Competitive Nonlinear Pricing," forthcoming, *Economics Letters*.

Esteban, Susanna, Eiichi Miyagawa and Matthew Shum (2003), " Nonlinear Pricing with Self-Control Preferences," working paper.

Eliaz, Kfir and Rani Spiegler (2004). "Contracting with Diversely Naïve Agents," mimeo, New York University.

Fudenberg, Drew and David K. Levine (2005). "A Dual Self Model of Impulse Control," working paper, UCLA.

Gauthier, David (1987). *Morals by Agreement*, Oxford University Press.

Gul, Faruk and Wolfgang Pesendorfer (2001). "Temptation and Self-Control," *Econometrica*, **69**, 1403–1435.

Kreps, David M. (1979). "A Representation Theorem for 'Preference for Flexibility'," *Econometrica*, **47**, 565–576.

Kreps, David M. (1988). *Notes on the Theory of Choice,* Westview Press, Co., U.S.A.

Luenberger, David G. (1969). *Optimization by Vector Space Methods*, Wiley, New York.

Parfit, Derek (1986). *Reasons and Persons,* Oxford University Press.

Rabin, Matthew (1998). "Psychology and Economics," *Journal of Economic Literature*, **36**, 11–46.

Samuelson, Larry and Jeroen Swinkels (2004). "Information, Evolution and Utility," mimeo, University of Wisconsin, Madison.

Schelling, Thomas C. (1978). "Egonomics, or The Art of Self-management," *American Economic Review*, **68(2)**, 290–294.

Schelling, Thomas C. (1984). "Self-command in Practice, in Policy and in a Theory of Rational Choice," (Richard T. Ely lecture), *American Economic Review*, **74(2)**, 1–11.