The Price of Flexibility: Towards a Theory of Thinking Aversion*

Pietro Ortoleva[†]

Job Market Paper

Abstract

The goal of this paper is to model an agent who dislikes large choice sets because of the "cost of thinking" involved in choosing from them. We take as a primitive a preference relation over lotteries of menus and impose novel axioms that allow us to separately identify the genuine preference over the content of menus, and the cost of choosing from them. Using this, we formally define the notion of *thinking aversion*, much in line with the definitions of risk or ambiguity aversion. We represent such preference as the difference between a monotone and affine evaluation of the content of the set and an anticipated thinking cost function that assigns to each set a thinking cost. We further extend this characterization to the case of monotonicity of the genuine rank and introduce a measure of comparative thinking aversion. Finally, we propose behavioral axioms that guarantee that the cost of thinking can be represented as the sum of the cost to *find* the optimal choice in a set and the cost to find out *which* is the optimal choice.

JEL classification: D81, D83, D84.

Keywords: Cost of Thinking, Contemplation Cost, Bounded Rationality, Preference Over Menus, Preference for Flexibility, Choice overload

*This version: December 2008. I would like to thank Ozgur Evren, Paolo Ghirardato, Alessandro Lizzeri, Massimo Marinacci, Leandro Nascimento, David Pearce, Debraj Ray, Anja Sautmann, Ennio Stacchetti, Gil Riella, Todd Sarver, the participants at seminars at NYU and the SED 2008 conference in Ann Arbor, and especially Efe Ok for useful comments and suggestions. Needless to say, all mistakes are mine.

[†]*Department of Economics, New York University* 19 W. 4th St., New York, 10012 NY. Email: *pietro.ortoleva@nyu.edu*.

Nothing is more difficult [...] than to be able to decide.
Napoléon Bonaparte

1. Introduction

1.1 Basic Idea: Thinking Aversion

Consider an individual who wants to buy a cell phone and can choose between providers A, B and C. The providers offer the same coverage, the same selection of phones, *etc.*, but different calling plans. Provider A offers three plans, B offers these three plans and three additional ones, and C offers not only these six but a total of 40 plans. Our agent appreciates the flexibility to pick a plan that better suits her needs, and consequently prefers provider B to A. However, she might also prefer B to C, despite C's larger selection. This might happen because C offers *too many* options: the agent is afraid of the cost involved in choosing the best plan in such a large set. She might therefore prefer to settle for B, which still offers a "good" selection without requiring her to exert too much effort in choosing.

The behavior of this agent is clearly incompatible with the standard paradigm in choice, *the more options the better*. In particular, our agent faces a tradeoff: on the one hand, she wants *more options* so that she will more likely find what's best for her; on the other hand, she wants *fewer options* since big sets make the decision process more costly. The first goal of this paper is to define rigorously the presence of such tradeoff: we call it *Thinking Aversion*. Then, we characterize this behavior axiomatically.

This problem could be analyzed in two ways. First, in a standard choice theoretic study, one could look at what an agent chooses when confronted with different sets, and look for violations of standard rationality (the WARP) due to the presence of a disutility from thinking.[1] Alternatively, one could take "a step back," and analyze the preference of an agent over menus: the preferences over the sets *she will choose from at a later stage*. This route has many advantages, and here in particular it allows us to observe if an agent dislikes sets that require a lot of "thinking" - the behavior that we are after. This is the route that we follow. (This could be seen as a first step to study this problem in general, which hopefully will leave us better suited for a future analysis of the choice behavior).

In particular, by looking at preferences over menus, we study the preferences of an agent who acts in two stages: first, at stage zero, she ranks the menus (and this is the ranking that we observe and analyze); then, at stage one, she chooses *from* a menu. In both stages the agent has to perform some "thinking." However, we are only interested

[1]See, for example, Cherepanov, Feddersen, and Sandroni (2008).

in the one involved with the choice *from* menus - call it stage-1-thinking. In particular, our analysis is meant to capture and analyze how the presence of a stage-1-thinking affects the way the agent ranks menus at time 0.[2]

1.2 Empirical evidence and other explanations

Our model is essentially motivated by introspection. At the same time, a number of studies in psychology and economics document how the presence of a large number of options might induce a disutility to individuals and affect their behavior. For psychology see, for example, Schwartz (2005). Within economics, the experiment in Salgado (2006) directly tests the existence of preferences like the ones we are after: subjects are given a large set (50) of lotteries to choose from, but before making a choice they can ask the computer to *randomly* select a subset of 5 lotteries, which replaces the original one and from which subjects will then make their choice. Notably, subjects are not shown the smaller set: they are simply told that it will consist of 5 elements randomly selected by the computer. (Accepting the subset is therefore risky.) In the experiment 48% of the subjects opt for this option. This shows that there is a sizable proportion of agents who are inclined to avoid complex choices and are willing to ask the computer to simplify their decision task, even if this means facing a potentially much worse set of alternatives. The experiment also tests the case in which the original set consists of only 25 lotteries: the percentage goes down to 32%, in line with the interpretation that the complexity of choice is lower when there are fewer alternatives and therefore fewer agents rely on the computer to simplify their choices.

Moreover, a strong empirical evidence suggest that agents tend to avoid choosing, or to choose the default option when confronted with larger or complicated sets - a phenomenon dubbed *choice overload*. This is documented in a variety of settings in papers like Tversky and Shafir (1992), Iyengar and Lepper (2000), Iyengar, Huberman, and Jiang (2004), Iyengar and Kamenica (2007).[3] For example, in Iyengar and Lepper (2000) the authors present the results of a field experiment about the purchase of jams in a gourmet grocery store in California. As customers would pass in front of a tasting

[2]Our analysis, however, is *not* able to capture the thinking that the agent had to sustain to choose *between* menus. This happens because we observe the preferences over menus that are the *outcome* of this thinking process, and therefore incorporate it. And since what we observe are well-behaved, transitive preferences, there will be no way to disentangle it. To do so, we would have to take another step back, and look at preferences over menus of menus, and so on. Instead, we focus on the point in time in which our agent has her well-formed preferences over menus, and study these to learn about the thinking at time 1 and our agent's attitude towards it.

[3]Kamenica (2008) suggests one equilibrium-based explanation for this phenomenon in a product differentiation model. He shows that if there are informational asymmetries, consumers can infer which good is optimal for them from the product line that is offered, and that consumer surplus is greater when there are fewer options. Consequently, fewer consumers will buy when the set of options is larger.

booth set up by the experimenters, they encountered a selection of either 6 or 24 jams. Their main finding is that only 3% of the customers who approached the booth did actually purchase a jam in the large selection case, against 30% in the small selection case. Other examples include the study of pattern of choice of the 401(k) plan, where similar behaviors are shown.

These experiments seem to support the relevance of the “disutility of thinking” to behavior. At the same time, however, we need to make sure that other, more standard, approaches cannot explain it. In particular, two candidate alternative explanations come to mind: the presence of some informational content in the set selection; or fear of regret over having made a wrong choice. (Later, we will present an axiomatic structure that will make this distinction more formal.)

First, one might argue that an agent prefers a smaller set because there is informational value in what is included in this smaller set. For example, she might prefer to go to a restaurant with a shorter wine list since she believes - correctly or not - that it is the outcome of a selection by an expert, conveying therefore some valuable information. Let us make two remarks. First, in most of the cases we are interested in, there seem to be no (relevant) informational value in the smaller sets - think about the mobile phones example. The same seems to be true for the experiments in the choice overload literature, like the cited one about jams, and certainly for the behavior in the experiment in Salgado (2006), where the subset is chosen *randomly* by a computer, and therefore has no informational value. Second, notice that a standard agent would not have a *strict* preference for a smaller set, even if it had informational value. This happens because in our setup the agent *sees* the smaller set at the time of choice between menus, and therefore she is exposed to the information and can incorporate it. A standard agent would then be indifferent between the larger and the smaller set, since she can focus on the elements of the smaller set inside the larger one.

The regret argument, suggested by Sarver (2007), assumes that a larger choice set gives the agent more “opportunity to be sorry” about their selection later on. Anticipating this, she might want to avoid it by restricting her own choice. Introspection suggests once again that, although possibly connected, our explanation is well distinguished from this one. Moreover, in most of the choices that we are trying to explain agents would never find out what the right choice was, making it harder to suggest that the main motivation is anticipated regret, and the same seems to be the case for the cited cases of choice overload. This is further confirmed with a direct test in the experiment in Salgado (2006), where it is shown that the behavior is essentially the same when subjects are given feedback about what was the best lottery (and are told beforehand), and when they are *not* given such feedback (and know that they will not be).

1.3 Related Theoretical Literature

In recent years a large number of papers have extended the realm of traditional decision theory to the framework of preferences over menus, and have obtained rigorous models of phenomena like preference for flexibility (Kreps (1979)), temptation and self control (Gul and Pesendorfer (2001), Dekel et al. (2007a)), regret (Sarver (2007)), or a potential combination of these elements (Dekel, Lipman, and Rustichini (2001), henceforth DLR01, Dekel et al. (2007b), henceforth DLRS). Some of these papers offer different explanations as to why an agent might prefer a *smaller* set: she might want to avoid the presence of a tempting item, or to feel regret had she made the "wrong" choice. The present paper fits into this literature as we analyze a different reason why an agent should prefer a smaller set: because she wishes to avoid the "cost of thinking" involved in the choice from a large one.

The idea that agents might have a cost of thinking is not new to economics, and neither to decision theory. A similar concept, dubbed "cost of contemplation," has been suggested in the framework of preferences over menus in two papers, Ergin (2003) and Ergin and Sarver (2008). Both present and justify axiomatically a model in which the agent chooses the optimal amount of contemplation to evaluate the sets she will have to choose from, where each act of contemplation is associated with a cost of performing it. Formally, their representation is of the form

$$W(A) = \max_{\mu\in\mathcal{M}}[\int_S \max_{p\in A} U(p,s)\mu(\mathrm{d}s) - c(\mu)] \tag{1}$$

where S is a set of states, U is an affine state dependent utility, and $\mathcal{M}$ is a set of signed Borel measures, which are interpreted as possible contemplation strategies, with c as their cost. These two models differ from ours in several aspects, which we will analyze in detail in Section 3. Let us for now point out that neither of them aim to capture the trade-off at the core of our analysis and, in particular, neither can model an agent who prefers a smaller set to avoid the cost of thinking connected to the bigger one. In Ergin (2003) the axioms simply impose that the agent always prefers larger sets - in fact, this is the only requirement. In Ergin and Sarver (2008), the agent might actually prefer a smaller set, but this can be due only to the role of other components like temptation. In fact, they prove that if we were to rule out these other components, then the agent always (weakly) prefers larger sets. By contrast, our work originates from the interest in preferences for smaller sets.

More generally, the concept of "cost of thinking" is connected to the broad notion of bounded rationality, understood as the presence of some form of constraints to the ability of the agent to process information: the cost of thinking could be seen as a way to represent such computational constraints. In this broad area, starting from Simon (1955), papers have focused on game theory (Abreu and Rubinstein (1988), Kalai and Stanford (1988), Rosenthal (1989), Rubinstein and Piccione (1993), Rubinstein and Osborne

(1998), Camerer et al. (2004)), individual decision making (Geanakoplos (1989), Dekel et al. (1998), Wilson (2004), Diasakos (2007)), bargaining, contracting and competitive equilbria (Sabourian (2004), Gale and Sabourian (2005),Tirole (2008)), macroeconomics (Sargent (1993), Sims (2003), Moscarini (2004), Sims (2006), Reis (2006)). A not so recent survey is offered in Rubinstein (1998). There are, however, two characteristics of our approach that distinguish it from the majority of the works in this literature. First, the agent we model is a standard agent who reacts to a non-standard cost, not a boundedly-rational agent. This implies that our agent can potentially think very hard if given the appropriate incentives, or very little otherwise; by contrast, a boundedly-rational agent's behavior is irresponsive to incentives. Second, most of the models in the literature are not defined axiomatically, but rather behaviorally.[4]

1.4 Our approach and preview of the main results

We now turn to describe our approach to this problem. We divide our analysis into four parts.

First, we introduce the central concept of the paper: "thinking aversion." The behavior that we are trying to characterize is that of an agent whose preferences over menus $\succeq$ (may) incorporate some considerations about how hard it will be to make a choice from each menu. This means that these preferences are a combination of two components: a "genuine" preference, which is how the agent would actually rank menus if there were no cost of thinking; and some measure about how hard it will be to actually choose from this menu. For a moment suppose that we could observe this genuine preferences over menus, and use $\succeq^*$ to denote them. (This is clearly not the case, but we will come to this later.) Then, we would like to say that an agent is *Thinking Averse* if for any set A and singleton $\{x\}$, we have

$$\{x\} \succ^* A \Rightarrow \{x\} \succ A.$$

The basic idea is the following. A singleton is a special set that requires *no thinking* - there is nothing to decide, while this might not be true for a generic set A. Then, if a singleton is better than a set according to the genuine ranking *and* it requires no thinking, then it should be preferred to this set by any agent who "dislikes thinking."

The problem is, however, that we do not directly observe this genuine preference $\succeq^*$, but only the general preference $\succeq$. We then have to develop an axiomatic framework

[4]Even more generally, a pattern in western philosophy has analyzed the preference for simplicity also from a normative point of view: from divine simplicity in Saint Thomas, to Occam's razor, to the early works of Wittgenstein (Wittgenstein (2001)). They underline what should or could be general aversion to complicated patterns and decisions - and complexity in general. A recent analysis of such preference for simpler *theories* and their consequences on learning is in Gilboa and Samuelson (2008).

that allows us to *elicit* this preference $\succeq^*$ from $\succeq$, and to do so uniquely, so that Thinking Aversion could be imposed behaviorally. To this end, we take as a primitive a preference relation $\succeq$ over lotteries of menus, and require that this lottery is performed after the agent has chosen from the menus. This means that, given two menus A and B, and $\alpha \in (0,1)$, when the agent faces the lottery $\alpha A \oplus (1-\alpha)B$ she has to form a *contingent plan*, and make a choice from both A and B. Then, she will receive her choice from A with probability α and her choice from B with probability $(1-\alpha)$. Using this structure, we develop novel axioms that allow us to elicit the genuine preference $\succeq^*$ from the general preference $\succeq$.

Our second goal is to characterize the behavior of an agent who exhibits thinking aversion. Let X be a finite subset of alternatives. Define $\mathcal{X}$ to be the set of non-empty subsets of X and $\Delta(\mathcal{X})$ to be the set of lotteries over $\mathcal{X}$. We obtain a representation of the following form. There exists a finite set S of states of the world, a state-dependent utility $u : X \times S \to \mathbb{R}$, a signed measure μ over S, and a function $\mathcal{C} : \Delta(\mathcal{X}) \to \mathbb{R}$, such that $\succeq$ is represented by

$$W(A) = \sum_{s\in S} \mu(s)[\max_{y\in A} u(y;s)] - \mathcal{C}(A), \tag{2}$$

where the first component represents $\succeq^*$, while the second component, $\mathcal{C}$, which we call an *Anticipated Thinking Cost* function, is concave, equal to zero for singletons or lotteries of singletons, and is weakly positive everywhere else.[5] We dub this representation the *Thinking-Averse representation*. (We also discuss the uniqueness properties of this representation, and introduce a notion of comparative thinking aversion.)

We interpret this representation as follows. The preferences of the agent consist of two components: 1) her evaluation of the *content* of the set, captured by the first part of the representation, which also represents her genuine preferences $\succeq^*$; 2) her evaluation of the *cost of thinking* about the set, captured by the anticipated thinking cost function $\mathcal{C}$. We represent the evaluation of the content of a set as follows. The agent's utility depends on the realization of the state of the world $s \in S$. She knows that she will discover the state of the world before choosing, and therefore she expects to pick the best option from the set given the realized state. At the time of ranking of menus, however, she doesn't know the state, and she forms an "expectation" of her future utility using the signed measure μ over the states.[6] This evaluation is reminiscent

[5]More precisely, equation 2 applies when A is a menu. If instead we have a lottery over menus $\bigoplus_i \alpha_i A_i$, where A_i are menus, $\alpha_i \in [0,1]$ and $\sum_i \alpha_i = 1$, the representation is:

$$W(\bigoplus_i \alpha_i A_i) = \sum_i \alpha_i [\sum_{s\in S} \mu(s)[\max_{y\in A_i} u(y;s)]] - \mathcal{C}(\bigoplus_i \alpha_i A_i).$$

[6]Just as in DLR01, this need not be a probability measure: rather, it could assume negative values, which are usually referred to as negative states. The interpretation is that these negative states capture the role of potentially negative components in a set, like tempting elements.

of what DLR01 call an Additive EU Representation, albeit translated into our setup. In fact, one could think of this representation as composed of *standard* preferences less the expected cost of thinking.

Our third goal is to show that, since we elicit the genuine preference $\succeq^*$, we can (almost) directly apply well known results in the literature to characterize it. In particular, we show how we can impose the axioms in Kreps (1979) *only* to $\succeq^*$, and obtain a characterization like the one in (2), but where μ is a probability measure.[7]

As a fourth and last goal, we further characterize the anticipated thinking cost function $\mathcal{C}$. Focusing on the case in which the genuine preference is monotone, we suggest that there could be two interpretations for this cost. First, it could be understood as the cost to *search* for the best option within a menu for an agent who knows her preferences. We call this the *search-cost* interpretation. Second, it could be the cost to figure out what her preferences actually are, i.e. the cost to determine *which* is the best choice in the set. In this latter case, we can understand the multiplicity of states in S as the multiplicity of preferences, and interpret the cost of thinking as the *cost to find out the state of the world*. We refer to this as the *introspection-cost* interpretation.

We offer behavioral axioms that guarantee that the cost of thinking is, in fact, well behaved under these two interpretations, and prove that these axioms are equivalent to a representation of the Anticipated Thinking Cost function as the *sum* of two functions: 1) an increasing function of the cardinality of the set - the search cost; 2) a function of the coarsest partition of the state space S necessary to select the optimal element from the set - the introspection cost. (The first function is monotone and the second is partition-monotone, i.e. assigns higher cost to finer partitions.) Finally, we offer behavioral axioms that separately identify when only one of the two interpretations apply, and characterize the cost of thinking in these two separate cases.

The rest of the paper is organized as follows. In Section 2 we present and characterize an axiomatic model that captures Thinking Aversion. Section 3 characterizes a more restrictive model in which the genuine preference $\succeq^*$ is monotone. Section 4 analyzes the two possible interpretations of the cost of thinking and provide stronger characterizations for it. Section 5 concludes. The proofs appear in the appendix, where we also have the extension of the model to the case of lotteries of menus of lotteries, with a characterization that guarantees the uniqueness of the state space.

[7] By extending the framework to the appropriate one, we also show that the same can be done with the results in Dekel, Lipman, and Rustichini (2001). In particular, we find a representation that inherits all the properties of theirs (including uniqueness, monotonicity, linearity, etc.). This part of the analysis appears in Appendix C.

2. A model for Thinking Aversion

2.1 Formal Setup

Consider a finite set X. Define by $\mathcal{X}$ its power set, that is, $\mathcal{X} := 2^X \backslash \{\emptyset\}$. By $\Delta(\mathcal{X})$ we understand the set of lotteries over $\mathcal{X}$, where $\alpha A \oplus (1-\alpha)B$ denotes the lottery that assigns probability $\alpha \in (0,1)$ to A and $(1-\alpha)$ to B for some $A, B \in \mathcal{X}$, and $\bigoplus_i \alpha_i A_i$ denotes the lottery that assigns weights $\alpha_i \in [0,1]$ to $A_i \in \mathcal{X}$, where $\sum_i \alpha_i = 1$. We use A, B, C to denote generic elements of $\Delta(\mathcal{X})$. We metrize $\Delta(\mathcal{X})$ in the standard way, with the corresponding Euclidean distance between the probability vectors understood as elements of $\mathbb{R}^N$, where $N = |\mathcal{X}|$. With a slight abuse of notation, we refer to $\Delta^S(\mathcal{X})$ as the set of elements of $\Delta(\mathcal{X})$ which contain only singletons in their support. We use p, q, r to indicate generic elements of $\Delta^S(\mathcal{X})$. Again abusing notation, denote by $\mathcal{X}$ the set of degenerate lotteries in $\Delta(\mathcal{X})$. Finally, for any function $F : \Delta(\mathcal{X}) \to \mathbb{R}$, we say that F is concave if, for any $A, B \in \Delta(\mathcal{X})$, $\alpha \in (0,1)$, we have $F(\alpha A \oplus (1-\alpha)B) \geq \alpha F(A) + (1-\alpha)F(B)$. Convexity is defined analogously.

The primitive of our analysis is a complete preference relation $\succeq$ over $\Delta(\mathcal{X})$.

As described in the introduction, we assume that a lottery over menus is performed *after* the agent chooses from each menu in the support. That is, given two menus $A, B \in \mathcal{X}$, the lottery $\frac{1}{2}A \oplus \frac{1}{2}B$ is the lottery that returns with probability $\frac{1}{2}$ the agent's choice from A and with probability $\frac{1}{2}$ her choice from B. When facing a lottery over menus, therefore, the agent needs to form a *contingent plan*, i.e. decide what to choose from each of menus in the support of the lottery. Figure 1 depicts the timing. To our knowledge, the idea to use contingent plans in this framework was introduced by Ergin and Sarver (2008).

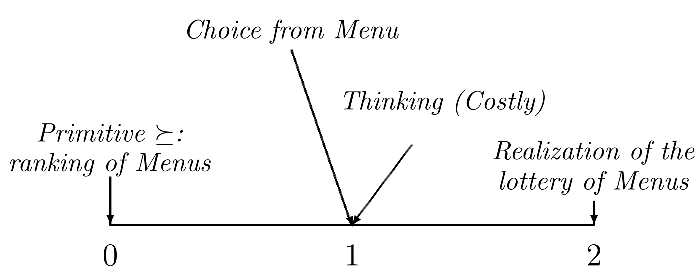

FIGURE 1 : TIMING OF THE SETUP

This setup is similar to the one used in Nehring (1996) and Epstein and Seo (2007), albeit with a different timing of the resolution of uncertainty. At the same time, it differs from the one used in most of the literature, where we usually find a finite set X, the set of lotteries on X, $\Delta(X)$, and a preference relation defined on the compact subsets of $\Delta(X)$.[8] That is, most papers in the literature look at menus of lotteries, while we look at lotteries over menus. We do so because we want agents to form contingent plans when they face lotteries over menus. But in the standard approach there is no "language" for lotteries of menus: instead, the standard set mixture operation in the sense of Minkowski is used to define postulates like independence. And since it wouldn't make sense for the agent to form a contingent plan when facing such menu of lotteries, we depart from the standard approach. (In Appendix C we extend our analysis to the case of lotteries of menus of lotteries in order to obtain stronger uniqueness results.)

2.2 Axioms and Definitions

We now introduce the axiomatic structure of our model.

A.1 (Singleton Independence). *For any $\gamma \in (0,1)$ and any $p, q, r \in \Delta^S(\mathcal{X})$,*

$$p \succeq q \Rightarrow \gamma p \oplus (1-\gamma) r \succeq \gamma q \oplus (1-\gamma) r.$$

This a standard postulate, imposed on a restricted set, lotteries of singletons. (It is standard practice to show that the other direction, $\Leftarrow$, is guaranteed by the continuity postulates that we will impose). As we argued, lotteries of singletons require no thinking, and we should therefore expect a standard behavior to take place when ranking them. At the same time, we do not want to impose linearity of $\succeq$ on the whole $\Delta(\mathcal{X})$. Remember that when an agent faces a lottery over menus she needs to form a contingent plan, and make a selection from all sets in the support. Then, she might be indifferent between two menus A and B, but at the same time strictly prefer A to $\frac{1}{2}A \oplus \frac{1}{2}B$, since in the latter case she needs to think about *both.* This is a clear violation of independence, due to the presence of the disutility of thinking. Accordingly, we do not impose independence on the whole $\Delta(\mathcal{X})$.

As we argued in the introduction, we want to be able to separate the *genuine* ranking of menus, that we would observe if there were no cost of thinking, from the *general* ranking of menus, which might also contain considerations of the costs of thinking. We now turn to this analysis. Consider an agent who is facing the lottery $\frac{1}{2}A \oplus \frac{1}{2}B$ for some $A, B \in \mathcal{X}$. In this case, the agent needs to form a contingent plan: she needs to make a choice from *both A and B*. Then, she needs to "think" about both sets. Suppose now that we increase by a *tiny* bit the probability that the agent receives her choice from A, and that we end up with the lottery $(\frac{1}{2}+\epsilon)A \oplus (\frac{1}{2}-\epsilon)B$ (where ϵ is small).

[8] For further reference, define as $\hat{\mathcal{X}}$ the set of closed and convex subsets of $\Delta(X)$. The two spaces $\Delta(\mathcal{X})$ and $\hat{\mathcal{X}}$ are in fact connected with each other. In Appendix A we show that there exists a continuous and linear bijection between our "world", $\Delta(\mathcal{X})$, and a compact, convex and finite-dimensional subset of $\hat{\mathcal{X}}$.

In this case the agent also has to think about both sets, which means that we have two problems that require basically the same amount of thinking. Assume now that this new mixture is preferred to the original one, which means that the agent liked this change in probabilities. That is, we have

$$(\tfrac{1}{2}+\epsilon)A \oplus (\tfrac{1}{2}-\epsilon)B \succ \tfrac{1}{2}A \oplus \tfrac{1}{2}B.$$

What does this mean? For both sets, the "amount of thinking" is approximately the same, and yet the agent prefers to receive her choice from A with a higher probability. This means that the agent likes her choice from A better than she likes her choice from B. In other words, a "genuine" evaluation, that looks only at the content of sets and disregards the cost of thinking, would say that the content of A is better than the content of B.

To simplify the notation in what follows, let us denote this "genuine" evaluation as the binary relation $\succeq^*$ on $\mathcal{X} \cup \Delta^S(\mathcal{X})$ defined as

$$A \succ^* B \Leftrightarrow (\tfrac{1}{2}+\epsilon)A \oplus (\tfrac{1}{2}-\epsilon)B \succ \tfrac{1}{2}A \oplus \tfrac{1}{2}B \succ (\tfrac{1}{2}-\epsilon)A \oplus (\tfrac{1}{2}+\epsilon)B$$

for all $\epsilon \in (0, \bar{\epsilon}]$, for some $\bar{\epsilon} > 0$. Correspondingly, define $A \sim^* B$ if neither $A \succ^* B$ nor $B \succ^* A$.[9] As argued, we interpret this relation $\succeq^*$ as the "genuine" preferences of the agent over menus. In accordance with this interpretation, we impose, as a postulate, that it must be transitive.

A.2 (Coherence). $\succeq^*$ *is transitive.*

Before we proceed, let us point out two features of the elicitation of $\succeq^*$. First, we have defined $\succeq^*$ only on $\mathcal{X} \cup \Delta^S(\mathcal{X})$, i.e. on degenerate lotteries and on lotteries of singletons, and not on the entire $\Delta(\mathcal{X})$, i.e. not on all lotteries of menus. This follows a precise rationale. We argued that the agent will actually think about both A and B when facing $\frac{1}{2}A \oplus \frac{1}{2}B$, but only as long as A and B are actually menus. If they were lotteries of menus, maybe with some common component, we do not know what "thinking about both" means. Therefore, we simply do not impose anything on those lotteries, making the axioms weaker, and define the relation $\succeq^*$ only on $\mathcal{X}$, degenerate lotteries, and on $\Delta^S(\mathcal{X})$, lotteries of singletons (which require no thinking).

Moreover, we have constructed and motivated this preference $\succeq^*$ arguing that our agent will in fact think about both A and B when she faces the lotteries $\frac{1}{2}A \oplus \frac{1}{2}B$ or

[9] In what follows we use this derived relation $\succeq^*$ in the statement of axioms and definitions, because we believe that it simplifies the notation and makes the statements easier to understand. Notably, however, this is done *just for convenience of exposition*: the same axioms can of course be stated using only the primitive $\succeq$ simply by replacing any statement involving $\succeq^*$ with its definition in terms of $\succeq$. They would simply be a bit longer to read.

$(\frac{1}{2}+\epsilon)A \oplus (\frac{1}{2}-\epsilon)B$. However, one might argue that she might not fully think about the sets, but rather only perform "some" of the thinking and make a suboptimal choice. In this latter case, one could still understand $\succeq^*$ as the preference over the content of menus fixing the cost of thinking, under the assumption that the thinking strategy does not "change abruptly" as we move from $\frac{1}{2}A \oplus \frac{1}{2}B$ to $(\frac{1}{2}+\epsilon)A \oplus (\frac{1}{2}-\epsilon)B$. Invoking an Envelope Theorem argument, we suggest that, as long as the change in the cost of thinking is a "second order effect" with respect to the change in utility as we vary the mixture around $\frac{1}{2}$, our interpretation follows through.

We are now ready to define the notion at the core of our analysis: *Thinking Aversion.*

Definition 1. Consider a preference $\succeq$ on $\Delta(\mathcal{X})$ that satisfies Coherence. Then, $\succeq$ satisfies **Thinking Aversion** if and only if for any $A \in \mathcal{X} \cup \Delta^S(\mathcal{X})$, $p \in \Delta^S(\mathcal{X})$, we have

$$p \succ^* A \Rightarrow p \succ A.$$

Suppose that we have a singleton p whose content is genuinely better than the content of a set A, or simply $p \succ^* A$. Then Thinking Aversion says that this singleton must be preferred *in general* to A: we must have $p \succ A$. This happens because p requires no thinking - it is a singleton, and there is nothing to decide.[10] Then any agent who dislikes thinking must prefer it to any set A that has a worse content and, moreover, might require some thinking.

Arguably, this notion parallels equivalent ways to define risk aversion by comparison with a risk-free option. (Or, ambiguity aversion by comparison with constant acts.) For example, one could define risk aversion for a monotone preference $\trianglerighteq$ on lotteries on $\mathbb{R}$ as follows ($\mathbb{E}[\cdot]$ is the expected value): $\trianglerighteq$ is risk averse if for any lottery p and a degenerate lottery x, if $\mathbb{E}[x] > \mathbb{E}[p]$, then $x \vartriangleright p$. In fact, our definition parallels this one, but instead of the expected value we compare our genuine preference $\succeq^*$, and instead of a risk-free alternative we use a "thinking-free" one, a singleton.

Since our focus is on preference relations that have this property, we impose it as a postulate.

A.3 (Thinking Aversion). $\succeq$ *satisfies Thinking-Aversion.*

The next axiom posits that agents dislike forming contingent plans, which in our setup turns out to be a form of concavity of the preferences. We call this axiom "Mixture Aversion."

[10]More precisely, our definition applies not only to the case of singletons, but also to the case of lotteries over singletons, since they require no thinking either, and the same argument applies.

A.4 (Mixture Aversion). *Take any $A, B \in \Delta(\mathcal{X})$, $p, q \in \Delta^S(\mathcal{X})$ such that $p \sim A$ and $q \sim B$, $\alpha \in (0,1)$. Then, the following must hold:*

$$\alpha p \oplus (1-\alpha)q \succeq \alpha A \oplus (1-\alpha)B.$$

We interpret the axiom as follows. Take two menus A and B and two lotteries of singletons p and q, and say that $p \sim A$ and $q \sim B$. Consider now a mixture of A and B and the same mixture of p and q. In the case of the mixture between A and B the agent must end up thinking (somehow) about *both* sets A and B. On the other hand, in the mixture of the two singletons, p and q, there is still no thinking involved. An agent who dislikes thinking, therefore, must weakly prefer this mixture of singletons to the mixture of the two sets, because in the first case the cost of thinking may increase, while in the second it is still zero.

Since we are after a representation theorem, we need to impose a continuity-type axiom. We can do this either by imposing full continuity of $\succeq$, or by restricting our attention to singletons and to their relations to sets. We consider the two cases separately.

A.5 (Weak Continuity). 1. *For any $A \in \Delta(\mathcal{X})$, the sets $\{p \in \Delta^S(\mathcal{X}) : p \succeq A\}$ and $\{p \in \Delta^S(\mathcal{X}) : A \succeq p\}$ are closed.*

2. *For any $A \in \mathcal{X}$, the sets $\{p \in \Delta^S(\mathcal{X}) : p \succeq^* A\}$ and $\{p \in \Delta^S(\mathcal{X}) : A \succeq^* p\}$ are closed.*

A. 5* (Full Continuity). *For any $A \in \Delta(\mathcal{X})$, the sets $\{B \in \Delta(\mathcal{X}) : B \succeq A\}$ and $\{B \in \Delta(\mathcal{X}) : A \succeq B\}$ are closed.*

As the names suggest, in our framework Weak Continuity is a weaker requirement than Full Continuity. (See Claim 5 in Appendix B.1 for a formal proof).

Finally, we impose two technical axioms. The first posits that there exist two elements x^*, x_* in X that are the best and worst elements in $\Delta(\mathcal{X})$ according to $\succeq$. The second, which we impose only to guarantee uniqueness of the representation, posits that there exists an element $x^* \in X$ that is the best element in $\mathcal{X}$ according to $\succeq^*$. Both of these postulates are technical and are not derived from any real world consideration, but it is quite easy to depict a situation in which they would exist.[11]

A.6 (Best/Worst). *There exist $p^*, p_* \in \Delta^S(\mathcal{X})$ such that $p^* \succeq A \succeq p_*$ for all $A \in \Delta(\mathcal{X})$.*

[11] For example, consider a set X composed of compact cars that the agent could receive for free. Add now to this set two options: a Ferrari and an old bike. Indeed, the first will be preferred to anything else in the set, while the second will be certainly the worst option.

A.7 (Best/Worst*). *For any $A \in \mathcal{X}$, there exists $p^* \in \Delta^S(\mathcal{X})$ such that $p^* \succeq^* A$.*

2.3 Representation

2.3.1 Anticipated Thinking Cost To express the notion of "thinking cost" we need a function that associates with every set a measure of the disutility caused by having to choose from it. This function should have some minimal properties that render it a real thinking cost: it should be null on singletons, or lotteries of singletons - since the choice is trivial; and weakly positive everywhere else. (Notice that this is not simply a matter of normalization: the main point is that the cost of thinking about any set cannot be below that of singletons.) In addition, it should be concave, to capture the fact that making a contingent plan is costly. This leads us to the following definition.[12]

Definition 2. A function $C : \Delta(\mathcal{X}) \to \mathbb{R}$ is an **Anticipated Thinking Cost function** if the following conditions hold:

1. $C(\{p\}) = 0$ for all $p \in \Delta^S(\mathcal{X})$.
2. $C(A) \geq 0$ for all $A \in \Delta(\mathcal{X})$.
3. C is concave.

One might also expect an Anticipated Thinking Cost function to be monotone, that is, to assign a higher cost of thinking to larger sets. This, however, might be too restrictive. For example, consider an agent who needs to choose from a menu of wines. Adding other wines to the list might make the problem harder. However, if we add the option "get ten million dollars," it would be so clearly dominant that this larger set becomes actually *easier* to think about, and not harder. (In Section 4.3.2 we provide behavioral axioms to guarantee that the cost is in fact an increasing function of the cardinality of the set.)

Finally, let us emphasize that this is an *anticipated* thinking cost function. That is, it represents the cost that the agent *expects* to endure when she *will* be choosing from a set (or when forming a contingent plan). In fact, the thinking effort is exerted not when *the menu* is chosen, but later, when the agent is choosing *from the menu*, or some time before that. (Refer to Figure 2 for the timing.)

2.3.2 Thinking-Averse representation We are now ready to introduce our first representation.

[12] Admittedly, to keep the analysis general, we have imposed only the minimal requirements on such a function. In Section 4 we suggest two possible interpretations of this cost, and provide behavioral axioms that allow us to find much stronger representations.

Definition 3. A preference relation $\succeq$ on $\Delta(\mathcal{X})$ has a **Thinking-Averse representation** if there exists a non-empty, finite set S of states of the world, a state-dependent utility $u : X \times S \to \mathbb{R}$, a signed measure μ over S and a function $\mathcal{C} : \Delta(\mathcal{X}) \to \mathbb{R}$ such that $\succeq$ is represented by

$$W(\bigoplus_i \alpha_i A_i) = \sum_i \alpha_i [\sum_{s \in S} \mu(s)[\max_{y \in A_i} u(y; s)]] - \mathcal{C}(\bigoplus_i \alpha_i A_i) \tag{3}$$

where:

1. $\mathcal{C}$ is an Anticipated Thinking Cost function;
2. $W + \mathcal{C}$ represents $\succeq^*$.

We interpret this representation as follows. The preferences of an agent are the difference of two components. First, her genuine evaluation of the content of the set. Second, her evaluation of the *cost of thinking* of the set. These two components are potentially pushing in different directions, in which case the agent faces a trade off between having a better content and facing a harder choice - which is the phenomenon we are after. In our cell-phone example, she weights the benefits of a large number of options with the (expected) cost of having to decide which one is the best.

The genuine evaluation of the content is modeled with a (finite) set of states of the world, a state-dependent utility function u and a signed measure μ over S. We interpret it as if our agent does not yet know what her preference will be at the time of choice - for each possible preference we have a state of the world s and a utility function $u(\cdot, s)$. However, she knows that she *will* discover the state of the world before making a decision: consequently, she expects to choose the best option and obtain a utility of $\max_{y \in A_i} u(y; s)$ for each state $s \in S$. Now, however, she does not know the state, and forms an "expectation" using the signed measure μ.[13] In addition, if the agent is evaluating a lottery of menus, she also doesn't know which part of the contingent plan will be put in place and consequently needs to further condition on the probabilities α_i of each realization of the lottery. Notice two more features of this first part of the representation. First, it represents the genuine preference $\succeq^*$, which is in fact the genuine evaluation that the agent would give to the a set *if there were no cost of thinking.* Second, it has the same intuition of what DLR01 call an Additive EU representation, although defined in a setup of menus of lotteries and not lotteries of menus. In a fact, a Thinking-Averse representation can be seen as the difference between a standard affine preference over menus from which a cost of thinking is subtracted.

The second component, the cost of thinking, is represented using what we defined as an Anticipated Thinking Cost function, with the properties that we have discussed.

[13]As standard in the literature μ need not be a probability measure, but rather contain negative components, usually referred to as negative states. We refer to DLR01, DLRS and Dekel, Lipman, and Rustichini (2007a) for further discussion on negative states.

Let us review the timing of this representation. First, at time zero, the agent chooses a menu. Then, before choosing *from* this menu, she discovers the state and thinks about what to choose. (Later we will discuss whether we should understand the revelation of the state as the *outcome* of the thinking process, or whether the thinking takes place after the revelation of the state). Then, at time 1, the agent chooses from the menu, or forms a contingent plan. In an additional *later* stage, the lottery is realized and the agent is given her choice as specified by the contingent plan. (We refer to Figure 2 for the timing).

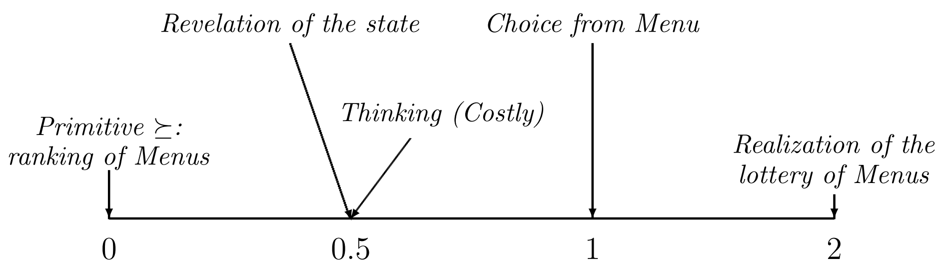

FIGURE 2 : TIMING OF THE REPRESENTATION

2.4 Representation Theorem

We are now ready to state the main representation theorem.

Theorem 1. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$ that satisfies Best/Worst. Then, the following two conditions are equivalent:*

(i) *$\succeq$ has a Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$;*

(ii) *$\succeq$ satisfies Thinking Aversion, Mixture Aversion, Singleton Independence, Weak Continuity and Coherence.*

Moreover, $\succeq$ has a Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ such that $\mathcal{C}$ is continuous if and only if $\succeq$ has a Thinking-Averse representation and satisfies Full Continuity.

The meaning of this representation and of the axioms has been discussed in the previous sections. (We postpone the comparison with other papers in the literature to Section 3, after we have discussed additional results.)

2.5 Uniqueness

First of all, we wish to establish the uniqueness properties of two components of the representation, the genuine evaluation of the content and the anticipated thinking cost function $\mathcal{C}$. In fact, such uniqueness is essential for this separation to be meaningful. The following Theorem shows that it is a feature of our model, provided that an additional technical axiom, Best/Worst*, is satisfied.

Theorem 2. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$ that satisfies Best/Worst and Best/Worst*. Then, if $< S, \mu, u, \mathcal{C} >$ and $< S', \mu', u', \mathcal{C}' >$ are both Thinking-Averse representations for $\succeq$, then there exists $\gamma \in \mathbb{R}_{++}$, $\beta \in \mathbb{R}$ such that*

$$\sum_{s \in S'} \mu'(s)[\max_{y \in A_i} u'(y; s)] = \gamma \Big[\sum_{s \in S} \mu(s)[\max_{y \in A_i} u(y; s)] \Big] + \beta$$

and

$$\mathcal{C}' = \gamma \mathcal{C}.$$

Theorem 2 shows that the evaluation of the content is unique up to a positive affine transformation, that the evaluation of the cost is unique up to a positive scalar multiplication, and that these two transformations must be the same (γ is the same for both). This implies that, if we fix the evaluation of the content, the representation of the cost is unique.

Moreover, we would also like the way we represent $\succeq^*$ to be unique: uniqueness of the endogenous state space S much in line with the analysis in DLR01. Unfortunately, however, this is not a feature of the model we have discussed here. This happens because in a sense our space is not "rich" enough: recall that, as opposed to DLR01, we do not work on the space of menus of lotteries, but rather on that of lotteries over menus, which is substantially smaller.[14] As a result, we do not have enough observations to identify the state space S uniquely. At the same time, if we extend our analysis to the case of lotteries of menus of lotteries (instead of lotteries of menus) we gain the full uniqueness of the state space and all the properties of the DLR01 representations for the characterization of $\succeq^*$. Such a framework would not be new to decision theory: it is used, for example, in Epstein, Marinacci, and Seo (2007). This part of the analysis appears in Appendix C.[15]

[14] By smaller we mean the following. In Appendix A we show that we are able to construct a bijection between our space and a subset of the space of menus of lotteries. This will be a strict, finite-dimensional subset of an infinite-dimensional space: in this sense we mean smaller.

[15] Alternatively, one might seek uniqueness of the representation in the sense of Epstein and Seo (2007): that is, require that any two representations generate an identical measure over the upper contour sets. As they argue, this could be seen as a more robust form of uniqueness than the one in DLR01. We refer to Epstein and Seo (2007) for a detailed discussion. It is easy to show that their uniqueness result (Theorem 3.1) applies here if the conditions of Theorem 2 hold and if μ is a probability measure (the conditions for which will be discussed in the next section).

2.6 Being more Thinking-Averse: a comparability result

We now introduce a comparability notion for thinking aversion, to be able to say when one agent is more "thinking averse" than another in a similar spirit to how we compare risk aversion or ambiguity aversion. In particular, we want to make such a comparison for agents that differ only in terms of Thinking Aversion, i.e. for two agents that *have the same genuine preference over the content of a set*, so that we can ascribe all the differences in their behavior to a different approach to thinking. Therefore, in what follows we consider two preference relations $\succeq_1$ and $\succeq_2$ such that $\succeq_1^* = \succeq_2^*$.

Definition 4. Consider two preference relations $\succeq_1$ and $\succeq_2$ on $\Delta(\mathcal{X})$ that satisfy Best/Worst, have a Thinking-Averse representation, and such that $\succeq_1^* = \succeq_2^*$. We say that $\succeq_1$ *is more Thinking Averse than* $\succeq_2$ if, for any $A \in \Delta(\mathcal{X})$ and $p \in \Delta^S(\mathcal{X})$, we have

$$A \succeq_1 p \Rightarrow A \succeq_2 p.$$

We have two agents with the same genuine evaluation of the content of sets. The two agents, however, might differ in the way they dislike "thinking." and we wish to say that the first dislikes thinking more than the second. Suppose that $A \succeq_1 p$ for some $A \in \Delta(\mathcal{X})$ and $p \in \Delta^S(\mathcal{X})$. This means that the first agent would rather think about A than take p, albeit the latter requires no thinking. Then, if the second agent has the same genuine evaluation of the content and an even lower dislike of thinking, she should do the same, and we should have $A \succeq_2 p$ as well. Notice that this definition parallels the one of comparative risk aversion and similar ones of comparative ambiguity aversion.[16] (Just like in the definition of Thinking Aversion, here we have a singleton as a thinking-free element, instead of a risk-free alternative (for risk aversion) or a constant act (for ambiguity aversion).)

Proposition 1. *Consider two preference relations $\succeq_1$ and $\succeq_2$ on $\Delta(\mathcal{X})$ that satisfy Best/Worst, have a Thinking-Averse representation, and such that $\succeq_1^* = \succeq_2^*$. Then, the following two statements are equivalent:*

(i) *$\succeq_1$ is more Thinking Averse than $\succeq_2$;*

(ii) *For any two Thinking-Averse representations $(S_1, \mu_1, u_1, \mathcal{C}_1)$ and $(S_2, \mu_2, u_2, \mathcal{C}_2)$ such that $S_1 = S_2$, $\mu_1 = \mu_2$ and $u_1 = u_2$, we have $\mathcal{C}_1 \geq \mathcal{C}_2$.*

3. Monotonicity in the content

The model we have discussed thus far allows the agent to prefer a smaller set independently of the cost of thinking. For example, she might have no cost of thinking at all

[16] For the latter, see in particular Ghirardato and Marinacci (2002).

($\mathcal{C}(A) = 0$ for all $A \in \Delta(\mathcal{X})$) and still prefer a smaller set to avoid temptation. Formally, this could be the case if $\mu(s) < 0$ for some $s \in S$: much in line with DLR01, our model allows for the presence of these negative states, which, together with the thinking cost, might induce the agent to prefer smaller sets. We now rule out this possibility and focus on the case of monotonicity of the genuine evaluation of the content: the case in which the cost of thinking is the *only* feature that might induce the agent to prefer a smaller set.

To do so, we make use of the fact that we elicit the preference $\succeq^*$, and that can therefore impose the required axioms only on this preference relation - one of the advantages of this approach. It turns out that we only need to impose the axioms in Kreps (1979) on $\succeq^*$. First, we want our agent to genuinely prefer the content of a larger set to that of a smaller one. We call this axiom "Content Monotonicity."

A.8 (Content Monotonicity). *For any $A, B \in \mathcal{X}$, $B \subseteq A \;\Rightarrow\; A \succeq^* B$.*

This axiom posits that, were it not for the cost of thinking, the agent would always prefer bigger sets. This corresponds to the case of the mobile phone plan example in the introduction. Genuinely she prefers more options, but more options imply a higher cost of thinking, and so she might not choose a larger set after all.

Following Kreps (1979), we add a property that guarantees that there is some consistency in the way the preference behaves for larger set.

A.9 (Content Submodularity). *For any $A, B, C \in \mathcal{X}$, $A \sim^* A \cup B \;\Rightarrow\; A \cup C \sim^* A \cup B \cup C$.*

The rationale of this axiom is the following. If adding B to A does not give any benefit, it must be the case that for any element in B there is an element in A that is at least as good. But then, adding B to $A \cup C$ should not give any benefit either. (We refer to Kreps (1979), where this axiom was introduced, for further discussion.)

Definition 5. A preference relation $\succeq$ on $\Delta(\mathcal{X})$ has a **Content-Monotone Thinking-Averse representation** if it has a Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ where μ is a probability distribution over S.

This representation differs from a generic Thinking-Averse representation exactly as we discussed: we are ruling out negative states, and have a monotone evaluation of the content.[17]

[17] The proof of Theorem 3 requires us to extend the results in Kreps (1979) to the case of lotteries over menus in a vNM sense. We refer to Appendix A.3 for a detailed discussion.

Theorem 3. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$ that satisfies Best/Worst. Then, the following two conditions are equivalent:*

*(*i*) $\succeq$ has a Content-Monotone Thinking-Averse representation;*

*(*ii*) $\succeq$ has a Thinking-Averse representation and satisfies Content Monotonicity and Content Submodularity.*

As we have mentioned in the introduction, the results in Theorem 1 and 3 are reminiscent of the ones in Ergin (2003) and Ergin and Sarver (2008).[18] There are, however, some important differences on the axioms, on the representations and on the goals. The differences between the Axioms are self evident: although both structures rely on the presence of contingent plans, we use them to separate the two components of the preference - the genuine evaluation of the content of the set and the anticipated thinking cost - which is the core of our structure. By contrast, Ergin and Sarver (2008) do not distinguish between the two components at all.[19] Moreover, at a formal level, our result is based on different primitives: while Ergin and Sarver (2008) use menus of lotteries, as we have discussed we use lotteries of menus.[20]

This difference in the axiomatic structure leads to representations that, although *look* similar, are in fact conceptually very different. What Ergin (2003) and Ergin and Sarver (2008) obtain is an agent that expects herself to choose the *optimal* thinking strategy from a pool of available ones: better strategies allow her to find better options in a menu, but at the same time have a higher cost. (See Equation 1 in Section 1.) This leads to a representation such that, if the agent has a monotone evaluation of the content of sets, then *the whole preference must be monotone*: in fact, by facing a bigger set the agent gets more content utility, and since she can at least use the same strategy she used for the smaller set, then she cannot be worse off. (This obvious in the case of Ergin (2003), since monotonicity is the only postulate, while in Ergin and Sarver (2008) it is proven in Theorem 1.B.) By contrast, in our representation this need not be true, as shown by Theorem 3. In particular, in our case the agent could dislike a bigger set since

[18]These two papers are deeply connected with each other: the latter is an extension of the former to the space of menus of lotteries. For simplicity we compare our result to this extended one, which is more similar to ours.

[19]A common feature is the presence of some form of concavity of the preferences: we impose A.4, Mixture Aversion, while they impose it in an axiom called Aversion to Contingent Planning. Nevertheless, it is easy to see that the two axiomatic structure are not nested in non degenerate cases. (However, as we will see, a direct comparison is not possible since the two papers are based on different primitives.)

[20]As argued, we believe the latter to be more appropriate to use with contingent plans, but as we have seen in this different setting we lose the uniqueness of the state space, one of the features of Ergin and Sarver (2008). At the same time, Ergin and Sarver (2008) make some very compelling arguments as to why one of the axioms that they use has a reasonable interpretation even with contingent plans in a standard setting of menus of lotteries.

she knows that *she will have to think harder to choose from it.* This is the core difference between the two representations. In a way, the two models stand at the opposite sides of an interpretation pole. On the one side, with Ergin and Sarver (2008), we have an agent who expect herself to rationally react to a computation limitation: she knows she will think just *as much as optimal.* This, as we have discussed, leads to monotonicity in the rank of sets if the evaluation of the content is monotone. On the other side of the interpretation pole, in our paper agents can expect themselves to think too much, possibly more than what they would consider optimal now. In this sense, we can view our agents as being "tempted" into *excessive* thinking. Anticipating that they will think so hard, our agents might then choose to have a smaller set to avoid this effort - which is the behavior that motivated our analysis.[21]

4. Characterizing cost

Our analysis so far has been almost silent on the form of the anticipated thinking cost function. We only required that it is zero for lotteries of singletons and that it is concave. The purpose of this section is to strengthen this characterization.

We suggest that there are two ways of interpreting this cost of thinking. First of all, an agent might incur in a cost in "reading" the menu. The idea is that the agent already knows what is best for her, but she needs to *find* the optimal option within a set, and this creates a cost of thinking. We refer to this interpretation as the "search-cost interpretation." This is the most standard interpretation of the cost of thinking.

Alternatively, we can interpret the cost of thinking as the cost that the agent has to incur to discover *what she wants.* That is, not the cost to locate the best option within a set, but the cost to understand *what* is the best option in a set. For example, an agent could have a high cost of thinking even when she faces only two alternatives, since it might be very hard for her to figure out which of the two is the best choice. In this sense, it is the cost the agent incurs to discover her own preferences. We refer to this interpretation as the "introspection-cost interpretation." This captures what we consider the most compelling view of the cost of thinking. In fact, one might consider the existence of such a cost as an (indirect) evidence of the fact that the agent has incomplete preferences, which she can "complete" by paying this cost.

[21] Let us emphasize that both Theorem 1 and the results in Ergin (2003) and Ergin and Sarver (2008) relate to how the agent expects to act *in the future*, when asked to make a choice from the set. That is, it would not be correct to say that in our representation agents think too much: rather, they *expect* themselves to possibly think too much when asked to choose from a set, and for this reason they prefer a smaller set now. (What they will do at the time of choice, we cannot say.) Moreover, notice that the axioms do not impose that the agent expects herself to think more than optimal. Rather, we have shown that we can represent her behavior *as if* she were tempted into excessive thinking.

It is easy to see that these two interpretations are conceptually well distinguished: in the first case the agent already knows what she wants, but she has to bear a cost to *locate* it; in the second case the agent has a cost to decide what she wants, but then has no difficulty locating it. At the same time, they are are not incompatible: one could easily depict a situation in which both emerge. The content of this section is to further characterize the anticipated thinking cost function in light of these two interpretations. In particular, first we offer a behavioral axiom that guarantees that the cost of thinking is, in fact, the sum of these two costs. Then, we strengthen this characterization to the case in which the search-cost is a function only of the cardinality of the set. Finally, we offer behavioral axioms that allow us to separately identify when the cost of thinking is only a search-cost or only an introspection-cost, and characterize each of the two cases. For simplicity, we carry out this analysis in the case of monotonicity of $\succeq^*$, that is, under the axiomatic structure of Theorem 3.

4.1 A general model of cost

To further strengthen our representation we need to guarantee that the cost of thinking is "well-behaved." To express these conditions, however, we want a way to express behaviorally that a set has a higher cost of thinking than another. To this end, let us introduce the notion of "thinking-free equivalent:" for any $A \in \mathcal{X}$, define p_A and p_A^* as the elements of $\Delta^S(\mathcal{X})$ s.t. $p_A \sim A$ and $p_A^* \sim^* A$. (Thinking-free equivalents serve in our setup the same purpose that certainty equivalents serve in a setup with risk.) Consider two sets $A, B \in \mathcal{X}$ and suppose that we want to express that the cost of thinking of A is higher than that of B. We know that the cost of thinking is the difference between the evaluation of A using $\succeq^*$ and using $\succeq$. Then, if W represents $\succeq$, the cost of thinking of A being higher than that of B implies $W(p_A^*) - W(p_A) \geq W(p_B^*) - W(p_B)$, which means $\frac{1}{2}W(p_A^*) + \frac{1}{2}W(p_B) \geq \frac{1}{2}W(p_B^*) + \frac{1}{2}W(p_A)$ and hence $\frac{1}{2}p_A^* \oplus \frac{1}{2}p_B \succeq \frac{1}{2}p_B^* \oplus \frac{1}{2}p_A$ (by A.1). We will therefore use $\frac{1}{2}p_A^* \oplus \frac{1}{2}p_B \succeq \frac{1}{2}p_B^* \oplus \frac{1}{2}p_A$ to express that the cost of thinking of A is higher than the cost of thinking of B.

Consider a menu A and $x \in A$ such that $A \sim^* \{x\}$: this means that the option x contains all the content-utility in A. From such a set our agent has no problem deciding *what* to choose: the choice of x is a *no brainer*.[22] Let us combine A with another set $C \in \mathcal{X}$, and look at the set $A \cup C$. Since in A we already had a no-brainer choice, the choice from $A \cup C$ cannot be simpler from this point of view. At the same time, $A \cup C$ is a strictly larger set and therefore finding the optimal choice is bound to be harder. We would then like to say that the cost of thinking about $A \cup C$ is higher then the cost of thinking about A. This leads us to the following axiom, which is stated using thinking-free equivalents.

[22] "If a thing can be done adequately by means of one, it is superfluous to do it by means of several," Saint Thomas Aquinas, Aquinas (1997, pg. 129).

A.10 (Cost Coherence). *Consider $A, C \in \mathcal{X}$ such that $\{x\} \sim^* A$ for some $x \in A$ and suppose that $p_A, p^*_A, p_{A\cup C}, p^*_{A\cup C}$ exist. Then,*

$$\tfrac{1}{2}p_A \oplus \tfrac{1}{2}p^*_{A\cup C} \succeq \tfrac{1}{2}p^*_A \oplus \tfrac{1}{2}p_{A\cup C}.$$

To state the new representation we need a few additional definitions. First, if S is a finite non-empty set (state space), denote by $\Pi(S)$ the set of partitions of S. Moreover, for any finite set S, state-dependent utility function $u : X \times S \to \mathbb{R}$ and any $A \in \mathcal{X}$, define $\mathcal{I}_{S,u}(A)$ as

$$\mathcal{I}_{S,u}(A) := \{\pi \in \Pi(S) : \text{for all } \pi_i \in \pi \ \exists x_i \in A \text{ s.t. } \max_{y\in A} u(y,s) = u(x_i,s) \text{ for all } s \in \pi_i\}.$$

We understand $\mathcal{I}_{S,u}(A)$ as the set of partitions that allow the agent to attain the full utility of a set A by choosing the same alternative in every state grouped by the partition. We now define a function that assigns to each set $A \in \mathcal{X}$ one partition in $\mathcal{I}_{S,u}(A)$ such that no coarser one is available.

Definition 6. For any non-empty set S and function $u : X \times S \to \mathbb{R}$, a function $\mathcal{P} : \mathcal{X} \to \Pi(S)$ is a **partition function** if for all $A \in \mathcal{X}$, $\mathcal{P}(A) \in \mathcal{I}_{S,u}(A)$ and there is no $\pi \in \mathcal{I}_{S,u}(A)$ s.t. $\pi \neq \mathcal{P}(A)$ and π is coarser than $\mathcal{P}(A)$.

Finally, for any finite set S, we need to define a function that indicates the cost of each partition. We focus on a specific class of these functions, that we call partition-monotone: cost functions that assign a (weakly) higher cost to finer partitions.

Definition 7. For any non-empty set S and function $f : \Pi(S) \to \mathbb{R}$, we say that f is **partition-monotone** if $f(\pi) \geq f(\pi')$ for any $\pi, \pi' \in \Pi(S)$ such that π is finer than π'.

We are now ready to state our representation theorem.

Theorem 4. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$ that has a Content-Monotone Thinking-Averse representation and satisfies Best/Worst and Best/Worst*. Then, the following two conditions are equivalent:*

(i) *$\succeq$ satisfies Cost Coherence;*

(ii) *there exist a Monotone Thinking Aversion Representation $< S, \mu, u, \mathcal{C} >$ of $\succeq$, a partition function $\mathcal{P} : \mathcal{X} \to \Pi(S)$, a partition-monotone function $c_I : \Pi(S) \to \mathbb{R}$ and a function $c_s : \mathcal{X} \to \mathbb{R}$ such that for all $A \in \mathcal{X}$,*

$$\mathcal{C}(A) = c_I(\mathcal{P}(A)) + c_s(A)$$

and:

1. $c_I(\{S, \emptyset\}) = 0$;
2. $c_s(\{x\}) = 0$ *for all* $x \in X$;
3. $c_s(A) \geq c_s(B)$ *for all* $A, B \in \mathcal{X}$ *s.t.* $A \supseteq B$.

Theorem 4 shows that, if Cost Coherence is satisfied, then the cost of thinking about a menu can be represented as the *sum* of two costs: a search cost c_s, and an introspection-cost c_I. These are the two forms of cost of thinking that we discussed above. The search cost c_s is the cost of *searching* for the best choice in a set: as such, it increases as the set gets larger, and it is zero when the set is a singleton. The introspection cost c_I is the cost to ascertain *what* is the best choice: we represent it as the cost incurred to discover the state of the world, but only insofar as required to make an optimal choice. In fact, our agent might not need to figure out the exact state of the world. For example, if all the alternatives in a set A that are optimal in state s_1 are also optimal in state s_2, then the agent has no need to distinguish between s_1 and s_2, and she can settle with a partition of the state space in which s_1 and s_2 are not distinguished. Theorem 4 shows that we can represent the introspection-cost as the cost of the coarsest partition necessary to make a choice. Moreover, this cost is partition-monotone (finer partitions require more thinking and are therefore more costly) and it is zero for the empty partition ($c_I(\{S, \emptyset\}) = 0$).

4.2 A general model of cost with cardinality

We now turn to strengthen the representation in Theorem 4 in order to obtain a stronger structure for the search-cost. Consider a menu A that contains a *no-brainer* choice, i.e. we have $x \in A$ s.t. $A \sim^* \{x\}$. Again, here our agent has no problem deciding *what* to choose. Let us now add to this set an element $y \in X$ such that this addition doesn't make the set genuinely better in content: we have $A \cup \{y\} \sim^* A \sim^* x$. That is, adding y to A doesn't add any utility: we refer to y as an *irrelevant alternative* in A. In a setup with a cost of thinking, we would then like to say that adding y to A cannot be beneficial for the agent: it does not add any utility, and it cannot simplify the choice (since there is already a no-brainer choice). Instead, it only adds "noise" to the set, and our agent should not like it. This is the content of the following axiom.

A.11 (Harm of Irrelevant Alternatives - HIA). *Consider* $A \in \mathcal{X}$, $x \in A$, $y \in X \backslash A$ *s.t.* $A \sim^* x$ *and* $A \cup \{y\} \sim^* A$. *Then,*

$$A \succeq A \cup \{y\}.$$

Consider now a set $A \in \mathcal{X}$ and replace one element x of A with some element y which was not in A. Call this new set B. Assume now that in this new set we have a no-brainer choice: there exists $z \in B$ s.t. $\{z\} \sim^* B$. This no-brainer choice could either be y, the option that we have added, or another option that was already in the set.[23] Then, we would like to conclude that the cost of thinking about B is (weakly) lower than the cost of thinking about A, because B has this one attractive option. This leads us to the following axiom (again expressed using thinking-free equivalents).

A.12 (Cost Reduction). *Consider $A \in \mathcal{X}$, and $B = (A \backslash \{x\}) \cup \{y\}$ for some $x \in A$, $y \in X \backslash A$ s.t. $B \sim^* \{z\}$ for some $z \in B$. Then, if p_A, p_B, p^*_A, p^*_B exist, we have*

$$\tfrac{1}{2}p^*_A \oplus \tfrac{1}{2}p_B \succeq \tfrac{1}{2}p^*_B \oplus \tfrac{1}{2}p_A.$$

It turns out that these two axioms are enough to guarantee a stronger characterization.

Theorem 5. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$ that has a Content-Monotone Thinking-Averse representation and satisfies Best/Worst and Best/Worst*. Then, the following two conditions are equivalent:*

(i) $\succeq$ satisfies Cost Reduction and HIA;

(ii) there exist a Monotone Thinking Aversion Representation $< S, \mu, u, \mathcal{C} >$ of $\succeq$, a partition function $\mathcal{P} : \mathcal{X} \to \Pi(S)$, a partition-monotone function $c_I : \Pi(S) \to \mathbb{R}$ and an increasing function $c_s : \mathbb{N} \to \mathbb{R}$ such that for all $A \in \mathcal{X}$

$$\mathcal{C}(A) = c_I(\mathcal{P}(A)) + c_s(|A|)$$

and:

1. $c_I(\{S, \emptyset\}) = 0$;
2. $c_s(1) = 0$.

Moreover, for any $\mathcal{C}$, c_s and $c_I \circ \mathcal{P}$ are unique.

4.3 Focusing on specific interpretations

Finally, we offer behavioral conditions that allow us to separately identify the cases in which the cost of thinking is only an introspection-cost from the case in which it is only a search-cost.

[23]For example, consider a menu A that contains possible meals, and an agent who likes lobster better than anything else. Let us assume that x and $z \in A$ are both lobster (prepared in different ways), while $y \notin A$ is some other entrée. When evaluating A the agent needs to choose between x and z, both lobster meals. However, if we construct a set B by replacing x with y, then in B the agent will have a no-brainer choice, since there is only one lobster dish to choose from.

4.3.1 Introspection-cost only As we argued earlier, if there is an option x in a set A that is better in all states, $\{x\} \sim^* A$, then the agent should have no cost in deciding what to choose, since she can simply pick x. In other words, the introspection-cost of this set A must be zero. And, if the only cost at play is the introspection cost, then the cost of thinking about A must be zero. But if A has the same genuine ranking of x, $A \sim^* x$, and has the same cost of thinking (zero), we must have $A \sim \{x\}$. This leads us to the following Axiom.

A.13 (Costly Flexibility). *For any $A \in \mathcal{X}$, $x \in A$, if $A \sim^* \{x\}$, then $A \sim \{x\}$.*

Notice that this postulate has some of the flavor of the standard Independence of Irrelevant Alternatives axiom, exactly in opposition to HIA: if irrelevant alternatives are added to a set which contain a no-brainer choice, the evaluation of the set remains untouched.

Theorem 6. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$ that has a Content-Monotone Thinking-Averse representation. Then, $\succeq$ satisfies Costly Flexibility if and only if there exist a Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ of $\succeq$, a partition function $\mathcal{P} : \mathcal{X} \to \Pi(S)$ and a partition-monotone function $c : \Pi(S) \to \mathbb{R}$ such that for all $A \in \mathcal{X}$*

$$\mathcal{C}(A) = c(\mathcal{P}(A))$$

and $c(\{S, \emptyset\}) = 0$.

4.3.2 Search-cost only In this case, the addition of an irrelevant alternative to a set is always *harmful* for the agent. In particular, such an addition will always (weakly) increase the cost of thinking, because if the agent needs to find her best option within a set, it will be harder for her to do so if the set is larger. And in the absence of an introspection cost the agent does not care if this addition simplifies the choice. Following this rationale, we have the following postulate.

A.14 (Strong Harm of Irrelevant Alternatives - SHIA). *For any $A \in \mathcal{X}$, $x \in X$ such that $A \sim^* A \cup \{x\}$ we have*

$$A \succeq A \cup \{x\}.$$

Next, we want to postulate that the cost of thinking is anonymous: if the cost of thinking is due to the difficulty of finding objects within a set, it should change in the same way when any element is added.

A.15 (Anonymity). *For any $A \in \mathcal{X}$, $x, y \in X$, $x, y \notin A$, if $p_{A\cup\{x\}}, p^*_{A\cup\{x\}}, p_{A\cup\{y\}}, p^*_{A\cup\{y\}}$*

exist, then

$$\tfrac{1}{2}p^*_{A\cup\{x\}} \oplus \tfrac{1}{2}p_{A\cup\{y\}} \sim \tfrac{1}{2}p^*_{A\cup\{y\}} \oplus \tfrac{1}{2}p_{A\cup\{x\}}.$$

Finally, one might want a testable postulate that guarantees that the agent's cost of thinking increases more then proportionally as the size of the set increases. This means that, if we add two "irrelevant" alternatives x, y to a set A - i.e. we have $A \sim^* A\cup\{x,y\}$ - then the difference between the evaluation of $A\cup\{x,y\}$ and $A\cup\{x\}$ should be higher than that between $A\cup\{x\}$ and A. This is precisely what we impose in the following axiom, again using think-free equivalents.

A.16 (Increasing Search Cost - ISC). *For any $A \in \mathcal{X}$, $x, y \in X$ such that $A \sim^* A\cup\{x,y\}$, if $p_{A\cup\{x,y\}}, p_A, p_{A\cup\{x\}}$ exist, then*

$$\tfrac{1}{2}p_{A\cup\{x,y\}} \oplus \tfrac{1}{2}p_A \succeq p_{A\cup\{x\}}.$$

Theorem 7. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$ that satisfies Best/Worst and Best/Worst*, and let $< S, \mu, u, \mathcal{C} >$ be a Content-Monotone Thinking-Averse representation of $\succeq$. Then, $\succeq$ satisfies Anonymity if and only if there exist a function $c : \mathbb{N} \to \mathbb{R}$ s.t. for all $A \in \mathcal{X}$*

$$\mathcal{C}(A) = c(|A|).$$

Moreover, $\succeq$ satisfies Anonymity and SHIA if and only if such a c is increasing. *Finally, $\succeq$ satisfies Anonymity and ISC if and only if such a c is* increasing and convex.

5. Conclusion

In this paper we analyze a preference relation on lotteries over menus characterized by the presence of a tradeoff. On the one hand, the agent prefers a larger set since it gives her more options to choose from. On the other hand, she prefers a smaller set since she wishes to avoid the disutility associated with having to think about what to choose from a larger set. We impose novel axioms that allow us to separate two distinct components of these preferences. The first component represents the agent's ranking of menus if she had no cost of thinking - which we called the genuine preference over the content of menus. The second component is the disutility of making a choice from that menu. We then formally define the notion of *thinking aversion* in a manner similar to the definitions of risk or ambiguity aversion.

We then turn to characterize a Thinking-Averse preference relation. We present an axiomatic structure built around Thinking Aversion, and show that it is equivalent to the existence of a representation characterized by the difference between a genuine evaluation of the content of the set, modeled in a way similar to DLR01, and an evaluation of the cost of thinking about the set, which is zero for singletons, positive everywhere else and concave.

We further strengthen this characterization in two ways. First, we show that simply adapting the axioms in Kreps (1979) to the genuine evaluation of the content of a menu allows us to obtain a characterization in which the content is evaluated monotonically. Then, we strengthen the characterization of the function that represents the cost of thinking about a set, focusing on two possible interpretations. First, the cost could be understood as the cost to *search* for the best option within a menu for an agent who knows her preferences. Second, it could be the cost of understanding what her preferences actually are. We offer behavioral axioms that allow us to characterize the cost as the *sum* of two components: 1) an increasing function of the cardinality of the set; 2) the cost to obtain the right partition of an endogenous state space necessary to make a choice.

Future research could analyze the possible implications of these preferences in a standard economic environment. One such analysis is presented in Ortoleva (2008), which shows that an adapted version of this model applied to portfolio choice could allow us to explain some behavioral anomalies observed in the financial market. In particular, it could explain the tendency of agents to avoid the stock market when they are facing too many options, or to choose naïve diversification strategies.

Appendix A: Preliminaries

A.1. A mapping result: connecting the two spaces

The content of this section is to build a connection between the space that we use in this paper, lotteries over menus, and the space used in most of the literature, menus of lotteries. (Recall that we denote by $\hat{\mathcal{X}}$ is the set of compact subsets of the simplex $\Delta(X)$, metrized with the Hausdorff metric.)

Lemma 1. *Let X be a finite set. Then, there exist an affine and continuous bijection between $\Delta(\mathcal{X})$ and a compact and convex subset of $\hat{\mathcal{X}}$.*

The proof goes as follows. Notice that any element of $\Delta(\mathcal{X})$ can be written as $\bigoplus_{i=0}^{N} \alpha_{C_i} C_i$ where $C_1, \dots, C_N \in \mathcal{X}$, $\alpha_{C_i} \in [0,1]$ for all i, and $\sum_{i=1}^{N} \alpha_i = 1$. Define $g : \Delta(\mathcal{X}) \to \hat{\mathcal{X}}$ as

$$g(\bigoplus_{i=0}^{N} \alpha_{C_i} C_i) := \sum \alpha_{C_i} \mathrm{conv}(C_i).$$

(Where $\sum$ in $\hat{\mathcal{X}}$ is understood in the standard sense of set mixing.) Define now H as the range of g, that is, $H := \{\hat{A} \in \hat{\mathcal{X}} : \hat{A} = g(A) \text{ for some } A \in \Delta(\mathcal{X})\}$.

Claim 1. g is a bijection between $\Delta(\mathcal{X})$ and H.

Proof. We only need to prove that for any $A, B \in \Delta(\mathcal{X})$, we have $g(A) \neq g(B)$. To prove it we shall proceed by induction on the cardinality n of X. First notice that when $n = 1$ the result is trivially true. Then, take X of cardinality n, take any $A, B \in \Delta(\mathcal{X})$, $A \neq B$ and say, by means of contradiction, that $g(A) = g(B)$. Further, as we argued before write $A = \bigoplus_{i=0}^{N} \alpha_{C_i} C_i$ and $B = \bigoplus_{i=0}^{N} \beta_{C_i} C_i$. First, notice that for all $x \in X$, we must have $\alpha_{\{x\}} = \beta_{\{x\}}$. To see why, say that, instead, we had $\alpha_{\{x\}} < \beta_{\{x\}}$. But then, it means that any lottery in $g(B)$ must give a minimum weight of $\beta_{\{x\}}$ to x, while there are lotteries in $g(A)$ which could give a lower weight, hence we would not have $g(A) = g(B)$. So, we have shown that we must have $\alpha_{\{x\}} = \beta_{\{x\}}$ for all $x \in X$.

Now, for any $x \in X$, define $C(x) := \{C \in \mathcal{X} : x \in C\}$. This is the class of subsets of X that contain x. Now, define $\hat{A}^x \in \Delta(\mathcal{X})$ as follows. (Again, we are defining only the corresponding weights, $\hat{\alpha}^x$.)

$$\hat{\alpha}_C^x = \begin{cases} 0, & \text{if} \quad x \in C \\ \frac{\alpha_C + \alpha_{C \cup \{x\}}}{1 - \alpha_{\{x\}}}, & \text{otherwise} \end{cases}$$

It is easy to see that we have $\sum_{i=1}^{N} \hat{\alpha}_{C_i}^x = 1$. Define $\hat{B}^x$ analogously. There are now two possible scenarios. Either there exists $x \in X$ such that $\hat{A}^x \neq \hat{B}^x$, or $\hat{A}^x = \hat{B}^x$ for all $x \in X$.

Say first that there exists $x \in X$ such that $\hat{A}^x \neq \hat{B}^x$. Notice that we find two sets in $\Delta(2^{X \setminus \{x\}} \setminus \{\emptyset\})$, call them $\bar{A}^x$ and $\bar{B}^x$, that assign the same distribution as $\hat{A}^x$ and $\hat{B}^x$. (The reason is that both $\hat{A}^x$ and $\hat{B}^x$ have in the support only sets that do not contain x). We can further define the function $\bar{g}$ analogous to g but defined on $\Delta(2^{X \setminus \{x\}} \setminus \{\emptyset\})$. Now, notice that the set $X \setminus \{x\}$ has cardinality $n - 1$, and hence the assumption of induction implies that we have $\bar{g}(\bar{A}^x) \neq \bar{g}(\bar{B}^x)$. This implies that we can say (without loss of generality), that there exists $p \in \bar{g}(\bar{A}^x) \setminus \bar{g}(\bar{B}^x)$. But then, notice that this immediately implies that $\alpha_{\{x\}}\{x\} + (1 - \alpha_{\{x\}})p \in g(A)$, since I can always replicate the lottery p in $g(A)$ provided that I assign enough weight to the singleton $\{x\}$.[24] But, notice that we cannot have that $\alpha_{\{x\}}\{x\} + (1 - \alpha_{\{x\}})p \in g(B)$, since it would imply that $p \in \bar{g}(\bar{B}^x)$ (since $\alpha_{\{x\}} = \beta_{\{x\}}$), which we know is not true. We have shown that there cannot exist $x \in X$ such that $\hat{A}^x \neq \hat{B}^x$.

We then must have $\hat{A}^x = \hat{B}^x$ for all $x \in X$. Notice that, since $\alpha_{\{x\}} = \beta_{\{x\}}$ for all $x \in X$, this implies that $\alpha_D + \alpha_{D \cup \{x\}} = \beta_D + \beta_{D \cup \{x\}}$ for all $x \in X$, $D \in \mathcal{X}$. Now, consider any $y \in X$, and notice that this implies that we have $\alpha_{\{y\}} + \alpha_{\{y\} \cup \{x\}} = \beta_{\{y\}} + \beta_{\{y\} \cup \{x\}}$, which in turns implies, since again $\alpha_{\{x\}} = \beta_{\{x\}}$ for all $x \in X$, that we have $\alpha_{\{x,y\}} = \beta_{\{x,y\}}$ for all $x, y \in X$. Then do the same for the set of three elements, and so on. This implies that $\alpha_D = \beta_D$ for all $D \in \mathcal{X}$, which means $A = B$, a contradiction. □

Claim 2. g is linear and continuous.

Proof. To prove continuity we need to prove that, for any $(A_n) \in \Delta(\mathcal{X})^\infty$, $A \in \Delta(\mathcal{X})$ with $A_n \to A$, we have $g(A_n) \to g(A)$. Denote $A_n = \bigoplus \alpha_i^n C_i$ and $A = \bigoplus \alpha_i C_i$ and notice that $A_n \to A$ implies $\alpha_i^n \to \alpha_i$ for all i.[25] But it is immediate to see that,in the Hausdorff topology of $\hat{\mathcal{X}}$, $\alpha_i^n \to \alpha_i$ for all i implies that $\sum \alpha_i^n \text{conv}(C_i) \to \sum \alpha_i \text{conv}(C_i)$.

To prove linearity, take any $A, B \in \Delta(X)$, $\gamma \in (0, 1)$. Write $A := \bigoplus \alpha_i C_i$ and $B := \bigoplus \beta_i C_i$, and notice that we must have $\gamma A \oplus (1 - \gamma)B = \bigoplus [\gamma \alpha_i + (1 - \gamma)\beta_i] C_i$. But notice that we

[24] To do so, simply consider, in the mixture, the same sets and elements we considered to create p.

[25] This comes from the way we metrize $\Delta(X)$.

must then have that $g(\gamma A + (1-\gamma)B) = \sum[\gamma\alpha_i + (1-\gamma)\beta_i]\mathrm{conv}(C_i) = \gamma[\sum \alpha_i \mathrm{conv}(C_i)] + (1-\gamma)[\sum \beta_i \mathrm{conv}(C_i)] = \gamma g(A) + (1-\gamma)g(B)$, which in turns proves linearity. □

Finally, notice that linearity and continuity of g guarantee the convexity and compactness of H. This concludes the proof of Lemma 1. *Q.E.D.*

A.2 Mapping of the representations

We now show that we can obtain a representation of $\succeq$, defined on $\Delta(\mathcal{X})$, that is reminiscent of what DLR01 call an Additive EU representation.(We lose, however, the main goal of DLR01: we no longer have uniqueness of the state space S.)

Definition 8. A preference relation $\succeq$ on $\Delta(\mathcal{X})$ satisfies independence if, for all $A, B, C \in \Delta(\mathcal{X})$, $\alpha \in [0,1]$

$$A \succeq B \Leftrightarrow \gamma A \oplus (1-\gamma)C \succeq \gamma B \oplus (1-\gamma)C.$$

Lemma 2. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$. Then, the following two conditions are equivalent:*

(i) $\succeq$ on $\Delta(\mathcal{X})$ satisfies Full Continuity (A.5) and independence;*

(ii) there exists a non-empty, finite set S of state of the world, a state-dependent utility $u : X \times S \to \mathbb{R}$ and a signed measure μ over S such that it is represented by

$$U(\bigoplus \alpha_i A_i) = \sum \alpha_i [\sum_{s \in S} \mu(s)[\max_{y \in A_i} u(y;s)]]$$

To prove this result we first "translate" our preference relation to the space used in DLR01, $\hat{\mathcal{X}}$. By Lemma 1, we know that we have a continuous bijection g between $\Delta(\mathcal{X})$ and a compact and convex subset H of $\hat{\mathcal{X}}$. Recall that we have $N = |X|$. Consider now the following set of utilities: $\bar{\mathcal{U}} := \{u \in \mathbb{R}^{\Delta(X)} : u$ is continuous, affine, $\max_{y \in \Delta(X)} u(y) = 1$, $\min_{y \in \Delta(X)} u(y) = 0$, $\exists x_1, x_2, \ldots, x_{N-1} \in X$ such that $u(x_1) = u(x_2) = \ldots = u(x_{N-1})$ $\}$. (Geometrically they are the utilities that generate indifference curves that are parallel to each of the faces of the simplex.) Notice that $|\bar{\mathcal{U}}| < \infty$. We will now show that these utilities are, in fact, enough to characterize our preference relation. Geometrically, we are simply going to show that we can separate every set $A \in H$ from every point outside of it (but still in the simplex) by means of one of those utilities. (Recall that we understand every $A \in H$ as $A \subseteq \mathbb{R}^{N-1}$.)

Claim 3. For any $A \in H$, $y \in \mathbb{R}^{N-1}$ with $y \notin A$, there exists $u \in \bar{\mathcal{U}}$ such that $\max_{x \in A} u(x) < u(y)$.

Proof. To prove the claim, notice that, by construction of H, the set of extreme points of H, denoted by $\mathrm{ext}(H)$, is the g-image of $\mathcal{X}$. Recall the geometrical intuition of the elements of $\bar{\mathcal{U}}$: they are the utilities whose indifferent curves are parallel to the face of the simplex. Given this intuition, it is trivial to show that the claim holds for $\Delta(X)$: simply, a point outside of it can be separated from $\Delta(X)$ by means of an hyperplane parallel to the appropriate face; but then, if this hyperplane is the indifference curve of a utility function u which increases in the direction of y, we must have $\max_{x \in \Delta(X)} u(x) < u(y)$. If $A \in \mathrm{ext}(H)$ and A is a face of the simplex (i.e., $|A| = N - 1$), then the same reasoning applies. If $A \in \text{ext}(H)$ but A is not a face of the simplex, we still know that A must be the $g - image$ of some element of $\mathcal{X}$, hence A must be the intersection of two or more faces of the simplex. But then again, one of the hyperplanes parallel to those faces must do. This proves that the claim is true for all $A \in \text{ext}(H)$.

Notice that any $A \in \text{ext}(H)$ is a polyhedron in $\mathbb{R}^{N-1}$. Moreover, notice that what we have just proven is *equivalent* to saying that, for any face F of A, there exist $u \in \bar{\mathcal{U}}$ such that $u(x) = u(y)$ for all $x, y \in F$. We now turn to prove that this is true for all $A \in H$. To do so, consider first two sets $B, C \in \text{ext}(H)$ and $\lambda \in (0, 1)$, and define $D := \lambda B + (1-\lambda)C$. Consider any face F of D, and notice that it must be either a subset of the mixture of a face F' of B and $x' \in C$, or of a face F'' of C and $x'' \in B$. Say that it is the first case (the second case is analogous). Then, we know that there exist $u \in \bar{\mathcal{U}}$ such that $u(y) = u(z)$ for all $y, z \in F'$. By linearity of u, it must be the case that the same is true if the elements are mixed with a fixed element $x' \in C$, which means that $u(r) = u(s)$ for all $r, s \in F$. This proves that the claim is true for $A \in H$ such that it is the convex combination of two elements in $\text{ext}(H)$. Repeat this argument to show that this is true for any $A \in H$ such that it is the convex combination of finitely many elements in $\text{ext}(H)$. But this is the entire H, and this concludes the proof. □

(Geometrically, what we have just proved is that we can separate all the sets in H by means of hyperplanes parallel to the face of the simplex.) By standard arguments, it is now trivial to show that we can therefore map any $A \in H$ *onto* a subset C of $\mathbb{R}^{|\bar{\mathcal{U}}|}$ (recall that $|\bar{\mathcal{U}}| < \infty$ by finiteness of X): simply associate every set to the vector that has the utility given by the set in every $u \in \bar{\mathcal{U}}$. Call this map h. Again, standard arguments show that h is a linear, continuous bijection. We have therefore a linear and continuous bijection $\gamma := g \circ h$ from $\Delta(X)$ to C. Now, define the preference relation $\hat{\succeq}$ on C by

$$\gamma(A) \hat{\succeq} \gamma(B) \Leftrightarrow A \succeq B.$$

Since γ is a linear and continuous bijection, $\hat{\succeq}$ preserves the affinity and continuity of $\succeq$. We have therefore a linear and continuous preference relation on a subset of $\mathbb{R}^{|\bar{\mathcal{U}}|}$. It is standard practice to show that there exist a set $\mathcal{U} \subseteq \bar{\mathcal{U}}$ (finite) and a signed measure μ on $\mathcal{U}$ such that, for any $x, y \in C$

$$x \hat{\succeq} y \Leftrightarrow \sum_{u_i \in \mathcal{U}} \mu(u_i) x_i \geq \sum_{u_i \in \mathcal{U}} \mu(u_i) y_i.$$

But then, by definition of $\hat{\succeq}$ and since γ is a bijection, we have

$$\sum_j \alpha_j A_j \succeq \sum_j \beta_j B_j \Leftrightarrow \sum_j \alpha_j [\sum_{u_i \in \mathcal{U}} \mu(u_i) \max_{x \in \text{conv}(A_j)} u_i(x)] \geq \sum_j \beta_j [\sum_{u_i \in \mathcal{U}} \mu(u_i) \max_{x \in \text{conv}(B_j)} u_i(x)].$$

Since, by affinity of u, we have $\max_{x \in \text{conv}(A_j)} u_i(x) = \max_{x \in A_j} u_i(x)$, this concludes the proof of Lemma 2. *Q.E.D.*

A.3 Extending Kreps (1979) to lotteries of menus

In order to prove Theorem 3 we need to extend the representation of Kreps (1979) to the case of lotteries over menus. In particular, we obtain a representation that is the extension of the representation in Kreps (1979) in a vNM sense. Of note, this very representation has been characterized in Nehring (1996) by means of one novel axiom (indirect stochastic dominance). By contrast, we prove here that the same representation can be derived imposing that the axioms of Kreps (1979) on the degenerate lotteries.

To prove this result, we use the following Lemma, which is an extension of Lemma 3 in Kreps (1979): for completeness we include the full proof. (The core idea is to show that the representation in Kreps (1979) is "so" not-unique that we can assign any utility value needed.)

Lemma 3. *Let Y be an arbitrary finite set endowed with two binary relation $\succeq$ and $\trianglerighteq$ such that:*

1. *$\succeq$ is complete and transitive;*
2. *$\trianglerighteq$ is reflexive;*
3. *$y \trianglerighteq y'$ and $y \neq y'$ imply not $y' \succeq y$.*

Then, for any utility representation U of $\succeq$ such that $U(y) < 0$ for any $y \in Y$, there exist negative numbers $a(y)$ such that $U(y') = \sum_{\{y: y \trianglerighteq y'\}} a(y)$.

Proof. To prove it, let $\sim$ and $\succ$ denote the symmetric and asymmetric parts of $\succeq$ and $\triangleright$ the asymmetric part of $\trianglerighteq$. Notice that $\succeq$ is a weak preference relation, and that (by (3)) we have $y \triangleright y' \Rightarrow y \succ y'$. Define w and w^* as $w(y') := \sum_{\{y: y \trianglerighteq y'\}} a(y)$ and $w^*(y') := \sum_{\{y: y \triangleright y'\}} a(y)$. Clearly we have $w(y') = a(y') + w^*(y')$. We now find the constants $a(y)$ inductively. First look at the $\sim$-equivalence class of the $\succeq$-preferred elements in Y. Define $a(y) = U(y)$ for any y in this equivalence class. (Since U represents $\succeq$, the value is always the same.) Now proceed downward in the $\sim$-equivalence classes. Note that once $a(y)$ are defined for all $y \succ y'$, $w * (y')$ is fixed. (This happens because $y \trianglerighteq y' \Rightarrow y \succ y'$. But we have already defined $a(y)$ for all for all $y \succ y'$.) Now assign $a(y') = U(y') - w^*(y')$. Notice that we must have that $a(y') + w^*(y')$ is the same for all y' in the same equivalence class, since $a(y') + w^*(y') = U(y')$ and U represents $\succeq$. For the same reason we have $a(y') + w^*(y') \leq a(y) + w^*(y)$ for any $y \succeq y'$. Since Y is finite, there are finitely many $\sim$-equivalence classes, and the induction procedure gives the representation. $\square$

We now need two additional definitions. Consider a preference relation $\succeq$ on $\Delta(\mathcal{X})$. (For any $A \in \mathcal{X}$, by δ_A we understand the dirac measure on A.)

Definition 9. A preference relation $\succeq$ on $\Delta(\mathcal{X})$ is **degenerate-monotone** if and only if for any $A, B \in \mathcal{X}$, $B \subseteq A$, we have $\delta_A \succeq \delta_B$.

Definition 10. A preference relation $\succeq$ on $\Delta(\mathcal{X})$ is **degenerate-submodular** if and only if for any $A, B, C \in \mathcal{X}$, $\delta_A \sim \delta_{A \cup B}$ implies $\delta_{A \cup B \cup C} \succeq \delta_{B \cup C}$.

We are now ready to state the main Lemma of the section.

Lemma 4. *Let $\succeq$ be a complete preference relation on $\Delta(\mathcal{X})$. Then, the following two conditions are equivalent:*

(i) *$\succeq$ satisfies continuity, independence, degenerate-monotonicity and degenerate-submodularity;*

(ii) *there exist a finite set S of state of the world, a state-dependent utility $u : X \times S \to \mathbb{R}$ and a probability measure μ over S such that it is represented by*

$$U(\bigoplus \alpha_i A_i) = \sum \alpha_i [\sum_{s \in S} \mu(s)[\max_{y \in A_i} u(y;s)]]$$

The if part is either standard or trivial. To prove the only if part, notice that by affinity and independence there exist $V : \mathcal{X} \to \mathbb{R}$ such that for any $\bigoplus \alpha_i A_i, \bigoplus \beta_i B_i \in \Delta(\mathcal{X})$, we have

$$\bigoplus \alpha_i A_i \succeq \bigoplus \beta_i B_i \Leftrightarrow \sum \alpha_i V(A_i) \geq \sum \beta_i V(B_i).$$

Now define by $\bar{\succeq}$ the restriction of $\succeq$ on $\mathcal{X}$. We now wish to show that there exist a finite non-empty set S an affine state-dependent utility $u : X \times S \to \mathbb{R}$ and a probability measure μ over S such that

$$\bar{U}(A) = \sum_{s \in S} \mu(s)[\max_{y \in A} u(y;s)]$$

represents $\bar{\succeq}$ and such that there exists $\beta \in \mathbb{R}$ such that $\bar{U}(A) = V(\delta_A) + \beta$ for all $A \in \mathcal{X}$. But notice that this claim is almost identical to Theorem 1 in Kreps (1979), with the exception of the last requirement. In fact, to prove it one could follow almost identical passages, but using Lemma 3 in this paper instead of Lemma 3 in Kreps (1979). This guarantee the fact that $\bar{U}(A) = V(A) + \beta$ for all $A \in \mathcal{X}$ for some $\beta \in \mathbb{R}$. The representation then follows immediately. *Q.E.D.*

Appendix B: Proofs of the results in the text

B.1. Proof of Theorem 1 and 3

Only if direction The proof of both theorems proceeds as follows: 1) we extend $\succeq^*$ to $\Delta(\mathcal{X})$ by linearity, and characterize such extension using Lemma 2 (for Theorem 1) or Lemma 4 (for Theorem 3); 2) we turn to characterize $\succeq$: first we characterize it with a function linear on the singletons, then we normalize this with the representation of $\succeq^*$ so that they coincide on singletons; 3) we show that emerging characterization must have the desired properties. For simplicity, we analyze the consequences of both continuity postulates (A. 5 and A. 5*) at the same time, emphasizing the differences whenever there is any.

We start from the characterization of $\succeq^*$. Notice that $\succeq^*$ is complete by construction, and that it is transitive by A. 2.

Claim 4. $\succeq^*$ agrees with $\succeq$ on $\Delta^S(\mathcal{X})$. That is, for any $p, q \in \Delta^S(\mathcal{X})$, $\{p\} \succeq^* \{q\} \Leftrightarrow \{p\} \succeq \{q\}$.

Proof. Take $p, q \in \Delta^S(\mathcal{X})$. Say that we have $(\frac{1}{2}+\epsilon)p + (\frac{1}{2}-\epsilon)q \succ \frac{1}{2}p + \frac{1}{2}q \succ (\frac{1}{2}-\epsilon)p + (\frac{1}{2}+\epsilon)q$ for some $\epsilon > 0$. By linearity of $\succeq$ on $\Delta^S(\mathcal{X})$ (A. 1), this is true if and only if $p \succeq q$, proving the claim. □

Claim 5. If $\succeq$ satisfies A. 5*. then it satisfies A. 5.

Proof. Assume that $\succeq$ satisfies A.5*. It is trivial to show that we must then have that for any $A \in \Delta(\mathcal{X})$, the sets $\{p \in \Delta^S(\mathcal{X}) : p \succeq A\}$ and $\{p \in \Delta^S(\mathcal{X}) : A \succeq p\}$ are closed. Now consider $A \in \mathcal{X}$ and $p^n \in (\Delta^S(X))^\infty$, $p \in \Delta^S(X)$ s.t. $p_n \to p$ and $p_n \succeq^* A$. We need to show that $p \succeq^* A$. (The proof that $p_n \to p$ and $p_n \preceq^* A$ imply $p \preceq^* A$ is analogous.) Say, by means of contradiction, that we have $A \succ^* p$. This means that there exist $\bar{\epsilon} > 0$ s.t. for all $\epsilon < \bar{\epsilon}$ we have $(\frac{1}{2}+\epsilon)p \oplus (\frac{1}{2}-\epsilon)A \prec \frac{1}{2}p \oplus \frac{1}{2}A \prec (\frac{1}{2}-\epsilon)p \oplus (\frac{1}{2}+\epsilon)A$. Notice also that we have $p_n \succeq^* A$ for all n, which implies that, for each each n, either we have that for each $\hat{\epsilon} > 0$ there exist $\epsilon_n \in (0, \hat{\epsilon})$ s.t. $(\frac{1}{2}+\epsilon_n)p_n \oplus (\frac{1}{2}-\epsilon_n)A \succeq \frac{1}{2}p_n \oplus \frac{1}{2}A$, or that, for each $\hat{\epsilon}' > 0$, there exist $\epsilon'_n \in (0, \hat{\epsilon}')$ s.t. $\frac{1}{2}p_n \oplus \frac{1}{2}A \succeq (\frac{1}{2}-\epsilon'_n)p_n \oplus (\frac{1}{2}+\epsilon'_n)A$ (or both). Notice that it is always possible to build a subsequence (p_m) in which one of the two conditions is true for all m, and say that it is possible to construct a subsequence that satisfies the first one (the proof in case it is the second one is analogous and is therefore omitted for brevity). That is, construct a subsequence (p_m) s.t., for all m, and for each $\hat{\epsilon} > 0$ there exist $\epsilon_m \in (0, \hat{\epsilon})$ s.t. $(\frac{1}{2}+\epsilon_m)p_m \oplus (\frac{1}{2}-\epsilon_m)A \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$. We now claim that, for every m, we have that for every $\epsilon \in (0, \frac{1}{2})$, $(\frac{1}{2}+\epsilon)p_m \oplus (\frac{1}{2}-\epsilon)A \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$. To see why, notice that, since $\succeq$ is linear on $\Delta^S(\mathcal{X})$ by A. 1, then by A. 4, $\succeq$ is concave, i.e. for every $A, B \in \Delta(\mathcal{X})$, $\alpha \in (0,1)$ and every continuous utility representation v of $\succeq$, we must have $v(\alpha A \oplus (1-\alpha)B) \leq \alpha v(A) + (1-\alpha)v(B)$ (that is, v is convex). But then, since there exist $\epsilon_m > 0$ s.t. $(\frac{1}{2}+\epsilon_m)p_m \oplus (\frac{1}{2}-\epsilon_m)A \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$, then we must have $p_m \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$. In turns, again by A. 4, this implies that if $\dot{\epsilon}_m \in (0, \frac{1}{2})$ and $(\frac{1}{2}+\dot{\epsilon}_m)p_m \oplus (\frac{1}{2}-\dot{\epsilon}_m)A \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$, then we have $(\frac{1}{2}+\epsilon)p_m \oplus (\frac{1}{2}-\epsilon)A \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$ for all $\epsilon \in (\dot{\epsilon}_m, \frac{1}{2})$. But since, $\dot{\epsilon}_m$ can be arbitrarily close to 0, then we have $(\frac{1}{2}+\epsilon)p_m \oplus (\frac{1}{2}-\epsilon)A \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$ for all $\epsilon \in (0, \frac{1}{2})$. Notice that this is true for all m. Consider now any $\epsilon \in (0, \bar{\epsilon})$. Notice that we have $(\frac{1}{2}+\epsilon)p_m \oplus (\frac{1}{2}-\epsilon)A \succeq \frac{1}{2}p_m \oplus \frac{1}{2}A$ for all m, and $(\frac{1}{2}+\epsilon)p \oplus (\frac{1}{2}-\epsilon)A \prec \frac{1}{2}p \oplus \frac{1}{2}A$. But since $\succeq$ is complete and $p_m \to p$ and hence, in our topology, $\frac{1}{2}p_m \oplus \frac{1}{2}A \to \frac{1}{2}p \oplus \frac{1}{2}A$ and $(\frac{1}{2}+\epsilon)p_m \oplus (\frac{1}{2}-\epsilon)A \to (\frac{1}{2}+\epsilon)p \oplus (\frac{1}{2}-\epsilon)A$, then this contradicts continuity of $\succeq$. □

Claim 6. $\succeq^*$ is continuous.

Proof. Notice that within $\mathcal{X} \cup \Delta^S(\mathcal{X})$, in the topology we are using the only possible convergence is for elements of $\Delta^S(\mathcal{X})$. Therefore, we only need to show that for any $A \in \mathcal{X}$ the sets $\{p \in \Delta^S(\mathcal{X}) : p \succeq^* A\}$ and $\{p \in \Delta^S(\mathcal{X}) : A \succeq^* p\}$ are closed. Now, if $\succeq$ satisfies A. 5, then this is trivially true. (By Claim 5 this is true even if $\succeq$ satisfies A. 5*). □

Claim 7. For any $A \in \Delta(\mathcal{X})$, there exist $p_A \in \Delta^S(\mathcal{X})$ s.t. $p_A \sim A$. Moreover, for any $B \in \mathcal{X}$ s.t. $x^* \succeq^* B \succeq^* x_*$, there exist $p \in \Delta^S(X)$ s.t. $p \sim^* B$.

Proof. Consider any $A \in \Delta(\mathcal{X})$ and any $x^*, x_* \in \Delta^S(\mathcal{X})$ s.t. $x^* \succeq A \succeq x_*$. (Their existence is guaranteed by A. 6). It suffices to show that there exist $\lambda \in [0,1]$ s.t. $\lambda x^* \oplus (1-\lambda)x_* \sim A$. Say, by means of contradiction, that this is not the case. Then, define $\lambda^*, \lambda_* \in [0,1]$ as $\lambda^* := \min\{\lambda \in [0,1] : \lambda x^* \oplus (1-\lambda)x_* \succeq A\}$ and $\lambda_* := \max\{\lambda \in [0,1] : A \succeq \lambda x^* \oplus (1-\lambda)x_*\}$, and notice that both are well-defined by A. 5. Notice that we cannot have $\lambda^* = \lambda_*$, since it would imply $\lambda^* x^* \oplus (1-\lambda^*)x_* \sim A$, which we know is not true. Notice also that we cannot have $\lambda^* > \lambda_*$. If this were the case, consider any $\lambda' \in (\lambda_*, \lambda^*)$, and notice that we could not have $\lambda' x^* \oplus (1-\lambda')x_* \succeq A$ since this violates the definition of λ^* (it is the minimum λ s.t. this is true), nor $A \succeq \lambda' x^* \oplus (1-\lambda')x_*$ since this violates the definition of λ_* (it is the maximum λ s.t. this is true). Therefore, we must have $\lambda_* > \lambda^*$. Notice that therefore we have

$\lambda^* x^* \oplus (1-\lambda^*)x_* \succ A \succ \lambda_* x^* \oplus (1-\lambda_*)x_*$, $\lambda_* > \lambda^*$, and $x^* \succeq x_*$. But this is a violation of independence of $\succeq$ on $\Delta^S(\mathcal{X})$, A. 1, since it implies that if $\lambda_* > \lambda^*$ and $x^* \succeq x_*$, then $\lambda_* x^* \oplus (1-\lambda_*)x_* \succ A \succeq \lambda^* x^* \oplus (1-\lambda^*)x_*$. The proof of the second part of the claim is analogous given the continuity of $\succeq^*$. □

Claim 8. There exist $\bar{U} : \mathcal{X} \cup \Delta^S(\mathcal{X}) \to \mathbb{R}$ that represents $\succeq^*$ and that is continuous and affine on $\Delta^S(\mathcal{X})$.

Proof. We will construct $\bar{U}$ as follows. Consider any continuous and affine representation v of $\succeq^*$ restricted to $\Delta^S(\mathcal{X})$. Notice that we have $x^* \succeq^* p \succeq^* x_*$. By Claim 7, for any $A \in \mathcal{X}$ s.t. $x^* \succeq^* A \succeq^* x_*$ there exist some $p^*_A \in \Delta^S(X)$ s.t. $p^*_A \sim^* A$. Now, for any $p \in \Delta^S(X)$, set $\bar{U}(p) = v(p)$, and for any $A \in \mathcal{X}$ s.t. that there exist $p^*_A \in \Delta^S(X)$ s.t. $p^*_A \sim^* A$, set $\bar{U}(p) = v(p^*_A)$. Finally, consider any $A \in \mathcal{X}$ s.t. $A \succ^* x^*$ for $x_* \succ^* A$. By finiteness of X, there are only finitely many of these sets: construct therefore any utility representation h for them normalized s.t. $h(A) > v(x^*)$ for any $A \in \mathcal{X}$ such that $A \succ^* x^*$, and s.t. $h(A) < v(x_*)$ for any $A \in \mathcal{X}$ such that $x_* \succ^* A$. Now set $\bar{U}(A) = h(A)$ for these elements. □

We now extend $\bar{U}$ to $\Delta(\mathcal{X})$ in an affine and continuous way. Define $U : \Delta(\mathcal{X}) \to \mathbb{R}$ as

$$U(\bigoplus_i \alpha_i A_i) := \sum \alpha_i \bar{U}(A_i).$$

It is easy to show, by means of standard arguments, that it will also be continuous and, by construction, affine. We only need to make sure that, for any $p \in \Delta^S(\mathcal{X})$ we have $\bar{U}(p) = U(p)$. But this is a trivial consequence of the affinity of U on $\Delta^S(\mathcal{X})$. This means that we have U defined on the whole $\Delta(\mathcal{X})$ such that $U(A) = \bar{U}(A)$ for any $A \in \mathcal{X} \cup \Delta^S(\mathcal{X})$.

Now, define $\hat{\succeq}^*$ as the preference relation induced on $\Delta(\mathcal{X})$ by U, that is

$$A \,\hat{\succeq}^*\, B \;\Leftrightarrow\; U(A) \geq U(B).$$

Notice that $\hat{\succeq}^*$ is a continuous and linear extension of $\succeq^*$ to $\Delta(\mathcal{X})$ (by construction of U).

We now turn to characterize $\hat{\succeq}^*$. For Theorem 1 we use Lemma 2, while for Theorem 3 we use Lemma 4. In both cases, there exist a nonempty, finite set S, a state-dependent utility function $u : X \times S \to \mathbb{R}$ and a signed measure μ over S such that

$$U(\bigoplus_i \alpha_i A_i) = \sum_i \alpha_i [\sum_{s \in S} \mu(s)[\max_{y \in A_i} u(y; s)]].$$

In the case of Theorem 3, μ is a probability measure.

We now turn to characterize the general preference relation $\succeq$.

Claim 9. There exists $W : \Delta(\mathcal{X}) \to \mathbb{R}$ such that:

1. W represents $\succeq$;
2. W is affine and continuous when restricted to $\Delta^S(\mathcal{X})$;
3. W is convex;
4. $W(A) \leq U(A)$ for all $A \in \Delta(\mathcal{X})$;
5. $W(p) = U(p)$ for all $p \in \Delta^S(\mathcal{X})$;

6. W is continuous if $\succeq$ satisfies Axiom 5*.

Proof. Consider first the restriction of $\succeq$ on $\Delta^S(\mathcal{X})$, and represent it with an affine and continuous $\hat{W} : \Delta^S(\mathcal{X}) \to \mathbb{R}$ (the existence of such $\hat{W}$ is guaranteed by Axioms 5 and 1). By Claim 7, for any $A \in \Delta(\mathcal{X})$ there exists $p_A \in \Delta^S(\mathcal{X})$ such that $A \sim p_A$. Now, define $W : \Delta(\mathcal{X}) \to \mathbb{R}$ as

$$W(A) := \hat{W}(p_A).$$

Notice that W is affine and continuous on $\Delta^S(\mathcal{X})$ and represent $\succeq$. Now, normalize W such that $W(p) = U(p)$ for all $p \in \Delta$. Notice that, since by Claim 4 $\succeq$ and $\succeq^*$ agree on $\Delta^S(\mathcal{X})$, such normalization exists.

We now turn to prove that W is convex. We need to show that, for any $A, B \in \Delta(\mathcal{X})$ and $\alpha \in (0, 1)$ we have $W(\alpha A \oplus (1-\alpha)B) \leq \alpha W(A) + (1-\alpha)W(B)$. Define p_A and p_B as above and notice that $\alpha W(A) + (1-\alpha)W(B) = \alpha W(p_A) + (1-\alpha)W(p_B) = W(p_A \oplus (1-\alpha)p_B) \geq W(\alpha A \oplus (1-\alpha)B)$, where the last inequality comes *directly* from A. 4.

We now prove that we have $W(A) \leq U(A)$ for all $A \in \Delta(\mathcal{X})$. To see this, notice that for any $A \in \mathcal{X} \cup \Delta^S(\mathcal{X})$, we have $A \sim p_A$ (where p_A has been defined above). Now, say that there exists $A \in \mathcal{X} \cup \Delta^S(\mathcal{X})$ such that $W(A) > U(A)$. Then, we have $W(A) = W(p_A) = U(p_A) > U(A)$. Hence, we have $p_A \succ^* A$ but $p_A \sim A$. But this clearly violates A. 3.

This implies that, for any $A \in \mathcal{X} \cup \Delta^S(\mathcal{X})$, we have $W(A) \leq U(A)$. But clearly convexity of W guarantees that the same is true for any $A \in \Delta(\mathcal{X})$. To see this, take $A, B \in \mathcal{X} \cup \Delta^S(\mathcal{X})$, $\alpha \in (0, 1)$. Notice that we have $U(\alpha A \oplus (1-\alpha)B) = \alpha U(A) + (1-\alpha)U(B) \geq \alpha W(A) + (1-\alpha)W(B) \geq W(\alpha A \oplus (1-\alpha)B)$, where the first equality comes from linearity of U, the last one from convexity of W. Since this clearly extends to larger mixtures, we have $W(A) \leq U(A)$ for all $A \in \Delta(\mathcal{X})$ as sought.

Finally, notice that if $\succeq$ is continuous (satisfies Axiom 5*), then the function W thus constructed must be continuous. To see this, notice that continuity of $\succeq$ implies that for any $A_n \in (\Delta(\mathcal{X}))^\infty$, $A \in \Delta(\mathcal{X})$, $A_n \to A$, and $p_{A_n} \to p_A$ for some $(p_{A_n}) \in (\Delta^S(\mathcal{X}))^\infty$ and $p_A \in \Delta^S(\mathcal{X})$, if $p_{A_n} \sim A_n$ for all n, then $p_A \sim A$. But then $W(A_n) = W(p_{A_n}) \to W(p_A) = W(A)$ as sought. □

Define now the function $C : \mathcal{X} \to \mathbb{R}$ as

$$C(A) := W(A) - U(A).$$

By construction we have $C(\{p\}) = 0$ for all $p \in \Delta(X)$, and $C(A) \geq 0$ for all $A \in \mathcal{X}$. Also, it is clearly concave (since W is convex and U is linear). And, in the case in which $\succeq$ is continuous (Axiom 5*), then $\mathcal{C}$ must be continuous, since W in this case is continuous.

If direction Take the representation as given. For simplicity, define $U : \Delta(\mathcal{X}) \to \mathbb{R}$ as

$$U(A) := \sum_i \alpha_i [\sum_{s \in S} \mu(s) [\max_{y \in A_i} u(y; s)]].$$

It is trivial to see that U is continuous and affine. A. 5 (Continuity) is immediate from the continuity of U and the existence of a utility representation W of $\succeq$. A. 1 (Independence over Singletons) is trivial since $\mathcal{C}(\{p\}) = 0$ for all $p \in \Delta^S(\mathcal{X})$ and since U is affine. A. 2 (Coherence), derives from the fact that U represents $\succeq^*$ (it implies that $\succeq^*$ is complete, transitive). To show A. 3 (Thinking Aversion), simply notice that, if $U(p) > U(A)$, since $\mathcal{C}(p) = 0$ and $\mathcal{C}(A) \geq 0$, we have $U(p) - \mathcal{C}(p) > U(A) - \mathcal{C}(A)$ as sought. A. 4 (Mixture Aversion) is an immediate consequence of the linearity of U and concavity of $\mathcal{C}$. Finally, if $\mathcal{C}$ is continuous, then W is continuous as well, which means that Axiom 5* is satisfied. The additional axioms for Theorem 3 are also trivially satisfied.

This concludes the proof of Theorem 1 and 3. *Q.E.D.*

B.2. Proof of Theorem 2

Let us consider the first part. First we need to show that there is a unique linear extension of $\succeq^*$ (defined on $\mathcal{X} \cup \Delta^S(\mathcal{X})$) to $\hat{\succeq}^*$ defined on $\Delta(\mathcal{X})$. To see the uniqueness, notice that, as proven in Claim 4, $\succeq^*$ will be affine on $\Delta^S(\mathcal{X})$. Furthermore, notice the following.

Claim 10. There exist $x^*, x_* \in \Delta(X)$ such that $x^* \succeq^* A \succeq^* x_*$ for all $A \in \mathcal{X} \cup \Delta^S(\mathcal{X})$.

Proof. The existence of x^* is directly assumed in A. 7. To show the existence of x_*, say, by means of contradiction, that there exist $A \in \mathcal{X}$ such that $y \succ^* A$ for all $y \in \Delta^S(\mathcal{X})$. (If such A belonged to $\Delta^S(\mathcal{X})$ it would prove the claim.) Then, by A. 3, we must also have $y \succ A$ for all $y \in \Delta^S(\mathcal{X})$. But this contradicts A. 6. □

Now notice that, since $\succeq^*$ is affine on $\Delta^S(\mathcal{X})$ and continuous (as proven in Claim 6), and given the existence of x^* and x_* proven in Claim 10, then it is standard to prove that for any $A \in \mathcal{X} \cup \Delta^S(\mathcal{X})$, there exist $\lambda \in [0,1]$ such that

$$A \sim^* \lambda x^* \oplus (1-\lambda)x_*.$$

Define this λ_A. But then, it is again standard practice to show that, for any linear extension $\hat{\succeq}^*$ of $\succeq^*$, we must be such that, for any $A, B \in \mathcal{X}$, $\alpha \in (0,1)$,

$$\alpha A \oplus (1-\alpha)B \hat{\sim}^* (\alpha\lambda_A + (1-\alpha)\lambda_B)x^* \oplus (\alpha(1-\lambda_A) + (1-\alpha)(1-\lambda_B))x_*.$$

But this immediately implies that this linear extension is bound to be unique. We have therefore a unique affine preference relation: standard results guarantee that it is unique up to a positive affine transformation. This proves the first part.

We now turn to prove the uniqueness of $\mathcal{C}$. Define $U : \Delta(\mathcal{X}) \to \mathbb{R}$

$$U(A) := \sum_i \alpha_i [\sum_{s \in S} \mu(s)[\max_{y \in A_i} u(y;s)]]$$

and $U'(A) = \sum_i \alpha_i [\sum_{s \in S'} \mu'(s)[\max_{y \in A_i} u'(y;s)]]$. Notice that we have just shown that we have $U' = \gamma U + \beta$. Now, consider $A \in \mathcal{X}$ such that $A \sim \{p\}$ for some $p \in \Delta(X)$. Clearly, we must have $U(\{p\}) = U(A) - \mathcal{C}(A)$ and $U'(\{p\}) = U'(A) - \mathcal{C}'(A)$. But given that $U' = \gamma U + \beta$, we have $\gamma U(\{p\}) + \beta = \gamma U(A) + \beta - \mathcal{C}'(A)$, which yields $\mathcal{C}'(A) = \mathcal{C}(A)$. Clearly, if A. 6 holds, then this is true for all $A \in \mathcal{X}$. This concludes the proof of Theorem 2. *Q.E.D.*

B.3. Proof of Proposition 1

First of all, consider two Thinking-Averse representations of $\succeq_1$ and $\succeq_2$, $(\mathcal{U}_1, \mu_1, \mathcal{C}_1)$ and $(\mathcal{U}_2, \mu_2, \mathcal{C}_2)$ such that $\mathcal{U}_1 = \mathcal{U}_2$ and $\mu_1 = \mu_2$. (Their existence is guaranteed by $\succeq_1^* = \succeq_2^*$.) For simplicity, define

$$U_1(A) := \sum_i \alpha_i [\sum_{s \in S_1} \mu_1(s) [\max_{y \in A_i} u_1(y; s)]]$$

and U_2 analogously. It is immediate to see that we must have $U_1 = U_2$. Now define $W_1 = U_1 - \mathcal{C}_1$ and $W_2 = U_2 - \mathcal{C}_2$. Notice that (2) is equivalent to saying that $W_1 \leq W_2$. Moreover, it is immediate to see that we must have $W_1(p) = W_2(p)$ for any $p \in \Delta^S(\mathcal{X})$

To prove (1) $\Rightarrow$ (2), consider $A \in \Delta(\mathcal{X})$. As shown in the proof of Claim 9, there must exist $p_A \in \Delta^S(\mathcal{X})$ such that $p_A \sim_1 A$. Hence, we have $W_1(p_A) = W_1(A)$. Furthermore, since $U_1 = U_2$ and $\mathcal{C}_1(p_A) = 0 = \mathcal{C}_2(p_A)$, we have $W_2(p_A) = W_1(p_A) = W_1(A)$. Now, condition (1) and $p_A \sim_1 A$ imply that we have $A \succeq_2 p_A$, hence $W_2(A) \geq W_2(p_A)$, which gives us $W_2(A) \geq W_1(A)$ as sought.

To prove (2) $\Rightarrow$ (1), take any $A \in \Delta(\mathcal{X})$ and $p \in \Delta^S(\mathcal{X})$ such that $A \succeq_1 p$. But then, this together with (2), implies $W_2(A) \geq W_1(A) \geq W_1(p) = W_2(p)$, hence $A \succeq_2 p$ as sought. This concludes the proof of Proposition 1. *Q.E.D.*

B.4. Proof of Theorem 4

Notice that for any Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ and any $A, B \in \mathcal{X}$ we have $\frac{1}{2}p_A \oplus \frac{1}{2}p_B^* \succeq \frac{1}{2}p_A^* \oplus \frac{1}{2}p_B$ if and only if $\mathcal{C}(A) \leq \mathcal{C}(B)$. This remark makes the if direction trivial. To prove the only if direction, let us define the set NT as

$$NT := \{A \in \mathcal{X} : A \sim^* \{x\} \text{ for some } x \in A\}.$$

Define $\hat{c_s} : NT \to \mathbb{R}$ as $\hat{c_s}(A) := \mathcal{C}(A)$. Now construct $c_s : \mathcal{X} \to \mathbb{R}$ as $c_s(A) := \max\{c_s(B) : B \in NT, B \subseteq A\}$. (This is well defined since $\mathcal{X}$ is a finite set and every singleton belongs to NT.) Notice that for every $A, B \in NT$ s.t. $A \subseteq B$ we have $\mathcal{C}(A) \leq \mathcal{C}(B)$ by A.10, and therefore $c_s(A) \leq c_s(B)$. In turns, this implies that for any $A, B \in \mathcal{X}$ s.t. $A \subseteq B$ we have $c_s(A) \leq c_s(B)$. Notice also that, for every $x \in X$, $\{x\} \in NT$, and therefore $c_s(\{x\}) = \mathcal{C}(\{x\}) = 0$.

We now turn to construct $c_I : \Pi(S) \to \mathbb{R}$. To do so, we make use of the following claim. In what follows, for brevity, we present a proof that is built upon the non-uniqueness of the state space. (Alternative, more instructive proofs are possible but, to our knowledge, they would not bring stronger results.) For any finite set S and partitions $\pi, \pi' \in \Pi(S)$, we say that π and π' are not comparable if we have that π is neither finer then coarser then π'.

Claim 11. There exist a Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ and partitioning function $\mathcal{P}$ s.t. for any $A, B \in \mathcal{X} \backslash NT$, $\mathcal{P}(A)$ and $\mathcal{P}(B)$ are not comparable.

Proof. If $|S| = 1$, then $\mathcal{X} = NT$ and the claim is trivially true. Consider now the case in which $|S| > 1$. Notice that, for any $A \in \mathcal{X}$, $\mathcal{I}_{S,u}(A)$ is non-empty. Notice also that we can add to our state space a dummy state $\bar{s}$ in which all elements are indifferent. That is, if there exist a Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$, then there exist a monotone thinking averse representation $< S \cup \{\bar{s}\}, \mu', u', \mathcal{C} >$, where $\mu'(\bar{s}) > 0$ and $u'(x, \bar{s}) = u'(y, \bar{s})$ for all $x, y \in S$.[26]

Consider now any Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ and any partition function $\mathcal{P}_0$ for this representation. If for any $A, B \in \mathcal{X} \backslash NT$ we have that $\mathcal{P}_0(A)$ and $\mathcal{P}_0(B)$ are not comparable, we are done. Say that this is not the case for some $A, B \in \mathcal{X} \backslash NT$. Since $A, B \notin NT$, we must have $\mathcal{P}_0(A) \neq \{S, \emptyset\} \neq \mathcal{P}_0(B)$. Take $s_1, s_2 \notin S$ and consider a Content-Monotone Thinking-Averse representation $< S \cup \{s_1, s_2\}, \mu', u', \mathcal{C} >$ s.t. $\mu'(s_1) > 0$, $\mu'(s_2) > 0$, $u'(x, s_1) = u'(y, s_1)$ and $u'(x, s_2) = u'(y, s_2)$ for all $x, y \in X$. Call $S' = S \cup \{s_1, s_2\}$. Fix any $\bar{s} \in S$ and construct a partition function $\mathcal{P}_1$ as any partition function for this representation such that: 1) for all $C \in \mathcal{X}$, $C \neq A, B$, $\mathcal{P}_1(C)$ is identical to $\mathcal{P}_0(C)$ with the only difference that s_1, s_2 are grouped together with $\bar{s}$; 2) $\mathcal{P}_1(A)$ is identical to $\mathcal{P}_0(A)$ but assigns s_1 to the same group of $\bar{s}$ and s_2 to any other group (we now that two groups exist since $\mathcal{P}_0(A) \neq \{S, \emptyset\}$); 3) $\mathcal{P}_1(B)$ is identical to $\mathcal{P}_0(B)$ with the only difference that assigns s_2 to the same group of $\bar{s}$ and s_1 to any other group (we now that two groups exist since $\mathcal{P}_0(B) \neq \{S, \emptyset\}$) That is, define $\mathcal{P}_1$ as any partition function s.t.

$$\mathcal{P}_1(C) := \{\beta \subseteq S' : \beta \backslash \{s_1, s_2\} \in \mathcal{P}_0(C) \text{ and } \bar{s} \in \beta\} \cup \{\beta \subseteq S' : \beta \in \mathcal{P}_0(C) \text{ and } \bar{s} \notin \beta\}.$$

for all $C \in \mathcal{X}$, $C \neq A, B$ and

$$\mathcal{P}_1(A) := \{\beta \subseteq S' : \beta \backslash \{s_1\} \in \mathcal{P}_0(A) \text{ and } \bar{s} \in \beta\} \cup \{\beta \subseteq S' : \beta \backslash \{s_2\} \in \mathcal{P}_0(A) \text{ and } \bar{s} \notin \beta\},$$

$$\mathcal{P}_1(B) := \{\beta \subseteq S' : \beta \backslash \{s_2\} \in \mathcal{P}_0(B) \text{ and } \bar{s} \in \beta\} \cup \{\beta \subseteq S' : \beta \backslash \{s_1\} \in \mathcal{P}_0(B) \text{ and } \bar{s} \notin \beta\}.$$

Notice that $\mathcal{P}_1(A)$ and $\mathcal{P}_1(B)$ are not comparable.

Repeat this procedure, defining $\mathcal{P}_2, \mathcal{P}_3$, until we obtain $\mathcal{P}_k$ such that $\mathcal{P}_k(A)$ and $\mathcal{P}_k(B)$ are not comparable for all $A, B \in \mathcal{X}$. (Define the state spaces $S_1, S_2, \ldots$, the utility functions $u_1, u_2, \ldots$, and the signed measures $\mu_1, \mu_2, \ldots$). Since $\mathcal{X}$ is finite, this is achievable in a finite number of steps k. Clearly $< S_k, \mu_k, u_k, \mathcal{C} >$ is a Content-Monotone Thinking-Averse representation, and $\mathcal{P}_k$ is a partition function with the desired properties. $\square$

Now construct a Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ and partitioning function $\mathcal{P}$ as of Claim 11 and notice that we have that for any $A, B \in \mathcal{X} \backslash NT$, $\mathcal{P}(A) \neq \mathcal{P}(B)$. Call $J := \{\pi \in \Pi(S) : \pi = \mathcal{P}(A) \text{ for some } A \in \mathcal{X} \backslash NT\}$, and $\mathcal{P}(C) = \{S, \emptyset\}$ for all $C \in NT$. Define $\hat{c}_I : J \cup \{S, \emptyset\} \to \mathbb{R}$ as $\hat{c}_I(\mathcal{P}(A)) = \mathcal{C}(A) - c_s(A)$ for all $A \in \mathcal{X} \backslash NT$ and $\hat{c}_I(\{S, \emptyset\}) = 0$ (Our previous discussion guarantees that it well defined.). We only need to show that $\hat{c_I}$ is partition-monotone, which in this case means $\hat{c}_I(\mathcal{P}(A)) = \mathcal{C}(A) - c_s(A)) \geq 0$ for all $A \in \mathcal{X}$. If $A \in NT$, this is trivially true. If $A \notin NT$, notice that we have $c_s(A) = c_s(B) = \mathcal{C}(B)$ for some $B \in NT$, $B \subseteq A$. By A.10, we have $\mathcal{C}(A) \geq \mathcal{C}(B)$, hence $\mathcal{C}(A) \geq c_s(A)$ as sought. Therefore, $\hat{c}$ is partition-monotone. Define now $c_I : \Pi(S) \to \mathbb{R}$ by extending $\hat{c_I}$ to $\Pi(S)$ preserving partition monotonicity. We only need to make sure that we have $\mathcal{C}(A) = c_s(|A|) + c_I(\mathcal{P}(A))$. But this is trivial from the definition of $\hat{c}_I$ and c_I. Finally, the uniqueness result is trivial if we notice that any partitioning function $\mathcal{P}$ must assign $\mathcal{P}(A) = \{S, \emptyset\}$ to any $A \in NT$. This concludes the proof of Theorem 4. *Q.E.D.*

[26] The addition of this state does not modify the representation. It is simply a scaling up of W in the same amount for all sets.

B.5. Proof of Theorem 5

The if direction is trivial. To prove the only if direction, define NT as in the proof of Theorem 4. For any $A, B \in NT$, we say that $(A, B) \in CO$ if $|A| = |B|$ and A and B have all but one element in common (i.e. $A \backslash \{x\} = B \backslash \{y\}$ for some $x \in A$, $y \in B$). Notice that, for any $A, B \in NT$ s.t. $(A, B) \in CO$, we must have $\mathcal{C}(A) = \mathcal{C}(B)$ by A.12. (Apply it once and obtain $\mathcal{C}(A) \geq \mathcal{C}(B)$, and once again and obtain $\mathcal{C}(A) \leq \mathcal{C}(B)$.) Notice that, for any $A, B \in NT$, $|A| = |B|$, there exist $C_1, \ldots, C_k \in NT$ s.t. $|A| = C_i$, $i = 1, \ldots, k$ and $(A, C_1), (B, C_k), (C_i, C_{i+1}) \in CO$ for $i = 1, \ldots (k-1)$. To construct it, if $x^* \in A$, then simply keep replacing all elements in A with elements in B one by one, leaving x^*. Since we have $x^* \sim^* X$ by A.7 and A. 8, then $C_i \in NT$. If $x^* \notin A$, construct C_1 replacing any element in A with x^*, and proceed as before until C_k. By our previous observation, this implies that for any $A, B \in NT$, $|A| = |B|$, $\mathcal{C}(A) = \mathcal{C}(B)$. Now, define $\bar{n} = |X|$, and notice that, since $x^* \sim^* X$, for any $n \leq \bar{n}$, there exist $A \in NT$ s.t.$|A| = n$. Construct $c_s : \{1, \ldots, \bar{n}\} \to \mathbb{R}$ as $c_s(n) := \mathcal{C}(A)$ for some $A \in NT$ s.t. $|A| = n$. Our discussion so far shows that it is well-defined. Moreover, A.11 immediately implies that it is increasing.

Now construct a Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ and partitioning function $\mathcal{P}$ as of Claim 11 and notice that we have that for any $A, B \in \mathcal{X} \backslash NT$, $\mathcal{P}(A) \neq \mathcal{P}(B)$. Call $J := \{\pi \in \Pi(S) : \pi = \mathcal{P}(A)$ for some $A \in \mathcal{X} \backslash NT\}$, and notice that, for all $C \in NT$ we have $\mathcal{P}(C) = \{S, \emptyset\}$. Define $\hat{c}_I : J \cup \{S, \emptyset\} \to \mathbb{R}$ as $\hat{c}_I(\mathcal{P}(A)) = \mathcal{C}(A) - c_s(|A|)$ for all $A \in \mathcal{X} \backslash NT$ and $\hat{c}_I(\{S, \emptyset\}) = 0$ (Our previous discussion guarantees that it well defined.). We only need to show that $\hat{c}_I$ is partition-monotone, which in this case means $\hat{c}_I(\mathcal{P}(A)) = \mathcal{C}(A) - c_s(|A|) \geq 0$ for all $A \in \mathcal{X}$. If $A \in NT$, this is trivially true. If $A \notin NT$, consider $A' = (A \backslash \{y\}) \cup \{x^*\}$ for some $y \in A$. Since $x^* \in A'$, then $A' \in NT$, which, by A.12, implies that $\mathcal{C}(A) \geq \mathcal{C}(A') = c_s(|A'|) = c_s(|A|)$ as sought (the last two equalities derive from the fact that $A' \in NT$ and $|A'| = |A|$.) Therefore, $\hat{c}$ is partition-monotone. Define now $c_I : \Pi(S) \to \mathbb{R}$ by extending $\hat{c_I}$ to $\Pi(S)$ preserving partition monotonicity. We only need to make sure that we have $\mathcal{C}(A) = c_s(|A|) + c_I(\mathcal{P}(A))$. But this is trivial from the definition of $\hat{c}_I$ and c_I.

To prove uniqueness, fix $\mathcal{C}$ and notice that every partitioning function $\mathcal{P}$ must assign $\mathcal{P}(A) = \{S, \emptyset\}$ if $A \in NT$, and therefore $c_I(\mathcal{P}(A)) = 0$ and so $\mathcal{C}(A) = c_s(A)$. Now remember that for every $n \leq \bar{n}$ we have $A \in NT$ s.t. $|A| = n$, which immediately implies that c_I is unique fixing $\mathcal{C}$, and in turns this implies that $c_C(\mathcal{P}(A)) = c'_c(\mathcal{P}'(A))$. This concludes the proof of Theorem 5. *Q.E.D.*

B.6. Proof of Theorem 6

The if direction is trivial. To prove the only if part, define NT as in the proof of Theorem 5 and notice that Claim 11 implies that there exist a Content-Monotone Thinking-Averse representation $< S, \mu, u, \mathcal{C} >$ and partitioning function $\mathcal{P}$ s.t. for any $A, B \in \mathcal{X} \backslash NT$ we have that $\mathcal{P}(A)$ and $\mathcal{P}(B)$ are not comparable. Call $J := \{\pi \in \Pi(S) : \pi = \mathcal{P}(A)$ for some $A \in \mathcal{X} \backslash NT\}$, and notice that, for all $C \in NT$ we must have $\mathcal{P}(C) = \{S, \emptyset\}$ since $\mathcal{P}$ is a partition function. Define $\hat{c} : J \cup \{S, \emptyset\} \to \mathbb{R}$ as $\hat{c}(\mathcal{P}(A)) = \mathcal{C}(A)$ for all $A \in \mathcal{X} \backslash NT$ and $\hat{c}(\{S, \emptyset\}) = 0$. It is well defined and partition-monotone since for all $A, B \in \mathcal{X} \backslash NT$, $\mathcal{P}(A)$ and $\mathcal{P}(B)$ are not comparable, and since $\mathcal{C}(A) \geq 0$. Define now $c_I : \Pi(S) \to \mathbb{R}$ by extending $\hat{c_I}$ to the whole $\Pi(S)$ preserving partition monotonicity. This concludes the proof of Theorem 6. *Q.E.D.*

B.7. Proof of Theorem 7

First notice that, as long as A. 6 and A. 7 are satisfied and there is a Content-Monotone Thinking-Averse representation, then it is trivial to show that for any $A \in \Delta(\mathcal{X})$, p_A and p^*_A exist. Moreover, by A.7, we have $\{x^*\} \sim^* X$, which implies that, if $|X| \geq 2$, there exist $x, y \in X$ s.t. $\{x\} \sim^* \{x, y\}$, and if $|X| \geq 3$, there exist $x, y, z \in X$ s.t. $\{x\} \sim^* \{x, y, z\}$. Now assume A. 15 and notice that to prove (1) we only need to show that for any $A, B \in \mathcal{X}$, $|A| = |B| \Rightarrow \mathcal{C}(A) = \mathcal{C}(B)$. If A, B are the same except for one element, then A. 15 implies $\frac{1}{2}p^*_A \oplus \frac{1}{2}p_B \sim \frac{1}{2}p^*_B \oplus \frac{1}{2}p_A$. Noticing that $\mathcal{C}$ is nothing but the difference between the representation of $\succeq$ and $\succeq^*$ proves $\mathcal{C}(A) = \mathcal{C}(B)$. If A and B are the same except for two elements, then there exists a set D of the same cardinality such that both A and D, and B and D differ of only one element (D is the set constructed replacing one of the non-common element in A with the corresponding one in B). But then, $\mathcal{C}(A) = \mathcal{C}(D) = \mathcal{C}(B)$. Clearly the same could be done however large the number of elements in which A and B differ: if this number is n, construct $n-1$ sets to form a chain that begins with A and ends with B, such that each set in this chain differ from the one before and the one after of only one element. This proves that there exist a function $c : \mathbb{N} \to \mathbb{R}$ such that $\mathcal{C}(A) = c(|A|)$ for all $A \in \mathcal{X}$.

To prove (2), we need to prove that c is increasing. Consider any $n \in 1, \ldots, (|X| - 1)$. If $|X| = 1$ this is trivially true, and otherwise we have argued that there exist $x, y \in X$ s.t. $x \sim \{x, y\}$, so we can always find $A \in \mathcal{X}$ and $z \in X$ s.t. $A \sim^* A \cup \{z\}$ and $|A| = n$: simply consider any $A \in \mathcal{X}$ s.t. $x \in A$ and $|A| = n$, and pick $z = y$. It is immediate to see from the representation that we must have $A \sim^* A \cup \{z\}$. (In fact, this is a direct consequence of A. 9, which implies submodularity of $\succeq^*$.). Now notice that A. 14 implies that $\mathcal{C}(A \cup \{x\}) \geq \mathcal{C}(A)$, and hence $c(|A \cup \{x\}|) \geq c(|A|)$. But since this is true for all $n = 1, \ldots, (|X| - 1)$, c must be increasing as sought.

To prove (3), we need to prove that c is increasing and convex. Consider any $n \in 1, \ldots, (|X| - 2)$. If $|X| = 1, 2$ this is trivially true, and otherwise we have argued that there exist $x, y, z \in X$ s.t. $x \sim \{x, y, z\}$, so we can always find $A \in \mathcal{X}$ and $z \in X$ s.t. $A \sim^* A \cup \{z\}$ and $|A| = n$ - just like before. Now notice that, by A. 16, we have $\frac{1}{2}p_{A\cup\{x,y\}} \oplus \frac{1}{2}p_A \succeq p_{A\cup\{x\}}$. But since $A \sim^* A \cup \{x, y\}$ (which implies $A \sim^* A \cup \{x\}$), then this means that $\frac{1}{2}\mathcal{C}(A \cup \{x, y\}) + \frac{1}{2}\mathcal{C}(A) \leq \mathcal{C}(A \cup \{x\})$, which in turn means $\frac{1}{2}c(|A| + 2) + \frac{1}{2}c(|A|) \leq c(|A| + 1)$. Since this is true for all $n \in 1, \ldots, (|X| - 2)$, c is convex. To show that c is monotone, notice that if we consider $A \in \mathcal{X}$, $x, y \in X$ such that $A \cup \{x, y\} \sim^* A$ with $x \in A$, then A. 16 implies $\mathcal{C}(A \cup \{y\}) \geq \mathcal{C}(A)$: following the same steps above implies that c must be monotone.

This concludes the proof of Theorem 7. *Q.E.D.*

Appendix C: Extension to menus of lotteries and the uniqueness of the state space

The analysis thus far has studied a preference relation over lotteries of menus. We have characterized what we defined a Thinking-Averse representation, and shown some uniqueness properties. However, we have not been able to show that the state space S in this representation is unique, because, as we have argued, our space was not *rich* enough. The content of this section is to extend the analysis to the case in which the menus themselves are menus of lotteries. A similar framework is used in Epstein, Marinacci, and Seo (2007). We show that this allows us to characterize the state space uniquely.

C.1. Formal Setup

Consider a finite set X. By $\Delta(X)$ we understand the set of probability distribution on X. By $\hat{\mathcal{X}}$ we understand the set of convex and compact subsets of $\Delta(X)$. We endow this collection by the Hausdorff topology, d_h. Further, we define the convex combinations of two sets to be the point-wise convex combination. That is, for any $\lambda \in (0,1)$ and $A, B \in \hat{\mathcal{X}}$, $\lambda A + (1-\lambda)B$ should be understood in the sense of Minkowski, that is, it is equal to the set $\{\lambda g + (1-\lambda)h : g \in A, h \in B\}$ (where $\lambda g + (1-\lambda)h$ is the probability distribution over X giving x with probability $\lambda g(x) + (1-\lambda)h(x)$). For any $A, B \in \hat{\mathcal{X}}$, we denote by $\alpha A \oplus (1-\alpha)B$ the lottery that assigns probability α to A and $(1-\alpha)$ to B. By $\Delta(\hat{\mathcal{X}})$ we understand the set of lotteries over $\hat{\mathcal{X}}$. We metrize $\Delta(\hat{\mathcal{X}})$ with the topology of weak convergence. By $\hat{\mathcal{X}}_S$ we understand the set of singletons in $\hat{\mathcal{X}}$. To keep the notation constant with the previous analysis, by $\Delta^S(\hat{\mathcal{X}})$ we understand the set of singletons or lotteries over singletons, $\Delta(\hat{\mathcal{X}}_S)$.

The primitive of our analysis is a complete preference relation $\succeq'$ defined over $\Delta(\hat{\mathcal{X}})$.

C.2. Axioms

We now introduce the axiomatic structure on the preference $\succeq$ on $\Delta(\hat{\mathcal{X}})$.

A.1' (Independence over singletons). *For any $\gamma \in (0,1)$ and any $p, q, r \in \Delta^S(\hat{\mathcal{X}})$,*

$$p \succeq q \Leftrightarrow \gamma p \oplus (1-\gamma)r \succeq' \gamma q \oplus (1-\gamma)r.$$

A.2' (Indifference between randomization for singletons). *For any $\alpha \in (0,1)$ and any $x, y \in \Delta(X)$,*

$$\alpha x + (1-\alpha)y \sim \alpha x \oplus (1-\alpha)y.$$

Define the binary relation $\succeq^*$ on $\hat{\mathcal{X}}$ as:

$$A \succ'^* B \Leftrightarrow (\tfrac{1}{2}+\epsilon)A \oplus (\tfrac{1}{2}-\epsilon)B \succ \tfrac{1}{2}A \oplus \tfrac{1}{2}B \succ (\tfrac{1}{2}-\epsilon)A \oplus (\tfrac{1}{2}+\epsilon)B.$$

for all $\epsilon < \bar{\epsilon}$, for some $\bar{\epsilon} > 0$. Define $A \sim'^* B$ when we have neither $A \succ'^* B$ nor $B \succ'^* A$.

A.3' (Coherence'). *$\succeq'^*$ is transitive.*

A.4' (Weak Continuity'). *For any $A \in \Delta(\hat{\mathcal{X}})$, the sets $\{p \in \Delta^S(\mathcal{X}) : p \succeq' A\}$ and $\{p \in \Delta^S(\mathcal{X}) : A \succeq' p\}$ are closed.*

A.4'* (Full Continuity'*). *For any $A \in \Delta(\hat{\mathcal{X}})$, the sets $\{B \in \Delta(\hat{\mathcal{X}}) : B \succeq' A\}$ and $\{B \in \Delta(\hat{\mathcal{X}}) : A \succeq' B\}$ are closed.*

A.5' (Content Continuity'). *For any $A \in \hat{\mathcal{X}}$, the sets $\{B \in \hat{\mathcal{X}} : B \succeq'^* A\}$ and $\{B \in \hat{\mathcal{X}} : A \succeq'^* B\}$ are closed.*

A.6' (Thinking Aversion'). *For any $A \in \hat{\mathcal{X}}$, $p \in \Delta^S(\hat{\mathcal{X}})$, we have*

$$p \succ'^* A \Rightarrow p \succ' A.$$

A.7' (Mixture Aversion'). *Take any $A, B \in \Delta(\hat{\mathcal{X}})$, $p, q \in \Delta^S(\hat{\mathcal{X}})$ such that $p \sim' A$ and $q \sim' B$, $\alpha \in (0,1)$. Then, the following must hold:*

$$\alpha p \oplus (1-\alpha) q \succeq' \alpha A \oplus (1-\alpha) B.$$

A.8' (Content Independence'). *$\succeq'^*$ satisfies independence, that is, for every $A, B, C \in \hat{\mathcal{X}}$, $\alpha \in (0,1)$, we have*

$$A \succeq^* B \Leftrightarrow \alpha A + (1-\alpha) C \succeq^* \alpha B + (1-\alpha) C.$$

A.9' (Best/Worst'). *There exist $x^*, x_* \in X$ such that $\{x^*\} \succeq' A \succeq' \{x_*\}$ for all $A \in \Delta(\hat{\mathcal{X}})$.*

A.10' (Content Monotonicity'). *For any $A, B \in \hat{\mathcal{X}}$, $B \subseteq A \;\Rightarrow\; A \succeq'^* B$.*

A.11' (Content L-continuity'). *There exist non-empty sets $A^*, A_* \in \hat{\mathcal{X}}$ and an $N > 0$ such that for every $\epsilon \in (0, \frac{1}{N})$, for every B and B with $d_h(B, C) \leq \epsilon$,*

$$(1 - N\epsilon) B + N\epsilon A^* \succeq'^* (1 - N\epsilon) C + N\epsilon A_*$$

Most of these requirements are either standard or identical to those in the previous analysis. The only differences are: A.2', which imposes that, for singletons, the agent is indifferent between the two randomizations; the continuity of $\succeq^*$, which is no longer a consequence of the continuity of $\succeq$; and L-continuity, which, from the analysis in DLRS we know that it is required to guarantee the existence of a representation when monotonicity is not satisfied. (We refer to their work, where it was introduced, for further discussion.)

C.3. Representation Theorem

We can define the analogous version of our notions of cost in this framework.

Definition 11. A function $\mathcal{C} : \Delta(\hat{\mathcal{X}}) \to \mathbb{R}$ is an **Anticipated Thinking Cost* function** if the following conditions hold:

1. $\mathcal{C}(\{p\}) = 0$ for all $p \in \Delta^S(\mathcal{X})$.
2. $\mathcal{C}(A) \geq 0$ for all $A \in \Delta(\hat{\mathcal{X}})$.
3. $\mathcal{C}$ is concave under $\oplus$, that is, for any $A, B \in \Delta(\hat{\mathcal{X}})$ and $\alpha \in (0,1)$, we have

$$\mathcal{C}(\alpha A \oplus (1-\alpha) B) \geq \alpha \mathcal{C}(A) + (1-\alpha) \mathcal{C}(B).$$

We can now state the representation.

Definition 12. A preference relation $\succeq'$ on $\Delta(\hat{\mathcal{X}})$ has a **Additive Thinking-Averse* Representation** if there exist a nonempty state space S, a continuous and affine state-dependent utility function $u : \Delta(X) \times S \to \mathbb{R}$, a signed measure μ over S and a function $\mathcal{C} : \Delta(\hat{\mathcal{X}}) \to \mathbb{R}$ such that $\succeq'$ is represented by

$$W(\alpha) = \int_{\hat{\mathcal{X}}} \int_S \max_{y \in A_i} u(y, s) \mu(\mathrm{d}s) \alpha(\mathrm{d}i) - \mathcal{C}(\alpha)$$

and:

1. each $u(\cdot, s)$ is an expected-utility function;
2. $\mathcal{C}$ is an Anticipated Thinking Cost* function;
3. $\int_S \max_{y \in A_i} u(y,s)\mu(\mathrm{d}s)$ represents $\succeq^*$.

Given S, u and $s \in S$, define the binary relation $\succ_s$ on $\hat{\mathcal{X}}$ as $A \succ_s B \Leftrightarrow u(A,s) > u(B,s)$, and $P(S,u)$ as $P(S,u) := \{\succ_s : s \in S\}$.

Definition 13. Given a Additive Thinking-Averse* Representation $< S, \mu, u, \mathcal{C} >$ s.t. $P(S,u)$ is finite, we say that a state $s \in S$ is relevant if there exist $A, B \in \hat{\mathcal{X}}$ such that $A \nsim'^* B$ and, for any $s' \in S$ with $\succ_s \neq \succ_{s'}$, $\max_{x \in A} u(x,s') = \max_{x \in B} u(x,s')$. If $P(S,u)$ is infinite, we say that a state $s \in S$ is relevant if for any neighborhood N of s, there exist $A, B \in \hat{\mathcal{X}}$ such that $A \nsim'^* B$ and, for any $s' \in S \backslash N$, $\max_{x \in A} u(x,s') = \max_{x \in B} u(x,s')$.

Theorem 8. *Consider a preference relation $\succeq'$ on $\Delta(\hat{\mathcal{X}})$ that satisfies A.9'. Then,*

1. *$\succeq'$ satisfies A. 1'-8' and 11' if and only if it has a Additive Thinking-Averse* Representation where every state is relevant;*
2. *$\succeq'$ satisfies A. 1'-8' and 10' if and only if it has a Additive Thinking-Averse* Representation $< S, \mu, u, \mathcal{C} >$ where μ is a probability measure over S and every state is relevant;*
3. *$\succeq'$ satisfies A. 1'-3', 4'* and 5'-8' and 11' if and only if it has a Additive Thinking-Averse* Representation where every state is relevant and $\mathcal{C}$ is continuous;*
4. *$\succeq'$ satisfies A. 1'-3', 4'* and 5'-8' and 10' if and only if it has a Additive Thinking-Averse* Representation $< S, \mu, u, \mathcal{C} >$ where μ is a probability measure over S, every state is relevant and $\mathcal{C}$ is continuous.*

Proof. The proof of this Theorem is the combination of the results in DLR01, DLRS and the intuition behind Theorem 1. For this reason, we will here only provide a sketch of of the proof. It is immediate to see that axioms A.5', 8', 10' and 11' guarantees that $\succeq^*$ has the properties desired to apply the results in either DLR01 or DLRS. This leads to different characterizations of $\succeq^*$ depending on the axioms we impose. Once we have a representation for $\succeq^*$, we can extend it to the whole $\Delta(\hat{\mathcal{X}})$ in a linear way to obtain the first component of the representation. Call this extension $\hat{\succeq}^*$. We need to show that such continuous and linear extension is in fact possible. But this is obvious since we are simply extending it to a mixture space which has as extreme points the elements of the original space $\hat{\mathcal{X}}$: linearity is trivial, and so is continuity given that we metrize this mixture space with the weak metric. (We are *not* extending it to a full linear space, which would require some Lipschitz continuity.) From the representation of $\succeq^*$, we then get the representation of $\hat{\succeq}^*$ using linearity. Call U the functional of this utility representation. Now, notice that we must have that $\succeq^*$ and $\hat{\succeq}^*$ coincide on $\Delta^S(\mathcal{X})$ (for the same reason this was the case in Theorem 1). Notice also that, because of A. 1', A.2' (and A.5') then there exist a representation W of $\succeq$ such that it is affine and continuous on $\Delta^S(\mathcal{X})$. Moreover, since $\succeq^*$ and $\hat{\succeq}^*$ coincide on $\Delta^S(\mathcal{X})$, it is trivial to see that W can be normalized so that $W(p) = U(p)$ for all $p \in \Delta^S(\mathcal{X})$. Then, simply mimic the remaining passages of the proof of Theorem 1 to conclude the proof: in particular, we can mimic the proofs of both Claim 7 and Claim 9 with minor modifications to obtain the desired representation. $\square$

C.4. Uniqueness

The main feature of this representation is the strong uniqueness that it entails. In fact, it combines the uniqueness result in Dekel, Lipman, and Rustichini (2001) and those in our Theorem 2.

Theorem 9. *Consider a preference relation $\succeq$ on $\Delta(\hat{\mathcal{X}})$ that satisfies A.9'. If $< S, \mu, u, \mathcal{C} >$ and $< S', \mu', u', \mathcal{C}' >$ are two Additive Thinking-Averse* Representation of $\succeq$ in which every state is relevant, then*

1. *If S is finite, then $S = S'$. If S is not finite, then $\bar{P}(S, u) = \bar{P}(S', u')$.*
2. *there exists $\gamma \in \mathbb{R}_{++}$, $\beta \in \mathbb{R}$ such that*

$$\int [\int_{S'} \max_{y \in A_i} u'(y, s)\mu'(\mathrm{d}s)]\alpha(\mathrm{d}i) = \gamma \Big[\int [\int_{S} \max_{y \in A_i} u(y, s)\mu(\mathrm{d}s)]\alpha(\mathrm{d}i) \Big] + \beta$$

   and

$$\mathcal{C}' = \gamma \mathcal{C}.$$

The proof is an immediate consequence of the uniqueness result in Dekel, Lipman, and Rustichini (2001) and the intuition behind Theorem 2. It is therefore left to the reader.

References

Abreu, D. and A. Rubinstein (1988): "The Structure of Nash Equilibria in Repeated Games with Finite Automata," *Econometrica*, 56, 1259–1282.

Aquinas, S. T. (1997): *Basic Writings of Saint Thomas Aquinas*, vol. 2, Hackett Publishing Company.

Camerer, C., T.-H. Ho, and J.-K. Chong (2004): "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119, 861–898.

Cherepanov, V., T. Feddersen, and A. Sandroni (2008): "Rationalization," Mimeo University of Pennsylvania and Northwestern University.

Dekel, E., B. L. Lipman, and A. Rustichini (1998): "Standard State-Space Models Preclude Unawareness," *Econometrica*, 66, 159–173.

——— (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891–934.

——— (2007a): "Temptation-Driven Preferences," Mimeo, Northwestern University.

Dekel, E., B. L. Lipman, A. Rustichini, and T. Sarver (2007b): "Representing Preferences with a Unique Subjective State Space: A Corrigendum," *Econometrica*, 75, 591–600.

Diasakos, T. (2007): "Bounded Rationality and Complexity in Individual Decision Problems," UC Berkeley.

Epstein, L., M. Marinacci, and K. Seo (2007): "Coarse contingencies and ambiguity," *Theoretical Economics*, 2, 355–394.

Epstein, L. and K. Seo (2007): "Subjective States: A More Robust Model," Mimeo, Boston University.

Ergin, H. (2003): "Costly Contemplation," Mimeo, MIT.

Ergin, H. and T. Sarver (2008): "A Unique Costly Contemplation Representation," Northwestern University.

Gale, D. and H. Sabourian (2005): "Complexity and Competition," *Econometrica*, 73, 739–769.

Geanakoplos, J. (1989): "Game Theory Without Partitions, and Applications to Speculation and Consensus," *Cowles Foundation Discussion Paper*, 914.

Ghirardato, P. and M. Marinacci (2002): "Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory*, 102, 251–289.

Gilboa, I. and L. Samuelson (2008): "Preferring Simplicity," Mimeo Yale University.

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

Iyengar, S. S., G. Huberman, and W. Jiang (2004): *How Much Choice is Too Much? Contributions to 401 (k) Retirement Plans*, Oxford University Press, chap. 5, 83–95.

Iyengar, S. S. and E. Kamenica (2007): "Choice Overload and Simplicity Seeking," University of Chicago, Graduate School of Business.

Iyengar, S. S. and M. L. Lepper (2000): "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, 79, 995–1006.

Kalai, E. and W. Stanford (1988): "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, 56, 397–410.

Kamenica, E. (2008): "Contextual Inference in Markets: On the Informational Content of Product Lines," *American Economic Review (forthcoming)*.

Kreps, D. (1979): "A Representation Theorem for "Preference for Flexibility"," *Econometrica*, 47, 565–578.

Moscarini, G. (2004): "Limited information capacity as a source of inertia," *Journal of Economic Dynamics and Control*, 28, 2003–2035.

Nehring, K. (1996): "Flexibility and Freedom of Choice in a Savage Framework," UC Davis.

Ortoleva, P. (2008): "Thinking Aversion and Portfolio Selection," Mimeo, New York University.

REIS, R. (2006): "Inattentive Consumers," *Journal of Monetary Economics*, 53, 1761–1800.

ROSENTHAL, R. W. (1989): "A Bounded-Rationality Approach to the Study of Noncooperative Games," *International Journal of Game Theory*, 18, 273–292.

RUBINSTEIN, A. (1998): *Modeling Bounded Rationality*, The MIT Press.

RUBINSTEIN, A. AND M. J. OSBORNE (1998): "Games with Procedurally Rational Players," *American Economic Review*, 88, 834–847.

RUBINSTEIN, A. AND M. PICCIONE (1993): "Finite Automata Play a Repeated Extensive Form Game," *Journal of Economic Theory*, 61, 160–168.

SABOURIAN, H. (2004): "Bargaining and Markets: Complexity and the Competitive Outcome," *Journal of Economic Theory*, 116, 189–228.

SALGADO, M. (2006): "Choosing to Have Less Choice," Fondazione Enrico Mattei, Nota di Lavoro 37.

SARGENT, T. J. (1993): *Bounded Rationality in Macroeconomics*, Oxford University Press, Oxford.

SARVER, T. (2007): "Anticipating Regret: Why Fewer Options May Be Better," *Econometrica (forthcoming).*

SCHWARTZ, B. (2005): *The Paradox of Choice: Why More is Less*, Harper Perennial.

SIMON, H. A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 99–118.

SIMS, C. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690.

——— (2006): "Rational Inattention: A Research Agenda," Mimeo Princeton University.

TIROLE, J. (2008): "Cognition and Incomplete Contracts," IDEI Working Papers 453, Toulouse.

TVERSKY, A. AND E. SHAFIR (1992): "Choice under conflict: The Dynamics of Deferred Decision." *Psychological Science*, 3, 358 – 361.

WILSON, A. (2004): "Bounded Memory and Biases in Information Processing," Mimeo.

WITTGENSTEIN, L. (2001): *Tractatus Logico-Philosophicus*, Routledge.