The price of flexibility: Towards a theory of Thinking Aversion

Pietro Ortoleva

Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E California Blvd., Pasadena, CA 91105, USA

Received 22 November 2011; final version received 2 October 2012; accepted 3 October 2012

Available online 8 January 2013

Abstract

We study the behavior of an agent who dislikes large choice sets because of the ‘cost of thinking’ involved in choosing from them. Focusing on preferences over lotteries of menus, we introduce the notion of Thinking Aversion. We characterize preferences as the difference between an affine evaluation of the content of the menu and a function that assigns to each menu a thinking cost. We provide conditions for which this cost can be seen as the cost that the agent has to sustain to figure out her preferences in order to make a choice.

© 2013 Elsevier Inc. All rights reserved.

JEL classification: D81; D83; D84

Keywords: Cost of thinking; Contemplation cost; Bounded rationality; Preference over menus; Preference for flexibility; Choice overload

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

I would like to thank the associate editor and three anonymous referees, Ozgur Evren, Paolo Ghirardato, Alessandro Lizzeri, Massimo Marinacci, Leandro Nascimento, David Pearce, Debraj Ray, Gil Riella, Todd Sarver, Anja Sautmann, Ennio Stacchetti, the participants at seminars at NYU, Society for Economic Design 2008, RUD 2009, Rice, UBC, Brown, Cornell, Caltech, Collegio Carlo Alberto, Bocconi, Luiss, Johns Hopkins, UCSD, and especially Efe Ok for useful comments and suggestions.

E-mail address: ortoleva@caltech.edu.

0022-0531/S – see front matter © 2013 Elsevier Inc. All rights reserved.

http://dx.doi.org/10.1016/j.jet.2012.10.009
1. Introduction

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 30 plans. Our agent appreciates the flexibility to pick a plan that better suits her needs, and consequently prefers provider B to A. At the same time, 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 trade-off: 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 trade-off: we call it Thinking Aversion. The second goal is to characterize this behavior axiomatically.

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, [38]. Within economics, some studies show how decision makers might prefer to face a strictly smaller set to simplify their choice: among these, [35] and [20]. Other studies provide evidence that suggests that agents dislike facing complicated choice sets: when confronted with them they tend to avoid choosing, or to choose the default option, a phenomenon dubbed choice overload. This is documented in a variety of settings: see, among others, [43,18,17,19].

These empirical findings suggest that agents’ behavior might be affected by the presence of a ‘disutility of thinking.’ This seems to be the case also because alternative, more standard explanations cannot readily account for such observations. For example, one might argue that an agent prefers a smaller set because there is informational value in what is included in this smaller set. She might then 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. However, in most of the cases we are interested in there seem to be no (relevant)

---

1 In [35] 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. (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, even if this means facing a potentially much worse set of alternatives. In [20] subjects are asked to choose the size of the choice set they will face in a later period: they could choose any even number between 2 and 20; once they have made their choice, the options are chosen randomly by the computer. The average number of columns chosen in all rounds was 11.4, and the median 10, against a maximum of 20; in addition, only 17.5% of the subjects choose 20 options in every case.

2 For example, [18] presents 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 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.

3 Kamenica [22] 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
informational value in the smaller sets, and this is especially true for most of the cited experiments. In particular, in [35] and in [20], the subset offered to the subjects is chosen \textit{randomly} by a computer, and therefore has no informational value. Another alternative explanation is that this behavior is due to fear of regret over having made a wrong choice. Introspection suggests that, although possibly connected, our explanation is well distinguished from regret. Moreover, in most of these cases agents would never find out what the right choice was, making it harder to suggest that the main motivation is anticipated regret.\footnote{This is further confirmed with a direct test in the experiment in [35], 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 \textit{not} given such feedback (and they know that they will not be).}

\subsection{Overview of the results}

We now turn to describe our approach to this problem. We take as a primitive the preference \(\succ\) of an agent over lotteries of menus, where by a menu we understand the set the agent will choose from at a later stage. The behavior that we aim to characterize is that of an agent whose preferences over menus may incorporate some considerations about how hard it will be to make a choice from each menu. This suggests that the preferences \(\succ\) are a combination of two components: 1) a ‘genuine’ preference, which depends on how good are the alternatives that each menu contains, and which represents how the agent would rank menus if there were no cost of thinking; 2) some measure about how hard it will be to actually choose from this menu. The first step of our analysis is to introduce an axiomatic framework that allows us to elicit the first component, the genuine preference, from the observed preference \(\succ\). To this end, we take as a primitive a preference relation \(\succ\) 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 \(\lambda \in (0, 1)\), when the agent faces the lottery \(\lambda A + (1 - \lambda) B\) 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 \(\lambda\) and her choice from \(B\) with probability \((1 - \lambda)\). Using this structure, we suggest one way to elicit the genuine preference, which we denote by \(\succ^*\), from the observed preference \(\succ\): we say that \(A\) is genuinely better than \(B\) if \(aA + (1 - a)B \succ bA + (1 - b)B\) for some \(a, b \in (0, 1)\) with \(a > b\), i.e., if the agent prefers to increase the probability of receiving her choice from \(A\) rather than her choice from \(B\) when she needs to make a choice from both.

We then use this elicited genuine preference \(\succ^*\) to define the main behavioral postulate of the paper: Thinking Aversion. To define it, notice first of all that singleton menus, i.e., menus with only one element, are special menus in this framework since they require no thinking – there is nothing to decide from a set of only one element. Consider then a lottery over menus \(\alpha\) and a singleton menu \(\{x\}\), and suppose that the singleton is ‘genuinely’ better than the lottery, i.e., \(\{x\} \succ^* \alpha\). Then, we would like to say that any agent who doesn’t like to think about what to choose should rank the singleton better than the lottery also according to the original ranking \(\succ\): we should have \(\{x\} \succ \alpha\). The reason is, the singleton not only is ‘genuinely’ better, but also it requires no thinking at all. (This definition parallels the one of risk aversion based on the comparison with a risk-free option, where instead of the expected value we compare the genuine evaluation \(\succ^*\), and instead of the risk-free option we have a ‘thinking-free’ option, a singleton.)
An identical intuition could be applied to the case in which we replace a menu in the support of a lottery with a singleton menu. Consider the lottery \( \lambda \alpha + (1 - \lambda) \beta \) for some \( \alpha, \beta \) and some \( \lambda \in [0, 1] \). Recall that in this case the agent needs to make a choice for all the menus in the support of \( \alpha \) and of \( \beta \). Let us compare this lottery with the identical one in which we have replaced \( \alpha \) with some singleton \( \{x\} \), i.e., \( \lambda \{x\} + (1 - \lambda) \beta \), and suppose, like before, that this singleton is genuinely better than \( \alpha \): we have \( \{x\} \succ^* \alpha \). Following the same intuition above, an agent who doesn’t like to think about what to choose should prefer \( \lambda \{x\} + (1 - \lambda) \beta \) to \( \lambda \alpha + (1 - \lambda) \beta \): the former has a weakly better content, as \( \{x\} \succ^* \alpha \), and it requires weakly less thinking, as the agent no longer needs to choose what she wants from the menus in the support of \( \alpha \) – they have been replaced with \( \{x\} \), that requires no decisions. We call this postulate Thinking Aversion.

We then turn to characterize the behavior of an agent whose preferences comply with the postulate above. Let \( X \) be a finite set 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 \( \mathcal{X} \), and denote by \( \alpha \) the lottery over menus that gives menu \( A \) with probability \( \alpha_A \in [0, 1] \). The first, more general representation that we obtain is 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 \( \mu \) over \( S \), and a function \( \mathcal{C} : 2^\mathcal{X} \to \mathbb{R}_+ \), such that \( \succeq \) is represented by

\[
W(\alpha) = \sum_{A \in \mathcal{X}} \alpha_A \left( \sum_{s \in S} \mu(s) \left[ \max_{y \in A} u(y; s) \right] \right) - \mathcal{C}(\text{supp}^*(\alpha)), \tag{1}
\]

where \( \text{supp}^*(\alpha) \) denote the support of the lottery minus the singletons, and \( \mathcal{C} \) is weakly positive everywhere, it is equal to zero on the empty set \( (\mathcal{C}(\emptyset) = 0) \), and it is monotone – larger supports require more thinking.

We interpret this representation as follows. The preferences of the agent consist of two components: 1) her evaluation of the expected quality of the content of the set, captured by the first part of the representation; 2) her evaluation of the cost of thinking about the set, captured by the anticipated thinking cost function \( \mathcal{C} \). The first component has a representation reminiscent of standard ones in the literature (although it is technically very different since we operate on a different primitive): the agent’s utility depends on the realization of a state of the world \( s \in S \), which she will discover before choosing from the menu; however, at the time of choice between menus she doesn’t know this state of the world yet, and she forms an “expectation” of her future utility using the signed measure \( \mu \) over the states.\(^5\) The second component, \( \mathcal{C} \), is a function that represents the cost of thinking about each lottery of menus. Following our interpretation that when the agent faces a lottery of menus she needs to make a choice for each lottery in the support – she needs a full contingent plan – the function \( \mathcal{C} \) assigns the same cost of thinking to all lotteries with the same \( \text{supp}^* \), i.e., to all those lotteries where the sets for which a decision is necessary are the same. (Singletons are not considered because decisions are not necessary there.) In addition, \( \mathcal{C} \) is weakly positive everywhere, it is equal to zero on the empty set, guaranteeing that the cost of thinking is zero for singletons and lotteries of singletons; and it is non-decreasing in the support – lotteries with a larger support have a (weakly) higher cost of thinking.

\(^5\) Notice that the fact that the agent computes this expectation does not mean that she already knows what she will choose – just like when people choose a restaurant they usually do not already choose what to order there. In fact, the very existence of multiple states of the world in \( S \) could be seen as representing the fact that the agent still doesn’t know what to choose – formally, she doesn’t know her preferences. She could discover them by thinking, but this might be costly, originating the cost of thinking. (See below.)
We then strengthen the representation above in two ways. First, we characterize the case in which the first component of the representation assigns a higher value to larger menus – the case in which \( \mu \) is a probability measure. (It turns out that this case can be characterized simply by imposing the axioms of [23] on how \( \preceq^* \) ranks degenerate lotteries over menus.) Second, and most importantly, we strengthen the representation above by providing a behavioral postulate that guarantees that the anticipated cost of thinking function \( C \) has a specific functional form, to which corresponds a specific interpretation: the cost of thinking about a menu is equal to the cost of the cheapest partition of the state space necessary to make the optimal choice. The idea is the following. The cost of thinking could be understood as the cost that the agent has to sustain to figure out her preferences, at least insofar as required to determine which is the best choice in the set. And since we can see the uncertainty over the preferences as represented by the uncertainty over the state of the world in \( S \) (because for each state we have a utility function), then the cost of thinking could be interpreted as the cost to find out the state of the world in order to make the optimal choice. In particular, the agent will have to partition the state space to make sure she identifies the state of the world with enough precision to make such choice. We then offer a behavioral axiom that guarantees that there exists a function that associates to each of these partitions a cost, and that the cost of thinking \( C \) must be equal to the cost of the cheapest partition of the state space that allows the agent to make the optimal choice in that set. That is, there exist some function \( c: \Pi(S) \to \mathbb{R}_+ \) (where \( \Pi(S) \) is the set of partitions of \( S \)) such that

\[
C(\{A\}) = \min_{\pi \in \mathcal{P}_{S,u}(A)} c(\pi)
\]

where \( \mathcal{P}_{S,u}(A) \) represents the set of all partitions that allow the agent to attain full utility from a set \( A \) by choosing the same alternative in every state not separated by the partition. This implies that, although our agent might potentially think too much, she does so efficiently – via the least costly strategy.

One feature of the representations above is that agents behave as if they expected themselves to choose the optimal option from a menu, instead of exerting only the optimal amount of thinking and possibly choosing a suboptimal alternative (balancing the quality of the choice with the effort of choosing it). That is, agents are represented as if they expected themselves to potentially think too much, more than optimal. While potentially unintuitive, in Section 3.4 we argue that the latter seems to be a feature of any representation of this class in which agents prefers smaller menus to avoid the cost of choosing from larger ones. The reason is, if an agent always chooses the optimal amount of thinking, then she would also (weakly) prefer larger menus, because they might contain better alternatives and the agent can simply disregard the additional options. By contrast, the agent in our model cannot disregard these additional options – it is as if she knew she’d be ‘tempted’ to consider them as well – and therefore prefer a smaller set.

1.2. Outline and related literature

The rest of the paper is organized as follows. In Section 2 we present the formal setup and the axioms. Section 3 presents the four representations and the representation theorems, discusses the uniqueness properties, and presents a comparative notion of Thinking Aversion. Section 4 concludes. The proofs appear in Appendix B.

The idea that agents might prefer to face fewer options is of course not new to economics. By looking at preferences over menus, a large number of papers have suggested models with this feature. Among these, we find models in which agents might want to avoid the presence of
a tempting alternative [16,5], or regret [37], or a potential combination of these elements ([4], henceforth DLR01, [6]). The present paper fits into this literature as we study preferences over menus and suggest 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 larger one.

In the framework of preferences of menus two papers, [10] and [12], introduce a similar concept, dubbed ‘cost of contemplation’. They present and justify axiomatically a model in which agents choose the optimal amount of contemplation to evaluate the sets they will have to choose from, where each act of contemplation is associated with a cost of performing it. These two models differ from ours in several aspects, which we will analyze in detail in Section 3.4. 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 simply to avoid the cost of thinking connected to the bigger one. In [10] the axioms simply impose that agents always prefer larger sets – in fact, this is the only requirement. In the model of [12] agents might actually prefer a smaller set, but this cannot be due to the cost of thinking alone, but rather only to other reasons (e.g., to avoid temptation). In fact, they prove that if we rule out these other reasons to prefer a smaller set, and we keep only the cost of thinking, then agents would always (weakly) prefer larger sets. By contrast, our work originates from the interest in preferences for smaller sets due to the presence of a cost of thinking.

More generally, the concept of ‘cost of thinking’ that we are interested in 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. However, while some of these models represent these constraints as a cost like we do, most of them study agents who have an actual bound in their computational abilities – which means that they cannot solve complex problems even if given the appropriate incentives. Furthermore, most of the models in this literature are not defined axiomatically, but rather behaviorally.

2. Setup and foundations

2.1. Formal setup

Consider a finite set $X$. Define by $\mathcal{X}$ its power set, that is, $\mathcal{X} := 2^X \setminus \{\emptyset\}$. We use $A, B, C$ to indicate generic elements on $\mathcal{X}$. By $\Delta(\mathcal{X})$ we understand the set of lotteries over $\mathcal{X}$, and we denote $aA + (1 - a)B$ as the lottery that assigns probability $a \in (0,1)$ to $A$ and $(1 - a)$ to $B$ for some $A, B \in \mathcal{X}$. We use $\alpha, \beta$ to indicate 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 $N = |\mathcal{X}|$. Abusing notation, we denote by $\mathcal{X}$ the set of degenerate lotteries in $\Delta(\mathcal{X})$. Also abusing notation, we refer to $\Delta^*(X)$ as the set of elements of $\Delta(\mathcal{X})$ the support of which includes only singletons – menus with only one element. We use $p, q, r$ to indicate generic elements of $\Delta^*(X)$. For any lottery $\alpha \in \Delta(\mathcal{X})$ we define by $\text{supp}(\alpha)$ its support, i.e., the menus to which $\alpha$ assigns a positive probability. By $\text{supp}^*(\alpha)$ we denote the set of menus to which $\alpha$ assigns a positive probabilities minus the singletons: that is, $\text{supp}^*(\alpha) = \{A \in \mathcal{X} : \alpha(A) > 0 \text{ and } |A| > 1\}$. This notation will be useful as non-singleton

---

6 In this broad area, starting from [39], papers have focused on game theory [1,21,30,33,32,2], individual decision making [14,24,3,44,7], bargaining, contracting and competitive equilibria [34,13,42], macroeconomics [36,40,25,41,29]. A not so recent survey is offered in [31].
menus are the only ones for which the agent needs to make a decision – there is nothing to decide from a singleton.

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

As described in the introduction, we wish to interpret a lottery over menus as performed after the agent chooses from each menu in the support, in an additional later stage. That is, given two menus \( A, B \in \mathcal{X} \), the lottery \( \frac{1}{2}A + \frac{1}{2}B \) 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 the menus in the support of the lottery. The timing is therefore the following: at time 0 the agent ranks the lotteries over menus; at time 1 she forms a contingent plan given the lottery over menus chosen at time 0; at time 2 the lottery over menus is realized, and the agent is given her choice according to the contingent plan. (See Fig. 1.)

Our interest on preferences over lotteries over menus is similar to the one of [26] and [8], although in their case the lottery over menus is resolved before the choice, and therefore there is no need to form contingent plans. At the same time, it differs from most of the literature, which focuses on menus of lotteries instead of lotteries of menus. While this different setup implies that we could not use standard representation results, we adopt it because it allows us to have agents that form contingent plans when they face lotteries over menus – which is not possible in the standard setup of menus of lotteries.\(^7\)

2.2. Foundations for a basic model

2.2.1. Basic postulate

We now introduce the axiomatic foundations of our model. As we discussed in the introduction, one of our goals is to be able to separate a genuine ranking of menus, that we would observe if there were no disutility from having to choose, from the observed ranking of menus \( \succeq \), which might also contain considerations about how hard it will be to choose. We now suggest a way to obtain such separation behaviorally.

Consider an agent who is facing the lottery \( \frac{1}{2}A + \frac{1}{2}B \) 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 \( B \); she needs to ‘think about’ both sets. Suppose now that we increase the probability that the agent receives her choice from \( A \), and that we end up with the lottery \( \left( \frac{1}{2} + \epsilon \right)A + \left( \frac{1}{2} - \epsilon \right)B \) for some \( \epsilon > 0 \).

\(^7\) The reason is, 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 a set mixture in the sense of Minkowski, we depart from the standard approach.
In this case as well the agent has to think about both sets. Assume now that this new mixture is preferred to the original one—the agent liked this change in probabilities, i.e.,

\[
\left(\frac{1}{2} + \epsilon\right)A + \left(\frac{1}{2} - \epsilon\right)B \succ \frac{1}{2}A + \frac{1}{2}B.
\]

What does this mean? In both cases the agent must still form a contingent plan, and yet the agent prefers to receive her choice from \(A\) with a higher probability. We interpret this to mean 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 \(B\). More in general, we will then say that the decision maker likes \(A\) ‘genuinely better’ than \(B\) if

\[
aA + (1-a)B \succ bA + (1-b)B
\]

for some \(a, b \in (0, 1)\) with \(a > b\). We can naturally extend this idea to generic lotteries over menus, and define the binary relation \(\succ^*\) on \(\Delta(\mathcal{X})\) by \(a \succ^* b\) if and only if

\[
aa \alpha + (1-a)\beta \succ ba + (1-b)\beta
\]

for some \(a, b \in (0, 1)\) with \(a > b\). We use \(\succ^*\) and \(\sim^*\) to represent the symmetric and asymmetric parts. 

We are now ready to define the notion at the core of our analysis: Thinking Aversion. We start from a basic postulate.

**A.1 (Weak Thinking Aversion).** For any \(\alpha \in \Delta(\mathcal{X})\) and \(p \in \Delta^S(X)\)

\[
p \succ^* \alpha \implies p \succ \alpha.
\]

Suppose that we have a lottery over singletons \(p\) the content of which is genuinely better than the content of some generic lottery \(\alpha\), or simply \(p \succ^* \alpha\). Then Weak Thinking Aversion says that we must also observe that this lottery over singletons is preferred to \(\alpha\): we must have \(p \succ \alpha\). This happens because \(p\) requires no thinking—since all the sets are singletons, there is nothing to decide. Then any agent who dislikes thinking must prefer \(p\) to any lottery \(\alpha\) that has a worse content and, moreover, might require some thinking. Arguably, this notion parallels equivalent ways to define risk aversion by a comparison with risk-free options, or of ambiguity aversion by a comparison with constant acts: following the same idea, we define Thinking Aversion by a comparison with ‘thinking-free’ objects, singletons.

While Weak Thinking Aversion regulates how lotteries compare to singleton menus, the same intuition can be applied to the case in which we replace a menu in the support of a lottery with a singleton menu. Consider \(\alpha, \beta \in \Delta(X)\), and the lottery \(\lambda \alpha + (1-\lambda)\beta\) for some \(\lambda \in (0, 1)\). We will now compare this lottery with the one in which we have replaced \(\alpha\) with some lottery over singletons \(p \in \Delta^S(X)\), obtaining the lottery \(\lambda p + (1-\lambda)\beta\), for some \(p\) which is genuinely better than \(\alpha\), i.e., \(p \succ^* \alpha\). Following the same intuition above, we argue that an agent who doesn’t

---

8 In what follows we use this derived relation \(\succ^*\) in the statement of axioms and definitions. While indeed the use of derived relations in axioms should be kept to a minimum, we use it here since we believe it significantly helps the exposition of the results. Naturally the same axioms can be stated using only the primitive \(\succ\) simply by replacing any statement involving \(\succ^*\) with its definition in terms of \(\succ\). See Section 2.2.3 for a discussion on the testability of these axioms.

9 In fact, it is standard practice to define risk aversion for a monotone preference \(\succ\) on lotteries on \(\mathbb{R}\) as follows (\(\mathbb{E}[\cdot]\) denotes the expected value): \(\succ\) is risk averse if for any lottery \(p\) and degenerate lottery \(x\), if \(x > \mathbb{E}[p]\), then \(x \succ p\).
like to think about what to choose should prefer \( \lambda p + (1 - \lambda)\beta \) to \( \lambda \alpha + (1 - \lambda)\beta \). First, because the former has a weakly better content, as we have replaced \( \alpha \) with some \( p \) such that \( p \succ^{*} \alpha \). Moreover, because with the latter the agent needs to consider fewer contingencies, i.e., perform less thinking: while with \( \lambda \alpha + (1 - \lambda)\beta \) the agent needs to think about what to choose from all the menus in the support of \( \alpha \) and \( \beta \), in \( \lambda p + (1 - \lambda)\beta \) she only needs to think about the menus in the support of \( \beta \). This leads us to our main axiom, Thinking Aversion.

A.2 (Thinking Aversion). For any \( \alpha, \beta \in \Delta(\mathcal{X}) \), \( p \in \Delta^{S}(X) \), and \( \lambda \in [0, 1] \) we have

\[
p \succ^{*} \alpha \quad \Rightarrow \quad \lambda p + (1 - \lambda)\beta \succ \lambda \alpha + (1 - \lambda)\beta.
\]

2.2.2. Linearity

It is standard practice in this literature to impose that preferences satisfy the independence axiom. In our case, however, postulating standard independence would be too strong. To wit, remember that when an agent faces a lottery over menus she needs to form a contingent plan: she need to make a choice from all the sets in the support. This means that our agent might be indifferent between two menus \( A \) and \( B \), but at the same time strictly prefer \( A \) to \( \frac{1}{2}A + \frac{1}{2}B \), a clear violation of independence, since in the latter case she needs to think about both sets. In what follows we will therefore posit forms of independence only for those cases in which such issue does not arise. First, we posit independence of \( \succ^{*} \), where the cost of thinking does not play a role. Second, we posit independence of \( \succ \) for those cases in which the act of mixing does not introduce new sets to be considered: this will be the case when we mix lotteries with the same \( \text{supp}^{*} \).

Finally, we consider an independence-like property on the mixture of two lotteries with a singleton menu. Consider some \( \alpha, \beta \), and the lotteries \( a(x) + (1 - a)\alpha \) and \( a(x) + (1 - a)\beta \) for some \( x \in X \) and \( a \in (0, 1) \). Suppose that we have \( \alpha \succeq \beta \). Should we then expect to have \( a(x) + (1 - a)\alpha \succeq a(x) + (1 - a)\beta \), as prescribed by independence? In general this might not be the case. For example, we may have \( \alpha \succeq \beta \) because \( \alpha \) is harder to think about but returned a better choice, but \( a(x) + (1 - a)\alpha < a(x) + (1 - a)\beta \) because, when we mix with \( \{x\} \), the sets for which a decision need to be taken have not changed, but the difference in what the lotteries return has shrunk – as with some probability both lotteries will return \( x \). To recover an independence-like ranking, we should then ‘compensate’ for the change in what the two lotteries return: instead of comparing \( a(x) + (1 - a)\alpha \) and \( a(x) + (1 - a)\beta \) directly, we compare them only after we have mixed them with singletons that contain the genuine evaluation of the original lotteries. To express this, for any \( \gamma \in \Delta(\mathcal{X}) \) denote by \( p_{\gamma} \) and \( p_{\gamma}^{*} \) any two elements of \( \Delta^{S}(X) \) s.t. \( p_{\gamma}^{*} \sim^{*} \gamma \) and \( p_{\gamma} \sim \gamma \). We can then posit that, if \( \alpha \succeq \beta \), we should have

\[
\frac{a}{1 + a}p_{\alpha}^{*} + \frac{1}{1 + a}p_{a(x) + (1 - a)\alpha} \succeq \frac{a}{1 + a}p_{\beta}^{*} + \frac{1}{1 + a}p_{a(x) + (1 - a)\beta}.
\]

Intuitively, by mixing with \( p_{\alpha}^{*} \) and \( p_{\beta}^{*} \) we are compensating for the change in quality of what is returned, and we can then expect an independence-like property to hold.

This leads us to the following axiom.

A.3 (Weak Independence). The following hold:

1. For any \( a \in (0, 1) \) and for any \( \alpha, \beta, \gamma \in \Delta(\mathcal{X}) \) such that \( \text{supp}^{*}(\alpha) = \text{supp}^{*}(\beta) = \text{supp}^{*}(\gamma) \),

\[
\alpha \succeq \beta \quad \Leftrightarrow \quad a\alpha + (1 - a)\gamma \succeq a\beta + (1 - a)\gamma.
\]
2. \( \succ^* \) satisfies independence, i.e., for any \( a \in (0, 1) \) and any \( \alpha, \beta, \gamma \in \Delta(\mathcal{X}) \), \[
\alpha \succ^* \beta \iff a\alpha + (1-a)\gamma \succ^* a\beta + (1-a)\gamma .
\]

3. For any \( \alpha, \beta \in \Delta(\mathcal{X}) \), \( a \in (0, 1) \), \( x \in X \), and \( p_\alpha^* \sim^* \alpha \), \( p_{a\{x\}+(1-a)\alpha} \sim a\{x\} + (1-a)\alpha \), \( p_\beta^* \sim^* \beta \), \( p_{a\{x\}+(1-a)\beta} \sim a\{x\} + (1-a)\beta \), \[
\alpha \succ^* \beta \iff \frac{a}{1+a} p_\alpha^* + \frac{1}{1+a} p_{a\{x\}+(1-a)\alpha} \succ^* \frac{a}{1+a} p_\beta^* + \frac{1}{1+a} p_{a\{x\}+(1-a)\beta} .
\]

2.2.3. Observability of \( \succ^* \) and testability of the axioms

In the previous section we have defined the genuine preference \( \succ^* \) by \( \alpha \succ^* \beta \) if and only if \( a\alpha + (1-a)\beta \succ b\alpha + (1-b)\beta \) for some \( a, b \in (0, 1) \) such that \( a > b \). At the same time, it is easy to see that under the first part of Weak Independence, this would hold if and only if \( a\alpha + (1-a)\beta \succ b\alpha + (1-b)\beta \) for all \( a, b \in (0, 1) \). The reason is, for any \( a \) or \( b \in (0,1) \) the supp* of this lottery is the same, and independence then prescribes that \( a\alpha + (1-a)\beta \succ b\alpha + (1-b)\beta \) holds for some \( a > b \) if and only if it holds for all \( a > b \). This means that, under weak independence, observing \( \succ^* \) is even easier: only one observation is enough not only to observe \( \succ^* \), but also to falsify any ranking of \( \succ^* \).

This is of particular importance for the testability of the axioms above. In fact, because both axioms are stated in terms of the derived relation \( \succ^* \), a natural concern is whether they are falsifiable with finite data. We will now argue that the axiomatic structure presented thus far is indeed falsifiable. It is straightforward to see that both Thinking Aversion, as well as the first and the last part of Weak Independence, could be easily falsified with finite data. We are left with the second part of Weak Independence. On the one hand, this part would not be falsifiable if we only considered the original definition of \( \succ^* \); to falsify it we would need to observe some \( \alpha, \beta \) such that \( \alpha \succ^* \beta \) holds, while \( a\alpha + (1-a)\gamma \succ^* a\beta + (1-a)\gamma \) does not hold for some \( \gamma \) – but the latter cannot be shown with finite data using the original definition of \( \succ^* \). However, the observation above shows that this axiom is in fact falsifiable assuming the first part of Weak Independence (itself falsifiable).\(^{10}\)

2.2.4. Technical postulates

Since we are after a representation theorem, we need to impose a continuity-type axiom. We impose such postulate both on the original preference relation \( \succ \), and on the derived on \( \succ^* \). For the former, however, we restrict our attention to singletons and to their relations to sets, because full continuity might be too strong of a requirement.\(^{11}\)

A.4 (Continuity). For any \( \alpha \in \Delta(\mathcal{X}) \):

1. the sets \( \{ p \in \Delta^S(X) : p \succ \alpha \} \) and \( \{ p \in \Delta^S(X) : \alpha \succ p \} \) are closed;
2. the sets \( \{ \beta \in \Delta(\mathcal{X}) : \beta \succ^* \alpha \} \) and \( \{ \beta \in \Delta(\mathcal{X}) : \alpha \succ^* \beta \} \) are closed.

We conclude with one last technical axioms. We posit that there exist two elements, \( x^* \) and \( x_* \) in \( X \) such that according to the genuine preference \( \succ^* \), the degenerate lottery that returns \( \{x_*\} \)

\(^{10}\) In particular, to guarantee that we don’t have \( a\alpha + (1-a)\gamma \succ^* a\beta + (1-a)\gamma \) all we need is to observe \( b(a\alpha + (1-a)\gamma) + (1-b)(a\beta + (1-a)\gamma) \prec (a\alpha + (1-a)\gamma) + (1-c)(a\beta + (1-a)\gamma) \) for some \( b > c \).

\(^{11}\) For example, we may have some \( \beta \succ^* aA + (1-a)B \) for all \( a \in (0, 1) \), but \( A \succ B \) (in violation of continuity), as when \( a > 0 \) the agent might wish to form a contingent plan after both \( A \) and \( B \), which is not necessary when \( a = 1 \).
is worst than any other lottery in Δ(Ω), while the degenerate lottery that returns the singleton \{x^*\} is better than any other element in Δ(Ω): we have \{x^*\} ≻_\alpha \succ_x \{x_\alpha\} for all \alpha. Moreover, if we now consider the original preference ≻, we posit that while a lottery \alpha can be either better than \{x^*\} or worse than \{x_\alpha\}, at the same time we can mix this lottery with some singleton menu and make its rank intermediate between that of \{x^*\} and \{x_\alpha\} – a continuity-type postulate. (This would naturally always hold if we had \{x^*\} ≻_\alpha \succ_x \{x_\alpha\} for all \alpha.) While this axiom is technical and is not derived from any real world consideration, it is quite easy to depict a situation in which it would hold.12

A.5 (Best/Worst). There exist x^*, x_\alpha \in X such that for all \alpha \in Δ(Ω), \{x^*\} ≻_\alpha \succ_x \{x_\alpha\} and \{x^*\} ≽_\lambda \alpha \succ \lambda \{x\} ≽_x \{x\} for some x \in X and \lambda \in [0, 1].

3. Models for Thinking Aversion

3.1. Basic representation

We are now ready to introduce our first representation.

Definition 1. A preference relation ≻ on Δ(Ω) 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 × S → R, a signed measure \mu over S and a function \mathcal{C} : 2^Ω → R such that ≻ is represented by

\[ W(\alpha) = \sum_{A \in \Omega} \alpha_A \left( \sum_{s \in S} \mu(s) \left[ \max_{y \in A} u(y, s) \right] \right) - \mathcal{C} \left( \text{supp}^*(\alpha) \right) \]

where \mathcal{C}(A) ≥ 0 for all A ∈ 2^Ω and \mathcal{C}(∅) = 0.

Theorem 1. Let ≻ be a complete preference relation on Δ(Ω) that satisfies Best/Worst. Then the following holds:

- ≻ satisfies Weak Thinking Aversion, Linearity and Continuity if and only if it admits a Thinking-Averse representation \langle S, \mu, u, \mathcal{C} \rangle.13

- ≻ satisfies Thinking Aversion, Linearity and Continuity if and only if it admits a Thinking-Averse representation \langle S, \mu, u, \mathcal{C} \rangle such that \mathcal{C} is monotone, i.e., \mathcal{C}(S) ≥ \mathcal{C}(T) for all S, T ∈ 2^Ω with S ⊇ T.

We interpret this representation as follows. We can see the preferences of an agent as the resultant of the difference between two components. First, her evaluation of the expected quality of the content of the set. Second, her evaluation of the cost of thinking of the set, represented

12 For example, consider a set X composed of economy 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 the worst option – both considering the cost of choosing, and not considering it.

13 While Best/Worst is an axiom with no conceptual value and it is therefore not included in the if and only if statement, it is easy to find conditions on the representation that would guarantee that it holds. If \mu is a probability measure, all we need is: that there exists some x^* such that u(x^*, s) ≥ u(x, s) for all x ∈ X and for all s ∈ S; and that \max_{H \in 2^\Omega} \mathcal{C}(H) < \max_{x \in X} (u(x^*) - u(x)), i.e., there exist two options the utility difference of which is higher than the cost of thinking about any lottery.
by \( C \). These two components are potentially pushing in different directions, in which case the agent faces a trade-off between a better content and a harder choice – which is the behavior we are after. Like in our initial cell phone example, the agent weights the benefits of a large number of options with the (expected) cost of having to decide which one is the best.

The first component of the representation can be interpreted as the evaluation of the expected quality of the content of a set. In line with this interpretation, it represents the genuine preference \( \succeq^* \), and it is modeled with a (finite) set of states of the world, a state-dependent utility function \( u \), and a signed measure \( \mu \) over \( S \). As it is standard in this literature, we interpret this as if our agent had some uncertainty over her preference at the time of choice: for each possible preference we have a state of the world \( s \), which the agent will discover before choosing. Consequently, she expects a utility of \( \max_{y \in A} u(y; s) \) for each state \( s \in S \). At the time of choice between menus, however, she does not know the state yet, and she forms an ‘expectation’ using the signed measure \( \mu \). In addition, if the agent is evaluating a lottery of menus, she also does not know which part of the contingent plan will be put in place, hence she needs to further condition on the probabilities \( \alpha_i \) of each realization of the lottery.

We should emphasize that even if the agent evaluates at time 0 the expected quality of the content of a menu, this does not mean that she already knows what she would choose from it. If this were the case, she would have already ‘solved the problem’ at time 0, which would render any thinking at time 1 unnecessary. Rather, it is as if at time 0 our agent only evaluated how good the options in a menu were, knowing that she would choose from them at a later stage – just like someone who evaluates which restaurant to go to dinner to without actually choosing what she will order there. In fact, we could also interpret the multiplicity of states in \( S \) as representing the fact that the agent has yet to decide what to choose – that she has not ‘solved the problem’ yet: each state of the world represents a possible preference of the agent, which she does not know yet, and the cost of thinking could be seen as the cost to figure out her preferences in order to make a choice from the menu. Following this interpretation, the cost of thinking would then be the cost of figuring out the state of the world. In Section 3.3 we formalize this intuition and characterize it axiomatically.

The second component of this representation is the function \( C \), which express the notion of ‘thinking cost’ by associating to every lottery a measure of the disutility caused by having to choose from it. It has some basic properties that allow us to interpret it as a cost of thinking. First of all, it only depends on the \( \text{supp}^* \) of a lottery: because the decision maker needs to make a choice for each menu in the support, if two lotteries have the same support, the cost of thinking is the same; and because there is nothing to decide from a singleton, only the \( \text{supp}^* \), and not the full support, matters for the cost of thinking. In addition, \( C \) is weakly positive everywhere, and

\[ To see why, consider \( \alpha, \beta \in \Delta(X) \), \( a, b \in (0, 1) \) s.t. \( a > b \). Because \( \text{supp}^*(a\alpha + (1-a)\beta) = \text{supp}^*(b\alpha + (1-b)\beta) \), then only the first component of the representation matters for the rank of these two lotteries. This means that we have \( \alpha \succeq^* \beta \) if \( a\alpha + (1-a)\beta \succeq b\alpha + (1-b)\beta \) which, in turn, hold iff the first component of the representation assigns a higher value to \( \alpha \) than to \( \beta \).

\[ As it is standard in the literature, \( \mu \) need not be a probability measure, but rather may contain negative components, usually referred to as negative states. This allows, for example, for the presence of temptation. We refer to DLR01, [6] and [5] for further discussion.

\[ This example also shows why the agent might be not able to solve the problem entirely at the time of the choice between menus: when she is choosing where to go to dinner, she might not know what she will feel like eating later when she will be actually seated in the restaurant (will she feel like eating fish or meat, or salad?). The fact that she doesn’t know her preferences yet is precisely what generates her preferences for flexibility, as in the original model of [23]; but it is also the reason why she cannot solve the problem at the time of the choice between menus.\]
it is equal to zero for the empty set, which guarantees a zero cost of thinking for the lotteries of singletons (whose supp* is empty). Finally, when Thinking Aversion holds, then the cost of thinking is monotone in the support of lotteries: lotteries of menus with strictly larger supports are weakly more costly to think about, because they require the agent to consider strictly more contingencies.

At the same time, we should stress that while the function $C$ is monotone in the support of lotteries, this does not mean that the cost of thinking of larger menus is larger: we could still have some $A, B \in \mathcal{X}$ such that $A \supset B$ but $C(\{A\}) < C(\{B\})$. While this form of monotonicity could appear natural, we believe it would be too restrictive, and that in general it might not hold: for example, a larger set could contain some option that is so obviously better than anything else, that the choice from the larger set is actually easier. For example, someone who really loves lobster might find it easier to choose from $\{\text{lobster, chicken, steak}\}$ than from $\{\text{chicken, steak}\}$, even though the former is a strictly larger menu.

Notice also that the function $C$ represents the anticipated thinking cost function: that is, it represents the cost that the agent expects to endure when she will have to choose from a set (or when she will need to form a contingent plan). In fact, $C$ captures a thinking effort that is exerted not when the menu is chosen, but later, when the agent is choosing from the menu, or some time before that. (Whatever thinking effort has been exerted in the choice between menus is already incorporated in the preferences $\succ$.)

The timing of this representation is therefore the following. First, at time zero, the agent chooses a lottery over menus. Then, before choosing from it, she discovers the state of the world and thinks about what to choose. (This is compatible with the interpretation that the discovery of the state is the outcome of the thinking process.) At time 1, the agent chooses from the menu, or forms a contingent plan if she is facing a lottery of menus. In an additional later stage, the lottery is realized and the agent is given her choice as specified by the contingent plan. (See Fig. 2.)

We conclude this discussion with a technical note. The first component of our representation is a generalization of the preferences in [23] to the case in which: 1) preferences are defined over lotteries over menus; 2) $\mu$ is not necessarily monotone, with a representation much reminiscent of the one in DLR01. One of the steps of the proof of Theorem 1 is to characterize such preference, a technical result which could be of independent interest: Lemma 2 in Appendix A.2 shows that any complete preference relation on $\Delta(\mathcal{X})$ admits a representation of this kind if and only it satisfies independence and continuity.17

A natural question is whether this could be obtained using the results in DLR01. We do not believe this is the case. The reason is, we operate on a different primitive, lotteries over menus, while DLR01 uses menus of lotteries. The key observation is then that our setup is, in some sense, ‘smaller:’ in Appendix A.1 we show that we are able to construct a bijection between our space and a finite-dimensional strict subset of the infinite-dimensional space of menus of lotteries. Results in DLR01 cannot therefore be applied.
3.2. Monotone content representation

Our assumptions thus far allowed the agent to prefer a smaller menu independently of the disutility she might receive from thinking about what she should choose. For example, she might have no cost of thinking at all ($\mathcal{C}(H) = 0$ for all $H \in 2^X$) and still prefer a smaller set to avoid temptation. Formally, this could be the case if $\mu(s) < 0$ for some $s \in S$, i.e., if there exist so-called negative states. We now introduce two postulates that rule out this possibility by guaranteeing that the presence of a disutility from thinking is the only feature that might induce the agent to prefer a smaller set.

A.6 (Content Monotonicity). For any $A, B \in \mathcal{X}$, $A \supseteq B$ implies $A \succ^* B$.

A.7 (Content Submodularity). For any $A, B, C \in \mathcal{X}$, $A \sim^* A \cup B$ implies $A \cup C \sim^* A \cup B \cup C$.

Axiom A.6 posits that $\succ^*$ is monotone: the agent genuinely prefers the content of a larger set to that of a smaller one, which is what we are after. (Naturally this does not imply that also $\succ$ is monotone.) Following [23], Axiom A.7 guarantees that there is a consistency in the way the preference behaves for larger set: 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 now define a representation in which the agent would always prefer a larger menu were it not for the cost of thinking.

Definition 2 (Content-Monotone Thinking-Averse representation). A preference relation $\succ$ on $\Delta_1(X)$ admits a Content-Monotone Thinking-Averse representation if it admits a Thinking-Averse representation $\langle S, \mu, u, \mathcal{C} \rangle$ where $\mu$ is a probability measure over $S$.

It turns out that to obtain this additional representation it is necessary and sufficient to add the two axioms above, Content Monotonicity and Content Submodularity.

Theorem 2. Let $\succ$ be a complete preference relation on $\Delta(\mathcal{X})$ that satisfies Best/Worst and that admits a Thinking-Averse representation. Then, $\succ$ satisfies Content Monotonicity and Content Submodularity if and only if it admits a Content-Monotone Thinking-Averse representation.

We should point out that the result above does not follow directly from applying the results in [23] to the preference $\succ^*$ and then using the representation in Theorem 1. The reason is, here preferences are defined on lotteries over menus, and not simply on menus. In particular, we need to obtain a representation which extends the one in [23] in a vNM sense, a result which could potentially be of independent interest: Lemma 4 in Appendix A.3 shows that a preference relation on $\Delta(\mathcal{X})$ admits a representation of this kind if and only if it satisfies continuity, independence, and the axioms of [23] on degenerate lotteries over menus.

3.3. Minimal cost representation

In our analysis thus far we have imposed the following requirements on the anticipated cost of thinking about a lottery: that it only depends on the $\text{supp}^*$ of the lottery; that it is weakly positive everywhere, and zero for singletons; and that it is monotone, i.e., that larger supports require more
thinking. We now strengthen this representation by means of an additional postulate that regulates the ‘disutility from thinking.’ Consider two menus \( A, B \in \mathcal{X} \) such that \( A \sim^* B \sim^* A \cup B \), and such that for each \( x \in A \) there exist some \( z_x \in B \) such that \( \{z_x\} \sim^* \{z_x, x\} \). That is, for each \( x \) in \( A \) there exist some \( z_x \) in \( B \) which is not only at least as good as \( x \), but that cannot be improved by adding \( x \) to it. At the same time, the aggregate utility of the content of \( A \) and \( B \) is the same, since \( A \sim^* B \sim^* A \cup B \). If this is the case, we would like to conclude that the cost of choosing from \( B \) cannot be higher than that of choosing from \( A \). The reason is that the agent could simply choose \( z_x \) from \( B \) whenever she choose \( x \) from \( A \), and by doing so she must obtain at least as much utility as she had from \( A \) (this happens because \( \{z_x\} \sim^* \{z_x, x\} \)). But since the utility of the content of \( A \) is the same as that of \( B \) (because \( A \sim^* B \)), then by following this strategy she would also obtain the full utility of the content of \( B \). That is, whatever rule the agent followed to choose from \( A \), it would work for \( B \) as well. However, there could be some more efficient rule that could be followed for \( B \), which means that the cost of choosing from \( B \) must be (weakly) lower. And since we have that: 1) \( A \sim^* B \), and 2) the choice from \( B \) is not harder than that from \( A \), then we should have \( B \succeq A \). This is the content of the following axiom.

A.8 (Cost Coherence). Consider \( A, B \in \mathcal{X} \) such that \( A \sim^* B \sim^* A \cup B \) and such that for each \( x \in A \) there exists some \( z \in B \) such that \( \{z\} \sim^* \{z, x\} \). Then \( B \succeq A \).

We will now show that this additional axiom allows us to strengthen our representation in light of one possible interpretation: the cost of thinking can be understood as the cost that the agent incurs to complete her preferences in order to make a choice. The idea of the representation is the following. As our agent faces a menu, e.g., enters a restaurant, she might be initially unable to compare two options, e.g., she might not know whether she feels like eating meat or fish. As we discussed, this incompleteness is already described in her preferences: the multiplicity of states can be seen as a way to represent the fact that the agent might be (currently) unable to compare two options.18

The agent will naturally have to resolve this incompleteness in order to make a choice – she will need to ‘complete’ these preference enough to decide what is the optimal option in the set. She will do so either by acquiring information about some of the options (‘how is the seabass prepared?’), or by simple introspection (‘do I feel like meat tonight?’). And just like the presence of incompleteness is represented by the multiplicity of states, the process of ‘completing’ the preferences could be understood as the process of finding out which is the true state – if the agent discovers the state, then her preferences become complete and are represented by \( u(\cdot, s) \) (where \( s \) is the realized state). That is, we can understand the process of thinking about what to choose as the process to figure out which is the state of world.19 In particular, it is as if the agent had to acquire some information, in the form of partitions of the state space, in order to make the optimal choice.

Which partition of the state space is required will naturally depend on the problem at hand. If some set \( A \) contains an option \( x \) that dominates all others in all possible states, then the agent doesn’t need to partition the state space at all – she can simply choose \( x \). Conversely, if in some

18 For example, there could exist some \( x, y \in X \) and \( s, s' \in S \) such that \( u(x, s) > u(y, s) \) and \( u(y, s') > u(x, s') \): in this case the agent is unable to compare \( x \) and \( y \) at the time of the preferences over menus. Some of this incompleteness might be (freely) resolved before she is asked to choose from the menu (she might receive some information). But some might not be, and the agent might have to think about which is her favorite option – which can be costly.

19 A similar interpretation is suggested in [12].
set \( B \) there is no element that is optimal both in \( s \) and in \( s' \) for some \( s, s' \in S \), then the agent will have to separate \( s \) and \( s' \): she will have to acquire information/perform introspection to figure out whether the state of the world is \( s \) or \( s' \). But this could be costly – hence the cost of thinking.

We now turn to write this intuition formally. To this end, we need a few additional definitions. If \( S \) is a finite non-empty set (state space), denote by \( \Pi(S) \) the set of partitions of \( S \). Moreover, for any state-dependent utility function \( u : X \times S \to \mathbb{R} \) and any \( A \in \mathcal{X} \), define \( \mathcal{P}_{S,u}(A) \) as

\[
\mathcal{P}_{S,u}(A) := \left\{ \pi \in \Pi(S) : \text{for all } E \in \pi \exists x \in A \text{ such that } \max_{y \in A} u(y, s) = u(x, s) \text{ for all } s \in E \right\}.
\]

We understand \( \mathcal{P}_{S,u}(A) \) as the set of all partitions that allow the agent to attain full utility from \( A \) by choosing the same alternative in every state not separated by the partition. That is, \( \mathcal{P}_{S,u}(A) \) is the set of partitions that the agent could use to choose from \( A \). Indeed there could be many of them: \( \mathcal{P}_{S,u}(A) \) is almost never a singleton – for example the finest partition will always belong to it. The idea of the representation is then the following. To each of these partitions there is a cost associated. This cost will be partition-monotone: finer partitions are more costly. Then, the cost of thinking about a menu is simply the cost of the least costly partition that the agent can use to choose from this menu.

**Definition 3.** 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 \( \pi \) is finer than \( \pi' \).

It turns out that adding Cost Coherence to our axiomatic structure gives us exactly the representation above.

**Theorem 3.** Let \( \succeq \) be a complete preference relation on \( \Delta(\mathcal{X}) \) that satisfies Best/Worst and that admits a Content-Monotone Thinking-Averse representation. Then, \( \succeq \) satisfies Cost Coherence if and only if it admits a Content-Monotone Thinking-Averse representation \( (S, \mu, u, \mathcal{C}) \) such that there exists a partition-monotone function \( c : \Pi(S) \to \mathbb{R}^+ \) such that \( c((S, \emptyset)) = 0 \) and for all \( A \in \mathcal{X} \):

\[
\mathcal{C}([A]) = \min_{\pi \in \mathcal{P}_{S,u}(A)} c(\pi).
\]

### 3.4. Discussion: thinking too much?

One of the features of the three representations above is that the agent ranks menus as if she expected to choose the best option from each of them, despite the cost that this might entail. In fact, this is evident from the fact that in the first part of the representation the agent expects to choose the element that maximizes her utility in every state. This might sound extreme: why should the agent always choose the best option, given that it is costly to do so, instead of just exerting the optimal amount of thinking?

The reason is the following: an excessive amount of thinking seems to be a feature of any model (of this general class) in which agents prefer a smaller set to avoid the cost of thinking involved in choosing from a larger one – the behavior that we are after. This happens because if an agent were to exert only the optimal amount of thinking, then she should always prefer to
have more options, since she can optimally choose not to look at them. It is only the ‘fear’ of excessive thinking that would induce her to prefer a smaller set.20

The presence of ‘excessive thinking’ constitutes the main difference between the present paper and [10] and [12].21 What the latter two papers model is an agent who expects herself to choose the optimal thinking strategy from a pool of available ones: better strategies are more costly, but at the same time allow the agent to find better options in a menu.22 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 (weakly) more utility from its content, and since to choose from it she can at least use the same strategy she used for the smaller set, then she cannot be worse off.23 By contrast, in our representation this need not be true, as shown by Theorem 2. In particular, in our case the agent could dislike a bigger set only because she knows that she will have to think harder to choose from it. From this point of view the two models stand at the opposite sides of an interpretation pole. On the one side, in [12] we have an agent who expect herself to rationally react to a computational limitation: she knows she will think just as much as optimal. This, as we have discussed, leads to a preference for larger sets (unless there are other reasons to prefer smaller ones, like an interest to avoid temptation). On the other side of the interpretation pole, in our paper agents behave as if they expected themselves to think too much, possibly more than what they would consider optimal now. In this sense, it is as if our agents knew that they would be ‘tempted’ into excessive thinking. Anticipating this, they might then choose to face smaller sets to avoid this effort – which is the behavior that motivated our analysis.24

3.5. Uniqueness properties

We now turn to discuss the uniqueness properties of the representations above. The first kind of uniqueness that we wish to establish is the ability to uniquely identify the two components of the representation, the genuine evaluation of the content and the anticipated thinking cost function $C$. Such uniqueness is essential for this separation to be meaningful. The following proposition shows that it is a feature of our model.

20 We should also point out that this ‘excessive thinking’ is a feature of the representation, and not of the axioms – we don’t impose that the agent expects to ‘solve’ the problem, but rather we find a representation in which the agent behaves as if she expected herself to solve the problem.
21 Moreover, there are also differences on the axioms and on the primitives: [10] looks at preferences over menus, [12] uses menus of lotteries, while we use lotteries of menus.
22 The representation in [12] is of the form

$$W(A) = \max_{\mu \in \mathcal{M}} \left[ \int_S \max_{p \in A} U(p, s) \mu(ds) - c(\mu) \right]$$

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. They show that each $\mu \in \mathcal{M}$ is positive if and only if the underlying preference relation is monotone.
23 This is obvious in the case of [10], since monotonicity is the only postulate. In [12] it is proven in their Theorem 1.B.
24 Notice that both results relate to how the agent expects to act in the future, when they will be asked to make a choice from the set. This means that in our representation agents expect themselves to possibly think too much in the future, and for this reason they prefer a smaller set now. What they will do at the time of choice, however, we cannot say.
Proposition 1. Let $≽$ be a complete preference relation on $\Delta(\mathcal{X})$ that satisfies Best/Worst. If $⟨S, μ, u, C⟩$ and $⟨S′, μ′, u′, C′⟩$ are both Thinking-Averse representations of $≽$, then there exist $a ∈ \mathbb{R}^+$, $b ∈ \mathbb{R}$ such that for all $α ∈ \Delta(\mathcal{X})$

$$\sum_{A ∈ \mathcal{X}} α_A \left( \sum_{s ∈ S'} μ'(s) \left[ \max_{y ∈ A} u'(y; s) \right] \right) = a \left[ \sum_{A ∈ \mathcal{X}} α_A \left( \sum_{s ∈ S} μ(s) \left[ \max_{y ∈ A} u(y; s) \right] \right) \right] + b$$

and

$$C' = aC.$$

Proposition 1 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 ($a$ is the same for both). This implies that, if we fix the evaluation of the content, the representation of the cost is unique.

Moreover, one might expect to have uniqueness of the endogenous state space $S$, much in line with the analysis in DLR01. Unfortunately, however, this is not a feature of our model. 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’. As a result, we do not have enough observations to identify the state space $S$ uniquely. At the same time, it is not hard to see that the full uniqueness of the state space, together with all the properties of the DLR01 representations for the characterization of $≽^*$, could be obtained if we extend our analysis to the case of lotteries of menus of lotteries, instead of lotteries of menus. (Such a framework would not be new to decision theory: it is used, for example, in [9] or [11].)

3.6. Being more Thinking-Averse: a comparability result

We now introduce a comparability notion for Thinking Aversion, much in line with similar notions for risk 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 cost of thinking. We shall therefore consider two preference relations $≽_1$ and $≽_2$ such that $≽_1^* = ≽_2^*$. This implies that their representations must be connected.

Observation 1. Consider two preference relations $≽_1$ and $≽_2$ on $\Delta(\mathcal{X})$ that have a Thinking-Averse representation and such that $≽_1^* = ≽_2^*$. Then, there exist a Thinking-Averse representation $(S_1, μ_1, u_1, C_1)$ that represents $≽_1$ and a Thinking-Averse representation $(S_2, μ_2, u_2, C_2)$ that represents $≽_2$ such that $S_1 = S_2$, $μ_1 = μ_2$ and $u_1 = u_2$.

[25] By smaller we mean the following. As we mentioned before, in Appendix A.1 we show that we are able to construct a bijection between our space and a finite-dimensional strict subset of the infinite-dimensional space of menus of lotteries. In this sense we mean smaller.

[26] Alternatively, one might seek uniqueness of the state space in the sense of [8]: 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 [8] for a detailed discussion.) It is easy to show that their uniqueness result (Theorem 3.1 in their paper) applies here to our Content-Monotone Thinking Aversion representation characterized in Theorem 2.
Definition 4. Consider two preference relations $\succsim_1$ and $\succsim_2$ on $\Delta(\mathcal{X})$ that satisfy Best/Worst, have a Thinking-Averse representation, and such that $\succsim^*_1 = \succsim^*_2$. We say that $\succsim_1$ is more Thinking Averse than $\succsim_2$ if, for any $\alpha \in \Delta(\mathcal{X})$ and $p \in \Delta^S(\mathcal{X})$, we have

$$\alpha \succsim_1 p \implies \alpha \succsim_2 p.$$ 

Consider two agents with the same genuine evaluation of the content of sets but such that the first ‘dislikes thinking’ more than the second. Suppose that for some $\alpha \in \Delta(\mathcal{X})$ and $p \in \Delta^S(\mathcal{X})$ we have $\alpha \succsim_1 p$. This means that the first agent would rather think about $\alpha$ than take $p$, albeit the latter requires no thinking. Then, if the second agent has the same genuine evaluation of the content of menus and an even lower dislike of thinking, she should do the same, and we should have $A \succsim_2 p$ as well. Notice that this definition parallels the one of comparative risk aversion and similar ones of comparative ambiguity aversion (like the one in [15]).

Proposition 2. Consider two preference relations $\succsim_1$ and $\succsim_2$ on $\Delta(\mathcal{X})$ that satisfy Best/Worst, have a Thinking-Averse representation, and such that $\succsim^*_1 = \succsim^*_2$. Then, the following two statements are equivalent:

(i) $\succsim_1$ is more Thinking Averse than $\succsim_2$;
(ii) for any two Thinking-Averse representations $(S_1, \mu_1, u_1, \mathcal{C}_1)$ of $\succsim_1$ and $(S_2, \mu_2, u_2, \mathcal{C}_2)$ of $\succsim_2$ such that $S_1 = S_2$, $\mu_1 = \mu_2$ and $u_1 = u_2$, we have $\mathcal{C}_1 \succeq \mathcal{C}_2$.

4. Conclusion

In this paper we analyze a preference relation on lotteries over menus characterized by the presence of a trade-off. On the one hand, the agent prefers larger sets since they give her more options to choose from. On the other hand, she prefers smaller sets since she does not like to think about what to choose from larger ones. 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 the menu. 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 standard representations in the literature, and an evaluation of the cost of thinking about the set, which depends only on the non-singleton elements of the support of the lottery, is weakly positive everywhere, assigns zero cost to thinking about singletons, and assigns a higher thinking cost to lotteries with a larger support.

We further strengthen this characterization in two ways. First, we show that simply adapting the axioms in [23] 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 provide a behavioral postulate that guarantees that we can interpret the cost of thinking as the cost of the cheapest partition of the state space that allows the agent to make the optimal choice from the set.

Future research could analyze the possible implications of these preferences in a standard economic environment. One such analysis is presented in [28], 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, like the tendency of investors 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. Let us denote by \( \mathcal{X} \) the set of compact and convex 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 \( \mathcal{X} \).

The proof goes as follows. Define \( g : \Delta(\mathcal{X}) \to \mathcal{X} \) as

\[
g(\alpha) := \sum \alpha_C \text{ conv}(C),
\]

where \( \sum \) in \( \mathcal{X} \) is understood in the standard sense of set mixing a’ la Minkowsky. Define now \( L \) as the range of \( g \), that is, \( L := \{ \hat{A} \in \mathcal{X} : \hat{A} = g(\alpha) \text{ for some } \alpha \in \Delta(\mathcal{X}) \} \).

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

**Proof.** We only need to prove that for any \( \alpha, \beta \in \Delta(\mathcal{X}) \), we have \( g(\alpha) \neq g(\beta) \). We 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 \( \alpha, \beta \in \Delta(\mathcal{X}) \), \( \alpha \neq \beta \) and say, by means of contradiction, that \( g(\alpha) = g(\beta) \). Notice that if there exists \( x \in X \) such that \( \alpha_{\{x\}} < \beta_{\{x\}} \), then any lottery in \( g(\beta) \) must give a minimum weight of \( \beta_{\{x\}} \) to \( x \), while there would be lotteries in \( g(\alpha) \) that give a weight lower than \( \beta_{\{x\}} \), hence we would not have \( g(\alpha) = g(\beta) \). So we have \( \alpha_{\{x\}} = \beta_{\{x\}} \) for all \( x \in X \).

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 \( x \). Define \( \hat{\alpha}^x \in \Delta(\mathcal{X}) \) as follows:

\[
\hat{\alpha}^x_C = \begin{cases} 0, & \text{if } x \in C, \\ \frac{\alpha_{\{x\}} + \alpha_{\{x\} \cup \{i\}}}{1 - \alpha_{\{x\}}}, & \text{otherwise}. \end{cases}
\]

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

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

We then must have \(\hat{\alpha}^x = \hat{\beta}^x\) for all \(x \in X\). Since \(\alpha_{\{x\}} = \beta_{\{x\}}\) for all \(x \in X\), then \(\alpha_D + \alpha_{D \cup \{x\}} = \beta_D + \beta_{D \cup \{x\}}\) 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 \(\alpha_{\{x\}} = \beta_{\{x\}}\) 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 \(\alpha = \beta\), a contradiction. □

Claim 2. \(g\) is linear and continuous.

Proof. To prove continuity we need to prove that, for any \((\alpha_n) \in \Delta(\mathcal{X})^\infty, \alpha \in \Delta(\mathcal{X})\) with \(\alpha_n \to \alpha\), we have \(g(\alpha_n) \to g(\alpha)\). It is immediate to see that, in the Hausdorff topology of \(\mathcal{X}\), \(a_i^n \to a_i\) for all \(i\) implies that \(\sum a_i^n \text{conv}(C_i) \to \sum a_i \text{conv}(C_i)\).

To prove linearity, take any \(\alpha, \beta \in \Delta(X), a \in (0, 1)\). Notice that we must have \(a\alpha + (1 - a)\beta = \sum[a\alpha_C + (1 - a)\beta_C]C\). But notice that we must then have \(g(a\alpha + (1 - a)\beta) = \sum[a\alpha_C + (1 - a)\beta_C]\text{conv}(C) = a[\sum a\alpha_C \text{conv}(C)] + (1 - a)[\sum \beta_C \text{conv}(C)] = ag(a) + (1 - a)g(\beta)\), which in turns proves linearity. □

Finally, notice that linearity and continuity of \(g\) guarantee the convexity and compactness of \(L\). This concludes the proof of Lemma 1. □

A.2. Mapping of the representations

We now show that if we consider a preference relation \(\succsim\) on \(\Delta(\mathcal{X})\) that satisfies independence and continuity, we can represent it in a way 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.)

Definition 5. A preference relation \(\succsim\) on \(\Delta(\mathcal{X})\) satisfies independence if, for all \(\alpha, \beta, \gamma \in \Delta(\mathcal{X}), a \in [0, 1]\),

\[\alpha \succsim \beta \iff a\alpha + (1 - a)\gamma \succsim a\beta + (1 - a)\gamma.\]

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

(i) \(\succsim\) on \(\Delta(\mathcal{X})\) satisfies continuity 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 \(\mu\) over \(S\) such that it is represented by

\[U(\alpha) = \sum \alpha_A \left(\sum_{s \in S} \mu(s) \left[\max_{y \in A} u(y, s)\right]\right),\]

\[27\] To do so, simply consider, in the mixture, the same sets and elements we considered to create \(p\).
To prove this result we first ‘translate’ our preference relation to $\mathcal{X}$ (the space used in DLR01). By Lemma 1, we know that we have a continuous bijection $g$ between $\Delta(\mathcal{X})$ and a compact and convex subset $H$ of $\mathcal{X}$. Recall that we have $N = |X|$. Consider now the following set of utilities: $\mathcal{U} := \{u \in \mathbb{R}^{\Delta(X)}: u$ 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 $u(x_1) = u(x_2) = \cdots = u(x_{N-1})\}$. (Geometrically these are the utilities that generate indifference curves that are parallel to each of the faces of the simplex.) Notice that $|\mathcal{U}| < \infty$. 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 $u \in \mathcal{U}$ 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 $\text{ext}(H)$, is the $g$-image of $\mathcal{X}$. Recall the geometrical intuition of the elements of $\mathcal{U}$: 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 $A \in \text{ext}(H)$ 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 $u \in \mathcal{U}$ 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 $u \in \mathcal{U}$ 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. \(\square\)

(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 $\mathbb{R}^{|\mathcal{U}|}$ (recall that $|\mathcal{U}| < \infty$ by finiteness of $X$): simply associate every set to the vector that has the utility given by the set in every $u \in \mathcal{U}$. Call this map $h$. Again, standard arguments show that $h$ is a linear, continuous bijection. We have therefore a linear and continuous bijection $l := g \circ h$ from $\Delta(X)$ to $C$. Now, define the preference relation $\succ$ on $C$ by

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

\[
x \hat{\succ} y \iff \sum_{u_i \in \mathcal{U}} \mu(u_i) x_i \geq \sum_{u_i \in \mathcal{U}} \mu(u_i) y_i.
\]

But then, by definition of \( \ hat{\succ} \) and since \( l \) is a bijection, we have

\[
\sum \alpha A A \succ \sum \beta A A \iff \sum \alpha A \left[ \sum_{u_i \in \mathcal{U}} \mu(u_i) \max_{x \in \text{conv}(A)} u_i(x) \right] \geq \sum \beta A \left[ \sum_{u_i \in \mathcal{U}} \mu(u_i) \max_{x \in \text{conv}(A)} u_i(x) \right].
\]

Since, by affinity of \( u \), we have \( \max_{x \in \text{conv}(A)} u_i(x) = \max_{x \in A} u_i(x) \), this concludes the proof of Lemma 2.

### A.3. Extending [23] to lotteries of menus

In order to prove Theorem 2 we need to extend the representation of [23] to the space of lotteries over menus, \( \Delta_1(X) \), in a vNM sense. Of note, this representation has been characterized in [26] (the working paper version of [27]) by means of one novel axiom, indirect stochastic dominance. By contrast, we prove here that the same representation can be derived imposing the axioms of [23] on the degenerate lotteries.

To prove this result, we use the following lemma, which is an extension of Lemma 3 in [23]: for completeness we include the full proof. (The core idea is to show that the representation in [23] 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 \( \succ \) and \( \succeq \) such that:

1. \( \succ \) is complete and transitive;
2. \( \succeq \) is reflexive;
3. \( y \succ y' \) and \( y \neq y' \) imply not \( y' \succ y \).

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

**Proof.** To prove it, let \( \sim \) and \( \succ \) denote the symmetric and asymmetric parts of \( \succ \) and \( \succeq \) the asymmetric part of \( \succeq \). Notice that \( \succ \) is a weak preference relation, and that (by (3)) we have \( y \succeq y' \implies y \succ y' \). Define \( w \) and \( w^* \) as \( w(y') := \sum_{y : y \succeq y'} a(y) \) and \( w^*(y') := \sum_{y : y \succeq y'} a(y) \). 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 \( \succ \), 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 \( y \succeq y' \implies y \succ y' \). But we have already defined \( a(y) \) 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 \( \succ \). For the
same reason we have \( a(y') + w^*(y') \leq a(y) + w^*(y) \) for any \( y \succ y' \). Since \( Y \) is finite, there are finitely many \( \sim \)-equivalence classes, and the induction procedure gives the representation. \( \square \)

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

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

(i) \( \succ \) satisfies continuity, independence, Content Monotonicity and Content 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 \( \mu \) over \( S \) such that it is represented by

\[
U(\alpha) = \sum \alpha_A \left( \sum_{s \in S} \mu(s) \left( \max_{y \in A} u(y; s) \right) \right).
\]

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 \( \alpha, \beta \in \Delta(\mathcal{X}) \), we have

\[
\alpha \succ \beta \iff \sum \alpha_i V(A_i) \geq \sum \beta_i V(B_i).
\]

Now define by \( \tilde{\succ} \) the restriction of \( \succ \) 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 \( \mu \) over \( S \) such that

\[
\tilde{U}(A) := \sum_{s \in S} \mu(s) \left[ \max_{y \in A} u(y; s) \right]
\]

represents \( \tilde{\succ} \), and such that there exists \( b \in \mathbb{R} \) such that \( \tilde{U}(A) = V(\delta_A) + b \) for all \( A \in \mathcal{X} \). But notice that this claim is almost identical to Theorem 1 in [23], with the exception of the last requirement. Following the proof, define \( f : \mathcal{X} \to \mathcal{X} \) as

\[
f(A) := \bigcup_{B \in \mathcal{X} : A \sim A \cup B} B.
\]

(2)

Define \( S := \{ A \in \mathcal{X} : A = f(A) \} \), and follow identical passages as the proof of Theorem 1 in [23], but using Lemma 3 in this paper instead of Lemma 3 in [23]. This guarantee the fact that \( \tilde{U}(A) = V(A) + b \) for all \( A \in \mathcal{X} \) for some \( b \in \mathbb{R} \). The representation then follows immediately.

**Appendix B. Proofs of the results in the text**

**B.1. Proof of Theorems 1 and 2**

**Sufficiency of the axioms**

To prove the sufficiency of the axioms, we proceed in the following steps: 1) We notice how \( \succ^* \) is a complete and transitive binary order that satisfies independence and continuity. 2) We then use Lemma 2 (or Lemma 4) to characterize it with a representation \( W^* \). 3) We notice that the restrictions of \( \succ \) and \( \succ^* \) on \( \Delta^S(X) \) coincide. 4) We then find an Expected-Utility representation \( \hat{W} \) of the restriction of \( \succ \) on \( \Delta^S(X) \) that coincides with \( W^* \) on \( \Delta^S(X) \). 5) We define a representation \( W \) for \( \succ \) on the subset of \( \Delta(\mathcal{X}) \) s.t. \( \alpha \sim p \) for some \( p \in \Delta^S(X) \), by defining
Proof. We then define a cost associated to each lottery $\alpha$ by the difference between $W^*$ and $W$. 6) We show how Axiom A.2 implies that if $\alpha$ has a larger $\sup^*$, then the cost is higher. 7) We extend to the general $\Delta(\mathcal{X})$ by Axiom A.3.

Claim 4. For every $\alpha, \beta \in \Delta(\mathcal{X})$, if $a\alpha + (1-a)\beta \succeq b\alpha + (1-b)\beta$ for some $a, b \in (0, 1)$, $a > b$, then $c\alpha + (1-c)\beta \succeq d\alpha + (1-d)\beta$ for all $c, d \in (0, 1)$, $c > d$.

Proof. This result follows immediately from Axiom A.3(1), because the lottery $a\alpha + (1-a)\beta$ has the same $\sup^*$ for any $a \in (0, 1)$. □

Claim 5. $\sup^*$ is transitive.

Proof. Consider $\alpha, \beta, \gamma \in \Delta(\mathcal{X})$ such that $\alpha \succ^* \beta \succ^* \gamma$. We need to show that we have $\alpha \succ^* \gamma$. Consider some $\epsilon > 0$ small enough, and notice that $\alpha \succ^* \beta$ implies $(\frac{1}{2} + 2\epsilon)\alpha + (\frac{1}{2} - 2\epsilon)\beta \succ^* (\frac{1}{2} - 2\epsilon)\alpha + (\frac{1}{2} + 2\epsilon)\beta$ by independence of $\sup^*$ (Axiom A.3). For the same reason, this implies $(\frac{1}{3} + 2\epsilon)\alpha + (\frac{1}{3} - 2\epsilon)\beta + \frac{1}{3} \gamma \succ^* (\frac{1}{3} - 2\epsilon)\alpha + (\frac{1}{3} + 2\epsilon)\beta + \frac{1}{3} \gamma$. Similarly, $\gamma \succ^* \beta$ implies $\frac{1}{3} \alpha + (\frac{1}{3} + 2\epsilon)\beta + (\frac{1}{3} - 2\epsilon)\gamma \succ^* \frac{1}{3} \alpha + (\frac{1}{3} - 2\epsilon)\beta + (\frac{1}{3} + 2\epsilon)\gamma$. By definition of $\succ^*$ and by Claim 4 these imply $(\frac{1}{4} + \epsilon)\alpha + (\frac{1}{4} - \epsilon)\beta + \frac{1}{4} \gamma \succ^* \frac{1}{4} \alpha + (\frac{1}{4} - \epsilon)\beta + \frac{1}{4} \gamma$ and $\frac{1}{3} \alpha + (\frac{1}{3} - \epsilon)\beta + \frac{1}{3} \gamma \succ^* \frac{1}{3} \alpha + (\frac{1}{3} - \epsilon)\beta + (\frac{1}{3} + \epsilon)\gamma$. By transitivity of $\succ^*$, we obtain $(\frac{1}{4} + \epsilon)\alpha + (\frac{1}{4} - \epsilon)\beta + \frac{1}{4} \gamma \succ^* \frac{1}{6} (\frac{3}{1 - 3\epsilon}) \alpha + (\frac{1}{3} - \epsilon)\beta + \frac{1}{2} (\frac{3}{1 - 3\epsilon}) \gamma$. By independence of $\sup^*$ this implies $\alpha \succ^* \gamma$ as sought. □

Claim 6. $\sup^*$ is complete.

Proof. Consider $\alpha, \beta \in \Delta(\mathcal{X})$. For them to be incomparable according to $\sup^*$, there must exist $a, b, c, d \in (0, 1)$ such that $a > b$, $c > d$, and $a\alpha + (1-a)\beta > b\alpha + (1-b)\beta$ but $c\alpha + (1-c)\beta < d\alpha + (1-d)\beta$. Let us assume, by means of contradiction, that they exist. Consider now some $\tilde{a} > \max \{a, b, c, d\}$ and $\tilde{a} < \min \{a, b, c, d\}$. Assume without loss of generality that we have $\tilde{a}\alpha + (1-\tilde{a})\beta \succ \tilde{a}\alpha + (1 - \tilde{a})\beta$. We will now argue that we must have that for any $\tilde{\lambda}, \tilde{\lambda}$ with $\tilde{\lambda} \geq \tilde{\lambda} > \tilde{\lambda} > a$, we have $\tilde{\lambda}\alpha + (1-\tilde{\lambda})\beta \succeq \tilde{\lambda}\alpha + (1-\tilde{\lambda})\beta$. To prove this, notice that we must have that the lottery $ha + (1-h)\beta$ has the same $\sup^*$ for all $h \in (0, 1)$. We can therefore apply Axiom A.3(1): $\tilde{a}\alpha + (1-\tilde{a})\beta \succeq \tilde{a}\alpha + (1 - \tilde{a})\beta$ implies $\lambda\alpha + (1-\lambda)\beta \succeq \lambda\alpha + (1-\lambda)\beta$; using it again, we get $\hat{\lambda}\alpha + (1-\hat{\lambda})\beta \succeq \lambda\alpha + (1-\lambda)\beta$. This means that we have $c\alpha + (1-c)\beta \succeq d\alpha + (1-d)\beta$, a contradiction. □

Consider now $x^*$ and $x_*$ from Axiom A.5, and define the set $H := \{\alpha \in \Delta(X): \{x^*\} \succeq \alpha \succeq \{x_*\}\}$.

Claim 7. For all $\alpha \in \Delta(X)$ there exist $p^* \in \Delta^S(X)$ such that $p^* \sim^* \alpha$. Similarly, for any $\alpha \in H$, there exists $p \in \Delta^S(X)$ s.t. $p \sim \alpha$.

Proof. Consider any $\alpha \in \Delta(\mathcal{X})$ and any $x^*, x_* \in X$ such that $x^* \succ^* \alpha \succ^* x_*$. (Their existence is guaranteed by Axiom A.5.) It suffices to show that there exists $\lambda \in [0, 1]$ such that $\lambda x^* + (1-\lambda)x_* \succeq^* \alpha$. Say, by means of contradiction, that this is not the case. Then, define $\lambda^* := \min \{\lambda \in [0, 1]: \lambda x^* + (1-\lambda)x_* \succeq^* \alpha\}$ and $\lambda_* := \max \{\lambda \in [0, 1]: \alpha \succeq^* \lambda x^* + (1-\lambda)x_*\}$, and notice that both are well defined by A.4. Notice that we cannot have $\lambda^* = \lambda_*$,
since it would imply \( \lambda^* x^* + (1 - \lambda^*) x_\ast \sim^* \alpha \), which we know is not true. Notice also that we cannot have \( \lambda^* > \lambda_\ast \). If this were the case, consider any \( \lambda' \in (\lambda_\ast, \lambda^*) \), and notice that we could not have \( \lambda' x^* + (1 - \lambda') x_\ast \nsim^* \alpha \) since this violates the definition of \( \lambda^* \) (it is the minimum \( \lambda \) such that this is true), nor \( \alpha \nsim^* \lambda' x^* + \lambda x_\ast \) since this violates the definition of \( \lambda_\ast \) (it is the maximum \( \lambda \) such that this is true). Therefore, we must have \( \lambda_\ast > \lambda^* \). Notice that therefore we have \( \lambda^* x^* + (1 - \lambda^*) x_\ast > \alpha > \lambda_\ast x^* + (1 - \lambda_\ast) x_\ast \), \( \lambda_\ast > \lambda^* \), and \( x^* \nsim^* x_\ast \). But this is a violation of independence of \( \nsim^* \) on \( \Delta^S(X) \), Axiom A.3, since it implies that if \( \lambda_\ast > \lambda^* \) and \( x^* \nsim^* x_\ast \), then \( \lambda_\ast x^* + (1 - \lambda_\ast) x_\ast > \alpha \nsim^* \lambda^* x^* + (1 - \lambda^*) x_\ast \). The proof of the second part is equivalent. \( \square \)

Therefore, we must have that \( \nsim^* \) is a complete, transitive preference relation that satisfies independence (Axiom A.3) and continuity (Axiom A.4). To proceed in the proof of Theorem 1 we could therefore apply Lemma 2 and obtain that 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 \( \mu \) over \( S \) such \( \nsim^* \) is represented by

\[
W^*(\alpha) := \sum_{s \in S} \alpha_A \left( \sum_{y \in A} \mu(s) \max \{u(y; s)\} \right).
\]

If, alternatively, we are proving Theorem 2, we also have that \( \nsim^* \) satisfies also Axioms A.6 and A.7. We can therefore apply Lemma 4 instead of Lemma 2 and obtain a representation as above, but in which \( \mu \) is a probability measure.

Now notice that the restriction of \( \nsim \) to \( \Delta^S(X) \) is a complete, transitive binary relation that satisfies standard continuity (by Axiom A.4) and independence (by Axiom A.3). There will therefore exist \( v : X \to \mathbb{R} \) such that \( \sum_x p(x) v(x) \) represents the restriction of \( \nsim \) to \( \Delta^S(X) \). Let us consider \( \bar{x}, \bar{x}' \) so that \( W^*(\delta_{\bar{x}}) \succeq W^*(\delta_{\bar{x}'}) \) for all \( x \in X \) (their existence is trivial by finiteness), and normalize \( v \) so that \( v(\bar{x}) = W^*(\delta_{\bar{x}}) \) and \( v(\bar{x}') = W^*(\delta_{\bar{x}'}) \). Now define \( \hat{W} : \Delta^S(X) \to \mathbb{R} \) by \( \hat{W}(p) := \sum_x p(x) v(x) \), and define \( W' : H \to \mathbb{R} \) by \( W'(\alpha) := \hat{W}(p_\alpha) \) for \( p_\alpha \sim \alpha \). (This is well defined by construction of \( H \).)

Claim 8. For any \( \alpha, \beta \in \Delta(\mathcal{X}, X) \) s.t. \( \text{supp}^*(\alpha) = \text{supp}^*(\beta) \), we have \( \alpha \nsim \beta \iff \alpha \nsim^* \beta \).

Proof. Consider \( \alpha, \beta \in \Delta(\mathcal{X}, X) \) s.t. \( \text{supp}^*(\alpha) = \text{supp}^*(\beta) \). By Axiom A.3, we have \( \alpha \nsim \beta \iff a\alpha + (1 - a)\beta \nsim b\alpha + (1 - b)\beta \) for all \( a, b \in (0, 1), a > b \). By construction of \( \nsim^* \), this holds iff \( \alpha \nsim^* \beta \) as sought. \( \square \)

Notice that the restrictions of \( \nsim \) and \( \nsim^* \) on \( \Delta^S(X) \) coincide, that is, for all \( p, q \in \Delta^S(X) \), we have \( p \nsim q \) if and only if \( p \nsim^* q \). To see why, it is standard practice to show that by Axiom A.3, we have \( p \nsim q \) if and only if \( ap + (1 - a)q \nsim bp + (1 - b)q \) for all \( a, b \in (0, 1) \) with \( a > b \), hence \( p \nsim^* q \). We will now argue that, since the restrictions of \( \nsim \) and \( \nsim^* \) on \( \Delta^S(X) \) coincide, and since we have normalized \( v \) so that \( v(\bar{x}) = W'(\delta_{\bar{x}}) = W^*(\delta_{\bar{x}}) \) and \( v(\bar{x}') = W'(\delta_{\bar{x}'}) = W^*(\delta_{\bar{x}'}) \), then we must have \( W^*(p) = W'(p) \) for all \( p \in \Delta^S(X) \). To see why, notice first of all that for all \( p, q \in \Delta^S(X) \) and \( a \in (0, 1) \) we must have that \( W'(ap + (1 - a)q) = aW'(p) + (1 - a)W'(q) \), and that \( W^*(ap + (1 - a)q) = aW^*(p) + (1 - a)W^*(q) \). Now consider any \( p \in \Delta^S(X) \), and notice that we must have \( \lambda \bar{x} + (1 - \lambda)\bar{x}' \sim p \) for some \( \lambda \in [0, 1] \) by standard arguments. We must then also have \( \lambda \bar{x} + (1 - \lambda)\bar{x}' \nsim^* p \). In turn, this means \( W'(p) = \lambda W'(\bar{x}) + (1 - \lambda)W'(\bar{x}') = \lambda \bar{W}^*(\bar{x}) + (1 - \lambda)W^*(\bar{x}') = W^*(p) \), where the central equality comes from the normalization of \( v \), while the first and the last equality come from the fact that \( W' \) and \( W^* \) represent, respectively, \( \nsim \) and \( \nsim^* \).
Now define \( \hat{C} : H \to \mathbb{R} \) as \( \hat{C}(\alpha) := W^*(\alpha) - W'(\alpha) \).

**Claim 9.** For all \( \alpha \in H, q \in \Delta^S(X) \), and \( \lambda \in (0, 1) \), we have \( \hat{C}(\alpha) = \hat{C}(\lambda \alpha + (1 - \lambda)q) \).

**Proof.** Consider \( \alpha \in H, q \in \Delta^S(X) \), and \( \lambda \in (0, 1) \). Now consider \( p_\alpha \in \Delta^S(X) \) such that \( p_\alpha \sim \alpha \) (the existence of which is guaranteed by definition of \( H \)). By Axiom A.3, this implies \( \frac{1}{1+x} p_\alpha + \frac{1}{1+x} p_{\lambda q} + (1-\lambda)\alpha \sim \frac{\lambda}{1+x} p_\alpha + \frac{1}{1+x} p_{\lambda q} + (1-\lambda)\alpha \). Because \( \sim \) is represented by \( W' = W^* - \hat{C} \), using the linearity of \( W^* \) this implies \( \hat{C}(\alpha) = \hat{C}(\lambda q + (1-\lambda)\alpha) \). \( \square \)

**Claim 10.** For all \( \alpha, \beta \in H \), if \( \text{supp}^*(\alpha) \supseteq \text{supp}^*(\beta) \) then \( \hat{C}(\alpha) \geq \hat{C}(\beta) \).

**Proof.** Consider \( \alpha, \beta \in H \) s.t. \( \text{supp}^*(\alpha) = \text{supp}^*(\beta) \). Consider now \( p_\alpha^*, p_\beta^* \in \Delta^S(X) \) s.t. \( \alpha \sim^* p_\alpha^*, \beta \sim^* p_\beta^* \) (the existence of which is guaranteed by Claim 7). Now notice that we must have \( \frac{1}{2} W^*(\alpha) + \frac{1}{2} W^*(\beta) = \frac{1}{2} W^*(\alpha) + \frac{1}{2} W^*(\beta) \), which implies \( \frac{1}{2} \alpha + \frac{1}{2} P_{\alpha}^* \sim^* \frac{1}{2} \beta + \frac{1}{2} \beta \). By Claim 8 this implies \( \frac{1}{2} \alpha + \frac{1}{2} P_{\alpha}^* \sim \frac{1}{2} \beta + \frac{1}{2} \beta \). In turn, this implies \( W^*(\frac{1}{2} \alpha + \frac{1}{2} P_{\alpha}^*) \sim \hat{C}(\frac{1}{2} \alpha + \frac{1}{2} \beta) \). By Claim 9, this implies \( \hat{C}(\alpha) \geq \hat{C}(\beta) \) as sought. \( \square \)

**Claim 11.** If \( \succeq \) satisfies Axiom A.2, then for all \( \alpha, \beta \in H \) if \( \text{supp}^*(\alpha) \supseteq \text{supp}^*(\beta) \) then \( \hat{C}(\alpha) \geq \hat{C}(\beta) \).

**Proof.** Consider \( \alpha, \beta \in \Delta(\mathcal{X}) \) such that \( \text{supp}^*(\alpha) \supseteq \text{supp}^*(\beta) \). If \( \text{supp}^*(\alpha) = \text{supp}^*(\beta) \), then the result holds by Claim 10. Suppose instead that we have \( \text{supp}^*(\alpha) \supseteq \text{supp}^*(\beta) \). Consider now first the case in which \( |\text{supp}^*(\alpha) \setminus \text{supp}^*(\beta)| = 1 \), and consider \( \lambda \in (0, 1) \). By construction we must have \( \alpha = \sum_{i=1}^{m} a_i A_i + a_m A \) for some \( A_1, \ldots, A_{m-1} \in \mathcal{X} \) and \( a_1, \ldots, a_m \in (0, 1) \) s.t. \( \sum a_i = 1 \). Consider now \( p_\lambda^* \in \Delta^S(X) \) s.t. \( p_\lambda^* \sim^* A \), and construct the lottery \( \alpha' = \sum_{i=1}^{m} a_i A_i + a_m p_\lambda^* \), and notice that we must have \( \alpha' \in H \). Then, Axiom A.2 implies \( \alpha' \succeq \alpha \). In turn, this implies \( \hat{C}(\alpha') \geq \hat{C}(\alpha) \). Since \( \text{supp}^*(\alpha') = \text{supp}^*(\beta) \), then by Claim 10 we have \( \hat{C}(\alpha) \geq \hat{C}(\beta) \) as sought. Finally, notice that for all \( \alpha, \beta \in \Delta(\mathcal{X}) \) s.t. \( \text{supp}^*(\alpha) \supseteq \text{supp}^*(\beta) \), there must exist some \( n \in \mathbb{N} \) s.t. \( |\text{supp}^*(\alpha) \setminus \text{supp}^*(\beta)| = n \). The claim can then be proved just repeating \( n \)-times the steps above. \( \square \)

Now define \( C : \Delta(\mathcal{X}) \to \mathbb{R} \) by \( C(M) := \hat{C}(\alpha) \) for some \( \alpha \in H \) such that \( \text{supp}^*(\alpha) = M \), where Axiom A.5 and Claim 10 guarantee this is well defined. Notice that, since \( W^*(p) = W'(p) \) for all \( p \in \Delta^S(X) \), then \( C(p) = 0 \), hence \( C(\emptyset) = 0 \). Notice that if \( \succeq \) satisfies Axiom A.1, then for all \( \alpha \in H \) we must have \( C(\alpha) \geq \emptyset \); the reason is, if \( p_\alpha \sim \alpha \) and \( p_\alpha^* \sim^* \alpha \), then we have \( p_\alpha \succeq \alpha \sim p_\alpha \) by Axiom A.1, hence \( W^*(\alpha) \succeq W'(\alpha) \). In addition, Claim 11 also guarantees that \( C \) is monotone if \( \succeq \) satisfies Axiom A.2. In turns, this means that we have \( W'(\alpha) = W^*(\alpha) - C(\text{supp}^*(\alpha)) \).

Our last step is to extend the representation to the whole \( \Delta(\mathcal{X}) \). Define \( W : \Delta(\mathcal{X}) \to \mathbb{R} \) by \( W(\alpha) := W^*(\alpha) - C(\text{supp}^*(\alpha)) \). (This naturally implies \( W(\alpha) = W'(\alpha) \) if \( \alpha \in H \).) We are left to show that \( W \) represents \( \Delta(\mathcal{X}) \).

**Claim 12.** For any \( \alpha, \beta \in \Delta(\mathcal{X}) \) there exist \( \lambda \in (0, 1) \) such that \( \{x^*\} \succeq \lambda \alpha + (1 - \lambda)\{x^*\} \supseteq \{x_+\} \) and \( \{x^*\} \succeq \lambda \beta + (1 - \lambda)\{x^*\} \supseteq \{x_+\} \).
Proof. Consider $\alpha, \beta \in \Delta(\mathcal{X})$ and notice that by Axiom A.5 there must exist $a, b \in (0, 1)$ and $x, y \in X$ such that $\{x^*\} \ni a\alpha + (1 - a)x \ni \{x_a\}$ and $\{x^*\} \ni b\beta + (1 - b)y \ni \{x_b\}$. Now find $c \in (0, 1)$ such that $a\alpha + (1 - a)x \sim \alpha (1 - c)\{x^*\} + (1 - c)x \ni \{x_c\}$. Since we also have $\{x_c\} \ni c\beta + (1 - c)\{x^*\} = \{x_a\}$. Notice that $\{x^*\} \ni c\beta + (1 - c)\{x^*\}$, which implies, by definition of $\geq^*$ and by Claim 4, that $d\alpha + (1 - d)\{x^*\} \ni c\beta + (1 - c)\{x^*\}$ for all $d \in (0, c)$. From the representation $W^* \nexists \alpha + (1 - c)\{x^*\}$, hence $\{x^*\} \ni c\beta + (1 - c)\{x^*\} \ni \{x_c\}$. Notice that $\{x_c\} \ni c\beta + (1 - c)\{x^*\}$, which implies, by definition of $\geq^*$ and by Claim 4, that $d\alpha + (1 - d)\{x^*\} \ni c\beta + (1 - c)\{x^*\}$ for all $d \in (0, c)$. Now consider $\lambda = \min\{c, e\}$, and notice that we have $\{x^*\} \ni (1 - \lambda)\alpha + \lambda\{x^*\} \ni \{x_{\lambda}\}$ and $\{x^*\} \ni (1 - \lambda)\beta + \lambda\{x^*\} \ni \{x_{\lambda}\}$. □

Consider $\alpha, \beta \in \Delta(\mathcal{X})$ such that $\alpha \nexists \beta$ and $\lambda$ as in the statement of Claim 12. Now by Axiom A.3 we must have $\alpha \nexists \beta$ iff $\frac{\lambda}{1 - \lambda}p^*_{\alpha} + \frac{1 - \lambda}{1 - \lambda}p^*_{\beta} + \frac{1}{1 - \lambda}p^*_{\alpha} + (1 - \lambda)\alpha \nexists \lambda x^* + (1 - \lambda)\beta$, where $p^*_{\alpha} \sim^* \alpha, p_{\alpha x^* + (1 - \lambda)\alpha} \sim \lambda x^* + (1 - \lambda)\beta$, $p_{\beta} \sim^* \beta, p_{\beta x^* + (1 - \lambda)\beta} \sim \lambda x^* + (1 - \lambda)\beta$. This holds iff $W^*(\frac{\lambda}{1 - \lambda}p^*_{\alpha} + \frac{1 - \lambda}{1 - \lambda}p^*_{\beta} + \frac{1}{1 - \lambda}p^*_{\alpha} + (1 - \lambda)\beta) \nexists W^*(\frac{\lambda}{1 - \lambda}p^*_{\alpha} + \frac{1 - \lambda}{1 - \lambda}p^*_{\beta} + \frac{1}{1 - \lambda}p^*_{\alpha} + (1 - \lambda)\beta)$. Solving the algebra (noticing $W^*(p^*_{\alpha}) = W^*(\alpha)$), this is true iff $W^*(\alpha - \mathcal{C}(\text{supp}^*(\alpha)) \nexists W^*(\beta - \mathcal{C}(\text{supp}^*(\alpha))$ as sought.

Necessity of the axioms

Consider a preference relation $\succeq$ that admits a Thinking-Averse representation $\langle S, \mu, u, \mathcal{C} \rangle$. Now define $W^*: \Delta(\mathcal{X}) \rightarrow \mathbb{R}$ as $W^*(\alpha) := \sum_{A \in \mathcal{C}} \alpha A(\sum_{s \in S} \mu(s))\{\max_{y \in \mathcal{A}} u(y; s)\}$. The Thinking-Averse representation thus becomes $W^*(\alpha) = \mathcal{C}(\text{supp}^*(\alpha))$. Notice that for all $\alpha, \beta \in \Delta(\mathcal{X})$ and $a \in (0, 1)$, we have $W^*(a\alpha + (1 - a)\beta) = aW^*(\alpha) + (1 - a)W^*(\beta)$. We now argue that $\succeq$ is represented by $W^*$. To see why, consider any $\alpha, \beta \in \Delta(\mathcal{X})$ and $a, b \in (0, 1), a > b$, and notice that we have $\text{supp}^*(a\alpha + (1 - a)\beta) = \text{supp}^*(b\alpha + (1 - b)\beta)$. This means $a\alpha + (1 - a)\beta \succeq b\alpha + (1 - b)\beta$ if and only if $W^*(\alpha) \geq W^*(\beta)$, which implies that $W^*$ represents $\succeq^*$. Axiom A.2 follows trivially since $\mathcal{C}$ is monotone. All other axioms also follow trivially once we notice that $W^*$ is affine and continuous.

B.2. Proof of Theorem 3

Notice first that the necessity of the axioms follows trivially. To prove the sufficiency of the axioms, construct a Content-Monotone Thinking-Averse representation $\langle u, S, \mu, \mathcal{C} \rangle$ following the procedure described in the proof of Theorem 2. (It is important that we construct the state space as defined above.) Define $M := \{x \in \mathbb{R}: \mathcal{C}(\{A\}) = x \text{ for some } A \in \mathcal{X}\}$. Define $n = |M|$ and enumerate $M$ as $M = \{m_1, \ldots, m_n\}$ where $m_i \geq m_{i-1}$ for $i = 2, \ldots, n$. Define now $C_i := \{A \in \mathcal{X}: \mathcal{C}(\{A\}) = m_i\}$ for $i = 1, \ldots, n$. Notice that we must have $m_1 = 0$. For every $A \in \mathcal{X}$ define now the set $D(A) := \{B \in \mathcal{X}: \mathcal{C}(\{B\}) \geq \mathcal{C}(\{A\})\}$.

**Claim 13.** For any Content-Monotone Thinking-Averse representation $\langle u, S, \mu, \mathcal{C} \rangle$ of $\succeq$, and for any $A \in \mathcal{X}$, if $\mathcal{C}(\{A\}) > 0$ then $\{S, \emptyset\} \notin \mathcal{P}_{u, S}(A)$ and there is no $z \in A$ such that $\{z\} \succeq^* A$.

**Proof.** Notice first of all that $\{S, \emptyset\} \notin \mathcal{P}_{u, S}(A)$ if and only if there exists some $z \in A$ such that $\{z\} \succeq^* A$. Say by contradiction that there exists some $A \in \mathcal{X}$ such that $\mathcal{C}(\{A\}) > 0$ and there exists some $z \in A$ such that $\{z\} \succeq^* A$. Notice that this can happen if and only if we have
for all \( z \sim^* \{z, x\} \) for all \( x \in A \). But then, by Axiom A.8 we have \( \mathcal{C}(\{A\}) \leq \mathcal{C}(\{z\}) \). But since \( \mathcal{C}(\{z\}) = 0 \), this is a contradiction. \qed

The main part of the proof is the following claim.

**Claim 14.** There exists a Content-Monotone Thinking-Averse representation \( \langle u, S, \mu, \mathcal{C} \rangle \) of \( \succeq \) such that, for any \( A \in \mathcal{X} \), \( \mathcal{P}_{u, S}(A) \not\subset \bigcup_{B \in D(A)} \mathcal{P}_{u, S}(B) \).

**Proof.** Consider first of all the Content-Monotone Thinking-Averse representation \( u, S, \mu, \mathcal{C} \) constructed following the procedure described in the proof of Theorem 2, which, in turn, follows from Lemma 4. 29 For every \( A \in \mathcal{X} \), define the sets \( D_1(A) := \{B \in D(A): A \cup B \sim^* A\} \), \( D_2(A) := \{B \in D(A): A \cup B \sim^* A\} \). Define now the set \( D := \{C \in \mathcal{X}: C \in D_2(A) \) for some \( A \in \mathcal{X} \}, \) and for each \( C \in D \), and \( x \in C \) create a new state identical to \( f(x) \) and call it \( f_C(x) \). Also, define \( m_x = |\{C \in D: x \in C\}| \) (\( m_x \) represents how many times the element \( x \) is repeated in the sets in \( D \)). Finally, define \( S' \) as the newly constructed state space. Define \( u'(\cdot, s) = u(\cdot, s) \) if \( s \in S \), and \( u'(\cdot, f_C(x)) = u(\cdot, f(x)) \) for all \( x \in C \) and \( C \in D \). Define the probability measure \( \mu' \) over \( S' \) as \( \mu'(s) = \mu(s) \) for all \( s \in S \) such that \( s \sim f(x) \) for some \( x \in C \), for some \( C \in D \); for all \( C \in D \), \( x \in C \), define \( \mu'(f_C(x)) = \frac{\mu(f(x))}{m_x+1} \).

Notice first of all that \( \langle u', S', \mu', \mathcal{C} \rangle \) must be a Content-Monotone Thinking-Averse representation of \( \succeq \): in fact, we only have duplicated the states at the original weight of that state. Notice now that we must have that \( \{S', \emptyset\} \not\in \mathcal{P}_{u', S'}(B) \) for all \( B \in D(A) \) for all \( A \in \mathcal{X} \). This follows directly from Claim 13 since \( \mathcal{C}(\{A\}) \geq 0 \) and \( \mathcal{C}(\{B\}) > \mathcal{C}(\{A\}) \), so \( \mathcal{C}(\{B\}) > 0 \). If \( \{S', \emptyset\} \in \mathcal{P}_{u', S'}(A) \) the claim is trivially true, so let us focus on the case in which \( \{S', \emptyset\} \not\in \mathcal{P}_{u', S'}(A) \).

Consider some \( B \in D_1(A) \), and then some \( z_B \) such that \( A \cup z_B \sim^* A \) and \( A \cup (B \setminus f(z)) \sim^* A \cup B \) (it is not hard to see that such \( z_B \) must exist for every \( B \in D_1(A) \)). Define now the set of partitions \( H \subset \Pi(S') \) that put each of the state \( f(z_B) \) together with the state \( f(A \cup (B \setminus f(z))) \), for all \( B \in D_1(A) \), and that also put together, for all \( C \in D_2(A) \), the states \( f_C(x) \) for all \( x \in C \).

For any \( A \in \mathcal{X} \) we will now argue that \( \mathcal{P}_{u', S'}(A) \cap H \not= \emptyset \) but that \( \mathcal{P}_{u', S'}(B) \cap H = \emptyset \) for all \( B \in D(A) \). To see why, notice that \( A \subset f(A \cup (B \setminus f(z))) \) for all \( B \in D_1(A) \), and hence for \( A \) the agent could avoid distinguishing between these two states. Consider now some \( C \in D_2(A) \), and notice that there are two possibilities: either \( A \sim^* C \), or \( A \sim^* C \). In the former case, there must exist some \( z_C \in A \) such that \( \{z_C, x\} \succ^* \{x\} \) for all \( x \in C \). But the same must be true also in the latter case by Axiom A.8. 30 This means that \( z_C \notin f_C(w) \) for all \( w \in C \), which implies that with \( A \) the agent does not need to separate the states \( f_C(w) \) for all \( w \in C \) (the agent could simply choose \( z_C \) in each of these states). Therefore, we must have \( \mathcal{P}_{u', S'}(A) \cap H \not= \emptyset \).

Consider now some \( B \in D_1(A) \), and notice that \( f(z_B) \cap B = B \setminus f(A \cup (B \setminus f(z))) \) for all \( B \in D_1(A) \), since, by construction, \( A \cup (B \setminus f(z)) \sim A \cup B \). At the same time, we cannot have \( B \subset f(z_B) \), otherwise it would mean that \( \{z_B\} \sim \{z_B\} \cup B \), which implies \( \{S', \emptyset\} \in \mathcal{P}_{u', S'}(B) \), a contradiction. Therefore, any partition in \( \mathcal{P}_{u', S'}(B) \) should distinguish \( f(z_B) \) and \( f(A \cup (B \setminus f(z))) \), which implies \( \mathcal{P}_{u', S'}(B) \cap H = \emptyset \) for all \( B \in D_1(A) \).

\[\text{28} \text{In fact, notice that for every } x \in \{z\}, \text{there exists some } z \in A \text{ (which is } z \text{ itself) such that } \{z\} \sim \{x, z\}, \text{hence } A \succ \{z\}, \text{and since we have } \{z\} \sim^* A, \text{then we must also have } \mathcal{C}(\{A\}) \leq \mathcal{C}(\{z\}). \]

\[\text{29} \text{Following that argument, we define the function } f \text{ as in Eq. (2) in the proof of Lemma 4.} \]

\[\text{30} \text{In fact, notice that we have } A \sim^* C \sim^* A \cup C \text{ and } \mathcal{C}(\{A\}) < \mathcal{C}(\{C\}). \text{Then, Axiom A.8 implies that there exists some } z_C \in A \text{ such that } \{z_C, x\} \succ^* \{x\}. \]
Consider now some \( C \in D_2(A) \), and notice that we must have that any partition in \( \mathcal{P}_{u',S}(B) \) must distinguish some \( f_C(x) \) from some \( f_C(y) \) for some \( x, y \in C \). To see why, notice that if this wasn’t the case, it would mean that the agent could choose only one option for all \( f_C(w) \), for all \( w \in C \). Call \( \tilde{w} \) this option. For \( \tilde{w} \) to be the optimal choice for \( C \) in state \( f_C(\tilde{w}) \), it must be that \( \{x, \tilde{w}\} \sim^a \{\tilde{w}\} \) for all \( x \in C \), which in turns means \( \{\tilde{w}\} \sim^a C \), hence \( \{S', \emptyset\} \in \mathcal{P}_{u',S}(C) \), which is a contradiction.

We have therefore proven that \( \mathcal{P}_{u',S}(A) \cap H \neq \emptyset \) but that \( \mathcal{P}_{u',S}(B) \cap H = \emptyset \) for all \( B \in D(A) \), which implies \( \mathcal{P}_{u',S}(A) \not\subset \bigcup_{B \in D(A)} \mathcal{P}_{u',S}(B) \) as sought. \( \Box \)

Now construct \( c : \Pi(S) \to \mathbb{R} \) as follows. Set \( c(\pi) = m_a \) for all \( \pi \in \mathcal{P}_{u,S}(A) \) for all \( A \in C_\pi \); for \( i = 1, \ldots, N - 1 \) set \( c(\pi) = m_i \) for all \( \pi \in \mathcal{P}_{u,S}(A)(\bigcup_{j=i}^n (\bigcup_{B \in C_j} \mathcal{P}_{u,S}(B))) \) for all \( A \in C_i \). Notice that we must have \( \mathcal{P}_{u,S}(A)(\bigcup_{j=i}^n (\bigcup_{B \in C_j} \mathcal{P}_{u,S}(B))) \neq \emptyset \) by Claim 14. Notice also that we must have \( c(S, \emptyset) = 0 \) since \( m_1 = 0 \). It is also easy to see that \( c \) must be partition-monotone, which concludes the proof.

**B.3. Proof of Proposition 1**

Let us consider the first part. In the proof of the necessity of the axioms of Theorem 1 we have shown that for any Thinking-Averse representation, the functional \( U : \Delta(\mathcal{X}) \to \mathbb{R} \) defined as

\[
U(\alpha) := \sum_{A \in \mathcal{X}} \alpha_A \left( \sum_{s \in S} \mu(s) \left[ \max_{y \in A} u(y; s) \right] \right)
\]

represents the binary relation \( \succeq^* \). Notice that \( U \) is affine. We have also shown that \( \succeq^* \) is a complete, transitive, preference relation that satisfies the independence axiom. It is standard practice to show that any affine representation is unique up to a positive affine transformation, leading to the first part of the proposition.

We now turn to prove the uniqueness of \( \mathcal{C} \). Consider two Thinking-Averse representations and construct the respective \( U \) and \( U' \) as above. We have just shown that \( U' = aU + b \) for some \( a \in \mathbb{R}_{++}, b \in \mathbb{R} \). Now consider any \( T \in 2^\mathcal{X} \) and any \( \alpha \in \Delta(\mathcal{X}) \) such that \( \text{supp}^*(\alpha) = T \) and \( \{x^*\} \succeq \alpha \succeq \{x_+\} \) (its existence is guaranteed by Axiom A.5). By Claim 7 there exists some \( p \in \Delta^2(X) \) such that \( \alpha \sim p \). This means that we have \( U(\{p\}) = U(\alpha, \mathcal{C}) \) and \( U'(\{p\}) = U'(\alpha, \mathcal{C}') \). But given that \( U' = aU + b \), we have \( \mathcal{C}'(\text{supp}^*(\alpha)) = a\mathcal{C}(\text{supp}^*(\alpha)) \) as sought.

**B.4. Proof of Proposition 2**

Consider two Thinking-Averse representations \( (S_1, \mu_1, \mathcal{C}_1) \) of \( \succeq_1 \) and \( (S_2, \mu_2, \mathcal{C}_2) \) of \( \succeq_2 \) such that \( S_1 = S_2 \) and \( \mu_1 = \mu_2 \). For each of these representations, define \( U_1 \) and \( U_2 \) as in the proof of Proposition 1. Notice that we 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 \( W_1 \leq W_2 \). Notice, moreover, that we have \( W_1(p) = W_2(p) \) for any \( p \in \Delta^2(X) \).

To prove (i) \( \Rightarrow \) (ii), consider any \( T \in 2^\mathcal{X} \) and any \( \alpha \in \Delta(\mathcal{X}) \) such that \( \text{supp}^*(\alpha) = T \) and \( \{x^*\} \succeq \alpha \succeq \{x_+\} \) (its existence is guaranteed by Axiom A.5). Claim 7 shows that there exists \( p_\alpha \in \Delta^2(X) \) such that \( p_\alpha \sim_1 \alpha \). Hence, we have \( W_1(p_\alpha) = W_1(\alpha) \). Furthermore, since \( U_1 = U_2 \) and since \( U_1(p_\alpha) = W_1(p_\alpha) \) and \( U_2(p_\alpha) = W_2(p_\alpha) \) (\( p_\alpha \) has a zero cost of thinking), this means \( W_2(p_\alpha) = W_1(p_\alpha) = W_1(\alpha) \). Now, condition (i) and \( p_\alpha \sim_1 \alpha \) imply \( \alpha \succeq_2 p_\alpha \), hence \( W_2(\alpha) \geq W_2(p_\alpha) \), which gives us \( W_2(\alpha) \geq W_1(\alpha) \) as sought.
To prove (ii) ⇒ (i), take any $\alpha \in \Delta(\mathcal{L})$ and $p \in \Delta^S(X)$ such that $\alpha \succcurlyeq_1 p$. Together with (ii), implies $W_2(\alpha) \geq W_1(\alpha) \geq W_1(p) = W_2(p)$, hence $\alpha \succcurlyeq_2 p$ as sought.

References


