

# Dynamic Consistency and Ambiguity: A Reappraisal

Brian Hill

GREGHEC, HEC Paris & CNRS\*

August 9, 2013<sup>†</sup>

## Abstract

Dynamic consistency demands that a decision maker's preferences over contingent plans agree with his preferences in the planned-for contingency. What counts are the contingencies the decision maker envisages – and plans for – rather than contingencies selected by a theorist. We show how this simple point resolves some crucial purported difficulties for ambiguity models in dynamic settings. Firstly, when properly formulated, dynamic consistency is compatible with consequentialism and non-expected utility. Secondly, the perspective provides a principled justification for the restrictions to certain families of beliefs (à la [Epstein and Schneider \(2003\)](#)) in applications to dynamic choice problems. Thirdly, the value of information under non-expected utility is revealed to be non-negative as long as the information offered does not compromise information that the decision maker had otherwise expected to receive. We also give behavioral foundations for the contingencies the decision maker envisages, in the form of a representation theorem.

**Keywords:** Decision under Uncertainty; Dynamic Consistency; Dynamic Choice; Value of Information; Epistemic Contingency.

**JEL classification:** D81, D83, D90.

---

\* 1 rue de la Libération, 78351 Jouy-en-Josas, France. Tel: + 33 1 39 67 72 65. Fax: + 33 1 39 67 71 09.  
E-mail: [hill@hec.fr](mailto:hill@hec.fr).

<sup>†</sup>This paper is a preliminary version, and may be subject to subsequent revisions. I would like to thank Itzhak Gilboa, and participants at the Cergy-Pontoise workshop 2012 and RUD 2013 for helpful comments. Please do not quote without the consent of the author.

# 1 Introduction

## 1.1 Motivation

One of the principal challenges in, and to, the literature on non-expected utility models for decision under uncertainty is that posed by application in dynamic situations. Violation of expected utility purportedly brings with it some problematic consequences, which have been at the base of criticism of the whole literature on ambiguity (for some examples, see [Al Najjar and Weinstein \(2009\)](#) and discussions in [Epstein and Le Breton \(1993\)](#), [Ghirardato \(2002\)](#)). Putting aside subtleties in the terminology and definitions, the gist of the problem lies in the argument that consequentialism – the decision maker ignores sunk costs (or the history in the decision tree) when deciding at any node – and dynamic consistency – the decision maker’s preferences over contingent plans agree with his preferences in the planned-for contingency – are incompatible with non-expected utility. Given the *prima facie* attraction of these properties of dynamic choice, not to mention their importance for economic applications, the option of abandoning one of them to leave space for non-expected utility is unappetizing. All the worse, some have concluded, for non-expected utility theories.

The main thesis of this paper is that this generally accepted state-of-play rests on a mistake. Standard formalizations of the dynamic consistency principle use sets of contingencies that are effectively imposed by the theorist. However, since the principle involves the decision maker’s plans, it is the contingencies that he himself envisages – and plans for – that are relevant under any reasonable version of the principle. Once the principle is formulated with the contingencies the decision maker himself envisages, the apparent incompatibility with non-expected utility is resolved.

This apparently innocuous point has wide-ranging conceptual and economic ramifications. First of all, it implies a reconceptualization in terms of what the decision maker envisages about his future states of belief. Dynamic consistency, as considered in the literature on decision under uncertainty, is revealed to be essentially a condition on the relationship between two sorts of belief: the decision maker’s current beliefs about the state of the world and his beliefs about what he will believe in the future about the state of the world.

Secondly, this relationship turns out to be particularly strong in certain cases: so strong, in fact, that one can draw conclusions, just on the basis of the decision maker’s current beliefs, about what he thinks he might believe in the future, and hence about the contin-

gencies he envisages. In particular, the standard modelling assumption that the decision maker is using the same decision tree as the theorist implies that his ex ante beliefs must be of a certain form. In perhaps the most significant economic contribution of the paper, this observation provides a principled justification for the restriction to a specific family of beliefs in the application of non-expected utility models to dynamic choice problems. In such problems, the use of a set of beliefs not in this family implies that the decision maker do not think that he is facing the decision tree used by the theorist.

Thirdly, the approach provides a new perspective on the purported propensity of non-expected decision makers to turn down free information. Information about the decision tree that the decision maker is faced with may impact upon his beliefs about his future beliefs, and hence, in the light of the implications mentioned above, his current beliefs. Incorporating this factor into his choice of whether to accept free information, it turns out that he will always accept an offer of free information, as long as it does not compromise information he had otherwise expected to receive.

In the rest of the Introduction, we present the notion of dynamic consistency, explain our challenge to the standard argument against non-expected utility, and set out the main contributions and plan of the paper, as well as some relations to existing literature.

## *1.2 Dynamic consistency, events and epistemic contingencies*

To present the notion of dynamic consistency for decision under uncertainty, let us adopt the standard setup in the literature, where a decision maker has ex post and ex ante preferences over acts – functions from a given (‘objective’) set of states of the world, called the state space and understood to represent all payoff-relevant factors, to a set of consequences. We focus on the dynamic consistency of preferences, rather than the dynamic consistency of behavior (see for example [Strotz \(1955\)](#); [Karni and Safra \(1989, 1990\)](#); [Siniscalchi \(2009, 2011\)](#)). Put succinctly, dynamic consistency demands harmony between ex ante preferences over contingent plans and preferences after the realisation of the planned-for contingency: the decision maker’s ex post preferences correspond to his ex ante preferences over plans involving the contingency in question. At first glance, application of this principle would seem to require determining which contingencies the decision maker envisages holding, and hence plans for. Somewhat surprisingly, there is little discussion of this question in the literature. Rather, it is generally assumed that one can associate

to any pair of acts two contingencies – one in which the acts necessarily yield the same consequence, and one in which this is not the case. The idea is that the acts can be thought of as plans for the contingency in which they differ. In this way, one avoids the question of which contingencies the decision maker himself envisages; the contingencies are, so to speak, imposed by the theorist in the pairs of acts considered.

An important consequence of this approach is that the contingencies are events, that is, sets of states of the world. It is clear on reflection that, strictly speaking, the contingencies that the decision maker plans for cannot simply be events, because they must factor in the fact that the decision maker knows them to hold. A decision maker does not plan for the contingency that event  $A$  holds but he does not know it, and it makes no sense to speak of his preference or state of belief after having learnt that such a contingency has arose, for that is not something that he can learn. Taking contingencies as events is shorthand: the contingency is not strictly speaking the fact that the event in question has occurred, but rather the fact that the decision maker learns that the event has occurred. Hence planned-for contingencies are more accurately represented not by events, but rather by future possible states of knowledge or belief (or by the preferences generated by these states of belief). We call these *epistemic contingencies*.

This pedantry would be of no interest if it did not undermine the practice described above of considering contingencies to be events on which acts yield the same or different consequences. Although, as we have seen, any event has a natural epistemic contingency corresponding to it (namely, the contingency that the decision maker learns only that the event in question has occurred), the converse is not true. A contingency in which the decision maker alters his probability for a certain event without giving it probability 0 or 1 (for example, by some form of non-Bayesian learning) is an epistemic contingency which cannot be naturally associated to an event. More importantly, what counts as overlapping, disjointness and exhaustivity for contingencies differs between events and epistemic contingencies. Whilst a (non-empty) event  $A$  and the event containing the whole state space  $S$  are not exclusive, the epistemic contingency in which the decision maker learns only that  $A$  and that in which he learns nothing (ie. where he learns  $S$ ) are: they cannot both concurrently occur. This has important consequences when one considers which contingencies may need to be planned for. While, when reasoning with events, planning for the eventualities  $A$  and  $A^c$  exhausts all possibilities, this is not the case when the contingencies

Figure 1: Dynamic consistency in the standard (static) Ellsberg urn (values in dollars)

	R	B	Y
$f_1$	10	0	10
$g_1$	0	10	10
$f_2$	10	0	0
$g_2$	0	10	0

are epistemic: learning only that  $A$ , learning only that  $A^c$  and learning nothing are three distinct epistemic contingencies, each of which occurs without any of the others occurring. Hence, even considering the decision maker's preference between acts that only differ on the event  $A$ , it does not follow that the only contingencies that he envisages – and plans for – are necessarily (learning that)  $A$  and (learning that)  $A^c$ : he could equally well envisage, and plan for, the contingency in which he learns nothing (or, indeed, the contingencies in which he learns only  $B$  or only  $B^c$  for a different event  $B$ ).

The set of epistemic contingencies that the decision maker envisages at a particular moment forms a decision tree; we shall call this the decision maker's *subjective* tree. The set of events that the theorist considers also forms a tree; since these trees can be defined independently of considerations regarding the decision maker's opinion on the relevant possible future states, we shall call these *objective* trees. As noted above, the class of objective trees can be embedded into the class of subjective trees, in the sense that for each objective tree, there could exist a decision maker who envisages all and only the (epistemic versions of) the events in that tree. However, there are subjective trees that correspond to no objective tree. For such trees, the argument that consequentialism and dynamic consistency rule out non-expected utility behavior does not go through.<sup>1</sup>

To illustrate this point, let us consider a standard example of violation of expected utility for decision under uncertainty, the Ellsberg one-urn example. Consider a decision maker who is told that an urn contains ninety balls, thirty of which are red (R), and the rest of which are black (B) and yellow (Y) in an unknown proportion, and is asked for his

<sup>1</sup>It might be objected that if the decision maker's envisaged future states of belief – the 'subjective information structure' he is using, if you will – are relevant, then they should be represented explicitly in the state space. As discussed in Section 2.2 (in particular, Remark 2), a setup involving such an extended state space is equivalent to the one used here, and all the points made below continue to hold when reformulated in these terms. See also Appendix A.

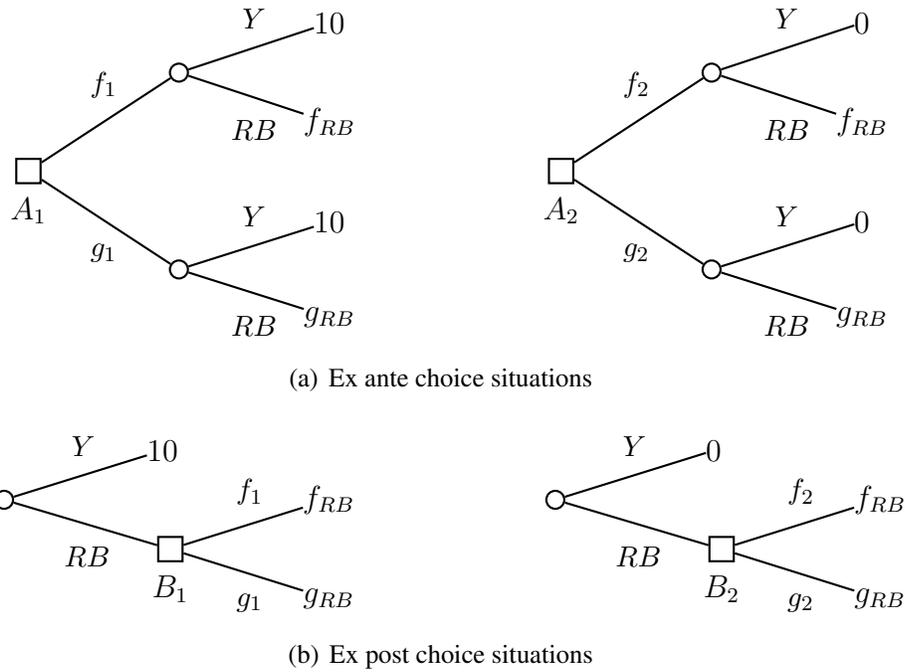


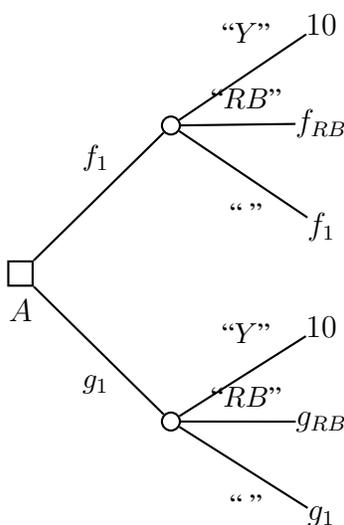
Figure 2: Dynamic consistency in the standard (dynamic) Ellsberg urn

preferences over the bets on the colour of a ball drawn from the urn shown in Figure 1. The standard preference pattern is  $f_1 < g_1$  and  $f_2 > g_2$ , violating the sure-thing principle (Savage’s P2 (1954)).

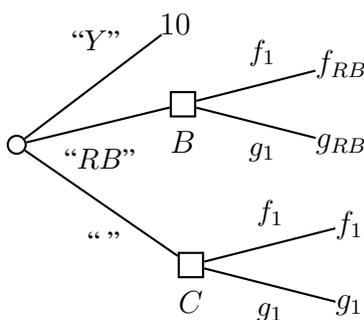
A standard formulation of the argument against these Ellsberg preferences using the principles of consequentialism and dynamic consistency is as follows. One considers not only the decision maker’s preferences over the bets in Figure 1, but one also imagines that, after the ball is drawn, the decision maker receives information as to whether the ball is yellow or not (and no other information), and is asked again for his preferences over the bets. Using  $f_{RB}$  to refer to the partial act, defined on the event  $\{R, B\}$ ,<sup>2</sup> and coinciding with  $f_1$  and  $f_2$  on these states, and similarly for  $g_{RB}$ , one can, the story goes, represent the decision maker’s choices before and after the information is revealed by the trees in Figure 2. Consequentialism demands that all that counts are the consequences of one’s choices; hence, in particular, it demands that the choices made at the  $B$  nodes in Figure 2(b) are independent of the values that could have been won if the ball were yellow (ie. the upper branches of the trees). So the choices made at  $B_1$  and  $B_2$  are the same. Dynamic consistency in preferences asks for coherence between one’s preferences over contingent

<sup>2</sup>A partial act is a partial function from states to consequences, defined on a subset of the state space.

plans concerning future options and one's preferences in the appropriate contingencies. So, if the decision maker prefers  $f_1$  to  $g_1$  at node  $A_1$  in Figure 2(a), then, since the only difference between the consequences yielded by these bets occurs when the ball is not yellow, this can be thought of as a preference over plans for the contingency that the ball is not yellow. Dynamic consistency thus demands that he prefer  $f_1$  to  $g_1$  at node  $B_1$  in Figure 2(b) (after the resolution of the uncertainty about whether the ball is yellow or not). Similar reasoning applies to the right hand trees (nodes  $A_2$  and  $B_2$ ) and the acts  $f_2$  and  $g_2$ . It follows that, contrary to the Ellsberg preferences, the decision maker must have the same preferences over  $f_2$  and  $g_2$  as over  $f_1$  and  $g_1$ .



(a) Ex ante choice situation



(b) Ex post choice situation

Figure 3: Dynamic consistency in the non-standard (dynamic) Ellsberg urn (choice between  $f_1$  and  $g_1$ )

This argument is most naturally read as involving objective trees, especially given that

the trees in Figure 2 are motivated by the choice of acts considered, independently of considerations about the contingencies the decision maker himself envisages. Certainly, if these trees were read as involving epistemic contingencies, then they involve a non-trivial assumption: that the decision maker envisages only two future states of belief – one in which he learns (only) that the ball is yellow and one in which he learns (only) that it is not. This assumption is far from innocent: it does not hold, for example, for a decision maker who envisages that he might learn whether the ball is yellow or not, but also that he might learn nothing.<sup>3</sup> In this case, the appropriate (subjective) decision trees are not those in Figure 2 but rather those given in Figure 3 (we only give the trees for the choice between  $f_1$  and  $g_1$ ; the trees for the choice between  $f_2$  and  $g_2$  are similar). There are three branches corresponding to the epistemic contingency in which the decision maker learns that the ball is yellow (denoted here with inverted commas to emphasise the difference from mere events), the epistemic contingency in which he learns that the ball is not yellow, and the contingency in which he learns nothing at all (denoted “ ”: the message is empty). On these trees, dynamic consistency can no longer be applied as in the argument presented above. It applied non-trivially on the trees in Figure 2 because the choice at the initial node in the ex ante case (node  $A_1$ ) corresponds to the choice at a single node ( $B_1$ ) in the ex post case, and hence can be interpreted as a plan for that node. By contrast, in Figure 3, the choice between the acts  $f_1$  and  $g_1$  at node  $A$  does not correspond to the choice at a single node in the ex post case; thus it cannot be thought of as a plan for a single ex post contingency. Hence dynamic consistency does not necessarily apply – for example, if the preferences over  $f_1$  and  $g_1$  differ at nodes  $B$  and  $C$  in Figure 3(b), then dynamic consistency does not imply anything about the choice in Figure 3(a). So the argument against Ellsberg-style behavior no longer goes through. This illustrates a general point that will be made precise in the sequel: consequentialism, non-expected utility preferences and dynamic consistency are not necessarily incompatible, if dynamic consistency is defined using subjective trees, with the contingencies the decision maker does in fact envisage.

---

<sup>3</sup>To give another example, it also does not hold for a decision maker who thinks that he might learn either whether the ball is yellow or not, or whether he has won his bet or not.

### 1.3 *Outline of the paper*

This paper has three aims. The first is to develop the rebuttal of the standard argument from dynamic consistency and consequentialism to expected utility in the context of decision under uncertainty. The second, and economically most important aim is to confront the proposed perspective with some of the toughest purported problems for non-expected utility models in dynamic contexts, such as application in dynamic choice problems and the question of value of information. The final aim is to fully defend the central notions of the proposal, and in particular that of epistemic contingency.

A first step will be to propose a formal representation of epistemic contingencies. Assuming, as is standard in this literature, that the decision maker's utility function remains fixed, we shall take contingencies to be states of belief, interpreted as the future states of belief that the decision maker considers possible *ex ante*. Although most of the main points continue to hold for other non-expected utility models, for concreteness we shall work with the maxmin expected utility model (Gilboa and Schmeidler, 1989),<sup>4</sup> and accordingly take epistemic contingencies to be sets of priors. We thus implicitly assume that the decision maker forms preferences according to the same decision rule in all contingencies as well as *ex ante* (this assumption has been explicitly discussed in Eichberger and Kelsey (1996); Sarin and Wakker (1998), for example).

We then go on to formulate a notion of dynamic consistency in terms of these epistemic contingencies, and show that it is possible in general for non-expected utility decision makers to satisfy this condition in conjunction with consequentialism; the point illustrated in the example above does indeed hold in general. The main conceptual contribution of the paper lies in the observation that dynamic consistency in the context of decision under uncertainty demands harmony between two sorts of belief: the beliefs the decision maker currently has about the state of the world (which determine his current preferences, including his preferences over plans), and the beliefs he has about what he will believe about the states of the world (which correspond to the possible future preferences he envisages). As an illustration, we characterise the constraints the decision maker's anticipations about his future beliefs place on his current beliefs, under dynamic consistency.

We then turn to the consequences of this refined version of dynamic consistency for

---

<sup>4</sup>It seems that versions of all results except the representation theorem (Theorem 1) are relatively easy to obtain for many other non-expected utility models.

economic applications. In standard applications to dynamic situations, it is often implicitly assumed that the decision maker knows the decision tree with which he is faced, and that this tree is as described by the theorist. In the current setup, these are cases where the decision maker's subjective tree corresponds to an objective one. We show that this particular form of the subjective decision tree has consequences for the decision maker's current beliefs, if he is to satisfy dynamic consistency with respect to epistemic contingencies. It implies that his ex ante set of priors satisfies a 'factorisability' property that is the equivalent in our framework of Epstein and Schneider's (2003) rectangularity. It is sometimes suggested that the need for a restriction to such rectangular sets of priors is a limitation of the Epstein and Schneider (2003) approach. Our result can be interpreted as showing that this criticism is unfounded: the specific shape of the set of priors is a consequence of the standard assumption that the decision maker is using the same decision tree as the theorist. This non-arbitrary vindication of the use of a restricted family of sets of priors of the sort suggested by Epstein and Schneider (2003) is perhaps the most significant economic contribution of the paper. Since this is the key to treating sequential choice problems, it shows that the approach developed here provides a solid, well-defended account of dynamic choice under non-expected utility.

Another pressing question concerning dynamic choice concerns the value of information, which is standardly argued to be possibly negative for non-expected utility decision makers (Al Najjar and Weinstein, 2009; Siniscalchi, 2011). The proposed approach implies a reconceptualisation of the issue, insofar as the decision maker's choice to accept an offer of information that he had not previously anticipated obtaining entails a change in his envisaged future states of belief. Given the relationship between envisaged future beliefs and current beliefs mentioned above, it follows that the latter may also change on learning that he will obtain the information; indeed, this change will occur even before the information is actually received. To the extent that this change is a result of his choice to accept the information, the situation has the flavour of a moral hazard problem. Incorporating these elements, we show that, when properly defined, the value of information is non-negative whenever the information offered does not compromise the reception of other information that the decision maker was previously expecting to receive.

An important difference between the objective and subjective trees on which the preceding considerations are based is the question of foundations. Whereas objective trees are

part of the setup, as posed by the theorists, subjective trees incorporate the contingencies that the decision maker envisages, and hence encapsulate elements of his state of mind. However, without some connection to observable choice behavior (beyond the apparent violation of dynamic consistency), the notion of epistemic contingency, and the previous considerations building on it, are on shaky ground. The final contribution of the paper is to treat this issue, by providing behavioral foundations for this notion.

The paper is organised as follows. Section 2 describes the framework and the representation of epistemic contingencies used. In Section 3, the notion of dynamic consistency on subjective trees is defined, and immediate consequences, including the possibility of satisfying dynamic consistency in tandem with consequentialism and non-expected utility, are explored. Section 4 examines the case of a decision maker whose subjective tree corresponds to an objective one, and the consequences for the application of non-expected utility models to dynamic problems. Section 5 analyses the consequences of the perspective developed here for the question of value of information. The final section, Section 6, provides choice-theoretical foundations for the notion of epistemic contingency that underlies the proposed approach. Proofs are collected in an Appendix.

#### ***1.4 Related Literature***

The literature on dynamic consistency is too large to be cited in its entirety. A thorough discussion of the question and the literature on dynamic arguments for expected utility under risk is given in Machina (1989); papers showing or discussing the inconsistency between dynamic consistency, consequentialism and non-expected utility in the case of uncertainty include Hammond (1988); Epstein and Le Breton (1993); Ghirardato (2002). Of the papers introducing update rules for non-expected utility models or considering dynamic choice in these models, the closest is without doubt Epstein and Schneider (2003). Indeed, as explained in Section 4, the notion of  $\mathcal{P}$ -rectangularity introduced here is essentially a version of their rectangularity condition adapted to our framework, and there are evident technical similarities between Proposition 4 and their results. However, as explained in the subsequent discussion, the conceptual contributions are different, so much so that our result can be read as a vindication of their restriction on sets of priors. Their approach has been adopted with other non-expected utility models (for example Maccheroni et al. (2006); Klibanoff et al. (2009)), and the perspective developed here applies similarly.

Conceptually, the closest suggestion to the one proposed here that we have been able to find was made in [Gilboa et al. \(2009\)](#), where it was suggested that some of the events required to exhibit violations of Savage’s Sure Thing Principle (P2) in some of the Ellsberg examples are ‘highly contrived’ and ‘will never be observed by the decision maker’. However, as they remark, this point does not hold for the Ellsberg one-urn example that we considered in Section 1.2. Indeed, the approach proposed here focusses on what the decision maker expects to learn, rather than what he can learn or observe.

The argument that non-expected utility decision makers may exhibit aversion for information has been proposed by [Wakker \(1988\)](#) for the case of decision under risk (see also [Machina \(1989\)](#) and references within); see [Eichberger et al. \(2007\)](#); [Al Najjar and Weinstein \(2009\)](#) for an analogous argument for the case of decision under uncertainty. [Siniscalchi \(2011\)](#) contains a thorough and insightful discussion of the issue, though arrives at a diagnosis that is clearly very different from ours.<sup>5</sup> For Siniscalchi, the nub of the matter lies in the trade-off between information acquisition and commitment: if commitment is more valuable than information acquisition, then the decision maker may turn down the information. The approach set out here, by contrast, draws on the fact that the choice to accept an offer of information will affect one’s envisaged possible future states of belief. If this change involves a loss of some other information that the decision maker had previously expected to receive, then he may turn down the offer.

As discussed in Section 6, the representation result proposed there can be understood (technically) as a contribution to the literature on preference for flexibility or decision making under unforeseen contingencies initiated by [Kreps \(1979, 1992\)](#); [Dekel et al. \(2001\)](#). The closest result in that literature to the current one is doubtless [Epstein et al. \(2007, Theorem 2\)](#), insofar as both our representation (5) and their representation (8) involve a set – of utilities in their case, of priors in ours – in each contingency, and use a maxmin rule. However, they use an expected utility ‘aggregator’ to form ex ante preferences, where we use a dominance aggregator (see Remark 7), and they use a richer framework (and, accordingly, different proof methods), involving lotteries over menus. Other related approaches (though not in the menu framework) include [Mukerji \(1997\)](#); [Ghirardato \(2001\)](#).

Other relations (notably technical ones) with existing literature are mentioned at appropriate points in the discussion.

---

<sup>5</sup>[Machina \(1989\)](#) attributes a similar approach to the issue for decision under risk to Edi Karni.

## 2 Preliminaries

### 2.1 Setup and technical notions

We use a version of the standard Anscombe-Aumann framework (Anscombe and Aumann, 1963), as adapted by Fishburn (1970). Let  $S$  be a non-empty finite set of *states*, and let  $\Sigma$  be the set of subsets of  $S$ , called *events*.  $\Delta(\Sigma)$  is the set of probability measures on  $(S, \Sigma)$ . Where necessary, we use the Euclidean topology on  $\Delta(\Sigma)$ .  $X$  is a nonempty set of *outcomes* endowed with a metric under which it is compact; a *consequence* is a Borel probability measure on  $X$ .  $\Delta(X)$  is the set of consequences; where necessary, we use the weak convergence topology, under which this space is compact metric (Billingsley, 2009, p 72). *Acts* are functions from states to consequences;  $\mathcal{A}$  is the set of acts, endowed with the product topology, under which it is compact metric.  $\mathcal{A}$  is a mixture set with the mixture relation defined pointwise: for  $f, h$  in  $\mathcal{A}$  and  $\alpha \in \mathfrak{R}$ ,  $0 \leq \alpha \leq 1$ , the mixture  $\alpha f + (1 - \alpha)h$  is defined by  $(\alpha f + (1 - \alpha)h)(s, x) = \alpha f(s, x) + (1 - \alpha)h(s, x)$ . The mixture relation is extended to sets of acts and acts pointwise: for  $A \subseteq \mathcal{A}$ ,  $h \in \mathcal{A}$  and  $0 \leq \alpha \leq 1$ ,  $\alpha A + (1 - \alpha)h = \{\alpha f + (1 - \alpha)h \mid f \in A\}$ . We write  $f_\alpha h$  as short for  $\alpha f + (1 - \alpha)h$  and similarly for  $A_\alpha h$ . With slight abuse of notation, a constant act taking consequence  $c$  for every state will be denoted  $c$  and the set of constant acts will be denoted  $\Delta(X)$ . We use  $\wp$  to denote the set of closed non-empty subsets of; hence, in particular,  $\wp(\mathcal{A})$  is the set of closed non-empty subsets of  $\mathcal{A}$ . Where required, we use the Hausdorff topology on  $\wp(\mathcal{A})$  (see for example Aliprantis and Border (2007, Section 3.17)). For any  $A \in \wp(\mathcal{A})$ ,  $\text{conv}(A)$  is the set of finite mixtures of elements of  $A$ :  $\text{conv}(A) = \{\sum_{i=1}^n \alpha_i f_i \mid \alpha_i \in [0, 1] \text{ with } \sum_{i=1}^n \alpha_i = 1, f_i \in A\}$ . It is obvious that  $\text{conv}(A) \in \wp(\mathcal{A})$ .

The symbol  $\leq$  (potentially with subscripts) will be used to denote preference relations over  $\mathcal{A}$ ; as standard  $<$  and  $\sim$  denote the asymmetric and symmetric parts of  $\leq$ . We adopt the standard notion of null event with respect to a preference relation  $\leq$ : an event  $A \subseteq S$  is  $\leq$ -null iff  $f \sim g$  whenever  $f(s) = g(s)$  for all  $s \in A^c$ . Throughout, we shall use  $\leq$  to denote the decision maker's current preference relation.

Finally, we introduce some standard properties of functions  $J : \mathfrak{R}^n \rightarrow \mathfrak{R}$ .  $J$  is: *constant additive* if  $J(\mathbf{x} + a\mathbf{e}) = J(\mathbf{x}) + a$  for all  $\mathbf{x} \in \mathfrak{R}^n$ ,  $a \in \mathfrak{R}$  and where  $\mathbf{e}$  is the unit vector; *positively homogeneous* if  $J(\alpha\mathbf{x}) = \alpha J(\mathbf{x})$  for all  $\mathbf{x} \in \mathfrak{R}^n$ ,  $\alpha \geq 0$ ; *monotonic* if  $J(\mathbf{x}) \geq J(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in \mathfrak{R}^n$  such that  $x_i \geq y_i$  for all  $i \in \{1, \dots, n\}$ , and *strongly monotonic* if

it is monotonic and  $J(\mathbf{x}) > J(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in \mathfrak{R}^n$  such that  $x_i \geq y_i$  for all  $i \in \{1, \dots, n\}$  with strict inequality for some  $i$ .

## 2.2 Epistemic contingencies

Until Section 5, we shall work under the assumption that the set of epistemic contingencies the decision maker envisages is given; in Section 6 we do away with this assumption and give a choice-theoretic foundation for this set. Moreover, as is standard, we shall implicitly assume a situation where there are two time periods: the current (ex ante) one and the future (ex post) one. In accordance with a significant number of non-expected utility models (Gilboa and Schmeidler, 1989; Ghirardato et al., 2004), we represent the decision maker's beliefs about the state of the world by sets of priors, that is convex, closed subsets of  $\Delta(\Sigma)$ .<sup>6</sup> The closed convex set  $\mathcal{C} \subseteq \Delta(\Sigma)$  represents the decision maker's current beliefs. The set of closed convex sets  $\{\mathcal{C}_i\}_{i \in I}$ , where  $I$  is a finite index set, represents the future beliefs the decision maker envisages himself as possibly having. These can be thought of as the epistemic contingencies the decision maker anticipates possibly being in. To focus on the issue of dynamic (in)consistency related to non-expected utility, we assume, as standard, that the decision maker uses the same affine utility function  $u : \Delta(X) \rightarrow \mathfrak{R}$  ex ante and ex post. Finally, we assume that the decision maker's preferences (both ex ante and ex post) can be represented according to the maxmin expected utility rule (Gilboa and Schmeidler, 1989). According to this rule, preferences  $\leq'$  corresponding to a set of priors  $\mathcal{C}'$  are such that, for every  $f, g \in \mathcal{A}$ ,  $f \leq' g$  iff

$$(1) \quad \min_{p \in \mathcal{C}'} \sum_{s \in S} u(f(s))p(s) \leq \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s)$$

Note that expected utility preferences correspond to the special case where the set of priors is a singleton.

It follows from these assumptions that  $\leq$ , the decision maker's current preference relation, satisfies (1) with respect to  $\mathcal{C}$ . Let  $\leq_i$  be the preference relation on  $\mathcal{A}$  satisfying (1) with respect to  $\mathcal{C}_i$  (for all  $i \in I$ ). Note that, under the aforementioned assumptions,

<sup>6</sup>For simplicity, we abstract from the debate about the degree to which, given the lack of separability of ambiguity and ambiguity attitude in some of these models, the sets of priors can be thought of as (purely) beliefs.

using the set of envisaged ex post beliefs  $\{\mathcal{C}_i\}_{i \in I}$  to represent the set of epistemic contingencies is equivalent to using a representation by the set of ex post preferences  $\{\preceq_i\}_{i \in I}$ , with each preference relation satisfying the [Gilboa and Schmeidler \(1989\)](#) axioms. For ease of presentation, we shall often use the representation in terms of preferences in the next few sections. Note that, whilst  $\preceq$  can be taken as an observable primitive, the set  $\{\preceq_i\}_{i \in I}$  cannot.

*Remark 1.* Consequentialism states that the decision maker’s ex post or conditional preference does not depend on ‘other branches’ in the decision tree. The standard formulation, which says that preferences conditional on an event  $A$  do not depend on what the acts deliver on  $A^c$  ([Ghirardato, 2002](#)), does not capture consequentialism in the current framework, since the  $\preceq_i$  do not necessarily correspond to updating on events. Nevertheless consequentialism is implied by the fact that the ex post preferences are determined entirely by the set of priors and the utility function: if other parts of the tree had an influence on the preferences under this contingency, then (1) would not be an appropriate representation of these preferences. Given that it is tacitly assumed in the framework, we shall not explicitly mention the condition in subsequent discussion.

*Remark 2.* An alternative setup would involve expanding the state space so that the decision maker’s ex post beliefs are explicitly represented in the states. As discussed in [Appendix A](#), under the appropriate assumptions, this framework is essentially equivalent to the one proposed above; moreover, all the points made below could equivalently have been in this ‘extended state space’ framework. Similar remarks apply to other possible setups, such as those involving a subjective state space in the style of [Kreps \(1992\)](#); [Dekel et al. \(2001\)](#) or a space of ‘signals’ or an information structure, beyond the objective state space  $S$ . See also [Remark 6](#) (in [Section 5](#)) on the relation to information structures, and [Remark 7](#) (in [Section 6](#)) on the relation to the literature on subjective state spaces.

### 3 Dynamic consistency on subjective trees

In [Section 1.2](#), we saw on a simple example how dynamic consistency on subjective trees, involving epistemic contingencies, does not imply expected utility preferences, even in the presence of consequentialism. In the current section, this point is made more rigorously: we formulate a version of dynamic consistency for subjective trees, and give an example to show that dynamic consistency, consequentialism and non-expected utility ex ante preferences are compatible, for all sets of anticipated ex post preferences (that is, for all subjective decision trees). In order to explore what dynamic consistency implies for non-

expected utility models, we shall also briefly consider its consequences for the relationship between ex ante and ex post preferences in the case of maxmin expected utility preferences.

### 3.1 *Introducing Dynamic Consistency*

To introduce the definition of dynamic consistency over subjective trees, it is reasonable to begin with the standard dynamic consistency condition over objective trees considered in the literature. Despite considerable differences between authors, the following condition seems to be fairly representative.

**Standard Dynamic Consistency (SDC).** For every  $f, g \in \mathcal{A}$  and partition  $\{A_j\}_{j \in J}$  of  $S$ , if  $f \leq_{A_j} g$  for every  $\leq$ -non-null  $A_j$ , then  $f \leq g$ , and moreover, if any of the  $\leq_{A_j}$  orderings are strict, then so is the  $\leq$  one.

Here  $\leq_{A_i}$  are the preferences conditional on  $A_i$ . Another standard condition used in the literature (see for example [Ghirardato \(2002\)](#)) is the special case of this condition applied to two-element partitions and acts which coincide on one element of the partition. It is straightforward to show that, given transitivity of preferences, this condition is equivalent to the one given above.

Standard Dynamic Consistency (SDC) captures the idea that when faced with any objective decision tree, comprising of a partition of events, if the decision maker will always prefer one act to another in all of the future eventualities, then this is the case under his current preferences. This corresponds to the requirement that his ex ante preferences should be coherent with his ex post preferences in the relevant contingencies, under the assumption that the relevant contingencies are the events of a partition. As argued previously, this assumption does not hold in general, and when it does not, then the ex post preferences that count are those in the (epistemic) contingencies that the decision maker does in fact envisage ex ante – that is the set of anticipated future preferences  $\leq_i$  – rather than those in the contingencies corresponding to the events imposed by the theorist – the  $\leq_{A_j}$ . It is straightforward to extend SDC to incorporate epistemic contingencies, and subjective trees.

**Dynamic Consistency (DC).** For all  $f, g \in \mathcal{A}$ , if  $f \leq_i g$  for all  $\leq_i$ , then  $f \leq g$ , and moreover, if any of the  $\leq_i$  orderings are strict, then so is the  $\leq$  one.

This is a straightforward modification of Standard Dynamic Consistency, where the preferences conditional on events in a partition are replaced by the anticipated future pref-

erences. Note that the non-nullness condition is not needed, because the set of preferences  $\{\leq_i \mid i \in I\}$  is assumed to contain only the future preferences that the decision maker envisages, and already leaves out future contingencies to which he gives ‘probability zero’. This is the refined notion of dynamic consistency that we propose, and with which we shall work throughout the rest of this paper.

*Remark 3.* Whilst SDC involves universal quantification over objective decision trees (in the form of quantification over the partitions of the state space), DC does not involve universal quantification over subjective decision trees (ie. over all possible sets of epistemic contingencies). One reason for this is that there is no normative justification for such quantification, least of all from the (English-language formulation of the) dynamic consistency principle itself. The principle is concerned with the decision maker’s preferences over contingent plans, and these involve the contingencies he actually envisages, not a set of contingencies that he *could have* but in fact does not envisage. Indeed, since he does not envisage such counterfactual sets of contingencies, there is no need for him to plan for them, so dynamic consistency has nothing to say about them. Moreover, adding quantification over sets of epistemic contingencies to DC would yield an empty condition, which can never be satisfied.<sup>7</sup> Finally, the apparent analogy between a quantified version of DC and SDC may be an artefact of the setup adopted here; as shown in Appendix A, when SDC is formulated appropriately in an extended state space with epistemic contingencies represented in the states, it turns out to be equivalent to DC.

### 3.2 *Dynamic Consistency and Non-expected utility*

As explained in the Introduction, a classic argument shows that Standard Dynamic Consistency and Consequentialism (in tandem with other basic assumptions on preferences) implies that the ex ante preferences are represented according to expected utility. This no longer goes through when SDC is replaced by Dynamic Consistency, which makes reference to epistemic contingencies rather than partitions of events. Indeed, no matter what epistemic contingencies the decision maker envisages, there exists a non-expected utility ex ante preference that satisfies Dynamic Consistency with respect to this set of contingencies (recall from Remark 1 that consequentialism is automatically satisfied in the framework

<sup>7</sup>A simple way to see this is to consider two sets  $\{\leq_i\}_{i \in I}, \{\leq_j\}_{j \in J}$  of epistemic contingencies, where an event  $A$  is null with respect to each preference relation in  $\{\leq_i\}_{i \in I}$  and  $A^c$  is null with respect to each preference relation in  $\{\leq_j\}_{j \in J}$ . No ex ante preference relation can satisfy DC with respect to both of these sets.

used here).

**Proposition 1.** *For any set  $\mathcal{D}$  of probability measures on  $I$  with  $p(i) > 0$  for all  $i \in I$  and  $p \in \mathcal{D}$ , let  $\leq_{\mathcal{D}}$  be the preference relation represented according to (1) with the set of priors  $\mathcal{C}_{\mathcal{D}} = \{\sum_{i \in I} p(i) \cdot q_i \mid p \in \mathcal{D}, q_i \in C_i\}$ . Then  $\leq_{\mathcal{D}}$  and  $\{\leq_i\}_{i \in I}$  satisfy Dynamic Consistency.*

This can be thought of as a possibility result, showing that once one recognises and adopts the contingencies which the decision maker himself envisages, dynamic consistency ceases to be inconsistent with non-expected utility theory. Rather, for any non-singleton set of anticipated ex post preferences (ie. for any non-trivial subjective decision tree), there is a non-expected utility ex ante preference satisfying dynamic consistency with respect to them: it suffices to take  $\mathcal{C}_{\mathcal{D}}$  for any non-singleton  $\mathcal{D}$ .

### 3.3 Respecting Dynamic Consistency

If Dynamic Consistency does not imply expected utility, then one might wonder what conditions it does place on ex ante preferences. Conceptually, this is basically a question concerning the relationship between two sorts of beliefs the decision maker has: his current beliefs about the state of the world – which determine his ex ante preferences – and his beliefs about what he will believe in the future about the state of the world – which correspond to the epistemic contingencies he envisages. What constraints do what the decision maker thinks about his future possible states of belief place on his current beliefs?

Technically, the question is analogous to the problem of aggregating non-expected utility decision makers' preferences. Indeed, dynamic consistency is evidently a strong dominance or Pareto condition on potential future preferences with respect to current preferences, of the sort commonly used in the literature on aggregation. Moreover, as noted by Crès et al. (2011), forming ex ante preferences with the set of priors given in Proposition 1 is equivalent to using a maxmin expected utility functional as the 'aggregator' of ex post maxmin EU preferences into ex ante ones.<sup>8</sup> However, as has been noted in literature on aggregating sets of priors, dynamic consistency does not imply that the decision maker must aggregate his future preferences according to this maxmin EU aggregator. (Crès et al.

---

<sup>8</sup>Their Proposition 1 shows that using the set of priors  $\mathcal{C}_{\mathcal{D}} = \{\sum_{i \in I} p(i)q_i \mid p \in \mathcal{D}, q_i \in C_i\}$  to form maxmin EU preferences over acts is equivalent to evaluating an act by its maxmin expected ex post valuation, taken with the set of priors  $\mathcal{D}$ , over the set of ex post functionals.

(2011, Section 2.3) give an example where this is not the case.) In fact, the aggregator may be more general, albeit not very illuminating, as the following proposition shows.

**Proposition 2.**  $\{\preceq_i\}_{i \in I}$  and  $\preceq$  satisfy Dynamic Consistency if and only if there exists a constant additive, positively homogeneous, monotonic function  $\phi : \mathfrak{R}^I \rightarrow \mathfrak{R}$  that is strongly monotonic on  $Ra((\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(\cdot)p(s))_{i \in I})$  such that  $\min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s) = \phi((\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s))_{i \in I})$  for all  $f \in \mathcal{A}$ .

Here  $Ra((\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(\cdot)p(s))_{i \in I}) \subseteq \mathfrak{R}^I$  is the range of the  $(\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(\cdot)p(s))_{i \in I}$ . Ghirardato et al. (2004) provide a representation of constant additive, positively homogeneous and monotonic functionals in terms of a generalised  $\alpha$ -maxmin expected utility representation, and this could be substituted into this proposition without changing its meaning.

Whilst technically useful below, this result is conceptually not particularly enlightening. It does have the virtue of showing that the refined version of Dynamic Consistency proposed here is not empty: it does impose non-trivial constraints on one's current beliefs ( $\mathcal{C}$ ) given one's beliefs about one's future beliefs ( $\{\mathcal{C}_i\}_{i \in I}$ ). Whilst, for largely technical reasons, there seems to be little more that can be said in general about the relationship between the ex ante set of priors and the ex post sets of priors under DC, in certain specific cases the relationship may be stronger and more interesting, as we shall see in the next section.

*Remark 4.* One can find supplementary conditions beyond DC that characterize the maxmin EU aggregator used in Proposition 1. One such condition draws its intuition from the observation that although there are cases where maxmin EU decision makers can satisfy dynamic consistency without taking a maxmin EU aggregator of their anticipated future preferences, such ways of aggregating future preferences are fragile in the sense that, if the set of ex post preferences happened to be different, the aggregated current preferences will generally not belong to the maxmin EU family. Hence the following notion of *robustness* of an aggregator.

*Definition 1.* An aggregator  $\phi : \mathfrak{R}^I \rightarrow \mathfrak{R}$  is *robust* if, for every set of maxmin EU preferences  $\{\preceq_i\}_{i \in I}$ , the derived preferences  $\preceq$ , defined by  $f \preceq g$  iff  $\phi((\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s))_{i \in I}) \leq \phi((\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(g(s))p(s))_{i \in I})$  for all  $f, g \in \mathcal{A}$ , is a maxmin EU preference.

A rule for aggregating one's anticipated future preferences into a current preference is robust if, irrespective of what one's future preferences are, whenever they are maxmin EU one is guaranteed to have maxmin EU current preferences. This notion is all that needs to be added to DC to characterize the maxmin EU aggregator.

*Proposition 3.*  $\phi$  is a robust aggregator satisfying DC iff there exists a unique set  $\mathcal{D}$  of probability measures on  $I$  with  $p(i) > 0$  for all  $i \in I$  and such that  $\phi(\mathbf{x}) = \min_{p \in \mathcal{D}} \sum_{i \in I} p(i)x_i$ .

## 4 Dynamic consistency and dynamic choice problems

We now turn to some economic consequences of the refined notion of dynamic consistency. Economists are often interested in cases where the decision maker is told which decision tree he is facing, and where this is the (objective) decision tree used by the theorist. At first glance, these cases would seem to pose a challenge for the approach developed above, given its reliance on the notion of subjective decision tree. Consider, for example, a maxmin EU decision maker considering the Ellsberg urn, with set of priors  $\{p \mid p(B) \in [0, \frac{2}{3}], p(R) = \frac{1}{3}\}$  (where  $B$  is the event that the black ball is drawn and  $R$  is the event that the red ball is drawn). Suppose that he is told that he will be faced with the ‘Dynamic Ellsberg’ decision tree given in Section 1.2 (Figure 2(a)). This decision maker cannot satisfy both dynamic consistency and consequentialism: herein lies the problem, and, it seems, the introduction of a refined notion of dynamic consistency involving epistemic contingencies does nothing to alleviate this difficulty. Moreover, in many dynamic problems, such as planning problems, intertemporal portfolio problems and applications to saving, these are precisely the sorts of implications of non-expected utility models that inhibit the use of standard techniques, such as backwards induction (or ‘folding back’) methods. In this section, we examine to what extent the proposed approach can handle such cases.

From the perspective of the framework introduced above, the situations just described concern decision makers whose subjective decision trees corresponds to objective ones. An objective decision tree involves as contingencies the events in a (finite) partition  $\mathcal{P}$  of the state space. In the situations described, each of the decision maker’s epistemic contingencies  $\{\leq_i\}_{i \in I}$  corresponds to learning exactly one of these events (and no two epistemic contingencies correspond to the same event). Hence, the subjective tree corresponds to the objective tree with partition  $\mathcal{P}$  precisely when, for each  $i \in I$ , there exists exactly one element  $A_i$  of  $\mathcal{P}$  such that  $A_i^c$  is  $\leq_i$ -null, and for any  $i, j \in I$ , if  $i \neq j$ , then  $A_i \neq A_j$ . If this holds for a set of epistemic contingencies  $\{\leq_i\}_{i \in I}$  and a partition  $\mathcal{P}$ , we say that  $\{\leq_i\}_{i \in I}$  is  $\mathcal{P}$ -ontic.

As the following result shows, under this assumption, dynamic consistency places a strong constraint on the form of the decision maker's ex ante set of priors.

**Proposition 4.** *Let  $\{\leq_i\}_{i \in I}$  and  $\leq$  satisfy DC. Suppose moreover that, for some partition  $\mathcal{P}$ ,  $\{\leq_i\}_{i \in I}$  are  $\mathcal{P}$ -ontic. Then there exists a unique set  $\mathcal{D}$  of probability functions on  $I$  with  $p(i) > 0$  for all  $i \in I$  and  $p \in \mathcal{D}$ , such that  $\mathcal{C} = \{\sum_{i \in I} p(i) \cdot q_i \mid p \in \mathcal{D}, q_i \in \mathcal{C}_i\}$ .*

For a partition  $\mathcal{P}$  and a set of priors  $\mathcal{C}$ , we say that  $\mathcal{C}$  is  $\mathcal{P}$ -rectangular if for every  $A_j \in \mathcal{P}$  there exists a set  $\mathcal{C}_j$  of probability measures with support in  $A_j$ , and there exists a set  $\mathcal{C}_0$  of probability measures on  $\mathcal{P}$  such that  $\mathcal{C} = \{\sum_{A_j \in \mathcal{P}} p(A_j) \cdot q \mid p \in \mathcal{C}_0, q \in \mathcal{C}_j\}$ . A  $\mathcal{P}$ -rectangular set of priors has a particular 'shape': it can be 'factorized' into a set of priors over  $\mathcal{P}$  multiplied by sets of priors on each element of the partition. This may be thought of as the equivalent in the present setup of the notion of rectangularity defined by [Epstein and Schneider \(2003\)](#).

Proposition 4 thus tells us that, under the assumption of Dynamic Consistency, whenever the decision maker's subjective decision tree corresponds to an objective tree (it is  $\mathcal{P}$ -ontic), then his ex ante set of priors has a special shape (it is  $\mathcal{P}$ -rectangular). It can be understood in the terms introduced in Section 3.3, namely, as concerning the relationship between the decision maker's current beliefs about the state of the world (the ex ante set of priors) and his beliefs about his possible future states of belief (the epistemic contingencies he envisages). By contrast with Proposition 2, which was relatively weak, Proposition 4 identifies a particularly strong relationship in a special case. First of all, it implies that whenever the agent's envisaged future beliefs have a particular form – the epistemic contingencies are  $\mathcal{P}$ -ontic – the aggregator by which the current beliefs are related to future ones is a maxmin EU one (of the sort discussed in Section 3). Furthermore, and more importantly, it tells us that this particular form of the envisaged ex post beliefs implies that the ex ante beliefs must also have a specific form: that is, they must be  $\mathcal{P}$ -rectangular.<sup>9</sup> Note that this form can be relatively straightforwardly verified on inspection of the ex ante set of priors. (It suffices to check that the set  $\{p(\bullet | A_i) \mid p \in \mathcal{C}, p(A_i) = x\}$  is the same for all  $x$ .) This renders the Proposition more interesting, because it makes the contrapositive of the implication meaningful.

---

<sup>9</sup>Note that the use of the maxmin EU aggregator alone does not imply the special rectangular shape of the ex ante set of priors: that also requires that the supports of the ex post sets of priors are disjoint.

Indeed, taking the contrapositive, Proposition 4 implies that (assuming Dynamic Consistency) if the decision maker's ex ante set of priors is *not*  $\mathcal{P}$ -rectangular – it does not have the particular shape – then his set of epistemic contingencies is not  $\mathcal{P}$ -ontic – he does *not* think that he will necessarily learn exactly one event in the partition  $\mathcal{P}$  and only that. In other words, it allows the theorist to draw a conclusion about the future beliefs the decision maker anticipates – and hence the subjective decision tree he is using – on the basis of a property of his current beliefs about the state of the world.

To see the consequences for the application of non-expected utility models in dynamic problems, consider the argument with which this section began. It is straightforward to check that the set of priors given above is not  $\{Y, Y^c\}$ -rectangular. As Proposition 4 makes clear, it does not follow that this decision maker is not Dynamically Consistent: under the assumption of Dynamic Consistency, it simply follows that he does not anticipate learning  $Y$  and learning  $Y^c$  as the only possible future states of belief. In other words, this ex ante set of priors implies that his subjective decision tree is not the objective decision tree given in Figure 2. Hence when he is told that he will be faced with the decision tree given in Figure 2, he is learning something: namely, something about the future contingencies that need to be planned for. Given that, as has been discussed in the previous section, DC implies that the decision maker's beliefs concerning his future beliefs place constraints on his current beliefs, it should not be surprising that when he updates his beliefs about his future beliefs – to incorporate the information that the only two future contingencies are learning  $Y$  and learning  $Y^c$  – this has consequences for his current beliefs. Indeed, Proposition 4 implies that when (or if) he updates his beliefs about his future beliefs to 'adopt' the decision tree in Figure 2, then he must also update his current beliefs so that they are  $\{Y, Y^c\}$ -rectangular. Since the subjective tree then coincides with an objective tree, the standard form of dynamic consistency will thus be satisfied, with respect to the ex ante beliefs the decision maker has *after* having updated on the information about the decision tree that he is facing. The decision maker has no trouble satisfying Dynamic Consistency in this example: it is always satisfied with respect to the subjective decision tree he thinks he is faced with and the ex ante beliefs he has when he thinks he is faced with that tree. The purported problem posed by the example is the result of the inappropriate comparison of his ex ante set of priors before learning that he is faced with the tree in Figure 2 and his subjective decision tree after having learnt that he will be faced with the tree in Figure 2.

We can thus draw two morals for the use of the maxmin expected utility model in dynamic problems.<sup>10</sup> First of all, Dynamic Consistency can be assumed to be universally satisfied without problem: indeed, given the reasonable intuition supporting it, it is a very natural assumption. Secondly, DC implies a particular relationship between the decision maker's current set of priors and the decision tree he considers himself to be faced with. Any theorist modelling decision makers must of course respect this relationship, if he is to avoid making inconsistent modelling assumptions.

The concrete advice for applications to dynamic problems is simple: when applying the maxmin expected utility model, the ex ante set of priors should be rectangular with respect to the partition formed by the nodes in the tree. It is known, since the seminal work of [Epstein and Schneider \(2003\)](#), that this is the only case in which the decision maker will be dynamically consistent in the classic sense, and hence where standard techniques such as folding back can be used. However, it is sometimes suggested that the restriction to rectangular sets of priors is at best a partial solution, for the problem of dynamic inconsistency remains when the set of priors is not rectangular with respect to an event of interest (see, for example, [Hanany and Klibanoff \(2007, p282\)](#) and [Al Najjar and Weinstein \(2009, §3.4\)](#)).<sup>11</sup> Proposition 4 provides a reply to this worry, in the form of a principled justification of the restriction to rectangular sets. If the set of priors is not of the appropriate rectangular form, this does not imply that the decision maker is dynamically inconsistent, in the refined sense given in Section 3. Rather, it implies that the decision maker is not using the same decision tree as the theorist. So the restriction to appropriately rectangular sets amounts to an assumption that the decision maker does indeed know the decision tree he is faced with, and that this is the same as the one the theorist is using. As such, there is nothing arbitrary or unnecessarily limitative about it; indeed, without such assumptions about the way decision makers conceive the problems they are faced with, economic modelling can hardly get off the ground!

We thus conclude that the proposed notion of dynamic consistency, and the subjective framework for thinking about such issues, has no trouble coping with the purported difficulties for non-expected utility models in dynamic choice problems. Quite to the contrary, it provides a reasoned defense of a certain approach to applying non-expected utility models

<sup>10</sup>These points hold *mutatis mutandis* for other non-expected utility models.

<sup>11</sup>Epstein and Schneider's own conclusion (p14) is that 'in *some settings*, ambiguity may render dynamic consistency problematic.'

to such problems, which has already been proposed in the literature. Non-expected utility models can be used in dynamic problems, and even using folding back reasoning: this is a consequence of the fact that there is no obstacle to non-expected utility decision makers being dynamically consistent, with respect to decision tree they consider themselves to be using. However, when using these models, one must respect the consequences of one's assumptions about what the decision maker takes the decision tree he is facing to be: the lesson of Proposition 4 is that many purported violations of dynamic consistency are just modelling errors on the part of the theorist.

*Remark 5.* Note that singleton sets of priors are the only sets that are  $\mathcal{P}$ -rectangular for every partition  $\mathcal{P}$ . Consequently, the only ex ante preferences such that, for every partition  $\mathcal{P}$ , there exists a  $\mathcal{P}$ -ontic set of epistemic contingencies with respect to which DC is satisfied, are expected utility preferences. This is a version of the standard result concerning SDC and expected utility, recovered in our framework. In the light of the preceding considerations, this means that expected utility preferences are the only ones from which one cannot immediately draw the sorts of implications just described about the decision tree that the decision maker is using. The potential for the sorts of modelling errors discussed above is thus reduced with expected utility preferences.

## 5 Value of Information

In this section, we consider another important economic issue raised in discussions of dynamic consistency, the attitude towards information. A commonly noted purported consequence of non-expected utility is that sophisticated decision makers may prefer not to obtain free information: they are *information averse* (Wakker, 1988; Al Najjar and Weinstein, 2009). This of course is in stark contrast with basic pre-theoretical intuition, as well as with the standard Bayesian expected utility model, and is often taken to be an argument, if not one of the strongest arguments, against non-expected utility theory. As a further illustration of the power of the approach proposed here, we analyse the issue of value of information in the framework introduced in the previous sections.

### 5.1 Information aversion: debunking the standard argument

The standard argument that non-expected utility decision makers are information averse can be formulated on a sequential extension of the dynamic Ellsberg example given in Section 1.2. Consider a decision maker with standard Ellsberg preferences – he prefers

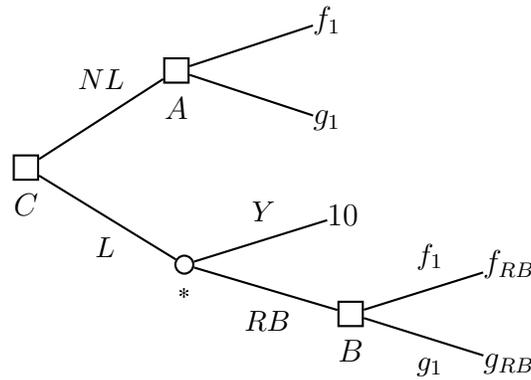


Figure 4: Information Aversion

$g_1$  to  $f_1$  at node  $A_1$  in Figure 2(a), but prefers  $f_{RB}$  to  $g_{RB}$  at node  $B_1$  in Figure 2(b) – and suppose that he is given the choice between facing these two decision trees.<sup>12</sup> That is, he is faced with the decision tree in Figure 4. Since he has the Ellsberg preferences just specified, he knows that if he reaches decision node  $A$  he will choose  $g_1$  and that if he reaches decision node  $B$  he will choose  $f_{RB}$ . Reasoning by backwards induction,<sup>13</sup> the decision maker at node  $C$  knows that if he takes the upper branch  $NL$  of the tree, he will essentially end up with  $g_1$ , and if he takes the lower branch  $L$  he will end up with  $f_1$  (given that the decision maker does not know the resolution of the uncertainty at the chance node \*, the choice of  $f_{RB}$  at node  $B$  essentially boils down to a choice of  $f_1$  from the point of view of a decision maker at  $C$ ). Hence, since he prefers  $g_1$  to  $f_1$ , he chooses  $NL$  at  $C$ . However, since the choice at  $C$  is essentially that between learning whether the ball is yellow before deciding to bet (option  $L$ ) or not learning whether it is yellow (option  $NL$ ), by choosing  $NL$ , the decision maker betrays a preference for not obtaining free information. This is the criticized information aversion.

In Section 4, we noted that the assumption of Dynamic Consistency implies a strong modelling constraint on the theorist, which undermines the standard practice of taking a decision maker with a given set of priors, subjecting him to a decision tree, and deducing that he is dynamically inconsistent. An essential insight is that, when the decision maker learns that he is faced with a particular decision tree, this may affect his current set of

<sup>12</sup>In this discussion, notations and numbering are taken from Section 1.2.

<sup>13</sup>Decision makers who reason in this way in situations of dynamic inconsistency are sometimes said to adopt ‘the strategy of consistent planning’, or to be ‘sophisticated’; see [Strotz \(1955\)](#); [Karni and Safra \(1990\)](#); [McClennen \(1990\)](#); [Siniscalchi \(2011\)](#).

beliefs. The same general point applies in the sequential situation described here. Let  $\mathcal{C}$  be the set of priors that decision maker has after having chosen not to learn – that is, at node  $A$ . As specified above, this set of priors yields the standard Ellsberg preferences, and so is not rectangular with respect to the partition  $\{Y, Y^c\}$ . However, if the decision maker chooses to learn, then *even before learning whether the ball is yellow or not* – that is, even at the chance node  $*$  – his set of priors will not be  $\mathcal{C}$ . This is a consequence of Proposition 4: at node  $*$  he knows that he is facing an objective tree which will determine which element of the partition  $\{Y, Y^c\}$  occurs, and so his set of priors must be rectangular with respect to this partition (assuming Dynamic Consistency). Since  $\mathcal{C}$  is not rectangular with respect to this partition, it follows that his set of priors at node  $*$  is not  $\mathcal{C}$ . In other words, his choice of whether to learn or not at node  $C$  affects the beliefs he has immediately after his decision, but before any (other) learning has taken place.

This is a natural consequence of the discussion in the previous sections. If aspects of the decision maker’s future beliefs are reflected in his current beliefs, then choices he makes that determine the future beliefs he envisages as possibly having – such as the choice of whether to obtain information or not – impact on his current beliefs. This impact occurs even before he acquires any of the promised information: it is, if you will, the information that he will acquire some information that induces this change in current beliefs, rather than the information that will be acquired itself. And this ‘second-order’ information comes directly from the choice he has made. As such, there is a strong analogy between the choice of whether to learn and moral hazard problems. The decision maker above is in a similar situation to an Agent whose chances of a particular reward under a given contract depend on the effort he chooses to expend. In both cases, the decision makers’ beliefs (about the colour of the ball, or about whether he will get the reward respectively) are directly determined by their choices.

This analogy suggests a diagnosis of the argument sketched above purporting to establish information aversion. The argument considers that the choice between  $NL$  and  $L$  at node  $C$  boils down to the choice between  $g_1$  and  $f_1$ . It assumes that the decision maker has the Ellsberg set of priors  $\mathcal{C}$  at node  $C$ , and evaluates the choice at that node on the basis of the preferences over  $f_1$  and  $g_1$  corresponding to these priors. However, this amounts to the assumption that the decision maker’s set of priors immediately after his choice at  $C$  are independent of the choice made, and we have just seen that this is false. Since the choice at

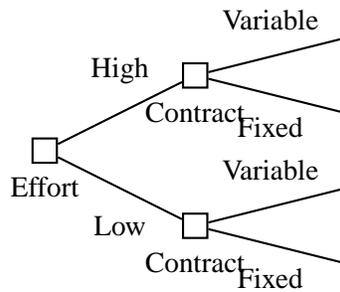


Figure 5: Agent's choice

node  $C$  alters his beliefs and hence his evaluations of acts, the decision maker should use the interim beliefs (or if you wish, the beliefs conditional upon the choice to learn or not) to evaluate the options  $NL$  and  $L$  at node  $C$ . This is the standard approach in sequential choice situations of this sort. Consider an Agent who has to choose a level of effort, and then will have the choice of which contract to sign, before observing the outcome of his work (eg. the profits); assuming two levels of effort and two contracts, the decision tree for this problem is given in Figure 5. Under the standard approach, one would not calculate ex ante (ie. before knowing the choice of effort) whether the variable or fixed contract is preferable, and then choose whatever effort level leads to that choice of contract at the Contract nodes. Such reasoning, which is the analogue of the above argument for information aversion, obviously errs by wrongly assuming that the optimal choice of contract is independent of the effort exerted.<sup>14</sup> Indeed, the standard approach is to evaluate each of the available strategies – (High effort, variable contract), (High effort, fixed contract) and so on – and these evaluations would use the probabilities (beliefs) of outcomes conditional on the choice of effort made. The analogue of this sort of calculation in the case of the choice of obtaining information does not immediately imply that non-expected utility decision makers are information averse.

<sup>14</sup>Another problem with the reasoning is that it relies on an ex ante probability distribution, which depends on the choice of effort and so requires the Agent to estimate the probability of making a particular choice in the decision he is currently deliberating upon. Similarly in the information case, properly determining the decision maker's set of priors at node  $C$  in the tree in Figure 4 would require his beliefs about whether he will choose to learn or not at that node. Doubts may be raised as to whether such beliefs can be formed or be economically meaningful.

## 5.2 Value of envisaged information

What, then, can be said about the value of information under the current proposal? Since the proposal incorporates the beliefs the decision maker envisages himself as possibly having, and hence implicitly the information he expects to receive, the question can be naturally divided into two. On the one hand, what is the value of the information that he expects to receive? On the other hand, what is the value of information that the decision maker is offered but had not necessarily expected to receive? For expositional purposes, it is convenient to treat the former question first.

The issue is essentially whether a decision maker would accept to delay his decision, on the (subjective) decision tree he envisages, until he arrives at the future nodes. It is sometimes suggested that this is immediately implied by dynamic consistency (see for example [Ghirardato \(2002\)](#)); however, such an interpretation involves identifying contingent plans with pairs of acts, and hence it does not easily transfer over to the case of subjective trees. Nevertheless it can be shown, by properly defining the value of information in a manner analogue to the standard Bayesian literature, that it is indeed non-negative.

The value of information over envisaged contingencies involves the comparison between the value of choosing now and the value of waiting for the realisation of the contingency before choosing.<sup>15</sup> Unlike the standard literature, the ex post value of the information and the ex ante value of the option of not receiving the information are calculated using the maxmin expected utility rule with the ex post and ex ante sets of priors. However, to calculate the ex ante value of obtaining the information, an ‘aggregator’ connecting ex post and ex ante preferences is required. (In the standard Bayesian case, the aggregator is the expected utility rule.) We thus assume that there is such an aggregator  $\phi : \mathfrak{R}^I \rightarrow \mathfrak{R}$ : i.e. that there exists  $\phi$  such that, for all  $f \in \mathcal{A}$ ,  $\min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s) = \phi((\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s))_{i \in I})$ . By [Proposition 2](#), such an aggregator exists whenever the decision maker satisfies DC. Under this assumption, the following definition of non-negative value of information is correctly formulated.

**Definition 2.** Let  $\mathcal{C}$  and  $\{\mathcal{C}_i\}_{i \in I}$ , for  $i \in I$  represent ex ante and ex post sets of priors and let  $\phi : \mathfrak{R}^I \rightarrow \mathfrak{R}$  be the aggregator. The *value of envisaged information for a menu*  $A \in \wp(\mathcal{A})$  is non-negative if and only if:

---

<sup>15</sup>We are following a standard approach to information value in, for example, [Marschak and Miyasawa \(1968\)](#); [Marschak and Radner \(1972\)](#); [Hilton \(1981\)](#).

$$(2) \quad \phi \left( \left( \max_{f \in A} \min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s) \right)_{i \in I} \right) \geq \max_{f \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s)$$

The *value of envisaged information is always non-negative* if it is non-negative for every  $A \in \wp(\mathcal{A})$ .

This notion of non-negative value of information is evidently analogous to the standard notion, the difference being that it does not involve the assumption of expected utility decision makers, and is specific to the information that the decision maker envisages acquiring ex ante – that is, to his subjective decision tree.

**Proposition 5.** *If the decision maker satisfies DC, then the value of envisaged information is always non-negative.*

For the information the decision maker envisages receiving, the refined notion of Dynamic Consistency implies that it has non-negative value. The standard point concerning SDC and objective trees thus continues to apply for DC and the decision maker's subjective tree: he weakly prefers to wait for the information before deciding. Here there is no possibility of information aversion.

### 5.3 Value of offered information

The situation is more subtle in cases where the decision maker had not expected to learn (just) the information on offer, and so learns something about his potential future beliefs – or perhaps more accurately, about the fact that they will be modified – on choosing to obtain the information. As explained above, such cases are analogous to moral hazard situations, and should be analyzed accordingly. In particular, the value of information involves the comparison of the value of choosing just after having rejected the offer of information, and the value of accepting the offer and choosing after having received the information. As is clear from the discussion in Section 5.1, this is not the standard way of reasoning about attitude to information in the literature on non-expected utility theory.

For the analysis, consider an offer of acquisition of information  $\mathcal{I}$ . Let  $\mathcal{C}$  be the decision maker's set of priors immediately after having chosen not to obtain the information, and let  $\mathcal{C}_{(\mathcal{I})}$  be his set of priors immediately after having chosen to obtain the information, but before actually receiving it. Being slightly slack with terminology, we shall continue to call

these the ex ante sets of priors in the two cases. Let  $\{\mathcal{C}_i\}_{i \in I}$ ,  $\phi : \mathfrak{R}^I \rightarrow \mathfrak{R}$  and  $\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}$ ,  $\phi_{(\mathcal{I})} : \mathfrak{R}^K \rightarrow \mathfrak{R}$  be the ex post sets of priors and aggregators just after having chosen not to turn down and acquire the information respectively.<sup>16</sup>

*Remark 6.* Note that this framework is broad enough to cover standard representations of information structures or systems. One could consider a set of signals, and understand each  $\mathcal{C}_{(\mathcal{I})k}$  to be a posterior set of beliefs after the reception of some signal, with  $\phi_{(\mathcal{I})}$  reflecting the decision maker's prior beliefs as to which signal will be received. As such, the framework essentially corresponds to a standard representation of information structures (for example [Gollier \(2004, Ch 24\)](#)), the only difference being that the beliefs and aggregator are not necessarily expected utility. Similarly, another standard representation of information structures as partitions (eg. [Marschak and Miyasawa \(1968\)](#); [LaValle \(1968\)](#)) can be incorporated, by adding assumptions that the support of each  $\mathcal{C}_{(\mathcal{I})k}$  belongs to exactly one element of the given partition.<sup>17</sup>

Moreover, this setup is silent on the question of the relationship between the beliefs in the case where the information is rejected and those in the case where it is accepted. The development of update rules that determine how the decision maker's envisaged future beliefs are revised upon learning that he will receive some hitherto unexpected information is a topic for future research.

The standard question of value of information involves comparison of the value of deciding immediately after having chosen not to receive the information – so the relevant set of priors is  $\mathcal{C}$  – with the anticipated value of deciding after having received the information – so the relevant elements are the sets of priors  $\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}$  and the opinion as to which will be realized, given by  $\phi_{(\mathcal{I})}$ . Hence the following definition of the non-negative value of offered information.

**Definition 3.** Let  $\mathcal{C}$ ,  $\{\mathcal{C}_i\}_{i \in I}$ ,  $\phi$ ,  $\mathcal{C}_{(\mathcal{I})}$ ,  $\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}$ ,  $\phi_{(\mathcal{I})}$  be the ex ante beliefs, envisaged ex post beliefs and aggregators in the case where the decision maker chooses to reject the information  $\mathcal{I}$  or to accept it respectively, for some offered information  $\mathcal{I}$ . The *value of*

<sup>16</sup>It is natural to assume that  $\mathcal{C}$ ,  $\{\mathcal{C}_i\}_{i \in I}$  and  $\phi$  are independent of the information  $\mathcal{I}$  offered, insofar as they can be thought of as representing the decision maker's state even before he envisaged the possibility of being offered the information. Nothing in the analysis below depends on this assumption.

<sup>17</sup>Note that there is not necessarily a one-to-one correspondence between the 'objective' signals or cells of the partition and the ex post sets of priors, because the decision maker may have expected to learn other information over and above the information offered. For example, if he expects to learn which of the events  $B$  or  $B^c$  hold, and is offered the information of which of  $A$  and  $A^c$  hold, then there may be up to four relevant ex post sets of priors – after learning  $A$  and  $B$ , after learning  $A$  and  $B^c$ , and so on – though the size of the information partition offered is two.

offered information  $\mathcal{I}$  for a menu  $A \in \wp(\mathcal{A})$  is non-negative if and only if:

$$(3) \quad \phi_{(\mathcal{I})} \left( \left( \max_{f \in A} \min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u(f(s))p(s) \right)_{k \in K} \right) \geq \max_{f \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s)$$

The value of offered information  $\mathcal{I}$  is always non-negative if it is non-negative for every menu  $A \in \wp(\mathcal{A})$ .

Envisaged information is recovered as the special case where the decision maker is offered information which he already expected to learn (ie. where  $\mathcal{C} = \mathcal{C}_{(\mathcal{I})}$ ,  $\{\mathcal{C}_i\}_{i \in I} = \{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}$ ,  $\phi = \phi_{(\mathcal{I})}$ ). In other cases, there are clear non-trivial necessary and sufficient conditions for the value of offered information to be non-negative.

**Proposition 6.** *Suppose that the decision maker satisfies DC. Then the value of information  $\mathcal{I}$  is always non-negative iff  $\mathcal{C}_{(\mathcal{I})} \subseteq \mathcal{C}$ .*

This proposition provides the final piece in the analysis of the allegation that non-expected utility decision makers necessarily exhibit information aversion. It tells us that the value of information for any maxmin expected utility decision maker will be non-negative as long as his set of priors immediately after having chosen to learn the information is contained in the set of priors he has immediately after having chosen not to acquire the information. Since the standard arguments, such as the one discussed in Section 5.1, involve objective trees, the information offered comes in the form of a partition and this adds another constraint: namely, that the set of priors immediately after accepting the information are  $\mathcal{P}$ -rectangular.<sup>18</sup> These two constraints are simultaneously satisfiable: for one, the singleton sets containing priors in  $\mathcal{C}$  satisfy both conditions. Hence, for any decision maker who has a set of priors  $\mathcal{C}$  if he decides not to accept the information, there exists a set of priors such that, if he holds this set of priors after having chosen to acquire the information, he will have non-negative value of information. In other words, a non-expected utility decision maker can always exhibit weak preference for information with respect to any information partition offered. Once the fact that the decision maker may change his beliefs after choosing to acquire information is properly incorporated into the calculation of the

<sup>18</sup>Since, as noted in footnote 17, there may be more than one ex post preference that is non-null on a given cell of the partition, Proposition 4 does not strictly speaking apply. However, it is straightforward to extend it to cover this case.

value of that information, it is clear that claim that non-expected utility necessarily implies information aversion is not only unfounded: it is false.

#### 5.4 *Negative value of information and opportunity costs*

One might nevertheless remain puzzled by the fact that it is possible, if the constraint in Proposition 6 is not met, for a decision maker to give a negative value to some information (for some menu). Is the mere possibility of turning down free information not itself an unacceptable consequence of non-expected utility? There is certainly a strong intuition that one should never have a strict preference for refusing free information; but it implicitly rests upon the assumption that the information acquisition does not compromise the reception of information one had otherwise expected to obtain. Suppose that James, a manager of a firm, has a spy in a competitor's board of directors, and expects to learn from the spy the competitor's strategy in a market B. He is offered the possibility to learn 'for free' the competitor's strategy in another market A. If this information can be acquired without consequence for the prospects of gaining information about market B, then it is evidently of non-negative value. Suppose, however, that acquiring the information about the strategy in market A involves blowing the spy's cover, so he will no longer be able to tell James anything about market B. In this case, it may well be reasonable to refuse the offer of information about market A, because acquiring this information jeopardizes other information that James had expected to receive. Of course, in this sort of case, the choice is essentially between two different pieces of information: the information the decision maker expected to get before the offer and the information he will get if he accepts the offer. If the value of the former piece of information is greater than that of the latter – if the opportunity cost of the offered information outweighs its value – then it is perfectly reasonable to turn down the offer of 'free' information. The condition in Proposition 6, we suggest, can be understood as reflecting this caveat.

A clear way to see this point is by considering what happens in situations where the assumptions underlying the standard intuition in favour of non-negative value of information are satisfied: that is, in cases where choosing to acquire the offered information does not imply forgoing information that one already expected to obtain. Note firstly that, as mentioned above (see Remark 6), the ex post priors and aggregators in the case where the information is declined ( $(\{C_i\}_{i \in I}, \phi)$ ) and the ex post priors and aggregators in the case where

the information is accepted ( $(\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}, \phi_{(\mathcal{I})})$ ) can be thought of as information structures, and the decision of whether to accept the information is effectively a choice between these two information structures. If no previously expected information is compromised on accepting the offer of information, then the latter information structure is more informative than the former one. According to a standard definition, informativeness can be defined in terms of decision makers' attitudes to the information: one information structure is more informative than another if every decision maker (no matter his utility function) would prefer to learn according to the former information structure (Marschak and Miyasawa, 1968; Gollier, 2004).<sup>19</sup> This definition can be adapted to the current framework as follows.

**Definition 4.** Let  $\mathcal{C}$ ,  $\{\mathcal{C}_i\}_{i \in I}$ ,  $\phi$ ,  $\mathcal{C}_{(\mathcal{I})}$ ,  $\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}$ ,  $\phi_{(\mathcal{I})}$  be the ex ante beliefs, envisaged ex post beliefs and aggregators in the case where the decision maker chooses to reject the information  $\mathcal{I}$  or to accept it respectively.  $(\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}, \phi_{(\mathcal{I})})$  is *at least as informative as*  $(\{\mathcal{C}_i\}_{i \in I}, \phi)$  if and only if, for every affine utility function  $u' : \Delta(X) \rightarrow \mathfrak{R}$ , and every menu  $A \in \wp(\mathcal{A})$ :

$$(4) \quad \phi_{(\mathcal{I})} \left( \left( \max_{f \in A} \min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u'(f(s))p(s) \right)_{k \in K} \right) \geq \phi \left( \left( \max_{f \in A} \min_{p \in \mathcal{C}_i} \sum_{s \in S} u'(f(s))p(s) \right)_{i \in I} \right)$$

It follows directly from this definition and Proposition 5 that, whenever the information structure one faces after accepting the information offered is at least as informative as the information structure faced when declining the offer, then the value of the offered information is non-negative.

**Proposition 7.** *Suppose that the decision maker satisfies DC. Then the value of information  $\mathcal{I}$  is always non-negative whenever  $(\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}, \phi_{(\mathcal{I})})$  is at least as informative as  $(\{\mathcal{C}_i\}_{i \in I}, \phi)$ .*

Hence, in accordance with the standard intuition, non-expected utility decision makers will not turn down free information when it does not compromise information they had previously expected to receive.

<sup>19</sup>Naturally, there are a battery of important results showing that this definition is equivalent to other possible definitions of informativeness under expected utility (Blackwell, 1953; Marschak and Miyasawa, 1968; Gollier, 2004). We are not aware of similar results for non-expected utility models; indeed, it is not even evident what the analogues for some of the other proposed definitions would be for such models.

We conclude that, even for non-expected utility decision makers, their attitude to information behaves exactly ‘as it should do’. On the one hand, the fact that they are non-expected utility does *not* imply that they will *necessarily* ever have negative value of information: Proposition 6 implies that it is always possible for their beliefs after having chosen to acquire the information to be such that the value of the information is non-negative. On the other hand, it is *possible* for them to exhibit negative value of information – the condition in Proposition 6 is non-trivial. However, when this happens, it does so for good reason. In particular, Proposition 7 tells us that, whenever the information offered does not compromise information that was expected to be acquired – whenever it is a simple addition of information, with no subtraction, so to speak – the value they assign to the information offered is non-negative. If the value of the information offered is negative, then this must be a case where the decision maker forgoes information he had expected to receive on accepting the information offered. In these cases, aversion towards the information offered is entirely acceptable: if the information lost by accepting the offer is worth more than the information acquired, then it is entirely reasonable to decline the offer.

## 6 Foundations

So far we have assumed a set of epistemic contingencies as given; however, they are evidently not directly observable. The approach depends on the assumption that the decision maker does envisage such epistemic contingencies; moreover, it is important, to ascertain whether the refined notion of Dynamic Consistency is respected or not, to determine which epistemic contingencies the decision maker envisages. The aim of this section is to present behavioral foundations for the notion of epistemic contingency in the context of the maxmin EU model. To this end, we shall introduce a choice-theoretic situation and propose axioms that are necessary and sufficient for a representation of choices involving envisaged sets of future priors  $\{\mathcal{C}_i\}_{i \in I}$ . The result in this section should relieve worries that the proposed reconceptualization of dynamic choice via the introduction of the notion of epistemic contingency relies on notions with no independent behavioral meaning. The notions have in-principle observable consequences (beyond the application to the question of dynamic consistency), and the set of epistemic contingencies envisaged by the decision maker can in principle be gleaned from choice behavior.

### 6.1 Setup and representation

We continue to work in the setup set out in Section 2.1. Rather than assuming (future) sets of priors or preference relations as primitives, we work now with choice functions, where a choice function is a function  $c : \wp(\mathcal{A}) \rightarrow \wp(\mathcal{A})$  such that, for any  $A \in \wp(\mathcal{A})$ ,  $c(A) \subseteq A$ .<sup>20</sup> The choice function  $c$  has the following interpretation, associated with the time line given in Figure 6. The decision maker knows that he will have to choose an act from a menu  $A$  at an ex post stage (before the realisation of the state of the world, but perhaps after receiving information), and has the opportunity in an ex ante stage of restricting the options left open to a subset of  $A$ , from which he will make his ex post choice. For an example of such a choice situation, consider a committee deciding on the allocation of a building contract or of a university post: they may in the first instance rule out some of the candidates, producing a shortlist, from which they will later choose the winner. For any  $A \in \wp(\mathcal{A})$ ,  $c$  yields the set of elements that the decision maker now wishes to keep as open alternatives for his future choice; the elements not in  $c(A)$  are those which the decision maker is willing to rule out now.

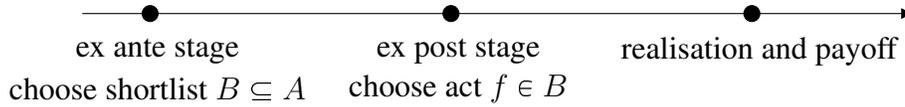
We shall consider the following representation:

$$(5) \quad c(A) = \{f \in A \mid f \in \arg \max_{g \in A} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) \text{ for some } \mathcal{C}' \in \mathcal{P}\}$$

where  $u$  is a continuous affine utility function on  $\Delta(X)$  and  $\mathcal{P}$  is a set of convex, closed subsets of  $\Delta(\Sigma)$ . The sets of priors in  $\mathcal{P}$  are interpreted as the future beliefs that the decision maker anticipates himself as possibly having at the moment when he will be faced with his final choice; they can be thought of as the epistemic contingencies he envisages. (5) represents a decision maker who envisages that he will hold one of the sets in  $\mathcal{P}$  as his set of priors at the specified future moment, and that he will form preferences with that set according to the maxmin EU rule. He retains as an open option any act which is optimal according to this rule with any of the sets of priors in  $\mathcal{P}$ . To see this more clearly, note that this representation can be reformulated in terms of the ex post preferences in the different epistemic contingencies, as follows. Define the set  $\mathcal{R}$  of preference relations on  $\mathcal{A}$  as follows:  $\leq_i \in \mathcal{R}$  if and only if there exists  $\mathcal{C}_i \in \mathcal{P}$  such that  $\leq_i$  is represented according to (1) by  $\mathcal{C}_i$  and  $u$ .  $\mathcal{R}$  contains the decision maker's anticipated ex post preferences, formed

<sup>20</sup>This is sometimes called a choice correspondence in the literature.

Figure 6: Time line



on the basis of the ex post beliefs according to the maxmin EU rule. Representation (5) is equivalent to the following representation of ex ante choice in terms of a set  $\mathcal{R}$  of preference relations each of which is represented according to (1):

$$(6) \quad c(A) = \{f \in A \mid \exists \preceq_i \in \mathcal{R}, \forall g \in A, f \succeq_i g\}$$

In other words, the decision maker keeps open any act which is optimal in the menu according to at least one of his anticipated ex post preferences, and rules out any act which is not optimal under any of the ex post preferences. The choice rule (6) yielding a choice function given a set of preferences has been studied in the choice-theoretic literature, where it has been called *pseudo-rationalisability* (Moulin, 1985).

*Remark 7.* Note that representation (5) could alternatively be formulated in terms of preferences over menus. To illustrate the relationship, define the preference relation  $\dot{<}$  on  $\wp(\mathcal{A})$  as follows: for all  $A, B \in \wp(\mathcal{A})$ ,

$$(7) \quad A \dot{<} B \quad \text{iff} \quad c(A \cup B) \cap A = \emptyset$$

It is straightforward to check that this preference over menus is represented by the  $u$  and  $\mathcal{P}$  featuring in representation (5) as follows: for all  $A, B \in \wp(\mathcal{A})$ ,  $A \dot{<} B$  iff

$$(8) \quad \max_{g \in A} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) < \max_{g \in B} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) \quad \text{for all } \mathcal{C}' \in \mathcal{P}$$

The relation  $\dot{<}$  is a preference relation over menus in the style of Kreps (1979); Dekel et al. (2001), but, unlike most preference relations studied in this literature, it will not satisfy the standard conditions on preferences: in particular, the weak version of the relation will not be complete. This is due to the representation. It is standard in the aforementioned literature to assume that ex ante

preferences are formed using an ‘aggregator’ that ensures completeness; by contrast, representation (8) combined with definition (7) involves a ‘neutral’ aggregator that only orders menus in cases of strict dominance. This difference is related to the goal of the exercise in this section. Whereas most of the representation theorems in the preference for flexibility literature deliver both ex post preferences (or states of a subjective state space) and ex ante preferences that are aggregated in a particular way with respect to the ex post ones, here the aim is to give foundations solely for ex post sets of priors or preferences (the epistemic contingencies), without making any assumptions about the ex ante preferences. Indeed, the issue of the relationship between ex post and ex ante preferences is essentially a question about dynamic consistency, and this should evidently be kept separate from the behavioral foundation of the notion of epistemic contingency.

## 6.2 Axioms and result

Consider the following axioms on the choice function  $c$ .

**Axiom A1** (Chernoff). For all  $A, B \in \wp(\mathcal{A})$ ,  $f \in \mathcal{A}$ , if  $A \subseteq B$  and  $f \in c(B)$ , then  $f \in c(A)$ .

**Axiom A2** (Aizerman). For all  $A, B \in \wp(\mathcal{A})$ , if  $c(B) \subseteq A \subseteq B$ , then  $c(A) \subseteq c(B)$ .

**Axiom A3** (Non-degeneracy). There exist  $d, e \in \Delta(X)$  such that  $d \in c(\{d, e\})$  and  $e \notin c(\{d, e\})$ .

**Axiom A4** (Fixed utilities). For all  $d, e \in \Delta(X)$  and  $A, B \in \wp(\Delta(X))$  with  $A \subseteq B$ , if  $d, e \in c(A)$ , then  $d \in c(B)$  if and only if  $e \in c(B)$ .

**Axiom A5** (Set C-Independence). For all  $A \in \wp(\mathcal{A})$ ,  $d \in \Delta(X)$ , and for all  $\alpha \in (0, 1)$ ,  $c(A_\alpha d) = c(A)_\alpha d$ .

**Axiom A6** (Union C-Independence). For all  $A \in \wp(\mathcal{A})$ ,  $\alpha \in (0, 1)$  and  $d \in \Delta(X)$  with  $d \in c(A)$ ,  $c(A) \subseteq c(A \cup A_\alpha d)$ .

**Axiom A7** (Monotonicity). For all  $A, B \in \wp(\mathcal{A})$  with  $A \subseteq B$ , if, for each  $g \in B$ , there exists  $f \in A$  with  $f(s) \in c(\{f(s), g(s)\})$  for all  $s \in S$ , then  $c(A) \subseteq c(B)$ . Moreover, for every  $g \in B$ , if there exists  $f \in B$  with  $g(s) \notin c(\{f(s), g(s)\})$  for all  $s \in S$ , then  $g \notin c(B)$ .

**Axiom A8** (Uncertainty aversion). For all  $A, B \in \wp(\mathcal{A})$  with  $B \subseteq c(A)$ ,  $f \in c(A \cup \{f\})$  for all  $f \in \text{conv}(B)$  whenever there exist  $d \in \Delta(X)$  and  $g \in B$  such that: i.  $f_\alpha d \in B$  for all  $\alpha \in [0, 1]$  and all  $f \in B$ ; and ii.  $g_\alpha d \notin c(A \cup \{f_\beta e\})$  for all  $f \in B$ , all  $e \in \Delta(X)$  with  $e > d$  and all  $\alpha, \beta \in (0, 1)$ .

**Axiom A9** (Continuity). For all sequences of menus  $(A_n)_{n \in N}$  and  $A \in \wp(\mathcal{A})$  with  $A_n \rightarrow A$  and all sequences of acts  $(f_n)_{n \in N}$  with  $f_n \in A_n$  for each  $n \in N$ , if  $f_n \rightarrow f$ , then  $f \in c(A)$ .

Chernoff (A1) and Aizerman (A2) are standard axioms in the choice-theoretical literature. The former is just Sen's axiom  $\alpha$ , and requires no further comment; the latter says that removing an alternative that does not belong to the choice set cannot lead to new alternatives being chosen. The conjunction of the two is weaker than the Weak Axiom of Revealed Preference, and equivalent to the pseudo-rationalisability of  $c$  (representation (6)); the reader is referred to Moulin (1985), from whom we borrow the nomenclature, for further discussion. Fixed utilities (A4) imposes Sen's axiom  $\beta$  on the restriction of  $c$  to menus containing only constant acts. It follows from standard choice theory results (see for example Sen (1971)) that the restriction to constant acts is represented by a single transitive complete preference relation. This axiom translates the assumption that the decision maker's preferences over constant acts are the same in all envisaged future contingencies; he only anticipates differences in beliefs. Non-degeneracy (A3) requires that the choice function is non-trivial on constant acts.

The remaining axioms can be thought of as choice-theoretical analogues of the Gilboa-Schmeidler axioms on preferences (1989). The C-independence axioms (A5 and A6) correspond to Gilboa and Schmeidler's C-independence. Recall that the idea behind their axiom is that mixing with a constant act does not 'change' the preference order. Similarly here, Set C-independence (A5) states that mixing a menu with a constant act does not 'change' the choice set: if the decision maker wanted to keep an act as an open option from a given menu, he would like to keep the mixture of the act as an open alternative from a mixture of the menu. A consequence of the Gilboa-Schmeidler C-independence axiom (in conjunction with other basic preference axioms) is that, for any act  $f$  and constant act  $d$ , whichever of the acts is weakly preferred between  $f$  and  $d$  remains weakly preferred over any mix  $f_\alpha d$  of the two acts. Union C-independence states the equivalent of this for menus: if the decision maker would keep open an act  $f$  and a constant act  $d$  from a menu, then adding  $f_\alpha d$ , or indeed any mixture of  $d$  with an element of the menu, does not 'change' his decision to keep  $f$  and  $d$  open. This translates the idea that if  $f$  or  $d$  will be possibly chosen in some future contingency when both are available, then  $f_\alpha d$  will not be chosen over it. The need for two C-independence axioms is due to the weak choice-theoretical properties that are assumed: if Chernoff (A1) and Aizerman (A2) are replaced by the Weak Axiom of

Revealed Preference, it is straightforward to check that Set C-independence implies Union C-independence.

Monotonicity (A7) is essentially the standard monotonicity or statewise dominance axiom formulated for choice sets. It includes both a weak dominance and a strict dominance clause, both of which are fairly intuitive. The first basically says that adding elements to a menu that are weakly dominated by some element already on the menu does not lead one to rule out any of the options that one initially left open. This translates the standard intuition that adding a weakly dominated option should not prevent a previously chosen option from being chosen.<sup>21</sup> The second clause just says that one does not leave strictly dominated acts open: this translates the intuition that strictly dominated options are never chosen. These conditions are in the spirit of the standard monotonicity condition.

Uncertainty Aversion (A8) can be thought of as a weakening of the standard Gilboa-Schmeidler axiom extended to the general menu setting. The standard axiom, formulated on preferences, states that if there is indifference between a pair of acts, then any mixture is weakly preferred to both. A natural extension to the case of general menus is obtained by replacing the pair of indifferent acts by a subset of the menu that is contained in its choice set, and the mixture by any mixture of the elements in the subset. That is, it states that, for all  $A, B \in \wp(\mathcal{A})$  with  $B \subseteq c(A)$ ,  $f \in c(A \cup \{f\})$  for all  $f \in \text{conv}(B)$ . The axiom A8 is evidently a weakening of this extension, stating that it holds under particular conditions. In fact, the interpretation of the extension requires considering acts in the choice set to be indifferent; however, whilst this is the case under WARP, it is no longer true under the weaker choice-theoretic axioms used here. The conditions in A8 guarantee that there is an ex post preference relation according to which the acts in  $B$  are indifferent.<sup>22</sup> So the axiom

<sup>21</sup>Note that the proposed interpretation of dominance is vindicated by the Fixed utilities axiom (A4).

<sup>22</sup>This can be seen as follows. Whenever a decision maker chooses to leave open the acts  $f, g$ , the constant act  $d$ , and mixtures  $f_\alpha d$  and  $g_\alpha d$  from a menu, this means that he envisages contingencies where each of these acts is optimal. Moreover, since, for any ex post maxmin EU preference,  $f_\alpha d$  is optimal from a menu containing  $f$  and  $d$  only if  $f$  and  $d$  are indifferent (this is a consequence of the Gilboa-Schmeidler axioms), there is a possible contingency in which  $f, d$  and  $f_\alpha d$  are all optimal. Similarly, there is a possible contingency in which  $g, d$  and  $g_\alpha d$  are all optimal. Finally, if, for any act dominating  $f_\beta d$ , adding such an act causes the mixtures  $g_\alpha d$  to no longer be left open, this implies that any act dominating  $f_\beta d$  is also strictly preferred to  $g, d$  and  $g_\alpha d$  according to the contingencies where they were all indifferent; so  $f_\beta d$ , and hence  $f$  and  $d$  are indifferent to  $g, d$  and  $g_\alpha d$  under some contingency. Extending this reasoning to more than two acts, the conditions in A8 imply that the decision maker considers it possible that he will be indifferent among the acts

can be understood as stating that if the decision maker anticipates that he will be indifferent between the acts in  $B$ , then he anticipates that he will be willing to choose any mixture – since mixtures may hedge the ambiguity in the acts – and so he leaves these mixtures open. As such, A8 captures the hedging intuition in the standard axiom, in the context of the specific interpretation of the choice function employed here.

The final axiom, A9, is the standard upper hemicontinuity property for choice functions. It states that whenever a sequence of menus tends to a menu, then the limit of any sequence of options left open in these menus is left open in the limit menu.

A foundation for the notion of epistemic contingency is given in the following representation theorem.

**Theorem 1.** *The following are equivalent:*

- (i)  $c$  satisfies A1–A9;
- (ii) *There exists a nonconstant continuous affine utility function  $u : \Delta(X) \rightarrow \Re$  and a set  $\mathcal{P}$  of closed convex sets of probability measures on  $S$  such that:*

$$(5) \quad c(A) = \{f \in A \mid f \in \arg \max_{g \in A} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) \text{ for some } \mathcal{C}' \in \mathcal{P}\}$$

*Moreover,  $u$  is unique up to positive affine transformation, and there is a unique minimal  $\mathcal{P}$ .*

This theorem shows that, under certain conditions, a set of sets of priors – or a set of epistemic contingencies – representing the choice function according to (5) exists. Moreover, there is a unique ‘canonical’ such set, namely the unique minimal set. We conclude that the notion of epistemic contingency does have solid behavioral foundations.

*Remark 8.* Note that the results in this section can be combined with those in Section 3 to provide a choice-theoretic axiomatization of a representation in terms of maxmin expected utility ex post preferences and an aggregator that yields complete ex ante preferences. One potential interest for this development would be to translate the discussion in Section 3 into purely choice-theoretic terms. One possibility would be to express the Dynamic Consistency condition purely as a relationship between two choice functions – the choice function represented according to (5) and interpreted as

---

in  $B$ .

discussed above, and a choice function represented according to the [Gilboa and Schmeidler \(1989\)](#) representation (translated into choice-function language) and interpreted in the standard way. With this, one would obtain a choice-theoretic formulation of the notion of dynamic consistency proposed in [Section 3](#), as well as a representation result for maxmin EU ex post preferences with an aggregator of the sort given in [Proposition 2](#).<sup>23</sup>

Since these formulations do not seem particularly illuminating, we do not present them here.

## 7 Conclusion

It is commonly held that dynamic consistency, consequentialism and non-neutral attitudes to ambiguity are incompatible. We have argued in this paper that this is not the case, if the dynamic consistency condition is properly formulated. The central idea is that one can only ask a decision maker to be dynamically consistent with respect to the contingencies that he in fact envisages – rather than those imposed by the theorist – and when these contingencies are properly taken into account, the apparent incompatibility is resolved.

To fully defend this rebuttal of the standard argument, we have given choice-theoretic foundations to the central notion of a contingency that the decision maker envisages, on the basis of the choices the decision maker makes to ‘leave options open’ for subsequent choice at a future time. The representation result can be considered as an independent contribution to the literature on preferences over menus and unforeseen contingencies, and is perhaps the first result of the sort to consider incomplete preferences over menus.

Perhaps even more significant than the defense of this perspective and its ability to resist the arguments against ambiguity models are its conceptual and economic consequences. First of all, it emphasizes that dynamic consistency is fundamentally concerned with the relationship between two sorts of belief: one’s current beliefs about the state of the world and one’s beliefs about one’s possible future beliefs about the state of the world.

Secondly, the perspective provides a principled justification for the use of a restricted family of sets of priors in applications to dynamic choice problems. Applications typically adopt the tacit assumption that the decision maker knows what the decision tree is and that it is the way the theorist describes it. It turns out that Dynamic Consistency, in the refined

---

<sup>23</sup>Along similar lines, replacing the Dynamic Consistency axiom by a choice-theoretic equivalent of the Expert Uncertainty Aversion axiom in [Crès et al. \(2011\)](#), one obtains a representation where the decision maker has maxmin ex post preferences and aggregates them using a maxmin functional.

sense introduced here, implies that this assumption can only hold if the decision maker's ex ante beliefs are of a specific form. A decision maker whose beliefs are not of this form may be perfectly dynamically consistent: he just will not consider himself to be facing the decision tree that the modeler is using. That, of course, is not a problem for the decision maker; it is a problem for the modeler.

Finally, the perspective reveals that the issue of information aversion has been wrongly analyzed in the standard arguments. Non-expected utility decision makers need not be information averse, and will in fact not be whenever the information on offer does not compromise information they had previously expected to receive. They may turn down an offer of 'free' information, but this only occurs when its opportunity cost – in terms of information they had otherwise expected to receive but that is jeopardized by accepting the offer – outweighs its value.

## Appendix A Incorporating the epistemic contingencies into the state space

It was claimed in Remark 2 that the analysis conducted in this paper goes through in other setups; in this Appendix, we illustrate this point by considering a framework where the epistemic contingencies (ie. future states of belief) are explicitly represented in the state space.

We adopt the terminology and assumptions in Section 2; recall that  $S$  is the ‘objective’ state space. Let  $\mathcal{E}$  be the set of closed convex subsets of  $\Delta(\Sigma)$ ; this is the set of sets of priors over  $S$ , and will represent the potential future states of belief (or epistemic contingencies) of the decision maker. The ‘extended state space’, incorporating the epistemic contingencies, is  $\Omega = S \times \mathcal{E}$ . The first ‘coordinate’  $s'$  of a state  $(s', \mathcal{C}') \in \Omega$  represents the state of the world that is realized; the second ‘coordinate’  $\mathcal{C}'$  is the decision maker’s future state of belief. Note that this state space is rich enough to represent the arrival of any piece of information: the information is represented by the ex post, or ‘conditional’, beliefs it induces. Although  $\Omega$  is rich enough to represent all ex ante uncertainty, both about the state of the world and the ex post state of belief,  $S$  is sufficient to represent all payoff-relevant uncertainty. As such, the domain of (ex ante and ex post) preferences is still the set  $\mathcal{A}$  defined in Section 2, which corresponds naturally to a subset of  $\Delta(X)^\Omega$ . Similar points hold for the sets of priors involved in the representation of preferences, which are defined over  $S$ , considered as the partition  $\{\{s\} \times \mathcal{E} \mid s \in S\}$  of  $\Omega$ .

In the setup used above (Section 2.1), an ex ante set of priors  $\mathcal{C}$  and a set of epistemic contingencies (ex post sets of priors)  $\{\mathcal{C}_i\}_{i \in I}$  were assumed. As noted, the former corresponds to a set of priors over the partition  $\{\{s\} \times \mathcal{E} \mid s \in S\}$  of  $\Omega$ . The set of epistemic contingencies corresponds in this setup to a subset of  $\Omega$ , namely the set  $\{(s', \mathcal{C}') \in \Omega \mid \exists i \in I, \mathcal{C}' = \mathcal{C}_i\}$ . We call this set  $EC$ .<sup>24</sup>

Given the lack of a preference relation over  $\Delta(X)^\Omega$ , some remarks are in order about what should count as (the equivalent of a) ‘null event’ in  $\Omega$ . On the one hand, states not belonging to  $EC$  can be thought of as ‘null’ insofar as the decision maker does not consider

---

<sup>24</sup>Whilst we do not assume a preference relation on  $\Delta(X)^\Omega$  or a set of priors over  $\Omega$ , the analysis undertaken below continues to hold under such an assumption, as long as the preference relation or set of priors is appropriately consistent with the ex ante set of priors ( $\mathcal{C}$ ) and the set of epistemic contingencies ( $EC$ ).

it possible for them to hold (he does not consider it possible for him to be in the future states of belief coded in these states). Moreover, any state  $(s', C')$  such that  $s'$  is null according to the ex ante preference relation  $\leq$  can naturally be thought of 'null'. We will say that  $EC \cap NN$  is the set of  $\Omega$ -non-null states, where  $NN = \{(s', C') \in \Omega \mid s' \leq_{ea} \text{-non-null}\}$ . Any event with non-empty intersect with  $EC \cap NN$  will be said to be  $\Omega$ -non-null.

Note that, given the interpretation of the extended state space, it does not make sense to consider learning or conditioning upon certain events of  $\Omega$ . To characterize the events that could conceivably be learnt, we introduce the following definition.

**Definition 5.** An event  $A \subseteq \Omega$  is *learnable* if it satisfies the following two conditions:

1.  $C' = C''$  for all  $(s', C'), (s'', C'') \in A$ .
2. If  $(s', C') \in A$ , then  $(s'', C') \in A$ , for every  $s''$  for which  $p(s'') > 0$  for some  $p \in C'$ .

Moreover, a partition of  $\Omega$  is *learnable* whenever it consists entirely of learnable events.

Learnable events are those which can conceivably be learnt; this motivates the conditions in the definition.<sup>25</sup> The first condition corresponds to the assumption that, in the ex post stage, the agent has no uncertainty about his ex post beliefs; these are represented by a single set of priors. So the only events in  $\Omega$  that can be learnt are those which have the same ex post beliefs for every state they contain. The second condition reflects the fact that the set of states of the world that the decision maker considers to be non-null in the ex post stage must respect the set of beliefs he has at this stage. In particular, a state  $s' \in S$  cannot be ruled out by the information  $A$  (ie. there is no state  $(s', C') \in A$ ) but nevertheless possibly have non-zero probability according to the ex post beliefs established by  $A$ .

Note that, for any learnable event  $A$ , the preferences conditional on  $A$ ,  $\leq_A$ , are generated by the set  $C^*$  such that  $C' = C^*$  for all  $(s', C') \in A$ .

We claimed that, with these constraints in mind, this framework yields the same analysis as that carried out above. One way to back up this claim is to show that the standard definition of dynamic consistency (SDC) applied on this framework is equivalent to the refined notion of dynamic consistency proposed in Section 3 (DC). Translated into this framework, SDC becomes:

---

<sup>25</sup>Recall that we are working in a simple setup with two time periods. So we are considering the learning of an event between the ex ante and ex post stage.

**Standard Dynamic Consistency on  $\Omega$ .** For every  $f, g \in \mathcal{A}$  and learnable partition  $\{A_j\}_{j \in J}$  of  $\Omega$ , if  $f \leq_{A_j} g$  for every  $\Omega$ -non-null  $A_j$ , then  $f \leq g$ , and moreover, if any of the  $\leq_{A_j}$  orderings are strict, then so is the  $\leq$  one.

**Proposition A.1.** *Standard Dynamic Consistency on  $\Omega$  is equivalent to DC.*

Note that this result can be thought of giving another argument that DC is the appropriate equivalent to SDC for the case of epistemic contingencies.

## Appendix B Proofs

Throughout the Appendix,  $B$  will denote the space of real-valued functions on  $S$ , and  $ba(S)$  will denote the set of real-valued set functions on  $S$ , both under the Euclidean topology.  $B$  is equipped with the standard order:  $a \leq b$  iff  $a(s) \leq b(s)$  for all  $s \in S$ . For  $x \in \mathfrak{R}$ ,  $x^*$  is the constant function taking value  $x$ ; we use  $\mathfrak{R}^*$  to denote the set of constant functions. Addition with acts and positive scalar multiplication is extended to sets as standard: for  $A \subseteq B$ ,  $a \in B$ ,  $\alpha > 0$ ,  $\alpha A = \{\alpha b \mid b \in A\}$  and  $A + a = \{b + a \mid b \in A\}$ . Finally, for any subset  $A \subseteq B$ , we shall use  $co(A)$  to denote the convex closure of  $A$ .

### B.1 Proofs of Propositions

For the proofs of Propositions 1–4, let  $J : \mathcal{A} \rightarrow \mathfrak{R}$  be the maxmin EU functional represented by  $\mathcal{C}$  (ie.  $J(f) = \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s)$  for all  $f \in \mathcal{A}$ ), and likewise for  $J_i$  and  $\mathcal{C}_i$ .  $Ra(J_I)$  is the range of the vector  $(J_i)_{i \in I}$  over  $\mathcal{A}$ , and  $Ra(J_i)$  the range of the function  $J_i$  for each  $i$ . Note that, since the utility functions are the same for all  $\leq_i$ ,  $Ra(J_i) = Ra(J_j)$  for all  $i, j$ ; call this set  $R$ .  $cone(Ra(J_I))$  is the cone spanned by  $Ra(J_I)$ . We use  $\mathbf{e}$  to denote the unit vector in  $\mathfrak{R}^I$ .

*Proofs of Proposition 1 and Proposition 2.* Proposition 1 and the ‘if’ direction Proposition 2 are straightforward, and their proofs are omitted.

The proof of the ‘only if’ direction of Proposition 2 draws on the developments in Crès et al. (2011). By their Lemmas 1–4, there exists a constant additive, positively homogeneous, monotonic real-valued function  $\phi$  on  $cone(Ra(J_I))$  such that  $J(f) = \phi((J_i(f))_{i \in I})$  for all  $f \in \mathcal{A}$ . Cerreia-Vioglio et al. (2013, Theorem 1) show that the real-valued function  $\bar{\phi}$  on  $\mathfrak{R}^I$ , defined by  $\hat{\phi}(\mathbf{y}) = \sup\{\phi(\mathbf{x}) + b \mid \mathbf{x} \in Ra(J_I), b \in \mathfrak{R}, \mathbf{x} + b\mathbf{e} \leq \mathbf{y}\}$  for all  $\mathbf{y} \in \mathfrak{R}^I$ ,

extends  $\phi$  and is constant additive and monotonic. It is clear from the definition and the positive homogeneity of  $\phi$  that  $\hat{\phi}$  is positively homogeneous. Finally, strong monotonicity on  $Ra(J_I)$  is a direct consequence of DC. □

*Proof of Proposition 3.* The right to left direction is essentially Proposition 1. As for the left to right direction, by Proposition 2,  $\phi$  is positively homogeneous, constant additive and monotonic. Applying (Hill, 2012, Lemmas 4 and 7), it follows that  $\phi$  is concave, so it satisfies all the properties required for the Gilboa and Schmeidler (1989) argument. Applying this reasoning (see also Crès et al. (2011)), we obtain a unique closed convex  $\Lambda$  representing  $\phi$  as described in the Proposition. Finally, since, by (Hill, 2012, Lemma 4),  $Ra(J_i)$  is fully dimensional, it follows from the strong monotonicity of  $\phi$  on  $Ra(J_I)$  and its positive homogeneity that  $\phi$  is strongly monotonic, so every  $p \in \mathcal{D}$  is such that  $p(i) > 0$  for all  $i \in I$ . □

*Proof of Proposition 4.* As shown by Crès et al. (2011, Lemmas 1–4), DC implies that  $J(f) = \phi((J_i(f))_{i \in I})$  where  $\phi$  is a constant additive, positively homogeneous, monotonic real-valued function on  $Ra(J_I)$ . Consider any vector  $\mathbf{x} \in R^I$ . For each  $i \in I$ , since  $x_i \in R = Ra(J_i)$ , there exists a constant act  $c_i^{\mathbf{x}} \in \Delta(X)$  with  $J_i(c_i^{\mathbf{x}}) = x_i$ . Hence, defining  $g^{\mathbf{x}} \in \mathcal{A}$  by  $g(s) = c_i^{\mathbf{x}}(s)$  whenever  $s \in A_i$ , it follows from the fact that  $\{\leq_i\}_{i \in I}$  is  $\mathcal{P}$ -ontic that  $((J_i(g^{\mathbf{x}}))_{i \in I}) = ((J_i(c_i^{\mathbf{x}}))_{i \in I}) = \mathbf{x}$ . It follows that  $R^I \subseteq Ra(J_I)$ ; by the definition of  $Ra(J_I)$ , this inclusion is in fact an equality. Now consider any pair of vectors  $\mathbf{x}, \mathbf{y} \in R^I$ . By definition,  $J_i(\alpha g^{\mathbf{x}} + (1 - \alpha)g^{\mathbf{y}}) = J_i(\alpha c_i^{\mathbf{x}} + (1 - \alpha)c_i^{\mathbf{y}}) = \alpha J_i(c_i^{\mathbf{x}}) + (1 - \alpha)J_i(c_i^{\mathbf{y}}) = \alpha J_i(g^{\mathbf{x}}) + (1 - \alpha)J_i(g^{\mathbf{y}}) = \alpha x_i + (1 - \alpha)y_i$ , where the middle equality holds by the definition of the maxmin EU functional. Hence  $\phi(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) = \phi((J_i(\alpha g^{\mathbf{x}} + (1 - \alpha)g^{\mathbf{y}}))_{i \in I}) = J(\alpha g^{\mathbf{x}} + (1 - \alpha)g^{\mathbf{y}}) \geq \alpha J(g^{\mathbf{x}}) + (1 - \alpha)J(g^{\mathbf{y}}) = \alpha \phi((J_i(g^{\mathbf{x}}))_{i \in I}) + (1 - \alpha)\phi((J_i(g^{\mathbf{y}}))_{i \in I}) = \alpha \phi(\mathbf{x}) + (1 - \alpha)\phi(\mathbf{y})$ , where the inequality in the middle holds because of the concavity of the maxmin EU functional  $J$ . Hence  $\phi$  is concave.

Using standard arguments,  $\phi$  can be extended to a monotonic, positively homogeneous, constant additive, concave function on  $\mathfrak{R}^I$ . Application of the argument in Gilboa and Schmeidler (1989) (see also (Crès et al., 2011, Lemma 8)) implies that there exists a closed convex  $\mathcal{D} \subseteq \Delta(I)$  such that  $\phi(\mathbf{x}) = \min_{p \in \mathcal{D}} \sum_{i \in I} p_i x_i$ . The strict positivity of the elements

in  $\mathcal{D}$  follows directly from DC and the fact that  $Ra(J_I) = R^n$ . The form of the set of priors  $\mathcal{C}$  representing  $J$  follows from Crès et al. (2011, Proposition 1).  $\square$

*Proof of Proposition 5.* By Proposition 2, DC implies that the aggregator  $\phi$  is monotonic. Consider any  $A \in \wp(\mathcal{A})$ , and let  $g = \arg \max_{f \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s) = \arg \max_{f \in A} \phi \left( \left( \min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s) \right)_{i \in I} \right)$ . For every  $i \in I$ , by definition,  $\max_{f \in A} \min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s) \geq \min_{p \in \mathcal{C}_i} \sum_{s \in S} u(g(s))p(s)$ ; it follows from the monotonicity of  $\phi$  that  $\phi \left( \left( \max_{f \in A} \min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s) \right)_{i \in I} \right) \geq \phi \left( \left( \min_{p \in \mathcal{C}_i} \sum_{s \in S} u(g(s))p(s) \right)_{i \in I} \right) = \max_{f \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s)$ , as required.  $\square$

*Proof of Proposition 6.* To show the right to left direction, suppose that  $\mathcal{C}_{(\mathcal{I})} \subseteq \mathcal{C}$ . Take any  $A \in \wp(\mathcal{A})$  let  $g = \arg \max_{f \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s)$ . By the definition of the maxmin EU rule (representation (1)), the containment of sets of priors implies that  $\min_{p \in \mathcal{C}_{(\mathcal{I})}} \sum_{s \in S} u(g(s))p(s) \geq \min_{p \in \mathcal{C}} \sum_{s \in S} u(g(s))p(s) = \max_{f \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s)$ ; however, it follows from Proposition 5 that  $\phi_{(\mathcal{I})} \left( \left( \max_{f \in A} \min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u(f(s))p(s) \right)_{k \in K} \right) \geq \min_{p \in \mathcal{C}_{(\mathcal{I})}} \sum_{s \in S} u(g(s))p(s)$ . Combining these two inequalities, one obtains that the value of information  $\mathcal{I}$  is always non-negative.

Now consider the other direction, and suppose that  $\mathcal{C}_{(\mathcal{I})} \not\subseteq \mathcal{C}$ ; we shall show that there exists  $A \in \wp(\mathcal{A})$  with  $\phi_{(\mathcal{I})} \left( \left( \max_{f \in A} \min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u(f(s))p(s) \right)_{k \in K} \right) < \max_{f \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s)$ . Since  $\mathcal{C}_{(\mathcal{I})} \not\subseteq \mathcal{C}$ , there exists  $p \in \Delta(\Sigma)$  with  $p \in \mathcal{C}_{(\mathcal{I})} \setminus \mathcal{C}$ . By a separation theorem (Aliprantis and Border, 2007, 5.80), there is a nonzero linear functional  $\phi$  on  $ba(S)$  and  $\alpha \in \mathfrak{R}$  such that  $\phi(p) \leq \alpha < \phi(q)$  for all  $q \in \mathcal{C}$ . Since  $S$  is finite (so  $B$  is finite-dimensional),  $B$  is reflexive, and, by the standard isomorphism between  $ba(S)$  and  $B^*$ , it follows that  $ba(S)^*$  is isometrically isomorphic to  $B$  (Dunford and Schwartz, 1958, IV.3); hence there is a real-valued function  $a \in B$  such that  $\phi(q) = \sum_{s \in S} a(s)p(s)$  for any  $q \in ba(S)$ . Without loss of generality  $\phi, a$  can be chosen so that  $a$  takes values in the range of  $u$ . Take  $g \in \mathcal{A}$  such that  $u \circ g = a$ , and consider the menu  $\{g\}$ .  $\phi_{(\mathcal{I})} \left( \left( \max_{f \in \{g\}} \min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u(f(s))p(s) \right)_{k \in K} \right) = \phi_{(\mathcal{I})} \left( \left( \min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u(g(s))p(s) \right)_{k \in K} \right) = \min_{p \in \mathcal{C}_{(\mathcal{I})}} \sum_{s \in S} u(g(s))p(s)$ . However, by the definition of  $g$ ,  $\min_{p \in \mathcal{C}_{(\mathcal{I})}} \sum_{s \in S} u(g(s))p(s) <$

$$\min_{p \in \mathcal{C}} \sum_{s \in S} u(g(s))p(s) = \max_{f \in \{g\}} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s). \quad \text{Hence}$$

$$\phi_{(\mathcal{I})} \left( \left( \max_{f \in \{g\}} \min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u(f(s))p(s) \right)_{k \in K} \right) < \max_{f \in \{g\}} \min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s),$$
 as required. □

*Proof of Proposition A.1.* By the definition of learnable events, for each  $A_j$ , there exists  $\mathcal{C}_j$  such that  $\leq_{A_j}$  is represented by  $\mathcal{C}_j$  according to (1). By the definition of  $\Omega$ -nullness, for any  $\Omega$ -non-null  $A_j$ ,  $\mathcal{C}_j \in \{\mathcal{C}_i\}_{i \in I}$  (it is one of the epistemic contingencies). Moreover, for any  $A_{j_1}, A_{j_2} \in \{A_j\}_{j \in J}$ ,  $\mathcal{C}_{j_1} \neq \mathcal{C}_{j_2}$ : if not, then for any state  $s$  such that  $p(s) > 0$  for some  $p \in \mathcal{C}_j = \mathcal{C}_{j_1} = \mathcal{C}_{j_2}$ ,  $(s, \mathcal{C}_j) \in A_{j_1} \cap A_{j_2}$ , contradicting the disjointness of the elements of the partition. So, for any learnable partition  $\{A_j\}_{j \in J}$ , there is an injective map from the non-null elements of  $\{A_j\}_{j \in J}$  to the set of epistemic contingencies  $\{\mathcal{C}_i\}_{i \in I}$ . Moreover, this map is surjective: for any epistemic contingency  $\mathcal{C}_i$ , there is a  $\Omega$ -non-null state  $(s, \mathcal{C}_i)$ , and hence there is a cell  $A_{j'}$  of the partition such that the associated set of priors  $\mathcal{C}_{j'} = \mathcal{C}_i$ . So, for any learnable partition, there is a bijection between the cells of this partition and the set of epistemic contingencies; moreover, this bijection is such that, for each cell of the partition, the preferences conditional on this cell coincide with the epistemic contingency related to it by the bijection. The equivalence between the two conditions follows immediately. □

## B.2 Proof of Theorem 1

We first consider the direction (i) to (ii). The proof proceeds as follows. Firstly, using monotonicity (A7) we show that  $c$  generates a choice function on a subset of the set of real-valued functions on  $S$ , which can be extended to a choice function on the whole set satisfying Chernoff, Aizerman, upper hemicontinuity and choice-function versions of constant linearity, superadditivity, monotonicity and union C-independence (Lemmas B.1 and B.2). The most important step in the proof is Proposition B.1, which shows that, for any set  $\hat{A}$  and any  $\hat{a} \in c(\hat{A})$ , there exists a closed convex set of finitely additive probability measures such that the maxmin expectation represents  $c$  in the sense that any maximal element over  $A'$  according to maxmin expectation with this set of priors is in  $c(A')$ , and such that  $\hat{a}$  maximises the maxmin expectation on  $\hat{A}$ . Taking the union of such sets of probability measures over all pairs  $a, A$  with  $a \in c(A)$  yields the required set  $\mathcal{P}$ .

### B.2.1 Preliminary Lemmas

By [A1](#) and [A4](#) and standard choice theory results (for example [Sen \(1971\)](#)), the restriction of  $c$  to sets containing only constant acts is represented in the standard way by a reflexive transitive complete order. By [A3](#), [A5](#) and [A9](#) and the Herstein-Milnor theorem, this order (and hence the restriction of the choice function) is represented by a non-degenerate affine utility function  $u$ ; by [A9](#),  $u$  is continuous. Let  $K = u(\Delta(X))$  and  $B(K)$  be the set of functions in  $B$  taking values in  $K$ . Without loss of generality, it can be assumed that  $0 \in K$ , and it is not on its boundary.

**Lemma B.1.** *For  $A, B \subseteq \mathcal{A}$ , suppose that there is a bijection  $\sigma : A \rightarrow B$  such that, for all  $f \in A$ ,  $c(\{f(s), \sigma(f)(s)\}) = \{f(s), \sigma(f)(s)\}$  for all  $s \in S$ . Then  $c(B) = \sigma(c(A))$ .*

*Proof.* Let  $A$  and  $B$  satisfy the properties specified, and let  $f \in c(A)$ . By applying [A7](#) on the sets  $A$  and  $A \cup B$ , we have that  $c(A) \subseteq c(A \cup B)$ . We distinguish two cases. If there exists  $d \in \Delta(X)$  with  $d \notin c(\{f(s), d\})$  for all  $s \in S$ , then consider  $c((A \cup B)_\alpha d \cup \{\sigma(f)\})$  for  $\alpha \in (0, 1)$ . By [A7](#),  $f_\alpha d \notin c((A \cup B)_\alpha d \cup \{\sigma(f)\})$ , whence  $c(A)_\alpha d \notin c((A \cup B)_\alpha d \cup \{\sigma(f)\})$ ; it thus follows from [A2](#) and [A5](#) that  $\sigma(f) \in c((A \cup B)_\alpha d \cup \{\sigma(f)\})$ . By [A9](#) and the fact that, as  $\alpha \rightarrow 1$ ,  $(A \cup B)_\alpha d \cup \{\sigma(f)\} \rightarrow A \cup B$ ,  $\sigma(f) \in c(A \cup B)$ . By [A1](#),  $\sigma(f) \in c(B)$  as required. If there is no  $d \in \Delta(X)$  with  $d \notin c(\{f(s), d\})$  for all  $s \in S$ , then take any  $e \in \Delta(X)$  with  $u(e)$  non-minimal in  $K$  and  $\beta \in (0, 1)$ . There exists  $d \in \Delta(X)$  with  $d \notin c(\{f_\beta e(s), d\})$  for all  $s \in S$ ; applying the previous argument to  $((A \cup B)_\beta e)_\alpha d \cup \{(\sigma(f))_\beta e\}$  yields the conclusion that  $(\sigma(f))_\beta e \in c(B_\beta e)$ . It follows by [A5](#) that  $\sigma(f) \in c(B)$  as required. By a similar argument on  $g \in c(B)$ , the result is obtained.  $\square$

There is thus a many-to-one mapping between acts in  $\mathcal{A}$  and elements of  $B(K)$ , given by  $a = u \circ f$ , for  $f \in \mathcal{A}$ . Define the choice function  $c_{B(K)}$ , on  $B(K)$ , as follows: for  $A \in \wp(B(K))$ , and  $A' \in \wp(\mathcal{A})$  such that  $A = u \circ A'$ ,  $c_{B(K)}(A) = u \circ c(A')$ . By [Lemma B.1](#),  $c_{B(K)}$  is well-defined. Let  $\wp_{bdd}(B)$  be the set of closed bounded subsets of  $B$ .

**Lemma B.2.** *There exists a choice function  $c : \wp_{bdd}(B) \rightarrow \wp_{bdd}(B)$  on  $B$  such that:*

- i. for all  $A \in \wp(B(K))$ ,  $c(A) = c_{B(K)}(A)$
- ii.  $c$  satisfies Chernoff and Aizerman (that is, [A1](#) and [A2](#))<sup>26</sup>

<sup>26</sup>Henceforth, we shall refer to these properties by these names.

- iii.  $c$  is constant linear: for all  $A \in \wp_{bdd}(B)$ ,  $\alpha > 0$  and  $x \in \mathfrak{R}$ ,  $c(\alpha A + x^*) = \alpha c(A) + x^*$
- iv.  $c$  is constant independent: for all  $A \in \wp_{bdd}(B)$ ,  $\alpha \in (0, 1)$  and  $x \in \mathfrak{R}$  with  $x^* \in c(A)$ ,  $c(A) \subseteq c(A \cup (\alpha A + (1 - \alpha)x^*))$
- v.  $c$  is monotonic: for all  $A, B \in \wp_{bdd}(B)$  with  $A \subseteq B$ , if for each  $b \in B$ , there exists  $a \in A$  with  $a \geq b$ , then  $c(A) \subseteq c(B)$ , and, for any  $b \in B$ , if there exists  $a \in B$  with  $a > b$ , then  $b \notin c(B)$ .
- vi.  $c$  is superadditive: for all  $A \in \wp_{bdd}(B)$ ,  $A' \subseteq c(A)$ ,  $a' \in c(A \cup \{a'\})$  for all  $a' \in co(A')$  whenever there exist  $x \in \mathfrak{R}$  and  $b \in A'$  such that: i.  $\alpha a + (1 - \alpha)x^* \in A'$  for all  $a \in A'$ ,  $\alpha \in (0, 1)$ ; and ii.  $\alpha b + (1 - \alpha)x^* \notin c(A' \cup \{a + y^*\})$  for all  $a \in A'$ ,  $y > 0$  and  $\alpha \in (0, 1)$ .
- vii.  $c$  is upper hemicontinuous: for all sequences  $(A_n)_{n \in \mathbb{N}}$ ,  $A_n \in \wp_{bdd}(B)$  and  $A \in \wp_{bdd}(B)$  with  $A_n \rightarrow A$  and for all sequences  $(a_n)_{n \in \mathbb{N}}$ ,  $a_n \in B$ , with  $a_n \in A_n$  for all  $n \in \mathbb{N}$ , if  $a_n \rightarrow a$ , then  $a \in c(A)$ .

*Proof.* Define  $c$  on  $B(K)$  by clause i. Note that, applying A5 on the inverse image of  $A$  and the inverse image of  $0^*$ , we have that, for any  $A \in \wp(B(K))$  and  $\alpha \in (0, 1)$ ,  $c(\alpha A) = \alpha c(A)$ . It follows that, for any  $A \in \wp(B(K))$  with  $\alpha A \in \wp(B(K))$  where  $\alpha > 1$ ,  $c(\alpha A) = \alpha c(A)$ .  $c$  can thus be coherently extended to  $\wp_{bdd}(B)$  as follows: for  $A \in \wp_{bdd}(B)$ ,  $c(A) = \frac{1}{\alpha}c(\alpha A)$ , where  $\alpha > 0$  is such that  $\alpha A \in \wp(B(K))$ . Note that  $c$  is positively homogeneous and satisfies the choice properties (point ii) by A1 and A2. Moreover, it is constant additive: applying A5 to the inverse image of  $2A$  and  $2x^*$  (or appropriate products with a sufficiently small  $\alpha$ ), we have that  $c(A + x^*) = c(\frac{1}{2}(2A) + \frac{1}{2}(2x^*)) = \frac{1}{2}c(2A) + \frac{1}{2}c(2x^*) = c(A) + x^*$ , as required. So  $c$  is constant linear. The remaining properties are a direct consequence of axioms A6–A9 (multiplying by a sufficiently small  $\alpha$  where appropriate).

□

### B.2.2 Proposition B.1

As stated above, the following proposition is the central part of the proof.

**Proposition B.1.** *Let  $c : \wp_{bdd}(B) \rightarrow \wp_{bdd}(B)$  be a choice function satisfying the properties in Lemma B.2, and suppose that, for some  $\hat{a} \in \hat{A} \in \wp_{bdd}(B)$ ,  $\hat{a} \in c(\hat{A})$ . Then there exists a closed convex set  $\mathcal{C}_{\hat{a}, \hat{A}}$  of finitely additive probability measures such that:*

(1) For all  $A \in \wp_{bdd}(B)$ ,

$$(9) \quad b \in \arg \max_{b' \in A} \min_{p \in \mathcal{C}_{\hat{a}, \hat{A}}} \sum_{s \in S} b'(s)p(s) \Rightarrow b \in c(A)$$

(2)  $\hat{a} \in \arg \max_{b' \in \hat{A}} \min_{p \in \mathcal{C}_{\hat{a}, \hat{A}}} \sum_{s \in S} b'(s)p(s)$

*Proof.* We consider the case where  $\hat{a}$  is not a constant act; the case where it is a constant act is treated similarly, by using the construction below with a non-constant act  $a$  such that  $\{\alpha a + (1 - \alpha)\hat{a} \mid \alpha \in [0, 1]\} = c(\{\alpha a + (1 - \alpha)\hat{a} \mid \alpha \in [0, 1]\})$ .

We begin with some notation. First, recall that  $B$  is (isomorphic to) Euclidean space, so for ease we may use Euclidean notation and intuition at points. In particular, let  $\|\cdot\|$  be the Euclidean norm. Let  $U_{\hat{a}} = \{b \in B \mid \inf_{x \in \mathfrak{R}} \|b - x^*\| \leq \inf_{x \in \mathfrak{R}} \|\hat{a} - x^*\|\}$ : this is the smallest ‘tube’ around the ray generated by the unit vector containing  $\hat{a}$ . By the assumption that  $\hat{a}$  is not constant, it is not a ray. Moreover, for any  $a \in B$ , let  $\bar{a} = \{\alpha a + x^* \mid \alpha \geq 0, x \in \mathfrak{R}\}$ ; this is the positive half-plane generated by  $a$  and the unit vector, with as boundary the ray generated by the unit vector. It is straightforward to check that the  $\{\bar{a} \setminus \mathfrak{R}^* \mid a \in B \setminus \mathfrak{R}^*\}$  form a partition of  $B \setminus \mathfrak{R}^*$ ; call the set of equivalence classes of the partition  $\mathcal{Q}$ .

For  $a \in B$ ,  $x \in \mathfrak{R}$ , we define  $[a, x^*] = \{\alpha a + (1 - \alpha)x^* \mid \alpha \in [0, 1]\}$ , and  $\Xi = \{[a, x^*] \mid a \in B, x \in \mathfrak{R}, c([a, x^*]) = [a, x^*]\}$ . Moreover, for any  $A \in \wp_{bdd}(B)$  and  $x \in \mathfrak{R}$ ,  $X_x^A = \bigcup_{a \in A} [a, x^*]$ . Note that, since  $A$  is closed and bounded, so is  $X_x^A$ .

Now consider the set  $\mathcal{Z}$  of pairs  $(O, I)$  where:

1.  $O$  is a non-empty convex closed subset of  $B$  such that:
  - a. for all  $a \in B$ , if  $a \in O$ , then  $\bar{a} \subseteq O$
  - b.  $\hat{a} \in O$
2.  $I : O \rightarrow \mathfrak{R}$  is a functional with the following properties:
  - a.  $I$  is monotonic: for all  $a, b \in B$ , if  $a \geq b$ ,  $I(a) \geq I(b)$
  - b.  $I$  is constant linear (ie. constant additive and positively homogeneous): for all  $a \in O$ ,  $x \in \mathfrak{R}$ ,  $\alpha > 0$ ,  $I(\alpha a + x^*) = \alpha I(a) + x$
  - c.  $I$  is superadditive: for any  $\alpha \in [0, 1]$ ,  $I(\alpha a + (1 - \alpha)b) \geq \alpha I(a) + (1 - \alpha)I(b)$

- d.  $I$  represents  $c$  on  $O$ : for all  $A \in \wp_{bdd}(O)$ ,  $\arg \max_A I \subseteq c(A)$
- e.  $\hat{a} \in \arg \max_{\hat{A} \cap O} I$
- f. for any  $A' \in \wp_{bdd}(O)$  with  $A' \subseteq \{a \in O \mid I(a) = I(\hat{a})\}$ ,  $A' \subseteq c(A' \cup X_{I(\hat{a})}^{\hat{A}})$
- g. for every  $A \in \wp_{bdd}(O)$  with  $(\{a \in O \mid I(a) = I(\hat{a})\} \cap U_{\hat{a}}) \cup X_{I(\hat{a})}^{\hat{A}} \subseteq A$ , and for every  $a' \in \{a \in O \mid I(a) = I(\hat{a})\} \cap U_{\hat{a}}$  and every  $z > 0$ ,  $\alpha \hat{a} + (1 - \alpha)(I(\hat{a}))^* \notin c(A \cup \{a' + z^*\})$  for every  $\alpha \in (0, 1)$ .

Note that, by property 2f, Chernoff, and the definition of  $U_{\hat{a}}$ ,  $[\hat{a}, I(\hat{a})^*] \subseteq \{a \in O \mid I(a) = I(\hat{a})\} \cap U_{\hat{a}}$ .

$\mathcal{Z}$  is equipped with the order  $\leq$ , defined as follows:  $(O_1, I_1) \leq (O_2, I_2)$  iff:

- $O_1 \subseteq O_2$
- $I_1 = I_2|_{O_1}$

This is evidently a partial order. In a pair of auxiliary lemmas (proved in Section B.3), we show that any element of this order not containing some  $b \in B$  can be extended to another element in  $\mathcal{Z}$  containing it (Lemma B.4) and that  $\mathcal{Z}$  is non-empty (Lemma B.12). Since, for each  $(O, I) \in \mathcal{Z}$ ,  $O$  is the convex hull of a set of half-planes, and  $B$  is finite dimensional, all chains of elements of  $\mathcal{Z}$  must be finite. Hence every chain has an upper bound, which is in fact its top element. Take any maximal chain, in the sense that there is no  $(O'', I'') \in \mathcal{Z}$  which is  $(O'', I'') > (O', I')$  for all  $(O', I')$  in the chain. For the top element of this chain,  $O = B$ ; if not, then there exists  $b \notin O$ , whence, by Lemma B.4, there is a  $(O', I') \in \mathcal{Z}$  with  $(O', I') > (O, I)$ . Hence there is an element  $(B, I) \in \mathcal{Z}$ , with  $I$  being a monotonic, constant linear, superadditive functional on  $B$  representing  $c$  (in the sense of property 2d) and satisfying property 2e.

By the result in Gilboa and Schmeidler (1989) (in particular Lemma 3.5 onwards), there exists a unique closed convex set of probability measures  $\mathcal{C}_{\hat{a}, \hat{A}}$  such that  $I(a) = \min_{c \in \mathcal{C}_{\hat{a}, \hat{A}}} \sum_{s \in S} a(s)p(s)$ . By the fact that  $I$  represents  $c$ , it follows that (9) holds for  $\mathcal{C}_{\hat{a}, \hat{A}}$ . Moreover, by property 2e,  $\hat{a} \in \arg \max_{\hat{A}} I$ , and so  $\hat{a} \in \arg \max_{b \in \hat{A}} \min_{p \in \mathcal{C}_{\hat{a}, \hat{A}}} \sum_{s \in S} b(s)p(s)$ , as required.

□

### B.2.3 Conclusion of the proof of Theorem 1

We conclude the direction (i) to (ii). Let  $\mathcal{P} = \{\mathcal{C}_{\hat{a}, \hat{A}} \mid \hat{A} \in \wp_{bdd}(B), \hat{a} \in c(\hat{A})\}$ , where the  $\mathcal{C}_{\hat{a}, \hat{A}}$  are as in Proposition B.1. By the definition of the choice function  $c$  on  $B$ , and clause 1 of Proposition B.1, for any  $\mathcal{C} \in \mathcal{P}$ , if  $f \in \arg \max_{g \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(g(s))p(s)$ , then  $f \in c(A)$ . Moreover, for any  $f \in \mathcal{A}$  and  $A \in \wp(\mathcal{A})$ , if  $f \in c(A)$ , then, by clause 2 of Proposition B.1,  $f \in \arg \max_{g \in A} \min_{p \in \mathcal{C}_{u \circ f, u \circ A}} \sum_{s \in S} u(g(s))p(s)$ . Hence  $\mathcal{P}$  and  $u$  represent  $c$  according to (5).

As concerns the necessity of the axioms (the (ii) to (i) direction), all axioms are evident or have been shown to be necessary elsewhere in the literature (see for example Moulin (1985) for the necessity of A1 and A2), except A8. To establish the necessity of this axiom, we first claim that the conditions of the axiom imply that there is a set  $\mathcal{C} \in \mathcal{P}$  such that  $B \subseteq \arg \max_{h \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(h(s))p(s)$ . Suppose that this is not the case and consider the act  $g \in B$  mentioned in the axiom. Take any set  $\mathcal{C}_1 \in \mathcal{P}$  such that  $g_\alpha d \in \arg \max_{h \in A} \min_{p \in \mathcal{C}_1} \sum_{s \in S} u(h(s))p(s)$  for some  $\alpha \in (0, 1)$ . By the reductio assumption, there must be an act  $f \in B$  such that  $f \notin \arg \max_{h \in A} \min_{p \in \mathcal{C}_1} \sum_{s \in S} u(h(s))p(s)$ ; take such an  $f$ . By the facts just established, and the fact that  $d \in A$ , we have that  $\min_{p \in \mathcal{C}_1} \sum_{s \in S} u(g_\alpha d(s))p(s) \geq u(d)$  and  $\min_{p \in \mathcal{C}_1} \sum_{s \in S} u(g_\alpha d(s))p(s) > \min_{p \in \mathcal{C}_1} \sum_{s \in S} u(f(s))p(s)$ . It thus follows that there exists  $e > d$  and  $\beta \in (0, 1)$  such that  $\min_{p \in \mathcal{C}_1} \sum_{s \in S} u(g_\alpha d(s))p(s) \geq \min_{p \in \mathcal{C}_1} \sum_{s \in S} u(f_\beta e(s))p(s)$ , and hence, by the representation (5), that  $g_\alpha d \in c(A \cup \{f_\beta e\})$ , contradicting A8. Hence there is a set  $\mathcal{C} \in \mathcal{P}$  such that  $B \subseteq \arg \max_{h \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(h(s))p(s)$ , as required. It follows by the concavity of the maxmin expected utility representation that, for any  $f \in \text{conv}(B)$ ,  $\min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s) \geq \max_{h \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(h(s))p(s)$ , and so by the representation (5), it follows that  $f \in c(A \cup \{f\})$  for all such  $f$ , as required.

The uniqueness of  $u$  follows from the standard von Neuman-Morgenstern result. We now show the existence of a unique minimal  $\mathcal{P}$  representing  $c$ . Let  $\{\mathcal{P}_m\}_{m \in M}$  be the sets of sets of priors representing  $c$  according to (5) and let  $\mathcal{P} = \bigcup_{m \in M} \mathcal{P}_m$ . Let  $I$  index  $\mathcal{P}$ . Note that  $\mathcal{P}$  also represents  $c$  according to (5). Pick any  $d \in \Delta(X)$  and for each  $\mathcal{C}_i \in \mathcal{P}$ , let  $A_i = \{f \in \mathcal{A} \mid \min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s) = u(d)\}$ . Since  $\mathcal{P}$  represents  $c$  according to (5),  $c(A_i) = A_i$  for all  $A_i$ . Define the following order on  $\{A_i \mid i \in I\}$ : for all  $i, j \in I$ ,  $A_i \sqsupseteq A_j$  iff there exists  $f \in A_i$ ,  $e \in \Delta(X)$  with  $e < d$  and  $\alpha \in (0, 1)$  such that  $f_\alpha e \in A_j$ . Note that, by the definition of  $A_i$ , and the fact that, for all  $i, j \in I$ ,  $A_i \neq A_j$ ,  $\sqsupseteq$  is complete, and the

strict relation is transitive. We say that a set  $A_i$  is *essential* if there exists no set  $J \subset I \setminus \{i\}$  such that  $A_i \not\supseteq A_j$  for all  $j \in J$  and  $\bigcup_{j \in J} A_j \supseteq A_i$ .

To get a preliminary understanding of these notions, note that, for any  $i \in I$ , if  $f \in A_i$  then  $f_\alpha d \in A_i$  for all  $\alpha \in [0, 1]$ . Moreover, it follows from the representation that, if  $A_i \supseteq A_j$ , then there exists  $f \in A_i$  such that  $f \succ_j d$ . Hence, by the properties of the maxmin EU representation, for all  $\alpha \in [0, 1)$ ,  $f_\alpha d \notin \{h \in A_i \mid h \succeq_j g, \forall g \in A_i\}$ . Moreover, once again by the properties of the EU maxmin representation, we have that, for any act  $f'$  with  $f'_\beta d \in A_i \cap A_j$  for all  $\beta \in [0, 1]$ ,  $f'_\beta d \notin \{h \in A_i \mid h \succeq_j g, \forall g \in A_i\}$  for all  $\beta \in [0, 1)$ . It follows that not only  $\{h \in A_i \mid h \succeq_j g, \forall g \in A_i\} \neq A_i$  for any  $A_j$  such that  $A_i \supseteq A_j$ , but also that  $\bigcup_{j \text{ s.t. } A_i \supseteq A_j} \{h \in A_i \mid h \succeq_j g, \forall g \in A_i\} \neq A_i$ . This motivates the definition of essential sets: they are those sets  $A_i$  for which the fact that  $A_i = c(A_i)$  cannot be ‘attained’ using a set of sets  $A_j$ . Uniqueness follows from Lemma B.3 below, which implies that there is a unique minimal subset of  $\mathcal{P}$  representing  $c$ ; it follows in particular that this is contained in every set of sets of priors  $\mathcal{P}_m$  representing  $c$ .

**Lemma B.3.**  $\{\mathcal{C}_i \mid i \in I, A_i \text{ essential}\}$  represents  $c$  according to (5), and for any set  $\{\mathcal{C}_k \mid k \in K\}$  with  $K \subseteq I$  that represents  $c$ ,  $K \supseteq \{i \in I \mid A_i \text{ essential}\}$ .

*Proof.* For ease of presentation, we reason on the preference relations  $\leq_i$  generated by the sets of priors  $\mathcal{C}_i$ ; recall that the corresponding representation in terms of preference orderings is (6), and that we use  $\mathcal{R}$  to denote the set of preference relations generated by  $\mathcal{P}$ . We first show that  $\{\leq_i \mid i \in I, A_i \text{ essential}\}$  represents  $c$ . Since  $\{\leq_i \mid i \in I\} = \mathcal{R}$  represents  $c$ , for each  $\leq_i \in \{\leq_i \mid i \in I, A_i \text{ essential}\}$  and each set  $A \in \wp(\mathcal{A})$ ,  $\{f \in A \mid f \geq_i g, \forall g \in A\} \subseteq c(A)$ . So for every  $A \in \wp(\mathcal{A})$ ,  $\bigcup_{i \text{ s.t. } A_i \text{ essential}} \{f \in A \mid f \geq_i g, \forall g \in A\} \subseteq c(A)$ . It remains to show the inverse inclusion. For reductio, suppose that it does not hold, that is, that there exists  $A \in \wp(\mathcal{A})$  and  $f' \in A$  such that  $f' \notin \bigcup_{i \text{ s.t. } A_i \text{ essential}} \{f \in A \mid f \geq_i g, \forall g \in A\}$  but  $f' \in c(A)$ . By the properties of the maxmin EU functional, the fact that  $\mathcal{R}$  represents  $c$ , and the definition of  $A_i$ , it follows that there exists  $f \in \bigcup_{i \in I} A_i \setminus \bigcup_{A_i \text{ essential}} A_i$ . Take any  $A_i$  with  $f \in A_i$  that is  $\supseteq$ -maximal – ie. for every  $A_k$  such that  $f \in A_k$ ,  $A_i \supseteq A_k$ . Since  $\supseteq$  is complete and the strict relation generated by it is transitive, such a maximal element exists. For such  $A_i$ , every set  $J \subset I \setminus \{i\}$  such that  $\bigcup_{j \in J} A_j \supseteq A_i$  must contain some  $k$  such that  $f \in A_k$ ; however, by the definition of  $A_i$ , it follows that  $A_i \supseteq A_k$ . Hence  $A_i$  is essential, contrary to the definition of  $f$ . So, by reductio, there exists no such  $f$ , and  $\bigcup_{i \text{ s.t. } A_i \text{ essential}} \{f \in A \mid f \geq_i g, \forall g \in A\} = c(A)$ ; so  $\{\leq_i \mid i \in I, A_i \text{ essential}\}$  represents

$c$ , as required.

We now show that any subset of  $\mathcal{R}$  representing  $c$  must contain all elements yielding essential  $A_i$ . For reductio, suppose that there exists  $K \subseteq I$  with  $\{\leq_k \mid k \in K \subseteq I\}$  representing  $c$  and  $K \not\supseteq \{i \in I \mid A_i \text{ essential}\}$ . Take any  $i \in \{i \in I \mid A_i \text{ essential}\} \setminus K$ , and consider  $A_i$ . Since  $c(A_i) = A_i$  by definition, there must exist  $J \subseteq K$  such that  $\bigcup_{j \in J} A_j \supseteq A_i$  and  $A_i \not\supseteq A_j$  for all  $j \in J$ . However, the existence of such a set contradicts the assumption that  $A_i$  is essential. Hence there exists no such  $K$  representing  $c$ , as required. Hence the claim is established. □

### B.3 Auxiliary Lemmas for the Proof of Proposition B.1

**Lemma B.4.** *Let  $(O, I) \in \mathcal{Z}$  be such that, for  $b \in B$ ,  $b \notin O$ . Then there exists  $(O', I') \supseteq (O, I)$  with  $b \in O'$ .*

*Proof of Lemma B.4.* Since  $b \notin O$ ,  $\bar{b} \cap O = \mathfrak{R}^*$ . Let  $O' = co(O \cup \bar{b})$ . We now extend  $I$  to  $I'$  on  $O'$ . Throughout the proof of this Lemma (and in particular Lemmas B.5–B.11),  $I$  will remain fixed; as a point of notation, for any  $a \in O$ , we let  $x_a \in \mathfrak{R}$  be such that  $I(a) = x_a$ .

Let  $[x_{\hat{a}}]^O = \{a \in O \mid I(a) = I(\hat{a})\} \cap U_{\hat{a}}$ . Note that, by the monotonicity and constant linearity of  $I$  and the fact that  $U_{\hat{a}}$  is closed,  $[x_{\hat{a}}]^O$  is closed and bounded. Note also that, by the constant linearity of  $I$ , for any  $a \in [x_{\hat{a}}]^O$  and  $\beta \in [0, 1]$ ,  $\beta a + (1 - \beta)x_{\hat{a}}^* \in [x_{\hat{a}}]^O$ ; hence, in particular,  $\beta[x_{\hat{a}}]^O + (1 - \beta)x_{\hat{a}}^* \subseteq [x_{\hat{a}}]^O$ .

Let  $F_b = \overline{(O' \setminus O)} \cap U_{\hat{a}^*} \cap \{a \in B \mid \sum_{s \in S} a(s) = x_{\hat{a}}\}$ . (Geometrically, this is the intersection between the closure of  $O' \setminus O$ ,  $U_{\hat{a}^*}$  and the hyperplane normal to the unit vector going through  $x_{\hat{a}}^*$ .)  $F_b$  is evidently closed and bounded and hence compact. Finally, for a compact subspace  $F \subseteq B$  and a continuous bounded function  $\sigma : F \rightarrow \mathfrak{R}$ , we let  $Y_\sigma = \{a + \sigma(a)^* \mid a \in F\}$ . Note that  $Y_\sigma$  is closed and bounded, because it is the image of a continuous map from a compact space to a Hausdorff one. We use  $\leq$  to denote the standard dominance order on functions  $\sigma$  ( $\sigma \leq \sigma'$  iff  $\sigma(a) \leq \sigma'(a)$  for all  $a \in F$ ).

Now consider the following set:

$$(10) \quad C_b = \left\{ \sigma : F_b \rightarrow \mathfrak{R} \left| \begin{array}{l} [x_{\hat{a}}]^O \cup Y_\sigma \subseteq c \left( [x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}} \right) \ \& \\ \forall d \in Y_\sigma, [d, x_{\hat{a}}^*] \subseteq Y_\sigma \ \& \\ \forall a' \in [x_{\hat{a}}]^O \cup Y_\sigma, \forall z > 0, \forall \alpha \in (0, 1), \\ \alpha \hat{a} + (1 - \alpha)x_{\hat{a}}^* \notin c \left( [x_{\hat{a}}]^O \cup Y_\sigma \cup \{a' + z^*\} \cup X_{x_{\hat{a}}}^{\hat{A}} \right) \end{array} \right. \right\}$$

Note that, since  $[x_{\hat{a}}]^O$ ,  $Y_\sigma$  and  $X_{x_{\hat{a}}}^{\hat{A}}$  are closed and bounded, so is  $[x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}}$ .

**Lemma B.5.**  $C_b \neq \emptyset$ .

*Proof.* Consider  $\{\sigma : F_b \rightarrow \mathfrak{R} \mid [x_{\hat{a}}]^O \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}})\}$ . This set is non-empty by property 2f of  $I$  and the monotonicity of  $c$  (property v in Lemma B.2). Moreover, by the monotonicity of  $c$  and Chernoff, for any  $\sigma', \sigma'' : F_b \rightarrow \mathfrak{R}$  with  $\sigma' \geq \sigma''$  if  $\sigma'$  is in this set, then so is  $\sigma''$ . It follows by continuity (property vii in Lemma B.2) that this set has a maximum element; let  $\sigma$  be any such element. By definition, we thus have that  $[x_{\hat{a}}]^O \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}})$ . Now consider any  $c \in F_b$  and  $\epsilon > 0$ . By the maximality of  $\sigma$ ,  $[x_{\hat{a}}]^O \not\subseteq c([x_{\hat{a}}]^O \cup Y_{\sigma_c^{+\epsilon}} \cup X_{x_{\hat{a}}}^{\hat{A}})$ , where  $\sigma_c^{+\epsilon}(c) = \sigma(c) + \epsilon$  and  $\sigma_c^{+\epsilon}(d) = \sigma(d)$  for  $d \neq c$ ; by Chernoff, it follows that  $[x_{\hat{a}}]^O \not\subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{c + (\sigma(c) + \epsilon)^*\} \cup X_{x_{\hat{a}}}^{\hat{A}})$ . It follows from Aizerman that  $c + (\sigma(c) + \epsilon)^* \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{c + (\sigma(c) + \epsilon)^*\} \cup X_{x_{\hat{a}}}^{\hat{A}})$ . Since this holds for any  $\epsilon > 0$ , it follows from continuity (property vii) that  $c + \sigma(c)^* \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}})$ . Since this holds for all  $c \in F_b$ , we have that  $[x_{\hat{a}}]^O \cup Y_\sigma \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}})$ . (Note that it follows in particular, using monotonicity, that for all  $a \in [x_{\hat{a}}]^O \cap F_b$ ,  $\sigma(a) = 0$ .)

We now show that  $[d, x_{\hat{a}}^*] \subseteq Y_\sigma$  for all  $d \in Y_\sigma$ . Take any  $d \in Y_\sigma$  and any  $\gamma \in (0, 1)$ , and consider  $c([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}} \cup (\gamma([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}}) + (1 - \gamma)x_{\hat{a}}^*))$ . Since, by their definition,  $\gamma[x_{\hat{a}}]^O + (1 - \gamma)x_{\hat{a}}^* \subseteq [x_{\hat{a}}]^O$  and  $\gamma X_{x_{\hat{a}}}^{\hat{A}} + (1 - \gamma)x_{\hat{a}}^* \subseteq X_{x_{\hat{a}}}^{\hat{A}}$ ,  $c([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}} \cup (\gamma([x_{\hat{a}}]^O \cup Y_\sigma \cup X_{x_{\hat{a}}}^{\hat{A}}) + (1 - \gamma)x_{\hat{a}}^*)) = c([x_{\hat{a}}]^O \cup Y_\sigma \cup (\gamma Y_\sigma + (1 - \gamma)x_{\hat{a}}^*) \cup X_{x_{\hat{a}}}^{\hat{A}})$ . Since  $x_{\hat{a}}^* \in [x_{\hat{a}}]^O$ , it follows from constant independence (property iv) that  $[x_{\hat{a}}]^O \cup Y_\sigma \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup (\gamma Y_\sigma + (1 - \gamma)x_{\hat{a}}^*) \cup X_{x_{\hat{a}}}^{\hat{A}})$ . Moreover, it follows from the maximality of  $Y_\sigma$  that, for every  $z > 0$ ,  $[x_{\hat{a}}]^O \not\subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{d + z^*\} \cup X_{x_{\hat{a}}}^{\hat{A}})$ , so, by constant linearity  $\gamma[x_{\hat{a}}]^O + (1 - \gamma)x_{\hat{a}}^* \not\subseteq c(\gamma([x_{\hat{a}}]^O \cup Y_\sigma \cup \{d + z^*\} \cup X_{x_{\hat{a}}}^{\hat{A}}) + (1 - \gamma)x_{\hat{a}}^*)$ , whence, by Chernoff  $\gamma[x_{\hat{a}}]^O + (1 - \gamma)x_{\hat{a}}^* \not\subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{\gamma(d + z^*) + (1 - \gamma)x_{\hat{a}}^*\} \cup (\gamma Y_\sigma + (1 - \gamma)x_{\hat{a}}^*) \cup X_{x_{\hat{a}}}^{\hat{A}})$ . By Aizerman it follows that  $\gamma(d + z^*) + (1 - \gamma)x_{\hat{a}}^* \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{\gamma(d + z^*) + (1 - \gamma)x_{\hat{a}}^*\} \cup (\gamma Y_\sigma + (1 - \gamma)x_{\hat{a}}^*) \cup X_{x_{\hat{a}}}^{\hat{A}})$  for every  $z > 0$ , so by continuity  $\gamma d + (1 - \gamma)x_{\hat{a}}^* \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup (\gamma Y_\sigma + (1 - \gamma)x_{\hat{a}}^*) \cup X_{x_{\hat{a}}}^{\hat{A}})$ . Since  $d \in Y_\sigma$ ,

it follows by the definition of  $Y_\sigma$  that there exists  $x \in \mathfrak{R}$  such that  $\gamma d + x^* \in Y_\sigma$ . Hence,  $\gamma d + x^*, \gamma d + (1 - \gamma)x_a^* \in c([x_a]^\circ \cup Y_\sigma \cup (\gamma Y_\sigma + (1 - \gamma)x_a^*) \cup X_{x_a}^\wedge)$ ; by monotonicity (property **v**), it follows that  $x = (1 - \gamma)x_a$ , and so  $\gamma d + (1 - \gamma)x_a^* \in Y_\sigma$ . Since this holds for every  $\gamma \in (0, 1)$ , we have that  $[d, x_a^*] \subseteq Y_\sigma$ , as required.

It remains to show that  $\alpha \hat{a} + (1 - \alpha)x_a^* \notin c\left([x_a]^\circ \cup Y_\sigma \cup \{a' + z^*\} \cup X_{x_a}^\wedge\right)$  for all  $a' \in [x_a]^\circ \cup Y_\sigma$ ,  $z > 0$  and  $\alpha \in (0, 1)$ . First note that, by property **2g**,  $\alpha \hat{a} + (1 - \alpha)x_a^* \notin c\left([x_a]^\circ \cup Y_\sigma \cup \{a' + z^*\} \cup X_{x_a}^\wedge\right)$  for all  $a' \in [x_a]^\circ$ ,  $z > 0$  and  $\alpha \in (0, 1)$ . We show that this is in fact the case for every  $a' \in [x_a]^\circ \cup Y_\sigma$ . For reductio, take any  $d \in Y_\sigma$  and  $z > 0$  and  $\alpha \in (0, 1)$  and suppose that  $\alpha \hat{a} + (1 - \alpha)x_a^* \in c\left([x_a]^\circ \cup Y_\sigma \cup \{d + z^*\} \cup X_{x_a}^\wedge\right)$ . Consider any  $a' \in [x_a]^\circ$  and  $\gamma \in (0, 1)$ ; by property **2g**,  $\alpha \hat{a} + (1 - \alpha)x_a^* \notin c\left([x_a]^\circ \cup Y_\sigma \cup \{\gamma(a' + y^*) + (1 - \gamma)x_a^*, d + z^*\} \cup X_{x_a}^\wedge\right)$  for any  $y > 0$ . It follows from Aizerman that  $\gamma(a' + y^*) + (1 - \gamma)x_a^* \in c\left([x_a]^\circ \cup Y_\sigma \cup \{\gamma(a' + y^*) + (1 - \gamma)x_a^*, d + z^*\} \cup X_{x_a}^\wedge\right)$ , and hence by continuity (property **vii**),  $\gamma a' + (1 - \gamma)x_a^* \in c\left([x_a]^\circ \cup Y_\sigma \cup \{d + z^*\} \cup X_{x_a}^\wedge\right)$ . Since this holds for every  $\gamma \in (0, 1)$  and  $a' \in [x_a]^\circ$ , it follows that  $[x_a]^\circ \subseteq c\left([x_a]^\circ \cup Y_\sigma \cup \{d + z^*\} \cup X_{x_a}^\wedge\right)$ , contradicting the maximality of  $\sigma$ . Hence  $\alpha \hat{a} + (1 - \alpha)x_a^* \notin c\left([x_a]^\circ \cup Y_\sigma \cup \{d + z^*\} \cup X_{x_a}^\wedge\right)$  for any  $d \in Y_\sigma$ ,  $z > 0$  and  $\alpha \in (0, 1)$ , and so  $\alpha \hat{a} + (1 - \alpha)x_a^* \notin c\left([x_a]^\circ \cup Y_\sigma \cup \{a' + z^*\} \cup X_{x_a}^\wedge\right)$  for all  $a' \in [x_a]^\circ \cup Y_\sigma$ ,  $z > 0$  and  $\alpha \in (0, 1)$ , as required. □

**Lemma B.6.** *For all  $d \in F_b$ ,  $\alpha > 0$  and  $x \in \mathfrak{R}$  such that  $\alpha d + x^* \in F_b$ , and all  $\sigma \in C_b$ ,  $\sigma(\alpha d + x^*) = (1 - \alpha)x_a^* - x^* + \alpha\sigma(d)$ .*

*Proof.* It suffices to show this for all  $\alpha \in (0, 1]$ ; the other cases follow immediately. Suppose that  $d, \alpha d + x^* \in F_b$  for  $\alpha \in (0, 1]$ , and  $\sigma \in C_b$ . Since  $d + \sigma(d)^* \in Y_\sigma$ , it follows from the definition of  $C_b$  that  $\alpha(d + \sigma(d)^*) + (1 - \alpha)x_a^* \in Y_\sigma$ . It thus follows by monotonicity (property **v**) and the definition  $C_b$ ,  $\sigma(\alpha d + x^*)$  is such that  $\alpha d + x^* + \sigma(\alpha d + x^*)^* = \alpha(d + \sigma(d)^*) + (1 - \alpha)x_a^*$ . So  $\sigma(\alpha d + x^*) = (1 - \alpha)x_a^* - x^* + \alpha\sigma(d)$ , as required. □

Take any element  $\sigma \in C_b$ ; Lemma **B.5** guarantees that such an element exists. Let  $I'' : O \cup F_b \rightarrow \mathfrak{R}$  to be the functional extending  $I$  and such that  $I''(d) = x_a - \sigma(d)$  for

$d \in F_b$ . By Lemma B.6,  $I''$  is constant linear where defined; hence, by the definition of  $F_b$  and  $O'$ , there exists a unique constant linear extension to  $O'$ , which we call  $I'$ . By the definition of  $C_b$ , the constant linearity of  $c$  and Chernoff,  $I'$  satisfies properties 2f and 2g. We now establish several other properties of  $I'$ .

**Lemma B.7.**  *$I'$  satisfies property 2e: for all  $d \in O'$ , if  $d \in \hat{A}$ , then  $I'(d) \leq x_{\hat{a}}$ .*

*Proof.* Since for  $d \in O$  this follows from the properties of  $I$ , it suffices to consider  $d \in O' \setminus O$ . Suppose for reductio that  $d \in \hat{A}$  and  $I'(d) > x_{\hat{a}}$ . By the definition of  $F_b$ , there exists  $\beta \in (0, 1]$ ,  $x \in \mathfrak{R}$  such that  $\beta d + x^* \in F_b$ ; hence  $\frac{1}{\beta}(x_{\hat{a}} - \sigma(\beta d + x^*) - x) = I'(d) > x_{\hat{a}}$ . By the definition of  $C_b$ , we have that  $\beta d + x^* + \sigma(\beta d + x^*)^* \in c([x_{\hat{a}}]^O \cup Y_{\sigma} \cup X_{x_{\hat{a}}}^{\hat{A}})$ ; however, by the monotonicity of  $c$  (property v of Lemma B.2), since  $\beta d + (1 - \beta)x_{\hat{a}}^* \in X_{x_{\hat{a}}}^{\hat{A}} \subseteq [x_{\hat{a}}]^O \cup Y_{\sigma} \cup X_{x_{\hat{a}}}^{\hat{A}}$  and  $x + \sigma(\beta d + x^*) < (1 - \beta)x_{\hat{a}}$ ,  $\beta d + x^* + \sigma(\beta d + x^*)^* \notin c([x_{\hat{a}}]^O \cup [b + z^*, x_{\hat{a}}^*] \cup X_{x_{\hat{a}}}^{\hat{A}})$ , which is a contradiction. So  $I'(d) \leq x_{\hat{a}}$ , as required.  $\square$

**Lemma B.8.**  *$I'$  is superadditive: for all  $a, a' \in O'$  and  $\alpha \in [0, 1]$ , then  $I'(\alpha a + (1 - \alpha)a') \geq \alpha I'(a) + (1 - \alpha)I'(a')$ .*

*Proof.* We show the result for  $a, a' \in U_{\hat{a}}$ ; it extends to other cases by the constant linearity of  $I'$ . Suppose for reductio that for some  $a, a' \in O' \cap U_{\hat{a}}$ ,  $\alpha \in [0, 1]$ ,  $I'(\alpha a + (1 - \alpha)a') < \alpha I'(a) + (1 - \alpha)I'(a')$ . By the convexity of  $U_{\hat{a}}$  and the definition of  $[x_{\hat{a}}]^O$ ,  $F_b$  and  $I'$ ,  $\alpha a + (1 - \alpha)a' + (x_{\hat{a}} - I'(\alpha a + (1 - \alpha)a'))^* \in [x_{\hat{a}}]^O \cup Y_{\sigma}$ . Similarly,  $a + (x_{\hat{a}} - I'(a))^*, a' + (x_{\hat{a}} - I'(a'))^* \in [x_{\hat{a}}]^O \cup Y_{\sigma}$ . It follows from the definition of  $C_b$  and superadditivity (property vi of Lemma B.2) that  $d \in c\left([x_{\hat{a}}]^O \cup Y_{\sigma} \cup \{d\} \cup X_{x_{\hat{a}}}^{\hat{A}}\right)$  for all  $d \in co([x_{\hat{a}}]^O \cup Y_{\sigma})$ ; hence  $\alpha(a + (x_{\hat{a}} - I'(a))^*) + (1 - \alpha)(a' + (x_{\hat{a}} - I'(a'))^*) \in c([x_{\hat{a}}]^O \cup \{\alpha(a + (x_{\hat{a}} - I'(a))^*) + (1 - \alpha)(a' + (x_{\hat{a}} - I'(a'))^*)\} \cup Y_{\sigma} \cup X_{x_{\hat{a}}}^{\hat{A}})$ . However, since  $I'(\alpha a + (1 - \alpha)a') < \alpha I'(a) + (1 - \alpha)I'(a')$ ,  $\alpha a + (1 - \alpha)a' + (x_{\hat{a}} - I'(\alpha a + (1 - \alpha)a'))^* \in Y_{\sigma}$  strictly dominates  $\alpha(a + (x_{\hat{a}} - I'(a))^*) + (1 - \alpha)(a' + (x_{\hat{a}} - I'(a'))^*)$ , and so it follows from that monotonicity of  $c$  (property v of Lemma B.2) that  $\alpha(a + (x_{\hat{a}} - I'(a))^*) + (1 - \alpha)(a' + (x_{\hat{a}} - I'(a'))^*) \notin c([x_{\hat{a}}]^O \cup \{\alpha(a + (x_{\hat{a}} - I'(a))^*) + (1 - \alpha)(a' + (x_{\hat{a}} - I'(a'))^*)\} \cup Y_{\sigma} \cup X_{x_{\hat{a}}}^{\hat{A}})$ , which is a contradiction. Hence  $I'(\alpha a + (1 - \alpha)a') \geq \alpha I'(a) + (1 - \alpha)I'(a')$ , as required.  $\square$

**Lemma B.9.**  *$I'$  is monotonic: for every  $a, d \in O'$ , if  $d \leq a$ , then  $I'(d) \leq I'(a)$ .*

*Proof.* We show the result for  $a, d \in U_{\hat{a}}$ ; it extends to other cases by the constant linearity of  $I'$ . Consider  $a, a' \in O' \cap U_{\hat{a}}$ ; by the definition of  $[x_{\hat{a}}]^O$ ,  $F_b$  and  $I'$ ,  $a + (x_{\hat{a}} - I'(a))^*$ ,  $d + (x_{\hat{a}} - I'(d))^* \in [x_{\hat{a}}]^O \cup Y_{\sigma}$ . If  $d \leq a$ , then, by the monotonicity of  $c$ , for each  $y > 0$ , since  $d + (x_{\hat{a}} - I'(a) - y)^* < a + (x_{\hat{a}} - I'(a))^*$ ,  $d + (x_{\hat{a}} - I'(a) - y)^* \notin c([x_{\hat{a}}]^O \cup Y_{\sigma} \cup \{d + (x_{\hat{a}} - I'(a) - y)^*\} \cup X_{x_{\hat{a}}}^{\hat{A}})$ , and so  $x_{\hat{a}} - I'(a) - y \neq x_{\hat{a}} - I'(d)$ . Hence  $I'(a) \geq I'(d)$ , as required.  $\square$

It remains to show that  $I'$  represents  $c$  on  $O'$  (property 2d). For this, we need the following preliminary lemma.

**Lemma B.10.** *Let  $A \in \wp_{bdd}(O')$ ,  $A \subset P$  for some  $P \in \mathcal{Q}$ , and let  $x \in \mathfrak{R}$  be such that  $x \geq I'(a'')$  for all  $a'' \in A$ . Then there exists  $a \in P$  such that  $c(A \cup [a, x^*]) = [a, x^*]$ .*

*Proof.* By the constant linearity of  $I'$ , there exists  $a' \in P$  with  $I'(a') = x$ . Moreover, for any  $d \in A$ , since  $a', d \in P$ , it follows from the definition of  $\mathcal{Q}$  that  $d = \alpha a' + z^*$  for some  $\alpha \geq 0$  and  $z \in \mathfrak{R}$ . Hence there exists  $\gamma_d \geq 1$ ,  $\beta' \in [0, 1]$  and  $y' \in \mathfrak{R}$  such that  $d = \beta'(\gamma_d a' + (1 - \gamma_d)x^*) + (1 - \beta')y'^*$ ; take any such  $\gamma_d$ . Since  $A$  is closed and bounded, there exists  $\gamma \geq \gamma_d$  for all  $d \in A$ ; let  $a = \gamma a' + (1 - \gamma)x^*$ . By construction (and the constant linearity of  $I'$ ),  $I'(a) = x$ , and, for each  $d \in A$ , there exists  $\beta \in [0, 1]$  and  $y \in \mathfrak{R}$  such that  $d = \beta a + (1 - \beta)y^*$ . It follows that, for any  $d \in A$  with  $I'(d) = x$ ,  $d \in [a, x^*]$ .

First we show that, for each  $d \in A$  with  $I'(d) < x$ ,  $d \notin c(A \cup [a, x^*])$ . Let  $d \in A$  be such that  $I'(d) < x$ . By the previous observation, there exists  $\beta \in [0, 1]$  and  $y \in \mathfrak{R}$  such that  $d = \beta a + (1 - \beta)y^*$ . Since, by constant linearity of  $I'$ ,  $I'(d) = \beta I'(a) + (1 - \beta)y < x = I'(a)$ , we have that  $y < x$ . Hence, by the monotonicity of  $c$ ,  $d = \beta a + (1 - \beta)y^* \notin c(\{\beta a + (1 - \beta)y^*, \beta a + (1 - \beta)x^*\})$ . Since  $\beta a + (1 - \beta)x^* \in [a, x^*]$ , it follows by Chernoff that  $d \notin c(A \cup [a, x^*])$ , as required.

It follows that  $c(A \cup [a, x^*]) \subseteq [a, x^*]$ . By Chernoff and Aizerman,<sup>27</sup> it follows that  $c(A \cup [a, x^*]) = c([a, x^*]) = [a, x^*]$ , where the second equality holds by property 2f of  $I'$ , the constant linearity of  $c$ , and Chernoff.  $\square$

**Lemma B.11.**  *$I'$  represents  $c$  on  $O'$ : for any  $A \in \wp_{bdd}(O')$ ,  $\arg \max_A I' \subseteq c(A)$ .*

<sup>27</sup>Note that, in the presence of Chernoff, Aizerman is equivalent to:  $c(B) \subseteq A \subseteq B \Rightarrow c(B) = c(A)$ . (See, for example, [Moulin \(1985\)](#).)

*Proof.* If  $A \subseteq O$ , the result follows from that fact that  $I'$  extends  $I$  and the fact that  $(O, I) \in \mathcal{Z}$ . So suppose that this is not the case, and consider  $a \in \arg \max_A I'$ . By Lemma B.10, for each  $P \in \mathcal{Q}$  with  $A \cap P \neq \emptyset$ , there exists  $a'_P \in P$  with  $I'(a'_P) = I'(a)$  and  $c((A \cap P) \cup [a'_P, x_a^*]) = [a'_P, x_a^*]$ . For such each  $P$ , let  $A_P^a = (A \cap P) \cup [a'_P, x_a^*]$ . By Chernoff and Aizerman,<sup>28</sup>  $c(\bigcup_{P \text{ s.t. } A \cap P \neq \emptyset} A_P^a) = c(\bigcup_{P \text{ s.t. } A \cap P \neq \emptyset} [a'_P, x_a^*])$ . Moreover,  $c(\bigcup_{P \text{ s.t. } A \cap P \neq \emptyset} [a'_P, x_a^*]) = \bigcup_{P \text{ s.t. } A \cap P \neq \emptyset} [a'_P, x_a^*]$ , by property 2f of  $I'$ , the constant linearity of  $c$ , and Chernoff. Hence, in particular,  $a \in c(\bigcup_{P \text{ s.t. } A \cap P \neq \emptyset} A_P^a)$ ; since  $A \subseteq \bigcup_{P \text{ s.t. } A \cap P \neq \emptyset} A_P^a$ , it follows by Chernoff that  $a \in c(A)$ , as required.  $\square$

So  $(O', I') \in \mathcal{Z}$  such that  $b \in O'$  and  $(O, I) \leq (O', I')$ . This concludes the proof of Lemma B.4.  $\square$

**Lemma B.12.**  $\mathcal{Z}$  is non-empty.

*Proof of Lemma B.12.* Define  $C_{\hat{a}}$  as follows:

$$(11) \quad C_{\hat{a}} = \left\{ y \in \mathfrak{R} \left| \begin{array}{l} [\hat{a}, y^*] \subseteq c(X_y^{\hat{A}}) \text{ \&} \\ \forall a' \in [\hat{a}, y^*], \forall z > 0, \forall \alpha \in (0, 1), \\ \alpha \hat{a} + (1 - \alpha)y^* \notin c(\{a' + z^*\} \cup X_y^{\hat{A}}) \end{array} \right. \right\}$$

We first establish the non-emptiness of  $C_{\hat{a}}$ , by an argument related to, but not identical to, that used in the proof of Lemma B.5. Consider  $\left\{ \sigma : [0, 1) \rightarrow \mathfrak{R} \mid \hat{a} \in c\left(\hat{A} \cup \{\alpha \hat{a} + (1 - \alpha)\sigma(\alpha)^* \mid \alpha \in [0, 1)\}\right) \right\}$ . By the monotonicity of  $c$  and the fact that  $\hat{a} \in c(\hat{A})$ , this set is non-empty. Moreover, by the monotonicity and continuity of  $c$ , it has a maximal element; let  $\sigma$  be such an element. By the maximality of  $\sigma$ ,  $\hat{a} \in c\left(\hat{A} \cup \{\alpha \hat{a} + (1 - \alpha)\sigma(\alpha)^* \mid \alpha \in [0, 1]\}\right)$ . For  $\alpha \in [0, 1)$  and  $\epsilon > 0$ , define  $\sigma_\alpha^{+\epsilon}$  by:  $\sigma_\alpha^{+\epsilon}(\alpha) = \sigma(\alpha) + \epsilon$  and  $\sigma_\alpha^{+\epsilon}(\beta) = \sigma(\beta)$  for  $\beta \neq \alpha$ . By the maximality of  $\sigma$ , for any  $\alpha \in [0, 1)$  and  $\epsilon > 0$ ,  $\hat{a} \notin c\left(\hat{A} \cup \{\alpha \hat{a} + (1 - \alpha)\sigma_\alpha^{+\epsilon}(\alpha)^* \mid \alpha \in [0, 1)\}\right)$ . It follows by Aizerman that  $\alpha \hat{a} + (1 - \alpha)(\sigma(\alpha) + \epsilon)^* \in c\left(\hat{A} \cup \{\alpha \hat{a} + (1 - \alpha)\sigma_\alpha^{+\epsilon}(\alpha)^* \mid \alpha \in [0, 1)\}\right)$ . Since this holds for all  $\epsilon > 0$  and  $\alpha \in [0, 1)$ , it follows from continuity that  $\{\alpha \hat{a} + (1 - \alpha)\sigma(\alpha)^* \mid \alpha \in [0, 1)\} \subseteq c\left(\hat{A} \cup \{\alpha \hat{a} + (1 - \alpha)\sigma(\alpha)^* \mid \alpha \in [0, 1]\}\right)$ . Let  $y = \sigma(0)$ ; in particular, we have

<sup>28</sup>Chernoff and Aizerman imply that  $c(\bigcup_{i \in I} A_i) = c(\bigcup_{i \in I} c(A_i))$ . See for example Moulin (1985, Lemma 6), whose proof for the case of finite unions is straightforwardly extended to infinite unions.

that  $y^* \in c\left(\hat{A} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in [0, 1]\}\right)$ . It follows, by repeated applications of constant independence and Chernoff, that  $\{\hat{a}, y^*\} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \subseteq c\left(\{y^*\} \cup \bigcup_{k=0}^n (\delta^k \hat{A} + (1-\delta^k)y^*) \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\}\right)$ . Taking the limit as  $\delta \rightarrow 1$  and  $n \rightarrow \infty$ , it follows from continuity that  $\{\hat{a}, y^*\} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \subseteq c\left(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\}\right)$ .

We now show that  $\sigma(\alpha) = y$  for all  $\alpha \in (0, 1)$ . Take any  $\beta \in (0, 1)$ . Since  $\beta X_y^{\hat{A}} + (1-\beta)y^* \subseteq X_y^{\hat{A}}, X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup (\beta(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\}) + (1-\beta)y^*) = X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup (\beta\{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} + (1-\beta)y^*)$ . It follows from the constant independence of  $c$  that  $\{\hat{a}, y^*\} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \subseteq c(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup (\beta\{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} + (1-\beta)y^*))$ . However, for any  $z > 0$ , by monotonicity,  $\hat{a} \notin c\left(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup \{\hat{a} + z^*\}\right)$ , so by constant linearity and Chernoff,  $\beta\hat{a} + (1-\beta)y^* \notin c(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup (\beta\{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} + (1-\beta)y^*) \cup \{\beta(\hat{a} + z^*) + (1-\beta)y^*\})$ . It follows from Aizerman that  $\beta(\hat{a} + z^*) + (1-\beta)y^* \in c(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup (\beta\{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} + (1-\beta)y^*) \cup \{\beta(\hat{a} + z^*) + (1-\beta)y^*\})$ ; since this holds for every  $z > 0$ , we have, by continuity that  $\beta\hat{a} + (1-\beta)y^* \in c(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup (\beta\{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} + (1-\beta)y^*))$ . However, as noted above,  $\beta\hat{a} + (1-\beta)\sigma(\beta)^* \in c(X_y^{\hat{A}} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} \cup (\beta\{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0, 1)\} + (1-\beta)y^*))$ ; it follows from the monotonicity of  $c$  that  $\sigma(\beta) = y$ . Applying this to all  $\beta \in (0, 1)$ , we have that the required conclusion. It follows in particular that  $[\hat{a}, y^*] \subseteq c(X_y^{\hat{A}})$ .

It remains to show that  $\alpha\hat{a} + (1-\alpha)y^* \notin c\left(\{a' + z^*\} \cup X_y^{\hat{A}}\right)$  for all  $\alpha \in (0, 1)$ ,  $a' \in [\hat{a}, y^*]$  and  $z > 0$ . We proceed by reductio: let  $\alpha \in (0, 1)$ ,  $a' \in [\hat{a}, y^*]$  and  $z > 0$  be such that  $\alpha\hat{a} + (1-\alpha)y^* \in c\left(\{a' + z^*\} \cup X_y^{\hat{A}}\right)$  for all  $\alpha \in (0, 1)$ . By definition  $a' = \beta\hat{a} + (1-\beta)y^*$ ; we distinguish cases according to whether  $\beta \leq \alpha$  or not. First suppose that  $\beta \leq \alpha$ . Chernoff implies that  $\alpha\hat{a} + (1-\alpha)y^* \in c\left(\{\beta\hat{a} + (1-\beta)y^* + z^*\} \cup \{\alpha d + (1-\alpha)y^* \mid d \in \hat{A}\}\right)$ . But since  $\{\beta\hat{a} + (1-\beta)y^* + z^*\} \cup \{\alpha d + (1-\alpha)y^* \mid d \in \hat{A}\} = \alpha\left(\left\{\frac{\beta}{\alpha}\hat{a} + \left(1 - \frac{\beta}{\alpha}\right)y^*\right\} + \frac{1}{\alpha}z^*\right) \cup X_y^{\hat{A}} + (1-\alpha)y^*$ , it follows from constant linearity of  $c$  that  $\hat{a} \in c\left(\left\{\frac{\beta}{\alpha}\hat{a} + \left(1 - \frac{\beta}{\alpha}\right)y^*\right\} + \frac{1}{\alpha}z^*\right) \cup X_y^{\hat{A}}$ , contradicting the maximality of  $\sigma\left(\frac{\beta}{\alpha}\right) = y$ . Now suppose that  $\beta > \alpha$ . Take any  $\gamma < \frac{\alpha}{\beta}$ ; by constant independence,  $\alpha\hat{a} + (1-\alpha)y^* \in c\left(\{a' + z^*, \gamma(a' + z^*) + (1-\gamma)y^*\} \cup X_y^{\hat{A}}\right)$ ,

and so, by Chernoff,  $\alpha\hat{a} + (1 - \alpha)y^* \in c\left(\{\gamma(a' + z^*) + (1 - \gamma)y^*\} \cup X_y^{\hat{A}}\right)$ . Since  $\gamma(a' + z^*) + (1 - \gamma)y^* = \gamma\beta\hat{a} + (1 - \gamma\beta)y^* + \gamma z^*$ , the conditions of the previous case are satisfied; the previous argument can thus be employed, yielding a contradiction. So  $\alpha\hat{a} + (1 - \alpha)y^* \notin c\left(\{a' + z^*\} \cup X_y^{\hat{A}}\right)$  for all  $\alpha \in (0, 1)$ ,  $a' \in [\hat{a}, y^*]$  and  $z > 0$ , as required.

The construction of a functional  $I$  on  $\bar{\hat{a}}$  proceeds in an analogous way to the proof of Lemma B.4; the proofs that it satisfies the appropriate conditions are either trivial or follow the same reasoning as used in the proof of Lemma B.4.

□

## References

- Al Najjar, N. and Weinstein, J. (2009). The Ambiguity Aversion Literature: A critical assessment. *Economics and Philosophy*, 25(3):249–284.
- Aliprantis, C. D. and Border, K. C. (2007). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, Berlin, 3rd edition.
- Anscombe, F. J. and Aumann, R. J. (1963). A Definition of Subjective Probability. *The Annals of Mathematical Statistics*, 34:199–205.
- Billingsley, P. (2009). *Convergence of Probability Measures*. John Wiley & Sons.
- Blackwell, D. (1953). Equivalent Comparisons of Experiments. *The Annals of Mathematical Statistics*, 24(2):265–272.
- Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., and Rustichini, A. (2013). Niveloids and Their Extensions: Risk Measures in Small Domains. Mimeo, Bocconi.
- Crès, H., Gilboa, I., and Vieille, N. (2011). Aggregation of multiple prior opinions. *Journal of Economic Theory*, 146:2563–2582.
- Dekel, E., Lipman, B. L., and Rustichini, A. (2001). Representing Preferences with a Unique Subjective State Space. *Econometrica*, 69:891–934.
- Dunford, N. and Schwartz, J. T. (1958). *Linear Operators. Part I: General Theory*. Interscience Publishers, New York.

- Eichberger, J., Grant, S., and Kelsey, D. (2007). Updating Choquet beliefs. *Journal of Mathematical Economics*, 43(7-8):888–899.
- Eichberger, J. and Kelsey, D. (1996). Uncertainty Aversion and Dynamic Consistency. *International Economic Review*, 37(3):625–640.
- Epstein, L. G. and Le Breton, M. (1993). Dynamically Consistent Beliefs Must Be Bayesian. *Journal of Economic Theory*, 61(1):1–22.
- Epstein, L. G., Marinacci, M., and Seo, K. (2007). Coarse contingencies and ambiguity. *Theoretical Economics*, 2:355–394.
- Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31.
- Fishburn, P. C. (1970). *Utility Theory for Decision Making*. Wiley, New York.
- Ghirardato, P. (2001). Coping with ignorance: unforeseen contingencies and non-additive uncertainty. *Economic Theory*, 17(2):247–276.
- Ghirardato, P. (2002). Revisiting Savage in a conditional world. *Economic Theory*, 20(1):83–92.
- Ghirardato, P., Maccheroni, F., and Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *J. Econ. Theory*, 118(2):133–173.
- Gilboa, I., Postlewaite, A., and Schmeidler, D. (2009). Is it always rational to satisfy Savage's axioms? *Economics and Philosophy*, 25(3):285–296.
- Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *J. Math. Econ.*, 18(2):141–153.
- Gollier, C. (2004). *The Economics Of Risk And Time*. MIT Press.
- Hammond, P. J. (1988). Consequentialist foundations for expected utility theory. *Theory and Decision*, 25:25–78.
- Hanany, E. and Klibanoff, P. (2007). Updating preferences with multiple priors. *Theoretical Economics*, 2:261–298.

- Hill, B. (2012). Unanimity and the aggregation of multiple prior opinions. Mimeo, GREGHEC.
- Hilton, R. W. (1981). The Determinants of Information Value: Synthesizing Some General Results. *Management Science*, 27(1):57–64.
- Karni, E. and Safra, Z. (1989). Dynamic Consistency, Revelations in Auctions and the Structure of Preferences. *The Review of Economic Studies*, 56(3):421–433.
- Karni, E. and Safra, Z. (1990). Behaviorally consistent optimal stopping rules. *Journal of Economic Theory*, 51(2):391–402.
- Klibanoff, P., Marinacci, M., and Mukerji, S. (2009). Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144(3):930–976.
- Kreps, D. M. (1979). A Representation Theorem for ‘Preference for Flexibility’. *Econometrica*, 47:565–576.
- Kreps, D. M. (1992). Static choice in the presence of unforeseen contingencies. In *Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn*, pages 258–281.
- LaValle, I. H. (1968). On Cash Equivalents and Information Evaluation in Decisions Under Uncertainty: Part I: Basic Theory. *Journal of the American Statistical Association*, 63(321):252–276.
- Maccheroni, F., Marinacci, M., and Rustichini, A. (2006). Dynamic variational preferences. *Journal of Economic Theory*, 128(1):4–44.
- Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, 27(4):1622–1668.
- Marschak, J. and Miyasawa, K. (1968). Economic Comparability of Information Systems. *International Economic Review*, 9(2):137–174.
- Marschak, J. and Radner, R. (1972). *Economic Theory of Teams*. Yale University Press, New Haven.
- McClennen, E. F. (1990). *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge University Press.

- Moulin, H. (1985). Choice Functions Over a Finite Set: A Summary. *Social Choice and Welfare*, 2:147–160.
- Mukerji, S. (1997). Understanding the nonadditive probability decision model. *Economic Theory*, 9(1):23–46.
- Sarin, R. and Wakker, P. P. (1998). Dynamic choice and nonexpected utility. *Journal of Risk and Uncertainty*, 17(2):87–120.
- Savage, L. J. (1954). *The Foundations of Statistics*. Dover, New York. 2nd edn 1971.
- Sen, A. K. (1971). Choice Functions and Revealed Preference. *Rev. Econ. Stud.*, 38(3):307–317.
- Siniscalchi, M. (2009). Two out of three ain't bad: a comment on "The Ambiguity Aversion Literature: A Critical Assessment". *Economics and Philosophy*, 25(3):335–356.
- Siniscalchi, M. (2011). Dynamic choice under ambiguity. *Theoretical Economics*, 6:379–421.
- Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *Rev. Econ. Stud.*, 23(3):165–180.
- Wakker, P. (1988). Nonexpected utility as aversion of information. *Journal of Behavioral Decision Making*, 1:169–175.