Dynamic Consistency and Ambiguity: A Reappraisal*

Brian Hill

GREGHEC, HEC Paris, CNRS†

First version: August 2013

This version: December 2019

Abstract

The famous conflict between dynamic consistency and ambiguity purportedly undermines these models' normative credibility, and challenges their use in economic applications. Dynamic consistency concerns preferences over contingent plans: so what counts are the contingencies the decision maker envisages – and plans for – rather than independently fixed contingencies, as implicitly assumed in standard formalisations. An appropriate formulation of dynamic consistency resolves the aforementioned conflict, hence undermining the criticisms of ambiguity models based on it. Moreover, it provides a principled justification for the restriction to certain families of beliefs in applications of these models in dynamic choice problems. Finally, it supports a new analysis of the value of information under ambiguity, showing that decision makers may only turn down information if it has an opportunity cost, in terms of the compromising of information they had otherwise expected to receive.

Keywords: Decision under Uncertainty; Dynamic Consistency; Dynamic Choice; Envisaged Contingency; Ambiguity; Value of Information.

JEL classification: D81.

---

*I would like to thank Alain Chateauneuf, Michèle Cohen, Larry Epstein, Itzhak Gilboa, Teddy Seidenfeld, Marciano Siniscalchi, Peter Wakker and workshop and seminar audiences at RUD 2013, DIET 2013, SAET 2014, ENS, Cergy-Pontoise, LMU and LSE for helpful comments. The author gratefully acknowledges support from the ANR Program DUSUCA (ANR-14-CE29-0003-01).

†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.

1 Introduction

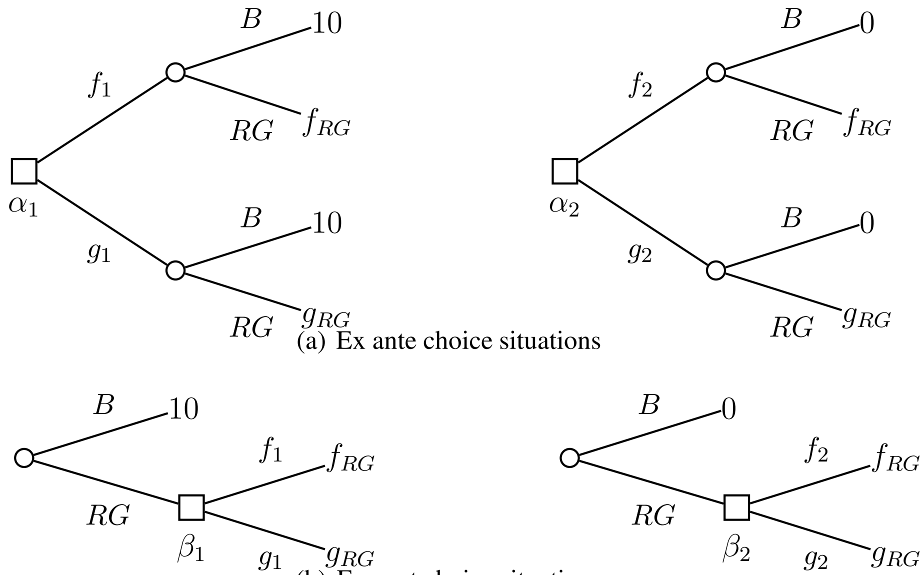
One of the principal challenges to, and in, the literature on non-expected utility models for decision under uncertainty – or 'ambiguity models' – is that posed by application in dynamic situations. In such situations, violation of expected utility purportedly has unpalatable consequences. Putting aside subtleties in terminology and definitions, the gist of the problem can be traced to an argument claiming to show that consequentialism – the decision maker ignores 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 contingencies – are incompatible with non-expected utility. Given the *prima facie* attraction of these properties of dynamic choice, the option of abandoning one of them to leave space for non-expected utility is unappetizing. All the worse, the argument concludes, for ambiguity models.

Such considerations are a central plank underlying the hegemony of the Bayesian approach. On the one hand, they cast doubt on the normative credentials of ambiguity models, insofar as the dynamic properties just mentioned appear to be sensible principles of rationality (Raiffa, 1968; Hammond, 1988; Machina, 1989; McClennen, 1990; Wakker, 1999). On the other hand, they have been suggested as problematic for the use of such models in economic modelling (Epstein and Le Breton, 1993; Al Najjar and Weinstein, 2009). As such, they probably constitute the biggest conceptual obstacle to the adoption of ambiguity models in normative economics, decision analysis and economic modelling, as recently defended in decision theory (Gilboa and Marinacci, 2013; Marinacci, 2015), mechanism design (Bose and Renou, 2014; Tillio et al., 2016), macroeconomics (Hansen, 2014; Hansen and Sargent, 2008) or climate economics (Millner et al., 2013; Kunreuther et al., 2013).

The main thesis of this paper is that this generally accepted state-of-play rests on a mistake. The dynamic consistency principle, we shall argue, has been misformalised in standard treatments; a more adequate formalisation dissolves the tension between non-expected utility and dynamic principles. Moreover, it provides solid guidance for the use of ambiguity models in economic

	R	G	B
f_1	10	0	10
g_1	0	10	10
f_2	10	0	0
g_2	0	10	0

Table 1: The investor / Ellsberg one-urn example (values in millions of dollars)



(a) Ex ante choice situations

(b) Ex post choice situations

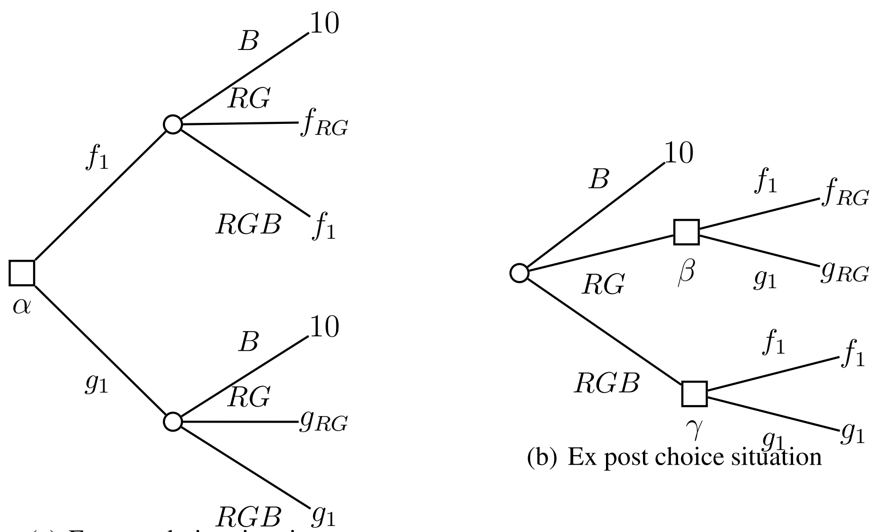
Figure 1: Dynamic consistency in the dynamic investor / Ellsberg example (standard version)
B: the information that B (i.e. the performance is bad) is received; RG: the information that R or G (i.e. the performance is not bad) is received. 0 and 10 are the outcomes, as indicated in Table 1. f_{RG} (respectively g_{RG}) is the bet conditional on RG (i.e. performance not being bad) that coincides with f_1 and f_2 (resp. g_1 and g_2; Table 1); e.g. f_{RG} yields 10 if R and 0 if G. Circles indicate nature nodes; squares are choice nodes.

applications. It thus removes the normative obstacle to the use of such models in guiding policy.

To illustrate the purported problem for non-expected utility and the basic insight behind our approach, consider an investor examining a start-up drug company that is currently running trials on its new, and only product. She has solid evidence that the product has probability $\frac{1}{3}$ of yielding regular performance in terms of cure rate (denoted R), but has no further information on the probabilities of good (G) or bad (B) performance. She can construct asset positions yielding performance-dependent returns as shown in Table 1. Note the structural similarity to the standard Ellsberg (1961) example involving bets on an urn with ninety balls, thirty of which are red (R) and the rest of which are black (B) and green (G) in an unknown proportion. Indeed, the investor exhibits the standard Ellsberg preferences – $f_1 \prec g_1$ and $f_2 \succ g_2$ – which violate expected utility. After the drug trials are complete, the investor is told whether the performance is bad or not (and receives no other information), and is asked again for her preferences. The choices she faces before and after receiving this information can be represented (the story goes) by the trees in Figure 1. Dynamic consistency demands harmony between ex ante preferences over contingent plans and preferences after the realisation of the planned-for contingencies. Interpreting the ex ante choice between f_1 and g_1 (node α_1 in Figure 1(a)) as a choice between plans for the contingency that the performance is not bad, it thus requires the investor to have the same preferences at α_1 and β_1. Similar reasoning applies to the right hand trees (nodes α_2 and β_2) and f_2 and g_2. Consequentialism demands that all that counts are the consequences of one's choices. It thus requires the investor to ignore the difference between the trees in Figure 1(b) and have the same preferences at the nodes β_1 and β_2. So these dynamic principles imply that she must have the same ex ante preferences over f_2 and g_2 as over f_1 and g_1 – in direct contradiction with the standard Ellsberg pattern. All the worse, the argument concludes, for the normative credentials of any account that allows such preferences.

But this argument rests upon a hidden assumption. The application of dynamic consistency to the trees in Figure 1 relies on identifying the ex ante choice between f_1 and g_1 with the choice between plans for the contingency in which the performance is not bad. However, the investor conceives of the ex ante choice in this way only if she envisages two possible contingencies to be planned for: one where she learns that the performance is bad, and the other where she learns that it is not. That is, she conceives of it like this only if she expects the company only to report whether the performance is bad or not (B or RG). The assumption that she conceives of the choice in this way is far from innocent. Equipped only with the information reported at the beginning of the example (i.e. Table 1, the probability of regular performance, and the fact that the company is conducting trials), she could reasonably entertain other possibilities, such as the company reporting whether the performance is good or not (G or RB), or disclosing the full details of the performance (R or G or B), or issuing a partially informative report (e.g. a probability distribution over $\{R, G, B\}$), or reporting nothing at all (RGB). In each of these cases, she will not conceive of the ex ante choice as portrayed in Figure 1: she will not consider it a choice between plans for the contingency in which the performance is not bad. For example, if she thinks that the company could report whether performance is bad or not or nothing at all (B or RG or RGB), then she envisages three contingencies, and will conceive the ex ante choice as involving plans for these three contingencies. In this case, the set of contingencies envisaged, and hence planned for, is more accurately represented by the trees in Figure 2. So the dynamic consistency principle should be applied to these trees, rather than those in Figure 1.

We call the tree representing the set of contingencies that a decision maker envisages in a given situation – or the information structure he believes that he is faced with – his *subjective tree*. Often a theorist adopts a decision tree that is determined independently of the decision maker's view on the relevant possible contingencies. For instance, the trees in Figure 1 are used because they involve the events on which the acts f and g differ (RG) and coincide (B) respectively. As this example



(a) Ex ante choice situation

(b) Ex post choice situation

Figure 2: Dynamic consistency in the dynamic investor / Ellsberg example (choice between f_1 and g_1) RGB: the decision maker learns nothing (i.e. the information is that R or G or B). Other notation is as in Figure 1.

illustrates, in studies of dynamic choice and in standard formalisations of dynamic consistency (Section 3), the trees adopted by theorists typically correspond to partitions (or more generally filtrations) of an 'objective' payoff-relevant state space. We call such trees *objective trees*. As is clear from Figure 2, subjective trees need not be objective. Information structures that do not correspond to partitions of the payoff-relevant state space are well known in the literature (e.g. Gollier, 2004), and subjective trees can be thought of as specific non-expected utility information structures of this sort (Section 2).

The central insight of this paper is that formulating dynamic consistency on objective trees is a mistake. Dynamic consistency concerns contingent plans, so any reasonable version of the principle should involve the contingencies that the decision maker in fact envisages, and hence plans for: it should be formulated on his subjective tree. No novelty is claimed for this observation, which will hopefully strike the reader as obvious. The paper's main contributions lie in showing that this apparently innocuous point has far-reaching consequences for the normative credentials of ambiguity models and their use in economic applications and decision analysis.

As a first conceptual contribution, formulating dynamic consistency on subjective trees resolves the purported conflict between non-expected utility and the aforementioned dynamic principles. The argument above relies crucially on an application of dynamic consistency where there is a single ex post choice (at node β_1 in Figure 1) corresponding to the ex ante choice (at node α_1): dynamic consistency in such cases implies that the investor has the same preferences at both nodes. However, this is not the case for the trees in Figure 2, where two ex post choice nodes (β and γ) correspond to the ex ante one (α). In these trees, the investor could prefer f_1 at node β, but g_1 at node γ, in which case the preferences in the planned-for contingencies differ. But then either ex ante preference (for f_1 or g_1) agrees with the preference in some planned-for contingency, and thus satisfies the harmony demanded by dynamic consistency applied on these trees. So it does not follow from a preference for f_1 at node β that the investor must prefer f_1 at α:[1] the standard argument does not hold when the decision maker's subjective tree differs from the theorist's objective one. Moreover, a technical result, pinpointing a family of cases in which dynamic consistency on subjective trees is satisfied (Proposition 1, Section 3), shows that not only is the argument fallacious, its conclusion is false. Non-expected utility is consistent with consequentialism and dynamic consistency formulated in terms of the contingencies the decision maker envisages.

Furthermore, and perhaps more surprisingly, we show that, even under the assumption that the decision maker's subjective tree corresponds to an objective one, the anti-ambiguity argument still fails. Under a correctly formulated notion of dynamic consistency, we prove that this assumption implies that the decision maker's ex ante beliefs must be of a certain form (Proposition 2, Section 4) – basically, the equivalent in our framework of Epstein and Schneider's (2003) rectangularity (see also Sarin and Wakker, 1998). It follows that, if the assumption is correct, he holds beliefs that do not generate the Ellsberg preferences, and hence he cannot be embarrassed by the argument (Section 4). Moreover, by the contrapositive implication, whenever he does exhibit the Ellsberg preferences, it follows that the assumption does not hold, so he is not using the tree in Figure 1, and the argument does not apply.

This latter observation yields the second main conceptual contribution of the paper, and the most important for applications in decision analysis, normative economics and economic modelling. It is standard to assume that the decision maker is using the same (typically objective) decision tree as the analyst or the theorist. But our approach shows that this assumption in itself implies that the decision maker's ex ante beliefs must be of a particular form. What if they are not of this form? Under the perspective developed here, this does not necessarily reveal any dynamic inconsistency on the decision maker's part, but simply tells us that the decision maker and the theorist may in fact be using different trees. For economic applications, our approach thus provides a new justification for the restriction to specific families of beliefs, such as rectangular sets of priors

[1] Clearly, similar points hold for any of the previously discussed cases: where, in addition to potentially learning whether the performance is bad or not, the investor thinks she might learn whether it is good or not, or full details of performance, or obtain a partially informative report. In all these cases, because of other nodes at the ex post stage, dynamic consistency does not imply, in general, that ex ante preferences coincide with preferences in the contingency where she learns that the performance is not bad (RG).

à la Epstein and Schneider, based on the standard assumption that the theorist has correctly modelled the problem the decision maker considers himself to be faced with. In decision analysis, such a restriction is justified by the assumption that the analyst and the decision maker have settled on a relevant decision tree for the problem, which they both understand. Since such restrictions are the key to treating sequential choice problems, the approach developed here thus upholds a solid, principled account of dynamic choice under non-expected utility.

Our final conceptual contribution concerns information-acquisition decisions, which are often taken to epitomise the difficulties facing ambiguity models, since they purportedly enjoin decision makers to turn down free information in certain situations (Wakker, 1988; Kadane et al., 2008; Al Najjar and Weinstein, 2009; Siniscalchi, 2011). Our approach provides a novel reconceptualisation of the issue, which defuses the purported difficulties by revealing the charge to rest implicitly on an incorrect calculation of the value of information. We characterise the value of information on subjective trees, and show that the only cases in which a non-expected utility decision maker will turn down information is if it has an opportunity cost – it compromises information that he otherwise expected to receive. Hence the information is in fact not really 'free', and its rejection no longer counts as unreasonable.

The paper is organised as follows. Section 2 describes the framework and the representation of envisaged contingencies. Section 3 defines the notion of dynamic consistency on subjective trees and draws some immediate consequences. Section 4 examines consequences of the assumption that the decision maker's subjective tree corresponds to an objective one, in particular for the application of ambiguity models to dynamic choice problems. Section 5 analyses information-acquisition decisions. Section 6 discusses outstanding issues, related arguments and literature not mentioned elsewhere. Proofs and other material are collected in the Appendices.

2 Preliminaries

2.1 Framework

We use a version of the Savage (1954) framework, in which most studies of dynamic consistency under uncertainty are formulated. Let S be a non-empty finite set of *states*; as standard, this 'objective' state space is assumed to be given, and represents all payoff-relevant factors. Subsets of S are called *events*. $\Delta(S)$ is the set of probability measures on S. The set of *consequences* shall be taken to be a real interval C (which may be interpreted as monetary payoffs).[2] *Acts* are state-contingent consequences (i.e. functions from S to C); $\mathcal{A}$ is the set of acts. With slight abuse

[2]The results continue to hold for richer consequence spaces, such as the space of lotteries à la Anscombe-Aumann.

of notation, a constant act taking consequence c in every state will be denoted c and the set of constant acts will be denoted C.

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

We implicitly restrict attention to two time periods: the current (ex ante) one and the future (ex post) one. To formalise contingencies, let M be a (grand) set of all possible messages or signals that could be received between the two time periods. We assume, as is standard in the literature on dynamic consistency, that signals are not directly payoff-relevant. One could define an 'extended state space' $S \times M$ comprising both payoff-relevant states and signals, and study acts over such a space whose consequences only depend on the first coordinate (S); all arguments and results presented below go through when the concepts involved are properly defined in such a context. Appendix A provides some technical details for the interested reader.

For each event $A \subseteq S$, the (unique) signal stating only that A holds is said to *correspond to* A. Let M^{event} be the set of signals corresponding to events in S; for each $m \in M^{event}$, we denote the corresponding event by A_m. We assume that $M^{event} \subseteq M$: for every event $A \subseteq S$, there exists $m \in M$ corresponding to A.

2.2 *Ex ante and ex post preferences*

We consider a single decision maker; $\geq$ denotes his current preference relation and, for each $m \in M$, $\geq_m$ denotes the ex post preferences he currently expects he would have in contingency m (i.e. after having received signal m). To respect the meaning of learning an event, we assume that, whenever a signal m corresponds to an event A_m, A_m^c is $\geq_m$-null: in the contingency where A_m is learnt, the decision maker's preferences do not depend on what happens off A_m.

For concreteness, 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).[3] According to this rule, a preference relation $\geq'$ is represented by a closed, convex set of priors $\mathcal{C}' \subseteq \Delta(S)$ and utility function u', with, for every $f, g \in \mathcal{A}$, $f \leq' g$ if and only if

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

[3] The main points and results continue to hold for many other non-expected utility models; see Section 6. Sarin and Wakker (1998) discuss the assumption that preferences are formed according to the same decision rule ex ante and ex post.

Expected utility preferences correspond to the special case where the set of priors is a singleton.

To focus on the issue of dynamic (in)consistency related to non-expected utility, we assume throughout that the same continuous increasing utility function $u : C \to \mathbb{R}$ is involved in the representation of all preferences, ex ante and ex post. We denote by $\mathcal{C}$ the closed convex set of priors representing the decision maker's current preferences according to (1); they can be thought of as capturing his current beliefs about the state of the world.[4] For each $m \in M$, the closed convex set $\mathcal{C}_m$ representing $\succeq_m$ according to (1) reflects the beliefs he currently expects to have after having received signal m – that is, the beliefs he currently expects to have in the contingency corresponding to m. So, as standard (Gollier, 2004, Ch 24), each signal $m \in M$ is associated with an ex post belief after receiving m, though, unlike the Bayesian approach, the ex post belief is represented by a set of priors.

2.3 Envisaged contingencies

The decision maker may envisage some contingencies – i.e. consider the contingencies, and hence the corresponding signals, as possible – whereas others he simply ignores or considers cannot occur. In the bulk of the paper, we work under the assumption that the set of contingencies the decision maker envisages is given. (For readers uncomfortable with this assumption, we provide a choice-theoretic foundation for this set in Appendix B.) Let $I \subseteq M$ be the finite set of the decision maker's 'envisaged signals'. This generates the set $\{\mathcal{C}_i\}_{i \in I}$ of beliefs the decision maker anticipates himself as possibly having at the ex post period. This set, or equivalently (under the previous assumptions) the set of envisaged ex post preferences $\{\succeq_i\}_{i \in I}$, fully characterises the decision maker's envisaged contingencies.

Note that the decision maker's 'qualitative' beliefs about the contingencies he could possibly be in at the ex post period, represented by I, are all that is required for the treatment in Sections 3 and 4 below. The representation here could be thought of as a 'reduced form' of a richer representation (e.g. in terms of the extended state space $S \times M$, the envisaged contingencies could be taken to be those receiving non-zero ex ante probability), though nothing in the formal development below rests on such an assumption.

In summary, the decision maker is represented by a triple $(\succeq, \{\succeq_m\}_{m \in M}, I)$ of ex ante preferences, ex post preferences after receiving any signal $m \in M$, and envisaged signals or contingencies I. Note that the notion of dynamic consistency defended below only involves ex post preferences in envisaged contingencies ($\{\succeq_i\}_{i \in I}$), and most of the discussion in the sequel will be

[4]For simplicity, we abstract from the debate about the extent to which, given the lack of separability of beliefs and ambiguity attitude in some of these models, the sets of priors can be thought of as (purely) beliefs.

conducted in terms of those. We introduce ex post preferences in contingencies that the decision maker expects not to be in (i.e. $\geq_m$ for $m \notin I$) only to formulate the standard dynamic consistency condition used in the literature (see Section 3).

2.4 *Assumptions and special cases*

The framework set out above is very general as concerns the information structure, the only assumption being the existence of (well-behaved) signals corresponding to events. In particular, it is not assumed that every signal corresponds to an event in the payoff-relevant state space S.[5] Indeed, many standard information structures (e.g. Blackwell, 1953; Gollier, 2004, Ch 24) do not involve such an assumption: for instance, they allow ex post probability distributions to be full support (so they cannot result from updating on an event). We could restrict the envisaged contingencies I to signals corresponding to learning events in S (i.e. specify that $I \subseteq M^{event}$) without affecting our results or arguments.

Furthermore, no specific assumptions have been made about the relationship between $\mathcal{C}$, $\mathcal{C}_m$ and I. One conceivable sort of assumption concerns update. For instance, for $m \in M^{event}$ and a given update rule for sets of priors, one could assume that the ex post set of priors $\mathcal{C}_m$ is the result of updating $\mathcal{C}$ on A_m. We make no such assumption here. On the one hand, it is unclear how to formulate such an assumption for signals that do not correspond to events ($m \notin M^{event}$): for instance, even Bayes' rule applied on such signals typically requires conditional probabilities of signals given states, and these are not provided in our framework. On the other hand, since the issue of dynamic consistency is related to the choice of update rule (e.g. Hanany and Klibanoff, 2007), it may be question-begging to assume an update rule for a general investigation on the compatibility between dynamic consistency and non-expected utility. Another sort of assumption concerns the way that $\mathcal{C}$ 'aggregates' the envisaged ex post beliefs, $\{\mathcal{C}_i\}_{i \in I}$. Following standard representations of information, one could have assumed a richer structure (e.g. a probability measure or set thereof) over the space of signals M, and some connection between it, $\mathcal{C}$ and $\{\mathcal{C}_i\}_{i \in I}$. Such assumptions are again not adopted below. Some of our results suggest that dynamic principles may have consequences for the relationship between $\mathcal{C}$ and $\{\mathcal{C}_i\}_{i \in I}$ (see Section 4 and Proposition C.2, Appendix C); systematic exploration is however beyond the scope of the current paper (Section 6).

2.5 *Subjective and objective trees*

As stated in the Introduction, the decision maker's subjective tree consists of the contingencies he envisages: it is characterised by the set I of envisaged signals. An objective tree is a tree

[5] By contrast, every signal evidently corresponds to an event in (i.e. subset of) the 'extended state space' $S \times M$.

corresponding to a partition of the payoff-relevant state space S. Formally, a finite subset $J \subseteq M$ characterises an *objective tree* if $J \subseteq M^{event}$ (each signal corresponds to an event) and $\{A_j : j \in J\}$ form a partition (the events learnt form a partition).[6] So, for the state space in Ellsberg-style example in the Introduction, the trees in Figure 1 form an objective tree, whereas those in Figure 2 do not. We denote the set of objective trees by $\mathcal{O}$. Moreover, for any partition $\mathcal{P}$ of S, $J_{\mathcal{P}} = \{j \in M^{event} : A_j \in \mathcal{P}\}$ characterises the corresponding objective tree.

As we shall see, objective trees underpin the standard formalisation of dynamic consistency, as well as more general discussions of dynamic choice in the context of ambiguity. However, it should be noted that they constitute a special class of the space of all possible trees – i.e. of all sets of signals that the decision could possibly envisage. For instance, if a decision maker is sure that he will learn some event of the state space, but does not know which one, then his information structure is not an objective tree: in terms of the previous example, the relevant set of signals corresponds to the set of events $\{R, B, G, RB, RG, BG, RGB\}$, which does not form a partition.[7] Moreover, if J_1 and J_2 are two objective trees, then their union is generally not an objective tree. An investor in the example who envisages learning only whether the performance is regular or not is using an objective tree (corresponding to the partition $\{R, BG\}$), as is one who envisages learning only whether it is bad or not ($\{B, RG\}$); however, a decision maker who thinks she might learn whether the performance is regular or not *or* whether it is bad or not is *not* using an objective tree (the union is not a partition). More generally, a decision maker who knows that the 'real' information structure corresponds to an objective tree $J \in \mathcal{O}$, but does not know which one, cannot be represented as facing an objective tree. As these examples illustrate, the restriction to objective trees is a strong one, even if one is only considering signals corresponding to events. Prima facie, there is no reason why a decision maker should restrict to such trees, unless, of course, he has been informed that these are the only possible signals he could receive.

2.6 *Consequentialism*

In its natural-language formulation, consequentialism states that the decision maker's ex post or conditional preference does not depend on branches in the decision tree that are no longer accessible. Since, in our framework, the ex post preferences depend only on the ex post set of priors and the utility function – and not, for instance, on the tree (subset of M) in which they are obtained – consequentialism is automatically satisfied here. A standard formalism of the principle (e.g. Ghirardato, 2002) states that, for every case of learning an event, $m \in M^{event}$, the event A^c_m is

[6]A set of events $\{A_j\}_{j \in J}$ in S form an partition if $A_{j_1} \cap A_{j_2} = \varnothing$ for all $j_1, j_2 \in J$ and $\bigcup_{j \in J} A_j = S$.

[7]Throughout, we use the notation introduced in Figures 1 and 2.

null according to the preferences conditional on A_m, $\succeq_m$. This notion only looks at contingencies corresponding to events, and is automatically satisfied in our setup for such contingencies (by the assumptions in Section 2.2). However, as noted, consequentialism is satisfied in our framework even in cases where the standard formalisation does not apply, and in particular in contingencies that do not correspond to learning events in S. The subsequent discussion will implicitly assume consequentialism (in its standard formalisation, where appropriate) without explicit mention.

3 Dynamic consistency on subjective trees

We now introduce the formal definition of dynamic consistency on subjective trees. It is reasonable to begin with the standard dynamic consistency condition considered in the literature, which is formulated over objective trees.[8] Despite considerable differences between authors, the following is the reformulation in our setup of a fairly representative condition.

Standard Dynamic Consistency (SDC). $(\succeq, \{\succeq_m\}_{m\in M}, I)$ satisfies *Standard Dynamic Consistency (SDC)* if, for every $f, g \in \mathcal{A}$ and $J \in \mathcal{O}$, if $f \preceq_j g$ for all $j \in J$ with A_j $\preceq$-non-null, then $f \preceq g$, and moreover, if any of the $\preceq_j$ orderings are strict, then so is the $\preceq$ one.

Another standard condition used in the literature (e.g. Ghirardato, 2002) is the special case of SDC applied to all $J \in \mathcal{O}$ containing only two signals $\{j_1, j_2\}$, with $f(s) = g(s)$ for all $s \in A_{j_2}$. It is straightforward to show that, under weak ordering, this condition is equivalent to SDC.[9]

Standard Dynamic Consistency (SDC) captures the idea that when faced with an objective tree (J), constituted by a partition of events ($\{A_j\}_{j\in J}$), if the decision maker prefers 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 over contingent plans should be coherent with his ex post preferences in the relevant contingencies, under the assumption that the relevant contingencies correspond to the events of a given partition. As argued previously, there is no guarantee that any such partition correctly represents the contingencies the decision maker in fact envisages, and hence plans for when forming his ex ante preferences. When it does not, this condition is unreasonable: the ex post preferences that count are those in the contingencies actually envisaged by the decision maker – they are the anticipated future preferences $\{\succeq_i\}_{i\in I}$ – rather than those in the

[8] We focus in this paper on the dynamic consistency of preferences, rather than the dynamic consistency of behaviour (Strotz, 1955; Karni and Safra, 1989, 1990; Siniscalchi, 2009, 2011).

[9] Some researchers use more restrictive notions of dynamic consistency, where the ex ante preferences are over plans or trees, rather than acts (McClennen, 1990; Sarin and Wakker, 1994, 1998; Siniscalchi, 2011); the SDC condition stated here is obtained from such notions by adding a complementary 'reduction' or 'invariance' property.

contingencies corresponding to the events imposed by the theorist – the $\{\succeq_j\}_{j\in J}$ for any particular $J \in \mathcal{O}$. It is straightforward to modify SDC to apply to envisaged contingencies, and subjective trees.

Dynamic Consistency (DC). $(\succeq, \{\succeq_m\}_{m\in M}, I)$ satisfies *Dynamic Consistency (DC)* if, for all $f, g \in \mathcal{A}$, if $f \preceq_i g$ for all $i \in I$, then $f \preceq g$, and moreover, if any of the $\preceq_i$ orderings are strict, then so is the $\preceq$ one.

This is a straightforward replacement, in Standard Dynamic Consistency, of the preferences conditional on events in a given partition by the anticipated future preferences. It corresponds closely to the English-language formulation of the dynamic consistency principle, which states that the decision maker's preferences over contingent plans agree with his preferences in the planned-for contingencies. The required harmony translates into a matching between the preferences he anticipates having in the contingencies he actually envisages – and plans for – and his ex ante preferences, which reflect his attitudes to plans for these contingencies.

The only real difference between DC and SDC is conceptual: the use of subjective rather than objective trees. Any other apparent difference between the conditions as formulated is technical, and largely an artefact of the setup. This is notably the case for the universal quantification over (objective) trees, which at first seems to appear in SDC but not in DC: as shown in Appendix A, when SDC (with its universal quantification) is formulated appropriately in the extended state space $S \times M$ mentioned in Section 2.1, it turns out to be equivalent to DC. Moreover, the non-nullness condition in SDC corresponds to the $i \in I$ condition in DC: both leave out future contingencies which the decision maker does not consider possible. See Appendix A for further discussion.

We claim that DC is the appropriate formalisation of the dynamic consistency principle, and hence more adequate for discussion of the consequences of non-expected utility in dynamic situations. Our aim is not to defend the principle itself, but rather to show that, once properly formulated, it ceases to cause any embarrassment for ambiguity models. We begin with the following fact (recall that $\{\succeq_i\}_{i\in I}$ characterises the contingencies envisaged by the decision maker).

Proposition 1. For any closed convex set $\mathcal{D}$ of probability measures on I with $p(i) > 0$ for all $i \in I$ and $p \in \mathcal{D}$, let $\succeq_{\mathcal{D}}$ be the preference relation represented according to (1) with the set of priors $\mathcal{C}_{\mathcal{D}} = \{\sum_{i\in I} p(i).q_i |\, p \in \mathcal{D},\ (q_i)_{i\in I} \in \prod_{i\in I} \mathcal{C}_i\}$. Then $(\succeq_{\mathcal{D}}, \{\succeq_m\}_{m\in M}, I)$ satisfies Dynamic Consistency.

As an illustration of this result, let us return to the example in the Introduction, and consider an investor with the subjective tree in Figure 2 and envisaged beliefs as given in Table 2, where $\mathcal{C}_B$

	Set of priors	Preferences	
$\mathcal{C}_B$	$\{(0,1,0)\}$	$f_1 \sim_B g_1$	$f_2 \sim_B g_2$
$\mathcal{C}_{RG}$	$co\{(\frac{1}{3},0,\frac{2}{3}),(1,0,0)\}$	$f_1 \succ_{RG} g_1$	$f_2 \succ_{RG} g_2$
$\mathcal{C}_{RGB}$	$co\{(\frac{1}{3},0,\frac{2}{3}),(\frac{1}{3},\frac{2}{3},0)\}$	$f_1 \prec_{RGB} g_1$	$f_2 \succ_{RGB} g_2$
$\mathcal{C}_{\mathcal{D}}$	$co\left\{\begin{array}{l}(\frac{5}{18},\frac{1}{6},\frac{10}{18}),(\frac{5}{18},\frac{11}{18},\frac{1}{9}),(\frac{7}{18},\frac{11}{18},0),\\(\frac{17}{36},\frac{19}{36},0),(\frac{17}{36},\frac{1}{12},\frac{4}{9}),(\frac{11}{36},\frac{1}{12},\frac{11}{18})\end{array}\right\}$	$f_1 \prec_{\mathcal{D}} g_1$	$f_2 \succ_{\mathcal{D}} g_2$

Table 2: Sets of priors and corresponding preferences.
We adopt the notation introduced in the Introduction (Figures 1 and 2). Priors are defined over the payoff-relevant state space $S = \{R, B, G\}$, with (r, r', r'') denoting the probability measure $p \in \Delta(S)$ such that $p(R) = r, p(B) = r'$ and $p(G) = r''$ and $co(\mathcal{C})$ for a set $\mathcal{C} \subseteq \Delta(S)$ denoting the convex closure of $\mathcal{C}$.[11] The preferences over the acts in Table 1 indicated here are those generated, according to (1), by the sets of priors and any utility function with $u(10) > u(0)$.

(respectively $\mathcal{C}_{RG}$, $\mathcal{C}_{RGB}$) is the set of priors after learning B (resp. RG, RGB).[10] The table also shows the generated ex post preferences. Note in particular that the investor prefers f_1 over g_1 at node β in Figure 2 (after learning RG), but prefers g_1 over f_1 at node γ (after RGB). Since the ex post preferences concerning f_1 and g_1 disagree, DC implies nothing about the ex ante preferences over these two acts. By contrast, SDC applied on the objective tree in Figure 1, containing only the ex post preferences $\succeq_B$ and $\succeq_{RG}$ implies that f_1 should be preferred ex ante.

The last row in the table shows the set $\mathcal{C}_{\mathcal{D}}$ defined as in Proposition 1 with $\mathcal{D} = \{p \in \Delta(\{B, RG, RGB\}) : p(RGB) = \frac{2}{3}, p(B) \in [\frac{1}{12}, \frac{1}{6}]\}$. The generated preference relation agrees with $\succeq_{RGB}$ on the acts in question (Table 2). So $\mathcal{C}_{\mathcal{D}}$ is a non-singleton set representing a non-expected utility preference relation $\succeq_{\mathcal{D}}$ exhibiting Ellsberg preferences, which, by Proposition 1, satisfies what we would argue is the properly formalised version of dynamic consistency, DC.

Proposition 1 provides the first take-home message of the paper: once formulated with the contingencies the decision maker himself envisages, dynamic consistency ceases to be inconsistent with non-expected utility and consequentialism (which, as noted in Section 2.6, is automatically satisfied in the framework used here). Rather, no matter the non-singleton set of contingencies the decision maker envisages, 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 $\mathcal{D}$. This is in stark contrast to the classic argument discussed in the Introduction, according to which (under basic

[10]Note that these sets are largely consistent with the interpretation of the signals, as well as with the details of the example provided in the Introduction: for instance, all members of $\mathcal{C}_B$ (resp. $\mathcal{C}_{RG}$) have support in B (resp. RG).

[11]In this context, the convex closure of $\mathcal{C}$ is the set of all mixtures of all members of $\mathcal{C}$.

assumptions on preferences) dynamic consistency, consequentialism and non-expected utility are incompatible. That argument relies on a formalisation of dynamic consistency on objective trees (SDC) that, we claim, inappropriately captures the sense of the principle. With a properly formalised version of the principle (DC), the alleged incompatibility – and associated embarrassment for ambiguity models – disappears.[12]

4 Dynamic consistency and dynamic choice problems

Whatever the conceptual significance of the decision maker's envisaged contingencies for dynamic consistency, the point might seem irrelevant for applications. It is standardly assumed that the decision maker *knows* which tree he is facing, and that this *is* the (typically) objective tree used by the economist or decision analyst – either because the analyst and decision maker have constructed the tree together, or because this is a standard modelling assumption in economics. Indeed, one might try to brush off the points made above by simply assuming that the decision maker knows that he is faced with an exogenously given objective tree. For instance, in our initial example, one might just assume that the investor knows that the company will only report whether the performance is bad or not – that is, she knows that she is facing the tree in Figure 1. Under this assumption, the standard argument appears to go through, so the conflict between non-expected utility and dynamic principles would seem to resurface, bringing with it all of the damning conclusions for the use of non-expected utility models. In this section, we first show that this revamped version of the standard argument still fails: the proposed approach copes comfortably with the mooted assumption. Moreover, our analysis brings out some important consequences for applications of ambiguity models in dynamic choice problems.

In the framework set out above, the assumption that the decision maker's subjective tree corresponds to an objective one with partition $\mathcal{P}$ is formalised as $I = J_{\mathcal{P}}$. (Recall from Section 2.5 that $J_{\mathcal{P}}$ characterises the objective tree corresponding to $\mathcal{P}$.) If this holds, we say that I is *$\mathcal{P}$-objective*.

Like any assumption, the assumption that the decision maker is using a particular sort of decision tree may have consequences in and of itself. As the following result shows, it has rather strong implications for the decision maker's ex ante beliefs.

Proposition 2. Let $(\geq, \{\geq_m\}_{m\in M}, I)$ satisfy DC. Suppose moreover that, for some partition $\mathcal{P}$, I is $\mathcal{P}$-objective. 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 $\mathcal{C} = \{\sum_{i\in I} p(i).q_i |\ p \in \mathcal{D}, (q_i)_{i\in I} \in \prod_{i\in I} \mathcal{C}_i\}$.

[12] Proposition C.2 and Remark C.1 (Appendix C) identify necessary and sufficient conditions for $(\geq, \{\geq_m\}_{m\in M}, I)$ to satisfy Dynamic Consistency and show that preferences of the form in Proposition 1 are not the only ones doing so.

For a partition $\mathcal{P}$ and a set of priors $\mathcal{C}$, we say that $\mathcal{C}$ is *$\mathcal{P}$-rectangular* if there exist a set $\mathcal{C}_0$ of probability measures on $\mathcal{P}$ and sets $\mathcal{C}_j$ of probability measures with support in A_j, one for each $A_j \in \mathcal{P}$, such that $\mathcal{C} = \{\sum_{A_j \in P} p(A_j).q \mid p \in \mathcal{C}_0,\ q \in \mathcal{C}_j\}$. 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 cell in $\mathcal{P}$. It is basically the equivalent in the present setup of the notion of rectangularity defined by Epstein and Schneider (2003) (see Section 6).[13]

Proposition 2 thus tells us that, under Dynamic Consistency, there is a strong relationship between the subjective tree the decision maker thinks he is faced with and his ex ante beliefs: whenever the former corresponds to an objective tree (it is $\mathcal{P}$-objective), the latter has a special shape (it is $\mathcal{P}$-rectangular). This may be understood conceptually as a 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 beliefs (the contingencies he envisages). As such, it is not surprising: one might expect one's current beliefs about an issue to be coherent with what one believes one will believe about it in the future.[14] It is also unsurprising that Dynamic Consistency – which requires a particular harmony between current and envisaged future preferences – implies a certain coherence between current and envisaged future beliefs.

Perhaps more important than the preceding direction of the implication is its contrapositive, which we state explicitly.

Corollary 1. Let $(\succeq, \{\succeq_m\}_{m \in M}, I)$ satisfy DC. For any partition $\mathcal{P}$, if $\mathcal{C}$ is not $\mathcal{P}$-rectangular, then I is not $\mathcal{P}$-objective.

In other words, if the decision maker's ex ante set of priors does *not* have the particular rectangular shape, then he does *not* think that he will necessarily learn exactly one event in the partition $\mathcal{P}$ and only that. This result provides a central conceptual insight: under DC, an analyst can draw conclusions about the decision maker's subjective tree on the basis of a property of his current beliefs about the state of the world.[15]

[13]Note that if $\mathcal{C}$ is $\mathcal{P}$-rectangular with $\mathcal{C}_0$ and $\mathcal{C}_j$ as specified in the text, then $\mathcal{C}_0 = \{p^{+1} | \ p \in \mathcal{C}\}$, where p^{+1} is the restriction of p to $\mathcal{P}$, and $\mathcal{C}_j = \{p(\bullet|A_j) | p \in \mathcal{C},\ p(A_j) > 0\}$. So $\mathcal{C}$ is rectangular over the partition $\mathcal{P}$ in the sense of Epstein and Schneider (2003, Definition 3.1). Note also that, since the ex post sets of priors in Proposition 1 do not necessarily have disjoint supports, the ex ante set of priors in that Proposition is not necessarily rectangular over any (non-trivial) partition.

[14]Indeed, such coherence is reminiscent of that defended by some philosophers under the name of the 'Reflection' principle (van Fraassen, 1984).

[15]Note that it can be straightforwardly verified on inspection whether a given set of priors is $\mathcal{P}$-rectangular or not: it suffices to check, for all $A_i \in \mathcal{P}$, whether the sets $\{p(\bullet|A_i) \mid p \in \mathcal{C},\ p(A_i) = x\}$ are the same for all x for which they are non-empty.

This insight exposes the fault in the anti-ambiguity argument at the beginning of this section. Note that no set of priors generating Ellsberg preferences can be $\{B, RG\}$-rectangular. So, by the Corollary, any decision maker satisfying the refined notion of dynamic consistency proposed here and exhibiting Ellsberg preferences does not think that he is facing the objective tree in Figure 1. Hence the anti-ambiguity argument, which assumes that the decision maker is using that tree, does not apply. In particular, the fact that the investor in the example has Ellsberg preferences may itself be an indication that she does not think she is facing the objective tree in Figure 1. So one *cannot* simply assume that she knows she is faced with this tree: for, under the very notion of dynamic consistency that one would like her to satisfy, this assumption is incompatible with her ex ante preferences. Herein lies the error in the argument: it is based precisely on such an assumption. To embarrass the investor with the standard argument, it needs to be established *not only* that she has Ellsberg preferences, *but also* that she simultaneously envisages precisely the contingencies depicted in the objective tree in question (see also Section 6).

A second contribution of this insight concerns economic applications using ambiguity models. For the maxmin expected utility model,[16] it provides a new justification for the use of sets of priors that are rectangular with respect to the partition formed by the nodes in the tree. Rectangular priors have been promoted since the work of Sarin and Wakker (1998); Epstein and Schneider (2003); Riedel (2004), but existing justifications are often considered partial at best, and *ad hoc* at worst. They generally operate by assuming a fixed partition (objective tree) and showing that, under some basic conditions, the standard version of dynamic consistency applied on it implies rectangularity of the ex ante set of priors. Such results are standardly read as tying the fate of rectangularity to that of dynamic consistency. But the standard version of dynamic consistency over particular partitions is violated by non-expected utility decision makers in some situations; in such cases, the justification fails and the set of priors may not be rectangular.[17] This point has been cited as both a motivation for new update rules for maxmin EU preferences and as a weakness of the rectangularity-based approach (e.g. Hanany and Klibanoff, 2007, p282; Al Najjar and Weinstein, 2009, §3.4). Indeed, many applications, including normative ones pertaining to monetary or environmental policy (Hansen and Sargent, 2008; Brock and Hansen, 2017) as well as studies on po-

[16] These points hold *mutatis mutandis* for other major non-expected utility models; see Section 6.

[17] Indeed, Epstein and Schneider's own conclusion (p14), in the context of the Ellsberg example discussed in the Introduction, is that 'in *some settings*, ambiguity may render dynamic consistency problematic.' Sarin and Wakker note that 'The main point of our approach is that a decision maker may not universally commit to any of the principles of dynamic consistency, consequentialism, or invariance, but may violate each of them in certain specific situations.' This still leaves the normative issue unresolved; indeed, as Wakker (1999) states: 'As a personal opinion, I find all dynamic principles [implying expected utility] normatively imperative.'

larization (Baliga et al., 2013) and mechanism design (Bose and Renou, 2014), employ alternative approaches to dynamic choice.

The justification of rectangularity derived from Proposition 2 involves the arguably more appropriate form of dynamic consistency proposed in Section 3, and gives pride of place to the previously neglected assumption that the decision maker is using the same tree as the theorist. If the set of priors is not rectangular, this does not necessarily mean that the decision maker is dynamically inconsistent, in the refined sense given in Section 3. Rather, it could simply imply that the decision maker and the theorist are not using the same decision tree. So the restriction to appropriately rectangular sets is justified by the assumption that the decision maker does indeed know the tree he is faced with, and that it is the same objective tree as the theorist is using. As such, there is nothing arbitrary or unnecessarily limitative about it. On the contrary, it is essential in normative or prescriptive applications: if the analyst and the decision maker have not agreed upon the relevant decision tree, the analysis is almost certainly doomed! Even in purely descriptive applications, without such assumptions about the way decision makers conceive the situations they are faced with, economic modelling can hardly get off the ground. Moreover, the question of what a decision analyst or economist should do when faced with a non-rectangular ex ante set of priors receives an equally simple answer: she should go back and recheck her model, because the non-rectangularity itself suggests the tree she is using is not necessarily the one the decision maker implicitly has in mind.

In summary, 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. On the contrary, it provides a novel, reasoned defence of rectangular sets of priors as the appropriate tool for modelling rational agents who can be assumed to know the (objective) tree they are facing.[18] As is well-known, this approach can be thought of as a non-expected utility analogue of standard Bayesian methods, with rectangular sets of priors being constructed from sets of priors on branches of a decision tree in much the same way as for Bayesian probability measures (Raiffa, 1968). More generally, since there is no obstacle to non-expected utility decision makers being dynamically consistent with respect to the tree they are using, ambiguity models can be used in dynamic problems with standard techniques such as backwards induction reasoning. However, when using these models, the theorist or analyst must respect the consequences of her assumptions about the tree the decision maker takes himself to be

[18]Of course, for applications where the theorist's tree does not correspond to a partition (e.g. where signals are probabilistically related to states) the assumption that the decision maker knows this tree does not imply rectangularity, although it may have other implications. For applications specifically involving agents who do not understand the tree they are faced with or have limited rationality, alternative approaches, such as those mentioned above, may be required.

facing: the lesson of Proposition 2 is that many of the purported problems for ambiguity models may just boil down to modelling errors on the part of the theorist.

5 Value of Information

Unlike the cases discussed above, many economic situations involve sequential decisions. An important subclass involve the choice of information acquisition prior to a decision. These are sometimes taken to pose the toughest normative challenge to ambiguity models, due to the argument that sophisticated non-expected utility decision makers necessarily display *information aversion*: they prefer to turn down an offer of free information in some situations (Wakker, 1988; Kadane et al., 2008; Al Najjar and Weinstein, 2009). We now apply the proposed approach to these situations.

5.1 Information aversion or moral hazard?

The argument can be formulated on a sequential extension of the example given in the Introduction. Consider the investor with standard Ellsberg preferences – she prefers g_1 to f_1 at node α_1 in Figure 1(a), but prefers f_{RG} to g_{RG} at node β_1 in Figure 1(b) – and suppose that she is offered the choice between facing these two decision trees.[19] That is, she is faced with the decision tree in Figure 3. Given the preferences just specified, she knows that if she reaches decision node α she will choose g_1 and that if she reaches decision node β she will choose f_{RG}. Reasoning by backwards induction,[20] at node δ she knows that if she takes the upper branch NL of the tree, she will end up with g_1, and if she takes the lower branch L she will essentially end up with f_1 (since she does not know the resolution of the uncertainty at the nature node $*$, the choice of f_{RG} at node β essentially boils down to a choice of f_1 from the point of view of node δ). Hence, since she prefers g_1 to f_1, she chooses NL at δ. However, since the choice at δ is essentially that between learning whether the performance of the drug is bad or not before deciding to invest (option L) or not learning (option NL), by choosing NL the investor betrays a preference for not obtaining free information. This is the alleged information aversion.

What does the proposed approach have to say about this argument? Since this is an information-acquisition problem, the investor knows whether she will receive the information *after* having made her decision; we focus on the simplest case and assume that her subjective tree at node $*$ coincides with the objective tree in Figure 3. For a decision maker satisfying DC, Proposition 2 thus applies,

[19] In this discussion, notations and numbering are taken from the Introduction.

[20] Decision makers who reason in this way 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).

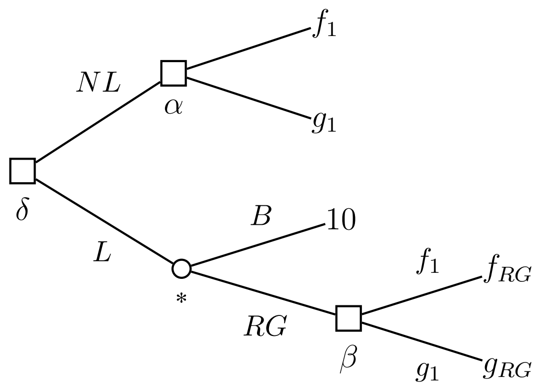


Figure 3: Information Aversion.
Notation as in Figure 1.

yielding conclusions about her sets of priors at nodes α and $*$ – that is, her beliefs immediately after having taken the information-acquisition decision (at node δ) but before any of the promised information has been received. On the one hand, since she has the standard Ellsberg preferences at node α, her set of priors at α *cannot* be $\{B, RG\}$-rectangular (Section 4). On the other hand, since she knows that she is facing the objective tree over the partition $\{B, RG\}$ at node $*$, her set of priors at $*$ *must* be $\{B, RG\}$-rectangular (Proposition 2). So she has different sets of priors at nodes α and $*$: her beliefs immediately after her information acquisition choice depend on the choice she makes.

Despite first appearances, such dependence is a natural consequence of the coherence between current and envisaged future beliefs that is imposed by dynamic consistency (Section 4). If any choice can directly determine one's beliefs, it is the choice whether to obtain information or not: it gives immediate 'second-order' information about what one will believe, and as such directly impacts one's current beliefs about one's possible future beliefs. Since, under dynamic consistency, envisaged future beliefs and current beliefs are strongly related, this may have a knock-on impact on current beliefs (about the state of the world). In demanding that a decision maker satisfy dynamic consistency, the potential dependence of post-choice beliefs on the choice made in information-acquisition decisions follows as a necessary consequence.

Pinpointing such dependence constitutes a first conceptual contribution of the proposed approach for information-acquisition decisions. In particular, it reveals the fault in the information aversion argument. Situations in which agents' choices can have an impact on their beliefs have been well-studied in economics, for instance in the literature on moral hazard. Such choices should be evaluated using the beliefs at the *interim* stage: in deciding whether to buy property insurance, an agent should use the probability of damage given the insurance or lack of it, and this probability may vary according to the policy purchased. So, in the previous example, the learning option

(L) should be evaluated using the beliefs at node $*$ and the NL option should be evaluated using those at node α. The information aversion argument does not do this: by using the Ellsberg preferences to evaluate both options, it proceeds as if they correctly reflected the investor's interim beliefs no matter what information-acquisition choice is made – and, as we have seen, they do not. The argument misanalyses the decision problem, relying on the erroneous assumption that the information-acquisition choice has no effect on beliefs.

5.2 *Value of information and opportunity costs*

To permit a more refined analysis, we now characterise the value of information for maxmin EU decision makers satisfying DC.

Consider a standard information-acquisition 3-period setup similar to that in the previous example. At period 0, there is a choice whether to acquire information $\mathcal{I}$, which would be delivered in period 1 for a subsequent decision in period 2; otherwise, the decision is made in period 1.[21] We use the framework set out in Section 2, with M the grand set of signals, and the decision maker using the maxmin EU rule at all periods. Let $\mathcal{C} \subseteq \Delta(S)$ be the decision maker's set of priors immediately after having chosen not to obtain the information, with $\mathcal{C}_m \subseteq \Delta(S)$ the ex post sets of priors he anticipates having after receiving signal $m \in M$ and I the set of 'envisaged' signals, both after having turned down the offered information. So $\{\mathcal{C}_i\}_{i \in I}$ are the decision maker's envisaged ex post (period 2) sets of priors just after having chosen to turn down the information $\mathcal{I}$. Similarly, let $\mathcal{C}_{(\mathcal{I})}$ be his set of priors, $\mathcal{C}_{(\mathcal{I})m} \subseteq \Delta(S)$ be the ex post sets of priors after having received $m \in M$, and K be the 'envisaged' signals, all immediately after having chosen to obtain the information but before actually receiving it (i.e. at the beginning of period 1). So $\{\mathcal{C}_{(\mathcal{I})k}\}_{k \in K}$ are the decision maker's envisaged ex post (period 2) sets of priors just after having chosen to acquire the information $\mathcal{I}$. As discussed, the two interim sets of priors $\mathcal{C}$ and $\mathcal{C}_{(\mathcal{I})}$ may differ. Moreover, I need not be a singleton, for the decision maker may envisage receiving information other than that offered; similarly, K need not coincide with the partition corresponding to $\mathcal{I}$. Extending the standard approach in the information literature (e.g. Marschak and Miyasawa, 1968) to maxmin EU preferences, we assume an 'aggregator' connecting ex post and ex ante preferences, under both information conditions.[22] That is, we assume $\phi_{(\mathcal{I})} : \mathbb{R}^K \to \mathbb{R}$ such that, for all $f \in \mathcal{A}$, $\min_{p \in \mathcal{C}_{(\mathcal{I})}} \sum_{s \in S} u(f(s))p(s) = \phi_{(\mathcal{I})}((\min_{p \in \mathcal{C}_{(\mathcal{I})k}} \sum_{s \in S} u(f(s))p(s))_{k \in K})$, and $\phi : \mathbb{R}^I \to \mathbb{R}$ such that, for all $f \in \mathcal{A}$, $\min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s) = \phi((\min_{p \in \mathcal{C}_i} \sum_{s \in S} u(f(s))p(s))_{i \in I})$.[23] As noted in Section 2, these are essentially non-expected utility information structures (Gollier, 2004, Ch 24), with

[21] If desired, one can assume that all uncertainty will be resolved and payments made in a subsequent fourth period.

[22] In the Bayesian case, the aggregator is the expected utility rule.

[23] By Proposition C.2 in Appendix C, such an aggregator exists whenever the decision maker satisfies DC.

the $\mathcal{C}_{(\mathcal{I})k}$ being the posterior set of beliefs after the reception of signal k, and $\phi_{(\mathcal{I})}$ reflecting the prior beliefs as to which signal will be received (and likewise for $\mathcal{C}_i$ and ϕ).

Calculating the value of information involves comparing 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 envisaged ex post sets of priors $\{\mathcal{C}_{(\mathcal{I})k}\}_{k\in K}$ and the opinion as to which will be realized, reflected in $\phi_{(\mathcal{I})}$. Hence the following definition of the non-negative value of information.

Definition 1. The *value of the information $\mathcal{I}$ for a (compact) menu $A \subseteq \mathcal{A}$ is non-negative* if and only if:

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

The *value of the information $\mathcal{I}$ is always non-negative* if it is non-negative for every (compact) menu $A \subseteq \mathcal{A}$.

The following result characterises when the value of information is non-negative.

Proposition 3. Suppose that the decision maker satisfies DC. Then the value of information $\mathcal{I}$ is always non-negative if and only if $\mathcal{C}_{(\mathcal{I})} \subseteq \mathcal{C}$.

This result shows that the proposed perspective can deliver non-trivial analysis of information value for non-expected utility decision makers. Moreover, on the conceptual front, it provides further insight into the alleged information aversion. For instance, in the special case where the decision maker *already expected* to receive information $\mathcal{I}$ – so $\mathcal{C} = \mathcal{C}_{(\mathcal{I})}$, $\{\mathcal{C}_i\}_{i\in I} = \{\mathcal{C}_{(\mathcal{I})k}\}_{k\in K}$, $\phi = \phi_{(\mathcal{I})}$ – the proposition implies that the information has non-negative value. So, in the simplest case of a choice between waiting for information he expects to receive or deciding before the information arrives, the non-expected utility decision maker behaves as one would expect: he always weakly prefers to wait. There is no possibility of shameful information aversion here.

More generally, note that whatever the information $\mathcal{I}$ on offer and the beliefs $\mathcal{C}$ if the offer is turned down, there exists a set of priors respecting the information structure $\mathcal{I}$ and satisfying the containment condition in Proposition 3: this is the case, for instance, for any singleton set $\{p\}$ with $p \in \mathcal{C}$.[24] So there is always a way of updating beliefs on the fact of having chosen to obtain the information under which the decision maker will assign non-negative value to it. This is a central

[24]Note that singletons are rectangular with respect to every partition, and hence satisfy any rectangularity condition imposed when $\mathcal{I}$ corresponds to a partition of S.

message of the result: the information aversion argument is not only fallacious, its conclusion – that non-expected utility decision makers are necessarily information averse – is false.

As an illustration, consider the following development of our running example. Suppose that initially the investor thinks that she will learn whether the performance is regular or not (i.e. R or BG), and is offered to learn whether B or not (as in Figure 3). Consider two cases (see Table 3 for the relevant information structures). In both cases, she thinks that, if she does not accept the information, there is a probability $\frac{1}{3}$ of learning R (first row, Table 3). In case i, she conceives the offered information about B as adding to the information she already expects to receive about R: so she anticipates learning precisely the state of the world (R or B or neither, i.e. G), and retains her probability of $\frac{1}{3}$ for learning R (second row, Table 3). In this case, the condition in Proposition 3 is satisfied ($\mathcal{C}_{*i} \subseteq \mathcal{C}_{\alpha}$) and she always weakly prefers accepting the information, for all menus. In case ii, she considers the receipt of information about B to come at the expense of the expected information about R: she expects, after accepting the offer, to learn whether B or not (B or RG), but nothing else. Moreover, she is relatively ignorant of the probability of learning B (final row, Table 3). In this case, the containment condition in the proposition is not satisfied: for instance, the probability $(\frac{4}{5}, \frac{1}{5}, 0)$ is in $\mathcal{C}_{*ii}$ but not in $\mathcal{C}_{\alpha}$. Indeed, the judgement that she would learn R with probability $\frac{1}{3}$ if she rejects the offered information, which translates to a precise probability for the state R in $\mathcal{C}_{\alpha}$, gives way a larger range of possible probabilities for this state under $\mathcal{C}_{*ii}$. By Proposition 3, the value of information will be negative for some menus in this case.[25]

As case ii illustrates and Proposition 3 confirms, there are cases where non-expected utility decision makers may turn down information. Our final conceptual contribution is to analyse them. By the Proposition, they can happen only when the set of priors after choosing to learn is *not* smaller than the set of priors after having chosen not to learn. To the extent that more informed agents are usually taken to be those with smaller sets of priors (e.g. Gajdos et al., 2008), this suggests that the decision maker does not consider the choice of learning to lead to a pure 'addition' of information with respect to the choice of not learning. In other words, he thinks that learning $\mathcal{I}$ may compromise information that he would otherwise have possessed; just as, in case ii above, the investor thinks that learning whether B would come at the expense of learning whether R. One way to verify this interpretation is by applying a comparative notion of informativeness borrowed from the literature. As already noted, the period 0 information-acquisition decision is effectively a choice between two information structures – $(\mathcal{C}, \{\mathcal{C}_i\}_{i\in I}, \phi)$ and $(\mathcal{C}_{(\mathcal{I})}, \{\mathcal{C}_{(\mathcal{I})k}\}_{k\in K}, \phi_{(\mathcal{I})})$. According to a standard definition, 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 one

[25] By straightforward calculation, one can check that this is the case for the menu $\{f_2, g_2\}$.

Ex ante set of priors		Ex post sets of priors		Envisaged contingencies	Aggregator
$\mathcal{C}_\alpha$	$co\{(\frac{1}{3},0,\frac{2}{3}),(\frac{1}{3},\frac{2}{3},0)\}$	$\mathcal{C}_R$	$\{(1,0,0)\}$	$\{R,BG\}$	$\left\{\begin{array}{l} p \in \Delta(R,BG): \\ p(R)=\frac{1}{3} \end{array}\right\}$
		$\mathcal{C}_{BG}$	$co\left\{\begin{array}{c}(0,0,1),\\ (0,1,0)\end{array}\right\}$		
$\mathcal{C}_{*i}$	$co\{(\frac{1}{3},\frac{2}{9},\frac{4}{9}),(\frac{1}{3},\frac{4}{9},\frac{2}{9})\}$	$\mathcal{C}_R$	$\{(1,0,0)\}$	$\{R,B,G\}$	$\left\{\begin{array}{l} p \in \Delta(R,B,G): \\ p(R)=\frac{1}{3}, \\ p(B) \in [\frac{2}{9},\frac{4}{9}] \end{array}\right\}$
		$\mathcal{C}_B$	$\{(0,1,0)\}$		
		$\mathcal{C}_G$	$\{(0,0,1)\}$		
$\mathcal{C}_{*ii}$	$co\left\{\begin{array}{c}(\frac{4}{15},\frac{1}{5},\frac{8}{15}),(\frac{4}{5},\frac{1}{5},0),\\ (\frac{2}{15},\frac{3}{5},\frac{4}{15}),(\frac{2}{5},\frac{3}{5},0)\end{array}\right\}$	$\mathcal{C}_B$	$\{(0,1,0)\}$	$\{B,RG\}$	$\left\{\begin{array}{l} p \in \Delta(B,RG): \\ p(B) \in [\frac{1}{5},\frac{3}{5}]\end{array}\right\}$
		$\mathcal{C}_{RG}$	$co\left\{\begin{array}{c}(\frac{1}{3},0,\frac{2}{3}),\\ (1,0,0)\end{array}\right\}$		

Table 3: Example information structures after rejecting and accepting the offer of learning whether B (i.e. at nodes α and $*$ in Figure 3 respectively).
The aggregators are generated according to $\phi(\mathbf{x}) = \min_{p\in\mathcal{D}}\left(\sum_{i\in I} p(i)x_i\right)$ for I as in the penultimate column of the table, and $\mathcal{D}$ in the final column. For the rest of the notation, see Table 2.

(Marschak and Miyasawa, 1968; Gollier, 2004).[26] This definition can be adapted to the current framework as follows.

Definition 2. $(\mathcal{C}_{(\mathcal{I})}, \{\mathcal{C}_{(\mathcal{I})k}\}_{k\in K}, \phi_{(\mathcal{I})})$ is *at least as informative as* $(\mathcal{C}, \{\mathcal{C}_i\}_{i\in I}, \phi)$ if and only if, for every utility function $u' : C \to \mathbb{R}$, and every compact menu $A \in \wp(\mathcal{A})$:

$$(3) \qquad \phi_{(\mathcal{I})}\left(\left(\max_{f\in A}\min_{p\in\mathcal{C}_{(\mathcal{I})k}}\sum_{s\in S} u'(f(s))p(s)\right)_{k\in K}\right) \geqslant \phi\left(\left(\max_{f\in A}\min_{p\in\mathcal{C}_i}\sum_{s\in S} u'(f(s))p(s)\right)_{i\in I}\right)$$

According to this definition, the information structure after accepting the offer to learn is at least as informative as that after rejecting the offer in case i of the previous example, but not in case ii. More generally, if no previously expected information is compromised on choosing to learn information $\mathcal{I}$, then $(\mathcal{C}_{(\mathcal{I})}, \{\mathcal{C}_{(\mathcal{I})k}\}_{k\in K}, \phi_{(\mathcal{I})})$ is at least as informative than $(\mathcal{C}, \{\mathcal{C}_i\}_{i\in I}, \phi)$. Proposition 3 implies that in such cases, information has non-negative value.

Corollary 2. Suppose that the decision maker satisfies DC. Then the value of information $\mathcal{I}$ is always non-negative whenever $(\mathcal{C}_{(\mathcal{I})}, \{\mathcal{C}_{(\mathcal{I})k}\}_{k\in K}, \phi_{(\mathcal{I})})$ is at least as informative as $(\mathcal{C}, \{\mathcal{C}_i\}_{i\in I}, \phi)$.

[26]This definition is equivalent to other possible definitions of informativeness under expected utility (Blackwell, 1953; Marschak and Miyasawa, 1968; Gollier, 2004); for non-expected utility, Li and Zhou (2016); Gensbittel et al. (2015) have obtained analogous results for similar notions to that in Definition 2.

So non-expected utility decision makers behave as the norms of rationality would recommend: they do not turn down information when its reception does not compromise information they had otherwise expected to receive – when it is a simple ‘addition’ of information, so to speak.

This result completes our conceptual analysis of information aversion under non-expected utility. It tells us that the only situation in which such a decision maker will refuse information is if, in his eyes, it is *not* free: it has an opportunity cost in the form of forgone information which he otherwise expected to obtain. There is nothing irrational in this: the effective cost of the information – foregoing other, expected information – could be too high to justify obtaining it. Even in cases where non-expected utility decision makers turn down ‘free’ information, the proposed perspective reveals that, when analysed properly, their behaviour is perfectly reasonable.

6 Discussion and Related Literature

Envisaged contingencies The proposed approach is based on the contingencies that the decision maker envisages: Dynamic Consistency requires consistency between the preferences in these contingencies and current preferences. As such, to identify the constraints it imposes on current preferences, one needs to determine which contingencies are envisaged. In economic applications, the relevant contingencies are often set as part of the modelling exercise: ideally, the theorist should correctly represent the tree the decision maker considers himself to be facing. On a conceptual score, however, those who adhere to the revealed preference paradigm may note that the envisaged contingencies are elements of the decision maker’s state of mind, like his beliefs and utilities, and as such require choice-theoretic foundations if they are to have independent behavioural meaning. Lack of such foundations could jeopardise the proposed normative defence of ambiguity models. For readers harbouring such theoretical concerns, Appendix B provides the required foundations, in the form of a representation theorem for the envisaged contingencies.

Interpretations of ex post preferences Some researchers (for example, Ghirardato, 2002) distinguish two possible interpretations of the ex post preferences involved in the dynamic consistency principle: one as the decision maker’s anticipated future preferences – those he thinks ex ante he will have in the relevant contingency – and the other as his actual ones – those he has when he finishes up in the contingency. Since this paper focusses on dynamic consistency understood as a rationality condition, it is couched in terms of the former interpretation. This is the one that lends dynamic consistency its strongest normative bite. There is certainly something abhorrent in a decision maker whose current preferences over plans do not match the preferences he thinks he will have in the future. By contrast, a decision maker whose actual preferences tomorrow do not correspond appropriately to his preferences today may be excused of the charge of irrationality (though perhaps not of foolishness) if his future preferences are not as he expected. Moreover, under this interpretation, the standard formalisation of dynamic consistency is most wanting conceptually, and the refinement proposed in Section 3 is most attractive: why should a decision maker be coherent with respect to a set of preference relations that he does not think correctly depict his own possible future preferences?

That said, the 'actual preference' interpretation of dynamic consistency coincides with the 'anticipated preference' one whenever the decision maker correctly anticipates his future preferences. So the points made in this paper can also be read as concerning actual future preferences, under the assumption that the decision maker correctly anticipates them – an assumption which, though seldom mentioned, is widespread in economic applications.

Other dynamic arguments Whilst we have only considered two among a variety of related arguments against ambiguity models, there is reason to suspect that the proposed perspective could be effective against others. For one, many are twists on the dynamic consistency argument discussed here, suggesting that the lessons from our analysis of that argument may apply. Moreover, several prominent arguments, such as the Dutch Book one (Raiffa, 1968), assume that the decision maker is naïve (Seidenfeld, 1988; Al Najjar and Weinstein, 2009) – he does not correctly anticipate and take into account what he will choose at future nodes. Since naïveté is itself criticizable on normative grounds (independently of the issue of ambiguity), such arguments are often considered less threatening to the rational credentials of non-expected utility than arguments assuming sophisticated decision makers – such as the information aversion one (Section 5) – or making no such assumption at all – such as the dynamic consistency argument (Sections 1, 3 and 4). In that sense, the arguments examined in this paper are among the most challenging for the normative credentials of ambiguity models.

Related literature There is a significant literature on dynamic choice, a full discussion of which is beyond the realm of this paper. For a thorough treatment of the issue and the literature on dynamic arguments for expected utility under risk (i.e. where probabilities are given), see 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); Siniscalchi (2011).

Of the papers introducing update rules or considering dynamic choice for non-expected utility models, the closest are without doubt Sarin and Wakker (1998); Epstein and Schneider (2003).

Indeed, as explained in Section 4, the notion of $\mathcal{P}$-rectangularity used here is essentially a version of the latter's rectangularity condition adapted to our framework. Sarin and Wakker (1998, Theorem 2.1) and one direction of Epstein and Schneider (2003, Theorem 3.2) establish that, on objective trees characterised by partitions (or more generally filtrations) of the state space, rectangular ex ante sets of priors satisfy dynamic consistency. Proposition 1 can be thought of as a simple generalisation to subjective trees – and hence information structures not corresponding to partitions – and ex ante sets of priors that are not necessarily rectangular. The other direction of the latter theorem – that dynamic consistency on objective trees can only be satisfied by rectangular priors – is technically related to Proposition 2. However, as explained in Section 4, the conceptual contributions are different, so much so that our result can be read as a new justification of their proposed restriction on sets of priors. Their approach has been adopted with other prominent ambiguity models (for example Maccheroni et al., 2006b; Klibanoff et al., 2009 for the variational preferences and smooth ambiguity models respectively; Maccheroni et al., 2006a; Klibanoff et al., 2005), and the perspective developed here applies similarly. It has also recently been adopted by Riedel et al. (2018) for the 'imprecise information' model due to Gajdos et al. (2008). This model takes 'information', modelled as a set of probability distributions, as a primitive in the objects of choice, and involves a representation where subjective beliefs are sets of priors suitably related to the information set. Riedel et al. (2018) provide a dynamic extension, following the approach cited above and in particular working on objective trees. So they adopt an objective, given set of contingencies, whilst accounting for the difference between information and subjective beliefs about the payoff-relevant state of the world. By contrast, the development here has only considered subjective beliefs about the state of the world at all stages (the sets of priors), but explicitly takes account of the distinction between subjective trees – reflecting subjective beliefs about the possible future contingencies – and objective trees imposed by the theorist. A potential direction for future research would be to combine these two perspectives to explore the relationship between the 'imprecise information' and subjective beliefs about possible future contingencies.

Other existing approaches include the update rule proposed by Hanany and Klibanoff (2007), which satisfies a version of Standard Dynamic Consistency but violates consequentialism, prior-by-prior Bayesian update (Pires, 2002) and maximum likelihood update (Gilboa and Schmeidler, 1993), both of which violate SDC. Given the difficulty in extending update rules to encompass contingencies that do not correspond to learning events (Section 2.4), DC cannot be neatly connected to any existing update rule. However, its general consequences for the relationship between ex ante and ex post beliefs can be brought out in our framework (see Proposition C.2 in Appendix C). Further investigation of consequences of DC for update is left as a topic for future research.

Conceptually, the closest suggestion to that 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 (1954) 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 consider in the Introduction. Indeed, the approach proposed here focusses on what the decision maker expects to learn, rather than what he can learn or observe. Li (2015) studies the relationship between ambiguity attitude and preferences for (partial) information, in a setup where the decision maker can choose which exogenously given partitional information structure (objective tree) he will face. By contrast, the central issue in this paper is the importance of recognising the information structure he actually thinks he is facing, and many of the results are about its consequences for choice.

7 Conclusion

It is commonly held that dynamic consistency, consequentialism and non-expected utility are incompatible. We have argued that this is not true, 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 a theorist. When these contingencies are properly taken into account, the apparent incompatibility is resolved.

The proposed perspective provides a principled justification for the use of a restricted family of sets of priors in applications to dynamic choice problems. In applications, one typically adopts the implicit assumption that the decision maker knows what the decision tree is and that it is the one the theorist or analyst is using. It turns out that dynamic consistency, in the refined 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 theorist or analyst is using. That, of course, is not necessarily a problem for the decision maker, but rather for the analyst.

Finally, the perspective provides a new analysis of information-acquisition decisions under non-expected utility, debunking the argument that such decision makers are information averse. Rather, it shows that non-expected utility decision makers will only turn down an offer of 'free' information when, in their eyes, it comes at a cost: it means foregoing information they had otherwise expected to receive.

Appendix A Incorporating the envisaged contingencies into the state space

In Section 2.1, we claimed 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 contingencies are explicitly represented in an 'extended' state space.

We adopt the terminology and assumptions in Section 2; recall that S is the 'objective' payoff-relevant state space and M the grand set of signals (possible contingencies). The 'extended state space', incorporating the payoff-relevant states and signals, is $\Omega = S \times M$. Although Ω is rich enough to represent all ex ante uncertainty, both about the state of the world and about the signal or contingency obtained ex post, S is sufficient to represent all payoff-relevant uncertainty. As such, the domain of preferences, $\geq$ and $\{\geq_m\}_{m \in M}$, is still the set $\mathcal{A}$ defined in Section 2, which corresponds naturally to a subset of C^{Ω}. Similar points hold for the sets of priors involved in the representation of preferences: the sets of priors from Section 2, $\mathcal{C}$ and $\{\mathcal{C}_m\}_{m \in M}$, correspond to sets of priors over the partition $\{\{s\} \times M \mid s \in S\}$ of Ω induced by S. A signal $m \in M$ corresponds to the set $S \times \{m\} \subseteq S \times M$; the ex post preferences conditional on this set are the preferences after having received m, $\geq_m$, introduced in Section 2. The set of envisaged contingencies corresponds in this setup to a subset of Ω, namely the set $S \times I$. We call this set EC.[27]

Given the lack of a preference relation over C^{Ω}, some remarks are in order about what should count as (the equivalent of a) 'null event' in Ω. On the one hand, 'extended' states not belonging to EC can be thought of as 'null' insofar as the decision maker does not consider it possible for them to hold (he does not consider it possible for him to receive the corresponding signals). Moreover, any state $(s', \mathcal{C}')$ such that s' is null according to the ex ante preference relation $\geq$ can naturally be thought of 'null'. We will say that $EC \cap NN$ is the set of Ω-*non-null states*, where $NN = \{(s', \mathcal{C}') \in \Omega \mid s' \text{ is } \geq \text{-non-null}\}$. Any event with non-empty intersect with $EC \cap NN$ will be said to be Ω-*non-null*.

Note that, given the representation of possible signals or 'learning events' in the extended state space, it does not make sense to consider learning or conditioning upon certain events of Ω. To characterize the events that can conceivably be learnt, we introduce the following definition.

Definition 3. An event $A \subseteq \Omega$ is *learnable* if it satisfies the following two conditions:

1. $m' = m''$ for all $(s', m'), (s'', m'') \in A$.

[27] Whilst we do not assume a preference relation on C^{Ω} or a set of priors over Ω, 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 $\mathcal{C}$, $\{\mathcal{C}\}_{m \in M}$ and EC.

2. If $(s', m') \in A$, then $(s'', m') \in A$, for every s'' for which $p(s'') > 0$ for some $p \in \mathcal{C}_{m'}$.

A partition of Ω 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.[28] The first condition corresponds to the assumption that, in the ex post stage, the agent has no uncertainty about the contingency he is in (or, equivalently, the signal he has received). So the only events in Ω that can be learnt are those which correspond to signals.[29] 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 beliefs he has at this stage. In particular, a state $s' \in S$ cannot be ruled out by the information A (i.e. there is no state $(s', m') \in A$) but nevertheless possibly have non-zero probability according to the ex post beliefs $\mathcal{C}_{m'}$ in a contingency permitted by A.

Note that each learnable event A corresponds to a (unique) signal $m^* \in M$, and the ex post preferences conditional on A, $\succeq_A$, coincide with $\succeq_{m^*}$.

We claim that this framework yields the same analysis as that carried out in the bulk of the paper. To establish this, it suffices to show that the standard definition of dynamic consistency (SDC) applied in this framework is equivalent to the refined notion of dynamic consistency proposed in Section 3 (DC). Translated into this framework, SDC becomes:

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

Proposition A.1. Standard Dynamic Consistency on Ω is equivalent to DC.

So all the results and points in the paper apply with DC replaced by Standard Dynamic Consistency on Ω. Note moreover that this result can be taken to support the claim that DC is the appropriate equivalent to SDC for subjective trees. Finally, it allows a reformulation of the discussion in this paper in terms of the difference between applying SDC to the 'objective' state space S and applying it to the 'extended' state space Ω. For example, the universal quantification in SDC can be understood accordingly: SDC on S allows the theorist to treat the decision maker as planning for any partition of S, whereas SDC on Ω allows the theorist to the treat the decision

[28] Recall that we are working in a simple setup with two time periods, and hence are considering the learning of an event between the ex ante and ex post stage.

[29] Of course, given the rich signal space, where the disjunction of any set of signals can be formalised as another signal, this does not restrict the signals that can be received.

maker as planning for any partition that is coherent with what he thinks he may learn. This is just another way of putting the basic insight of this paper: in conceiving the decision maker as choosing between plans, the theorist must respect what he expects to learn.

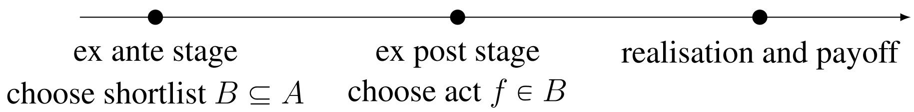
Appendix B Foundations

In the bulk of the paper we have assumed the decision maker's current beliefs (ex ante set of priors) and envisaged contingencies – and in particular his anticipated ex post sets of priors – are given. Whilst the former can in principle be gleaned from behaviour (Gilboa and Schmeidler, 1989), so much cannot be said yet for the latter. However, the approach set out above relies on the contingencies envisaged by the decision maker, and the derived normative defence of non-expected utility models supposes this notion to have independent meaning. Under the dominant revealed preference paradigm, behavioural foundations for the notion of contingency envisaged by the decision maker are thus required. Since these contingencies are needed in the context of a discussion about, and assessment of, dynamic consistency, the foundations should avoid relying on assumptions about whether or not it is satisfied, and therefore about the relationship between the envisaged contingencies and the decision maker's ex ante preferences. The objective of this Appendix is to present such behavioural foundations in the context of the maxmin EU model. The aim is simply to settle the conceptual issue of whether they can be provided with behavioural foundations, not to provide a practical method for eliciting them. This latter question is left for future research.

B.1 Setup and representation

We adopt an Anscombe-Aumann-style refinement of the framework set out in Section 2.1, which is common in axiomatic work. Henceforth, we let the consequences $C = \Delta(X)$, the set of Borel probability measures over a nonempty set of (perhaps, but not necessarily, monetary) *outcomes* X. X is endowed with a metric under which it is compact, and we adopt the weak convergence topology on C, under which it is compact metric (Billingsley, 2009, p 72). Under the product topology, $\mathcal{A}$, the set of acts (i.e. functions from S to C) is compact metric. Moreover, it is a mixture set with the mixture relation defined pointwise: for $f, h \in \mathcal{A}$ and $\alpha \in \mathbb{R}$, $0 \leqslant \alpha \leqslant 1$, 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 $0 \leqslant \alpha \leqslant 1$, $\alpha A + (1-\alpha)h = \{\alpha f + (1-\alpha)h \mid f \in A\}$. We write $f_\alpha h$ as short for $\alpha f + (1-\alpha)h$ and $A_\alpha h$ for $\alpha A + (1-\alpha)h$. $\wp(\bullet)$ denotes the set of closed non-empty subsets of $\bullet$; hence, in particular, $\wp(\mathcal{A})$ is

Figure 4: Time line



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 $A \in \wp(\mathcal{A})$, $co(A)$ is the set of finite mixtures of elements of A: $co(A) = \{\sum_{i=1}^{n} \alpha_i f_i |\ \alpha_i \in [0,1] \text{ with } \sum_{i=1}^{n} \alpha_i = 1,\ f_i \in A\}$. Note that $co(A) \in \wp(\mathcal{A})$.

By contrast to the bulk of the paper, where (future) sets of priors or preference relations were taken as primitive, we now assume a choice correspondence on $\mathcal{A}$: a function $c : \wp(\mathcal{A}) \to \wp(\mathcal{A})$ such that, for any $A \in \wp(\mathcal{A})$, $c(A) \subseteq A$. It has the following interpretation, associated with the time line given in Figure 4. 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). He 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 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 $A \in \wp(\mathcal{A})$, $c(A)$ is the set of acts that the decision maker wishes to keep as open alternatives for his future choice; the elements not in $c(A)$ are those that he is willing to rule out now.[30]

We consider the following representation:

$$c(A) = \{f \in A \mid f \in \arg\max_{g \in A} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) \ \text{ for some } \mathcal{C}' \in \mathcal{K}\} \tag{4}$$

where u is a continuous affine utility function on $\Delta(X)$ and $\mathcal{K}$ is a set of convex, closed subsets of $\Delta(S)$. The sets of priors in $\mathcal{K}$ 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 contingencies he envisages: that is, $\mathcal{K}$ is basically the set $\{\mathcal{C}_i\}_{i \in I}$ of envisaged contingencies introduced in Section 2. (2) represents a decision maker who anticipates

[30]Note that nothing is assumed about the relationship between $c(A)$ (which will reveal the envisaged contingencies) and the acts the decision maker would choose from A if he were asked to choose now (which, as standard, can be represented by his ex ante preferences).

that, in each envisaged contingency, he will form preferences according to the maxmin EU rule with the set of priors corresponding to that contingency. He retains as an open option any act that is optimal according to this rule with at least one of the sets of priors in $\mathcal{K}$, and rules out any act that is not optimal under any of the sets.

B.2 Axioms and result

Consider the following axioms on the choice correspondence 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 $f \in co(B)$ 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 $d \notin c(\{d, e\})$ 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 $A_n \to A$ and all sequences of acts $(f_n)_{n \in N}$ with $f_n \in c(A_n)$ for each $n \in N$, if $f_n \to f$, then $f \in c(A)$.

Chernoff (A1) and Aizerman (A2) are standard axioms in the choice-theoretical literature. The conjunction of the two is weaker than the Weak Axiom of Revealed Preference, and equivalent to a notion of rationalisability discussed in Moulin (1985) (from whom we also borrow the nomenclature). Fixed utilities (A4) imposes Sen's axiom β on the restriction of c to menus containing only constant acts. It follows from standard choice theory results (Sen, 1971) that the restriction to constant acts is represented by a single complete transitive 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) and Continuity (A9) are fairly standard.

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. 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' which acts are kept open: 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 mixture $f_\alpha d$. 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.

Monotonicity (A7) is essentially the standard monotonicity or statewise dominance axiom formulated for menus. It includes both a weak and a strict dominance clause. The first basically says that adding elements to a menu that are weakly dominated by some element already present 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.[31] 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.

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 $c(A)$, 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 $f \in co(B)$. The axiom A8 is evidently a weakening of this extension, stating that it holds under particular

[31] The proposed interpretation of dominance is vindicated by the Fixed utilities axiom (A4).

conditions. In fact, the interpretation of the extension requires considering the acts in $c(A)$ to be indifferent; however, whilst this is the case under WARP and the standard interpretation of choice correspondences, it is no longer true under the weaker choice-theoretic axioms and alternative interpretation used here. The conditions in A8 guarantee that there is an ex post preference relation according to which the acts in B are indifferent.[32] So the axiom 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 correspondence employed here.

A foundation for the notion of contingency envisaged by the decision maker is given by the following representation theorem.

Theorem B.1. Let c be a choice correspondence on $\mathcal{A}$. The following are equivalent:

(i) c satisfies A1–A9;

(ii) There exists a nonconstant continuous affine utility function $u : \Delta(X) \to \mathbb{R}$ and a set $\mathcal{K}$ of closed convex sets of probability measures on S such that:

$$c(A) = \left\{ f \in A \mid f \in \arg\max_{g \in A} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) \ \text{ for some } \mathcal{C}' \in \mathcal{K} \right\} \tag{2}$$

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

This theorem shows that, under certain conditions, a set of sets of priors – or a set of envisaged contingencies – representing the choice correspondence according to (2) exists. Moreover, there is a unique 'canonical' such set, namely the unique minimal set. We conclude that the notion of contingency envisaged by the decision maker introduced in Section 2 does have solid behavioural foundations.

[32]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, there is an 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 in B.

Remark B.1. Technically, Theorem B.1 is related to the representation result in Seidenfeld et al. (2010), which, in the same formal framework, studies the case where the decision maker's ex post preferences are expected utility. It may be also considered – technically, again – as a contribution to the literature on preference for flexibility or unforeseen contingencies initiated by Kreps (1979, 1992); Dekel et al. (2001). To see this, define the preference relation $\dot{<}$ on $\wp(\mathcal{A})$ as follows: for all $A, B \in \wp(\mathcal{A})$, $A \dot{<} B$ if and only if $c(A \cup B) \cap A = \varnothing$. It is straightforward to check that this preference relation over menus is represented by the u and $\mathcal{K}$ featuring in representation (2) as follows: for all $A, B \in \wp(\mathcal{A})$, $A \dot{<} B$ if and only if

$$\max_{g \in A} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) < \max_{g \in B} \min_{p \in \mathcal{C}'} \sum_{s \in S} u(g(s))p(s) \text{ for all } \mathcal{C}' \in \mathcal{K} \tag{3}$$

It is standard in this literature to form menu preferences using an 'aggregator' that ensures completeness: if menu A is better than B under one ex post preference and B is better under a different one, the decision maker is assumed to 'weigh off' the two ex post preferences and order the menus. Representation (3) involves no aggregation of this sort: menus are ordered only when there is strict (ex post preference-wise) dominance. So in the sort of example just given, it is not assumed that the decision maker 'weighs off' the ex post preferences, and the weak version of $\dot{<}$ is incomplete. This difference is crucial for the goal in this Appendix. A standard aggregator fully determines preferences over singleton menus, which are naturally interpreted as reflecting the decision maker's ex ante preferences over acts. To this extent, it embodies a particular relationship between ex ante and ex post preferences. So any theory that delivers ex post preferences by relying on such an aggregator effectively incorporates an assumption on the relationship with ex ante preferences. But, as explained previously, foundations for the ex post sets of priors (the envisaged contingencies) are needed to ascertain whether dynamic consistency is satisfied, and so they should avoid invoking assumptions about the relationship between ex post and ex ante preferences. The result above, combined with the previous interpretation of the choice function, involves no such assumption.[33]

Appendix C Proofs of results in the paper

Proof of Proposition 1. Let $\mathcal{D}$ be as specified, and consider $f, g \in \mathcal{A}$ with $f \preceq_i g$ for all $i \in I$. By representation (1), it follows that $\min_{q_i \in \mathcal{C}_i} \sum_{s \in S} u(f(s))q_i(s) \leqslant \min_{q_i \in \mathcal{C}_i} \sum_{s \in S} u(g(s))q_i(s)$ for all $i \in I$. By assumption,

[33] See also footnote 30 above.

$$\begin{aligned}
\min_{\hat{p}\in\mathcal{C}_{\mathcal{D}}}\sum_{s\in S}u(f(s))\hat{p}(s) &= \min_{p\in\mathcal{D};(q_i)_{i\in I}\in\prod_{i\in I}\mathcal{C}_i}\sum_{s\in S}u(f(s))\left(\sum_{i\in I}p(i)q_i(s)\right)\\
&= \min_{p\in\mathcal{D}}\sum_{i\in I}p(i)\left(\min_{q_i\in\mathcal{C}_i}\sum_{s\in S}u(f(s))q_i(s)\right)\\
&\leqslant \min_{p\in\mathcal{D}}\sum_{i\in I}p(i)\left(\min_{q_i\in\mathcal{C}_i}\sum_{s\in S}u(g(s))q_i(s)\right)\\
&= \min_{\hat{p}\in\mathcal{C}_{\mathcal{D}}}\sum_{s\in S}u(g(s))\hat{p}(s)
\end{aligned}$$

where the second and last equalities follow from the fact that $p(i) \geqslant 0$ for all $i \in I$. So $f \preceq_{\mathcal{D}} g$, as required. Since $p(i) > 0$ for all $i \in I$ and $p \in \mathcal{D}$, whenever one of the $\succeq_i$ preferences are strict, so is the $\succeq_{\mathcal{D}}$ one; hence DC is satisfied.

□

Proof of Proposition 2. Since u is continuous and increasing and C is a real interval, for every $c_1, c_2 \in C$ and $\alpha \in (0,1)$, there exists a $d \in C$ such that $u(d) = \alpha u(c_1) + (1-\alpha)u(c_2)$. Henceforth, for any $c_1, c_2 \in C$ and $\alpha \in (0,1)$, we use $\alpha c_1 + (1-\alpha)c_2$ to refer to a $d \in C$ with this property. This notion of mixture is extended to acts pointwise: for $g_1, g_2 \in \mathcal{A}$ and $\alpha \in (0,1)$, $\alpha g_1 + (1-\alpha)g_2 \in \mathcal{A}$ is defined by: $(\alpha g_1 + (1-\alpha)g_2)(s) = \alpha g_1(s) + (1-\alpha)g_2(s)$ for all $s \in S$.

Let $J : \mathcal{A} \to \mathbb{R}$ be the maxmin EU functional represented by $\mathcal{C}$ (i.e. $J(f) = \min_{p\in\mathcal{C}}\sum_{s\in S}u(f(s))p(s)$ 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 $\preceq_i$, $Ra(J_i) = Ra(J_j)$ for all i, j; call this set R. Since the utility function is continuous, this is an interval. $cone(Ra(J_I))$ is the cone spanned by $Ra(J_I)$. We use $\mathbf{e}$ to denote the unit vector in $\mathbb{R}^I$.

We say that a function $\phi : \mathbb{R}^n \to \mathbb{R}$ is: *constant additive* if $\phi(\mathbf{x} + a\mathbf{e}) = \phi(\mathbf{x}) + a$ for all $\mathbf{x} \in \mathbb{R}^n$, $a \in \mathbb{R}$; *positively homogeneous* if $\phi(\alpha\mathbf{x}) = \alpha\phi(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$, $\alpha \geqslant 0$; *monotonic* if $\phi(\mathbf{x}) \geqslant \phi(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $x_i \geqslant y_i$ for all $i \in \{1, \ldots, n\}$; and *strongly monotonic* if it is monotonic and $\phi(\mathbf{x}) > \phi(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $x_i \geqslant y_i$ for all $i \in \{1, \ldots, n\}$ with strict inequality for some i.

As shown by Crès et al. (2011, Lemmas 1–4), DC implies that $J(f) = \phi((J_i(f))_{i\in I})$ where ϕ 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 $x_i \in R = Ra(J_i)$, it follows from the fact that the utility function is continuous and increasing that there exists a constant act $c_i^{\mathbf{x}} \in C$ with $J_i(c_i^{\mathbf{x}}) = x_i$. Hence, defining $g^{\mathbf{x}} \in \mathcal{A}$ by $g^{\mathbf{x}}(s) = c_i^{\mathbf{x}}(s)$ whenever $s \in A_i$, it follows from the fact that $\{\preceq_i\}_{i\in I}$ is

$\mathcal{P}$-objective 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 fact that the maxmin EU functional coincides with EU on constant acts (and the definition of $\alpha c_i^{\mathbf{x}} + (1-\alpha)c_i^{\mathbf{y}}$). Hence $\phi(\alpha\mathbf{x}+(1-\alpha)\mathbf{y}) = \phi((J_i(\alpha g^{\mathbf{x}}+(1-\alpha)g^{\mathbf{y}}))_{i\in I}) = J(\alpha g^{\mathbf{x}}+(1-\alpha)g^{\mathbf{y}}) \geqslant \alpha J(g^{\mathbf{x}})+(1-\alpha)J(g^{\mathbf{y}}) = \alpha\phi((J_i(g^{\mathbf{x}}))_{i\in I})+(1-\alpha)\phi((J_i(g^{\mathbf{y}})_{i\in I}) = \alpha\phi(\mathbf{x})+(1-\alpha)\phi(\mathbf{y})$, where the inequality in the middle holds because of the concavity of the maxmin EU functional J. Hence ϕ is concave.

Using standard arguments, ϕ can be extended to a monotonic, positively homogeneous, constant additive, concave function on $\mathbb{R}^I$. 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 $Ra(J_I) = R^I$. The form of the set of priors $\mathcal{C}$ representing J follows from Crès et al. (2011, Proposition 1).

□

Proof of Proposition 3. To show the right to left direction, suppose that $\mathcal{C}_{(\mathcal{I})} \subseteq \mathcal{C}$. Take any (compact) $A \subseteq \mathcal{A}$ 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 $\min_{p\in\mathcal{C}_{(\mathcal{I})}}\sum_{s\in S} u(g(s))p(s) \geqslant \min_{p\in\mathcal{C}}\sum_{s\in S} u(g(s))p(s) = \max_{f\in A}\min_{p\in\mathcal{C}}\sum_{s\in S} u(f(s))p(s)$. For every $k \in K$, by definition, $\max_{f\in A}\min_{p\in\mathcal{C}_{(\mathcal{I})k}}\sum_{s\in S} u(f(s))p(s) \geqslant \min_{p\in\mathcal{C}_{(\mathcal{I})k}}\sum_{s\in S} u(g(s))p(s)$. By Proposition C.2, DC implies that the aggregator $\phi_{(\mathcal{I})}$ is monotonic, and hence that $\phi_{(\mathcal{I})}\left(\left(\max_{f\in A}\min_{p\in\mathcal{C}_{(\mathcal{I})k}}\sum_{s\in S} u(f(s))p(s)\right)_{k\in K}\right) \geqslant \phi_{(\mathcal{I})}\left(\left(\min_{p\in\mathcal{C}_{(\mathcal{I})k}}\sum_{s\in S} u(g(s))p(s)\right)_{k\in K}\right) = \min_{p\in\mathcal{C}_{(\mathcal{I})}}\sum_{s\in S} u(g(s))p(s)$ (the last equality by the definition of $\phi_{(\mathcal{I})}$). 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 $\mathcal{C}_{(\mathcal{I})} \nsubseteq \mathcal{C}$; 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 $\mathcal{C}_{(\mathcal{I})} \nsubseteq \mathcal{C}$, there exists $p \in \Delta(\Sigma)$ with $p \in \mathcal{C}_{(\mathcal{I})}\backslash\mathcal{C}$. By a separation theorem (Aliprantis and Border, 2007, 5.80), there is a nonzero linear functional ϕ on $ba(S)$ and $\alpha \in \mathbb{R}$ such that $\phi(p) \leqslant \alpha < \phi(q)$ 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 ϕ, 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, $\min_{p\in\mathcal{C}_{(\mathcal{I})}}\sum_{s\in S}u(g(s))p(s) < \min_{p\in\mathcal{C}}\sum_{s\in S}u(g(s))p(s) = \max_{f\in\{g\}}\min_{p\in\mathcal{C}}\sum_{s\in S}u(f(s))p(s)$. 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.

□

Proposition C.2. $(\succeq, \{\succeq_m\}_{m\in M}, I)$ satisfies Dynamic Consistency if and only if there exists a constant additive, positively homogeneous, monotonic function $\phi : \mathbb{R}^I \to \mathbb{R}$ 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 $f\in\mathcal{A}$.[34]

Proof. The 'if' direction is straightforward. The proof of the 'only if' direction 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 ϕ 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 $\hat{\phi}$ on $\mathbb{R}^I$, defined by $\hat{\phi}(\mathbf{y}) = \sup\{\phi(\mathbf{x})+b \mid \mathbf{x}\in Ra(J_I),\ b\in\mathbb{R},\ \mathbf{x}+b\mathbf{e}\leqslant\mathbf{y}\}$ for all $\mathbf{y}\in\mathbb{R}^I$, extends ϕ and is constant additive and monotonic. It is clear from the definition and the positive homogeneity of ϕ that $\hat{\phi}$ is positively homogeneous. Finally, strong monotonicity on $Ra(J_I)$ is a direct consequence of DC.

□

Remark C.1. It is straightforward to check that the preferences in Proposition 1 correspond to the special case of those characterised here where ϕ is concave. The following counterexample (inspired by Crès et al., 2011) shows that there are $(\succeq, \{\succeq_m\}_{m\in M}, I)$ satisfying DC, where the ϕ in Proposition C.2 is not concave, and hence where preferences are not of the sort in Proposition 1.

Let there be two states, $S = \{s, t\}$. Suppose that the decision maker envisages two contingencies with ex post sets of priors $\mathcal{C}_1 = \left\{p\in\Delta(S) : p(s)\geqslant\frac{1}{2}\right\}$ and $\mathcal{C}_2 = \left\{p\in\Delta(S) : p(s)\leqslant\frac{1}{2}\right\}$, and that his ex ante set of priors is the singleton $\mathcal{C} = \left\{p\in\Delta(S) : p(s) = \frac{1}{2}\right\}$. Note that, for all $f\in\mathcal{A}$,

$$\begin{aligned}\min_{p\in\mathcal{C}}\sum_{s\in S}u(f(s))p(s) &= \frac{1}{2}u(f(s)) + \frac{1}{2}u(f(t))\\ &= \max\left(\min\left(\frac{1}{2}u(f(s)) + \frac{1}{2}u(f(t)), u(f(s))\right), \min\left(\frac{1}{2}u(f(s)) + \frac{1}{2}u(f(t)), u(f(t))\right)\right)\\ &= \max\left(\min_{p\in\mathcal{C}_1}\sum_{s\in S}u(f(s))p(s), \min_{p\in\mathcal{C}_2}\sum_{s\in S}u(f(s))p(s)\right)\end{aligned}$$

[34] $Ra((\min_{p\in\mathcal{C}_i}\sum_{s\in S}u(\cdot)p(s))_{i\in I})\subseteq\mathbb{R}^I$ is the range of the $(\min_{p\in\mathcal{C}_i}\sum_{s\in S}u(\cdot)p(s))_{i\in I}$.

So the relation in Proposition C.2 holds with $\phi(x, y) = \max(x, y)$. Since ϕ is constant additive, positively homogenous and strongly monotonic, DC is satisfied; however, ϕ is clearly not concave, as required.

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.

Baliga, S., Hanany, E., and Klibanoff, P. (2013). Polarization and ambiguity. *American Economic Review*, 103(7):3071–83.

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.

Bose, S. and Renou, L. (2014). Mechanism design with ambiguous communication devices. *Econometrica*, 82(5):1853–1872.

Brock, W. A. and Hansen, L. P. (2017). Wrestling with uncertainty in climate economic models.

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.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *Quart. J. Econ.*, 75:643–669.

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. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31.

Gajdos, T., Hayashi, T., Tallon, J.-M., and Vergnaud, J.-C. (2008). Attitude toward imprecise information. *J. Econ. Theory*, 140(1):27–65.

Gensbittel, F., Renou, L., and Tomala, T. (2015). Comparisons of ambiguous experiments.

Ghirardato, P. (2002). Revisiting Savage in a conditional world. *Economic Theory*, 20(1):83–92.

Gilboa, I. and Marinacci, M. (2013). Ambiguity and the Bayesian Paradigm. In *Advances in Economics and Econometrics: Volume 1, Economic Theory: Tenth World Congress*, volume 49, page 179. Cambridge University Press.

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.

Gilboa, I. and Schmeidler, D. (1993). Updating Ambiguous Beliefs. *Journal of Economic Theory*, 59(1):33–49.

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.

Hansen, L. P. (2014). Nobel lecture: Uncertainty outside and inside economic models. *Journal of Political Economy*, 122(5):945–987.

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton university press.

Kadane, J. B., Schervish, M. J., and Seidenfeld, T. (2008). Is ignorance bliss? *Journal of Philosophy*, CV(1):5–36.

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. (2005). A Smooth Model of Decision Making under Ambiguity. *Econometrica*, 73(6):1849–1892.

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.

Kunreuther, H., Heal, G., Allen, M., Edenhofer, O., Field, C. B., and Yohe, G. (2013). Risk management and climate change. *Nature Climate Change*, 3(5):447–450.

Li, J. (2015). Preferences for Partial Information and Ambiguity. Technical report, McGill.

Li, J. and Zhou, J. (2016). Blackwell's informativeness ranking with uncertainty-averse preferences. *Games and Economic Behavior*, 96(Supplement C):18–29.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006a). Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica*, 74(6):1447–1498.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006b). 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.

Marinacci, M. (2015). Model uncertainty. *Journal of the European Economic Association*, 13(6):1022–1100.

Marschak, J. and Miyasawa, K. (1968). Economic Comparability of Information Systems. *International Economic Review*, 9(2):137–174.

McClennen, E. F. (1990). *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge University Press.

Millner, A., Dietz, S., and Heal, G. (2013). Scientific ambiguity and climate policy. *Environmental and Resource Economics*, pages 1–26.

Moulin, H. (1985). Choice Functions Over a Finite Set: A Summary. *Social Choice and Welfare*, 2:147–160.

Pires, C. P. (2002). A Rule For Updating Ambiguous Beliefs. *Theory and Decision*, 53(2):137–152.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choice under uncertainty*. Addison-Wesley.

Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic processes and their applications*, 112(2):185–200.

Riedel, F., Tallon, J.-M., and Vergopoulos, V. (2018). Dynamically consistent preferences under imprecise probabilistic information. *Journal of Mathematical Economics*.

Sarin, R. and Wakker, P. (1994). Folding back in decision tree analysis. *Management Science*, 40(5):625–628.

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.

Seidenfeld, T. (1988). Decision Theory Without "Independence" or Without "Ordering". *Economics and Philosophy*, 4:267–290.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2010). Coherent choice functions under uncertainty. *Synthese*, 172(1):157–176.

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.

Tillio, A. d., Kos, N., and Messner, M. (2016). The design of ambiguous mechanisms. *The Review of Economic Studies*, 84(1):237–276.

van Fraassen, B. (1984). Belief and the Will. *The Journal of Philosophy*, 81(5):235–256.

Wakker, P. (1988). Nonexpected utility as aversion of information. *Journal of Behavioral Decision Making*, I:169–175.

Wakker, P. P. (1999). Justifying Bayesianism by Dynamic Decision Principles. Technical report, Medical Decision Making Unit, Leiden University Medical Center.

Appendix D Supplementary Material. Technical Appendix: Auxiliary Proofs

Throughout this 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: $a \leqslant b$ iff $a(s) \leqslant b(s)$ for all $s \in S$. For $x \in \mathbb{R}$, x^* is the constant function taking value x; we use $\mathbb{R}^*$ 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 $A + a = \{b + a \mid b \in A\}$. Recall that, for any subset $A \subseteq B$, we use $co(A)$ to denote the convex closure of A.

D.1 Proofs of Results in the Appendices

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

□

Proof of Theorem B.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 correspondence on a subset of the set of real-valued functions on S, which can be extended to a choice correspondence on the whole set satisfying Chernoff, Aizerman, upper hemicontinuity and choice-correspondence versions of constant linearity, superadditivity, monotonicity and union C-independence (Lemmas D.2 and D.1). The most important step in the proof is Proposition D.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 expctation 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 $\mathcal{K}$. Lemmas used below are proved in Appendix D.2.

We now begin the proof. 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 correspondence) 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.

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 correspondence $c_{B(K)}$, 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 D.2 (Appendix D.2), $c_{B(K)}$ is well-defined. Let $\wp_{bbd}(B)$ be the set of closed bounded subsets of B.

Lemma D.1. There exists a choice correspondence $c : \wp_{bdd}(B) \to \wp_{bdd}(B)$ 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)[35]

iii. c is constant linear: for all $A \in \wp_{bdd}(B)$, $\alpha > 0$ and $x \in \mathbb{R}$, $c(\alpha A + x^*) = \alpha c(A) + x^*$

iv. c is constant independent: for all $A \in \wp_{bdd}(B)$, $\alpha \in (0, 1)$ and $x \in \mathbb{R}$ 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 $a \geqslant b$, 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 $x \in \mathbb{R}$ 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 $(A_n)_{n \in N}$, $A_n \in \wp_{bdd}(B)$ and $A \in \wp_{bdd}(B)$ with $A_n \to A$ and for all sequences $(a_n)_{n \in N}$, $a_n \in B$, with $a_n \in c(A_n)$ for all $n \in N$, if $a_n \to a$, then $a \in c(A)$.

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

[35] Henceforth, we shall refer to these properties by these names.

Proposition D.1. Let $c : \wp_{bdd}(B) \to \wp_{bdd}(B)$ be a choice correspondence satisfying the properties in Lemma D.1, 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 probability measures on S such that:

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

$$b \in \arg\max_{b' \in A} \min_{p \in \mathcal{C}_{\hat{a},\hat{A}}} \sum_{s \in S} b'(s)p(s) \Rightarrow b \in c(A) \tag{4}$$

(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 is it a constant act is treated similarly, by using the construction below with a non-constant act a such that $\{\alpha a + (1-\alpha)\hat{a} |\ \alpha \in [0,1]\} = c(\{\alpha a + (1-\alpha)\hat{a} |\ \alpha \in [0,1]\})$.

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 $U_{\hat{a}} = \{b \in B \mid \inf_{x \in \mathbb{R}} \|b - x^*\| \leqslant \inf_{x \in \mathbb{R}} \|\hat{a} - x^*\|\}$: 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 $\overline{a} = \{\alpha a + x^* |\ \alpha \geqslant 0,\ x \in \mathbb{R}\}$; 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 $\{\overline{a} \backslash \mathbb{R}^* |\ a \in B \backslash \mathbb{R}^*\}$ form a partition of $B \backslash \mathbb{R}^*$; call the set of equivalence classes of the partition $\mathcal{Q}$.

For $a \in B$, $x \in \mathbb{R}$, we define $[a, x^*] = \{\alpha a + (1-\alpha)x^* |\ \alpha \in [0,1]\}$, and $\Xi = \{[a, x^*] |\ a \in B,\ x \in \mathbb{R},\ c([a, x^*]) = [a, x^*]\}$. Moreover, for any $A \in \wp_{bdd}(B)$ and $x \in \mathbb{R}$, $X_x^A = \bigcup_{a \in A} [a, x^*]$. 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 $\overline{a} \subseteq O$
   b. $\hat{a} \in O$
2. $I : O \to \mathbb{R}$ is a functional with the following properties:
   a. I is monotonic: for all $a, b \in B$, if $a \geqslant b$, $I(a) \geqslant I(b)$
   b. I is constant linear (i.e. constant additive and positively homogeneous): for all $a \in O$, $x \in \mathbb{R}$, $\alpha > 0$, $I(\alpha a + x^*) = \alpha I(a) + x$

c. I is superadditive: for any $\alpha \in [0,1]$, $I(\alpha a + (1-\alpha)b) \geqslant \alpha I(a) + (1-\alpha)I(b)$

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 $A' \subseteq \{a \in O | I(a) = I(\hat{a})\}$, $A' \subseteq c(A' \cup X^{\hat{A}}_{I(\hat{a})})$

g. for every $A \in \wp_{bdd}(O)$ with $(\{a \in O | I(a) = I(\hat{a})\} \cap U_{\hat{a}}) \cup X^{\hat{A}}_{I(\hat{a})} \subseteq A$, and for every $a' \in \{a \in O | I(a) = I(\hat{a})\} \cap U_{\hat{a}}$ 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 $U_{\hat{a}}$, $[\hat{a}, I(\hat{a})^*] \subseteq \{a \in O | I(a) = I(\hat{a})\} \cap U_{\hat{a}}$.

$\mathcal{Z}$ is equipped with the order $\leqslant$, defined as follows: $(O_1, I_1) \leqslant (O_2, I_2)$ 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 Appendix D.2), 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 D.5) and that $\mathcal{Z}$ is non-empty (Lemma D.13). 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 D.5, 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 (4) 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.

□

We conclude the direction (i) to (ii). Let $\mathcal{K} = \{\mathcal{C}_{\hat{a},\hat{A}} \mid \hat{A} \in \wp_{bdd}(B),\ \hat{a} \in c(\hat{A})\}$, where the $\mathcal{C}_{\hat{a},\hat{A}}$ are as in Proposition D.1. By the definition of the choice correspondence c on B, and clause 1 of Proposition D.1, for any $\mathcal{C} \in \mathcal{K}$, 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 D.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 $\mathcal{K}$ and u represent c according to (2).

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 $\mathcal{C} \in \mathcal{K}$ 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 $\mathcal{C}_1 \in \mathcal{K}$ 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 $\min_{p \in \mathcal{C}_1} \sum_{s \in S} u(g_\alpha d(s))p(s) \geqslant u(d)$ 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 $\min_{p \in \mathcal{C}_1} \sum_{s \in S} u(g_\alpha d(s))p(s) \geqslant \min_{p \in \mathcal{C}_1} \sum_{s \in S} u(f_\beta e(s))p(s)$, and hence, by the representation (2), that $g_\alpha d \in c(A \cup \{f_\beta e\})$, contradicting A8. Hence there is a set $\mathcal{C} \in \mathcal{K}$ 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 $f \in co(B)$, $\min_{p \in \mathcal{C}} \sum_{s \in S} u(f(s))p(s) \geqslant \max_{h \in A} \min_{p \in \mathcal{C}} \sum_{s \in S} u(h(s))p(s)$, and so by the representation (2), 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. The existence of a unique minimal $\mathcal{K}$ representing c is established by Lemma D.3.

□

D.2 Lemmas used in the Proof of Theorem B.1

Lemma D.2. For $A, B \subseteq \mathcal{A}$, suppose that there is a bijection $\sigma : A \to B$ 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 $c(A)_\alpha d \nsubseteq c((A \cup B)_\alpha d \cup \{\sigma(f)\})$; 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 $\alpha \to 1$, $(A \cup B)_\alpha d \cup \{\sigma(f)\} \to A \cup B$, $\sigma(f) \in c(A \cup B)$. 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.

□

Proof of Lemma D.1. 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 α), we have that $c(A + x^*) = c(\frac{1}{2}(2A) + \frac{1}{2}(2x^*)) = \frac{1}{2}c(2A) + \frac{1}{2}(2x^*) = c(A) + x^*$, as required. So c is constant linear. The remaining properties are a direct consequence of axioms A6–A9 (multiplying by a sufficiently small α where appropriate).

□

Lemma D.3. Let c be representable according to (2). Then there exists a unique minimal $\mathcal{K}$ representing c.

Proof. Let u be any utility function involved in a representation of c according to (2); as established in Appendix C, it is unique up to positive affine transformation. Let $\{\mathcal{K}_m\}_{m\in M}$ be the sets of sets of priors representing c along with u according to (2) and let $\mathcal{K} = \bigcup_{m\in M}\mathcal{K}_m$. Let I index $\mathcal{K}$. Note that $\mathcal{K}$ also represents c according to (2). Pick any $d \in \Delta(X)$ and for each $\mathcal{C}_i \in \mathcal{K}$, 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 $\mathcal{K}$ represents c according to (2), $c(A_i) = A_i$ 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 $e \prec d$ 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 $i, j \in I$, $A_i \neq A_j$, $\sqsubseteq$ is complete, and the strict relation is transitive. We say that a set A_i is *essential* if there exists no set $J \subset I\backslash\{i\}$ such that $A_i \not\sqsupseteq A_j$ 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 $A_i \sqsupseteq A_j$, 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 maxmin EU 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 $A_i \sqsupseteq A_j$, but also that $\bigcup_{j \text{ s.t. } A_i \sqsupseteq A_j}\{h \in A_i \mid h \succeq_j g,\ \forall g \in A_i\} \neq A_i$. 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 D.4 below, which implies that there is a unique minimal subset of $\mathcal{K}$ representing c; it follows in particular that this is contained in every set of sets of priors $\mathcal{K}_m$ representing c.

□

Lemma D.4. $\{\mathcal{C}_i \mid i \in I,\ A_i \text{ essential}\}$ represents c according to (2), 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 $\preceq_i$ generated by the sets of priors $\mathcal{C}_i$. Let $\mathcal{R}$ be the set of preference relations on $\mathcal{A}$ defined as follows: $\preceq_i \in \mathcal{R}$ if and only if there exists $\mathcal{C}_i \in \mathcal{K}$ such that $\preceq_i$ is represented according to (1) by $\mathcal{C}_i$ and u. We say that a set of preference relations $\mathcal{R}'$ represents c if the following holds: $c(A) = \{f \in A \mid \exists \preceq_i \in \mathcal{R}',\ \forall g \in A,\ f \succeq_i g\}$ for all $A \in \wp(\mathcal{A})$. Note that $\mathcal{R}$ represents c in this sense.

We first show that $\{\preceq_i \mid i \in I,\ A_i \text{ essential}\}$ represents c. Since $\{\preceq_i \mid i \in I\} = \mathcal{R}$ represents c, for each $\preceq_i \in \{\preceq_i \mid i \in I,\ A_i \text{ essential}\}$ and each set $A \in \wp(\mathcal{A})$, $\{f \in A \mid f \succeq_i g,\ \forall g \in A\} \subseteq c(A)$. So for every $A \in \wp(\mathcal{A})$, $\bigcup_{i\ s.t. A_i \text{ essential}} \{f \in A \mid f \succeq_i g,\ \forall g \in A\} \subseteq c(A)$. 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 $f' \notin \bigcup_{i\ s.t. A_i \text{ essential}} \{f \in A \mid f \succeq_i g,\ \forall g \in A\}$ 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 $f \in \bigcup_{i \in I} A_i \backslash \bigcup_{A_i \text{ essential}} A_i$. Take any A_i with $f \in A_i$ that is $\sqsupseteq$-maximal – i.e. for every A_k such that $f \in A_k$, $A_i \sqsupseteq A_k$. Since $\sqsupseteq$ is complete and the strict relation generated by it is transitive, such a maximal element exists. For such A_i, every set $J \subset I \backslash \{i\}$ 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 $A_i \sqsupseteq A_k$. Hence A_i is essential, contrary to the definition of f. So, by reductio, there exists no such f, and $\bigcup_{i\ s.t. A_i \text{ essential}} \{f \in A \mid f \succeq_i g,\ \forall g \in A\} = c(A)$; so $\{\preceq_i \mid i \in I,\ A_i \text{ essential}\}$ 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 $\{\preceq_k \mid k \in K \subseteq I\}$ representing c and $K \not\supseteq \{i \in I \mid A_i \text{ essential}\}$. Take any $i \in \{i \in I \mid A_i \text{ essential}\} \backslash K$, 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 $A_i \not\sqsupseteq A_j$ 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.

□

D.2.1 Auxiliary Lemmas for the Proof of Proposition D.1

Lemma D.5. Let $(O, I) \in \mathcal{Z}$ be such that, for $b \in B$, $b \notin O$. Then there exists $(O', I') \geqslant (O, I)$ with $b \in O'$.

Proof of Lemma D.5. Since $b \notin O$, $\bar{b} \cap O = \mathbb{R}^*$. 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 D.6–D.12), I will remain fixed; as a point of notation, for any $a \in O$, we let $x_a \in \mathbb{R}$ be such that $I(a) = x_a$.

Let $[x_{\hat{a}}]^O = \{a \in O | I(a) = I(\hat{a})\} \cap U_{\hat{a}}$. 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 $F_b = \overline{(O' \backslash O)} \cap U_{\hat{a}*} \cap \{a \in B \mid \sum_{s \in S} a(s) = x_{\hat{a}}\}$. (Geometrically, this is the intersection between the closure of $O' \backslash O$, $U_{\hat{a}*}$ 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 $\sigma : F \to \mathbb{R}$, we let $Y_\sigma = \{a + \sigma(a)^* \mid a \in F\}$. Note that Y_σ is closed and bounded, because it is the image of a continuous map from a compact space to a Hausdorff one. We use $\leqslant$ to denote the standard dominance order on functions σ ($\sigma \leqslant \sigma'$ iff $\sigma(a) \leqslant \sigma'(a)$ for all $a \in F$).

Now consider the following set:

$$
(5) \qquad C_b = \left\{ \sigma : F_b \to \mathbb{R} \;\middle|\; \begin{array}{l} [x_{\hat{a}}]^O \cup Y_\sigma \subseteq c\left([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}}\right) \; \& \\ \forall d \in Y_\sigma,\ [d, x^*_{\hat{a}}] \subseteq Y_\sigma \; \& \\ \forall a' \in [x_{\hat{a}}]^O \cup Y_\sigma,\ \forall z > 0, \forall \alpha \in (0,1), \\ \alpha\hat{a} + (1-\alpha)x^*_{\hat{a}} \notin c\left([x_{\hat{a}}]^O \cup Y_\sigma \cup \{a' + z^*\} \cup X^{\hat{A}}_{x_{\hat{a}}}\right) \end{array} \right\}
$$

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

Lemma D.6. $C_b \neq \varnothing$.

Proof. Consider $\{\sigma : F_b \to \mathbb{R} | \; [x_{\hat{a}}]^O \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}})\}$. This set is non-empty by property 2f of I and the monotonicity of c (property v in Lemma D.1). Moreover, by the monotonicity of c and Chernoff, for any $\sigma', \sigma'' : F_b \to \mathbb{R}$ with $\sigma' \geqslant \sigma''$ if σ' is in this set, then so is σ''. It follows by continuity (property vii in Lemma D.1) that this set has a maximum element; let σ be any such element. By definition, we thus have that $[x_{\hat{a}}]^O \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}})$. Now consider any $c \in F_b$ and $\epsilon > 0$. By the maximality of σ, $[x_{\hat{a}}]^O \nsubseteq c([x_{\hat{a}}]^O \cup Y_{\sigma^{+\epsilon}_c} \cup X^{\hat{A}}_{x_{\hat{a}}})$, where $\sigma^{+\epsilon}_c(c) = \sigma(c) + \epsilon$ and $\sigma^{+\epsilon}_c(d) = \sigma(d)$ for $d \neq c$; by Chernoff, it follows that $[x_{\hat{a}}]^O \nsubseteq$

$c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{c + (\sigma(c) + \epsilon)^*\} \cup X^{\hat{A}}_{x_{\hat{a}}})$. It follows from Aizerman that $c + (\sigma(c) + \epsilon)^* \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{c + (\sigma(c) + \epsilon)^*\} \cup X^{\hat{A}}_{x_{\hat{a}}})$. Since this holds for any $\epsilon > 0$, it follows from continuity (property vii) that $c + \sigma(c)^* \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}})$. Since this holds for all $c \in F_b$, we have that $[x_{\hat{a}}]^O \cup Y_\sigma \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}})$. (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 $c([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}} \cup (\gamma \left([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}}\right) + (1-\gamma)x^*_{\hat{a}}))$. Since, by their definition, $\gamma[x_{\hat{a}}]^O + (1-\gamma)x^*_{\hat{a}} \subseteq [x_{\hat{a}}]^O$ and $\gamma X^{\hat{A}}_{x_{\hat{a}}} + (1-\gamma)x^*_{\hat{a}} \subseteq X^{\hat{A}}_{x_{\hat{a}}}$, $c([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}} \cup (\gamma \left([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}}\right) + (1-\gamma)x^*_{\hat{a}})) = c([x_{\hat{a}}]^O \cup Y_\sigma \cup (\gamma Y_\sigma + (1-\gamma)x^*_{\hat{a}}) \cup X^{\hat{A}}_{x_{\hat{a}}})$. Since $x^*_{\hat{a}} \in [x_{\hat{a}}]^O$, it follows from constant independence (property iv) that $[x_{\hat{a}}]^O \cup Y_\sigma \subseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup (\gamma Y_\sigma + (1-\gamma)x^*_{\hat{a}}) \cup X^{\hat{A}}_{x_{\hat{a}}})$. Moreover, it follows from the maximality of Y_σ that, for every $z > 0$, $[x_{\hat{a}}]^O \nsubseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{d + z^*\} \cup X^{\hat{A}}_{x_{\hat{a}}})$, so, by constant linearity $\gamma[x_{\hat{a}}]^O + (1-\gamma)x^*_{\hat{a}} \nsubseteq c(\gamma \left([x_{\hat{a}}]^O \cup Y_\sigma \cup \{d + z^*\} \cup X^{\hat{A}}_{x_{\hat{a}}}\right) + (1-\gamma)x^*_{\hat{a}})$, whence, by Chernoff $\gamma[x_{\hat{a}}]^O + (1-\gamma)x^*_{\hat{a}} \nsubseteq c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{\gamma(d+z^*) + (1-\gamma)x_{\hat{a}}\} \cup (\gamma Y_\sigma + (1-\gamma)x^*_{\hat{a}}) \cup X^{\hat{A}}_{x_{\hat{a}}})$. By Aizerman it follows that $\gamma(d+z^*) + (1-\gamma)x^*_{\hat{a}} \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{\gamma(d+z^*) + (1-\gamma)x^*_{\hat{a}}\} \cup (\gamma Y_\sigma + (1-\gamma)x^*_{\hat{a}}) \cup X^{\hat{A}}_{x_{\hat{a}}})$ for every $z > 0$, so by continuity $\gamma d + (1-\gamma)x^*_{\hat{a}} \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup (\gamma Y_\sigma + (1-\gamma)x^*_{\hat{a}}) \cup X^{\hat{A}}_{x_{\hat{a}}})$. Since $d \in Y_\sigma$, it follows by the definition of Y_σ that there exists $x \in \mathbb{R}$ such that $\gamma d + x^* \in Y_\sigma$. Hence, $\gamma d + x^*, \gamma d + (1-\gamma)x^*_{\hat{a}} \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup (\gamma Y_\sigma + (1-\gamma)x^*_{\hat{a}}) \cup X^{\hat{A}}_{x_{\hat{a}}})$; by monotonicity (property v), it follows that $x = (1-\gamma)x_{\hat{a}}$, and so $\gamma d + (1-\gamma)x^*_{\hat{a}} \in Y_\sigma$. Since this holds for every $\gamma \in (0,1)$, we have that $[d, x^*_{\hat{a}}] \subseteq Y_\sigma$, as required.

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

□

Lemma D.7. For all $d \in F_b$, $\alpha > 0$ and $x \in \mathbb{R}$ such that $\alpha d + x^* \in F_b$, and all $\sigma \in C_b$, $\sigma(\alpha d + x^*) = (1-\alpha)x^*_{\hat{a}} - x^* + \alpha\sigma(d)$.

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 $\alpha(d + \sigma(d)^*) + (1-\alpha)x^*_{\hat{a}} \in Y_\sigma$. It thus follows by monotonicity (property v) and the definition C_b, $\sigma(\alpha d + x^*)$ is such that $\alpha d + x^* + \sigma(\alpha d + x^*)^* = \alpha(d + \sigma(d)^*) + (1-\alpha)x^*_{\hat{a}}$. So $\sigma(\alpha d + x^*) = (1-\alpha)x^*_{\hat{a}} - x^* + \alpha\sigma(d)$, as required.

□

Take any element $\sigma \in C_b$; Lemma D.6 guarantees that such an element exists. Let $I'' : O \cup F_b \to \mathbb{R}$ to be the functional extending I and such that $I''(d) = x_{\hat{a}} - \sigma(d)$ for $d \in F_b$. By Lemma D.7, 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 D.8. I' satisfies property 2e: for all $d \in O'$, if $d \in \hat{A}$, then $I'(d) \leqslant x_{\hat{a}}$.

Proof. Since for $d \in O$ this follows from the properties of I, it suffices to consider $d \in O' \backslash O$. Suppose for reductio that $d \in \hat{A}$ and $I'(d) > x_{\hat{a}}$. By the definition of F_b, there exists $\beta \in (0,1]$, $x \in \mathbb{R}$ 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 $\beta d + x^* + \sigma(\beta d + x^*)^* \in c([x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}})$; however, by the monotonicity of c (property v of Lemma D.1), since $\beta d + (1-\beta)x^*_{\hat{a}} \in X^{\hat{A}}_{x_{\hat{a}}} \subseteq [x_{\hat{a}}]^O \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}}$ and $x + \sigma(\beta d + x^*) < (1-\beta)x_{\hat{a}}$, $\beta d + x^* + \sigma(\beta d + x^*)^* \notin c([x_{\hat{a}}]^O \cup [b + z^*, x^*_{\hat{a}}] \cup X^{\hat{A}}_{x_{\hat{a}}})$, which is a contradiction. So $I'(d) \leqslant x_{\hat{a}}$, as required.

□

Lemma D.9. I' is superadditive: for all $a, a' \in O'$ and $\alpha \in [0,1]$, then $I'(\alpha a + (1-\alpha)a') \geqslant \alpha I'(a) + (1-\alpha)I(a')$.

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 D.1) that $d \in c\left([x_{\hat{a}}]^O \cup Y_\sigma \cup \{d\} \cup X^{\hat{A}}_{x_{\hat{a}}}\right)$ for all $d \in co([x_{\hat{a}}]^O \cup Y_\sigma)$; hence $\alpha(a + (x_{\hat{a}} - I'(a))^*) + (1-\alpha)(a' + (x_{\hat{a}} - I'(a'))^*) \in c([x_{\hat{a}}]^O \cup \{\alpha(a + (x_{\hat{a}} - I'(a))^*) + (1-\alpha)(a' + (x_{\hat{a}} - I'(a'))^*)\} \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}})$. 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 D.1) that $\alpha(a+(x_{\hat{a}}-I'(a))^*)+(1-\alpha)(a'+(x_{\hat{a}}-I'(a'))^*) \notin c([x_{\hat{a}}]^O \cup \{\alpha(a+(x_{\hat{a}}-I'(a)))^*)+(1-\alpha)(a'+(x_{\hat{a}}-I'(a')))^*)\} \cup Y_\sigma \cup X^{\hat{A}}_{x_{\hat{a}}})$, which is a contradiction. Hence $I'(\alpha a + (1 - \alpha)a')) \geqslant \alpha I'(a) + (1 - \alpha)I(a')$, as required.

□

Lemma D.10. I' is monotonic: for every $a, d \in O'$, if $d \leqslant a$, then $I'(d) \leqslant I'(a)$.

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 $d \leqslant a$, then, by the monotonicity of c, for each $y > 0$, since $d + (x_{\hat{a}} - I'(a) - y)^* < a + (x_{\hat{a}} - I'(a))^*$, $d + (x_{\hat{a}} - I'(a) - y)^* \notin c([x_{\hat{a}}]^O \cup Y_\sigma \cup \{d + (x_{\hat{a}} - I'(a) - y)^*\} \cup X^{\hat{A}}_{x_{\hat{a}}})$, and so $x_{\hat{a}} - I'(a) - y \neq x_{\hat{a}} - I'(d)$. Hence $I'(a) \geqslant I'(d)$, as required.

□

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

Lemma D.11. Let $A \in \wp_{bdd}(O')$, $A \subset P$ for some $P \in \mathcal{Q}$, and let $x \in \mathbb{R}$ be such that $x \geqslant I'(a'')$ 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 $\alpha \geqslant 0$ and $z \in \mathbb{R}$. Hence there exists $\gamma_d \geqslant 1$, $\beta' \in [0, 1]$ and $y' \in \mathbb{R}$ such that $d = \beta'(\gamma_d a' + (1 - \gamma_d)x^*) + (1 - \beta')y'^*$; take any such γ_d. Since A is closed and bounded, there exists $\gamma \geqslant \gamma_d$ 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 $y \in \mathbb{R}$ 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 $y \in \mathbb{R}$ 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,[36] 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.

[36]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).)

□

Lemma D.12. I' represents c on O': for any $A \in \wp_{bdd}(O')$, $\arg\max_A I' \subseteq c(A)$.

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 D.11, 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 $A^a_P = (A \cap P) \cup [a'_P, x^*_a]$. By Chernoff and Aizerman,[37] $c(\bigcup_{P \text{ s.t } A\cap P \neq \emptyset} A^a_P) = c(\bigcup_{P \text{ s.t } A\cap P \neq \emptyset}[a'_P, x^*_a])$. 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, $a \in c(\bigcup_{P \text{ s.t } A\cap P \neq \emptyset} A^a_P)$; since $A \subseteq \bigcup_{P \text{ s.t } A\cap P \neq \emptyset} A^a_P$, it follows by Chernoff that $a \in c(A)$, as required.

□

So $(O', I') \in \mathcal{Z}$ such that $b \in O'$ and $(O, I) \leqslant (O', I')$. This concludes the proof of Lemma D.5. □

Lemma D.13. $\mathcal{Z}$ is non-empty.

Proof of Lemma D.13. Define $C_{\hat{a}}$ as follows:

$$C_{\hat{a}} = \left\{ y \in \mathbb{R} \;\middle|\; \begin{array}{l} [\hat{a}, y^*] \subseteq c\left(X^{\hat{A}}_y\right) \;\& \\ \forall a' \in [\hat{a}, y^*],\ \forall z > 0, \forall \alpha \in (0,1), \\ \alpha\hat{a} + (1-\alpha)y^* \notin c\left(\{a' + z^*\} \cup X^{\hat{A}}_y\right) \end{array} \right\} \tag{6}$$

We first establish the non-emptyness of $C_{\hat{a}}$, by an argument related to, but not identical to, that used in the proof of Lemma D.6. Consider $\left\{\sigma : [0,1) \to \mathbb{R} \;\middle|\; \hat{a} \in c\left(\hat{A} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in [0,1)\}\right)\right\}$. 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 σ be such an element. By the maximality of σ, $\hat{a} \in c\left(\hat{A} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in [0,1]\}\right)$. For $\alpha \in [0,1)$ and $\epsilon > 0$, define $\sigma^{+\epsilon}_\alpha$ by: $\sigma^{+\epsilon}_\alpha(\alpha) = \sigma(\alpha) + \epsilon$ and $\sigma^{+\epsilon}_\alpha(\beta) = \sigma(\beta)$ for $\beta \neq \alpha$. By the maximality of σ, for any $\alpha \in [0,1)$ and $\epsilon > 0$, $\hat{a} \notin c\left(\hat{A} \cup \{\alpha\hat{a} + (1-\alpha)\sigma^{+\epsilon}_\alpha(\alpha)^* \mid \alpha \in [0,1)\}\right)$. It follows by Aizerman that $\alpha\hat{a} + (1-\alpha)(\sigma(\alpha)+\epsilon)^* \in c\left(\hat{A} \cup \{\alpha\hat{a} + (1-\alpha)\sigma^{+\epsilon}_\alpha(\alpha)^* \mid \alpha \in [0,1)\}\right)$. Since this holds for all $\epsilon > 0$ and $\alpha \in [0,1)$, it follows from continuity that

[37] 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.

$\{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in [0,1)\} \subseteq c\left(\hat{A} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in [0,1]\}\right)$. Let $y = \sigma(0)$; in particular, we have 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 $\{\hat{\alpha}, y^*\} \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0,1)\} \subseteq c\left(\{y^*\} \cup \bigcup_{k=0}^{n}(\delta^k\hat{A} + (1-\delta^k)y^*) \cup \{\alpha\hat{a} + (1-\alpha)\sigma(\alpha)^* \mid \alpha \in (0,1)\}\right)$. Taking the limit as $\delta \to 1$ and $n \to \infty$, 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 $\sigma(b) = y$. 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 $\beta \leqslant \alpha$ or not. First suppose that $\beta \leqslant \alpha$. 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 $\{\beta\hat{a} + (1-\beta)y^* + z^*\} \cup \{\alpha d + (1-\alpha)y^* \mid d \in \hat{A}\} = \alpha(\{(\frac{\beta}{\alpha}\hat{a} + (1-\frac{\beta}{\alpha})y^*) + \frac{1}{\alpha}z^*)\} \cup X_y^{\hat{A}}) + (1-\alpha)y^*$, it follows from constant linearity of c that $\hat{a} \in c(\{(\frac{\beta}{\alpha}\hat{a} + (1-\frac{\beta}{\alpha})y^*) + \frac{1}{\alpha}z^*)\} \cup X_y^{\hat{A}})$, contradicting the maximality of $\sigma(\frac{\beta}{\alpha}) = y$. 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 $\overline{\hat{a}}$ proceeds in an analogous way to the proof of Lemma D.5; the proofs that it satisfies the appropriate conditions are either trivial or follow the same reasoning as used in the proof of Lemma D.5.

□