Hypothesis Testing and Ambiguity Aversion*

Pietro Ortoleva[†]

August 2, 2014

Abstract

We study a model of non-Bayesian updating, based on the Hypothesis Testing model of Ortoleva (2012), for ambiguity averse agents. Agents ranks acts following the MaxMin Expected Utility model of Gilboa and Schmeidler (1989) and when they receive new information they update their set of priors as follows: If the information is such that all priors in the original set of priors assign to it a probability above a threshold, then the agent updates every prior in the set using Bayes' rule. Otherwise: she looks at a prior over sets of priors; she updates it using a rule similar to Bayes' rule for second order beliefs over sets; finally, she chooses the set of priors to which the updated prior over sets priors assigns the highest likelihood.

JEL: *C61, D80, D81*

Keywords: *Bayes' Rule, Updating, Dynamic Coherence, Dynamic Consistency, Unambiguous Preferences, Ambiguity Aversion.*

1 Introduction

In recent years a large literature has studied the case in which decision-makers are instead *ambiguity averse*, i.e., depart from Expected Utility maximization in line with the Ellsberg paradox, a pattern robustly documented in many experimental settings. (See Gilboa and Marinacci 2011 for a survey of this vast literature.) Many papers have

*A previous version of this paper circulated under the title "Hypothesis Testing and Multiple Priors." I would like to thank Mark Dean, Paolo Ghirardato, Larry Epstein, Jawwad Noor, Efe Ok, Leonardo Pejsachowicz, Gil Riella, and the participants at various seminars and conferences for useful comments and discussions.

[†]Department of Economics, Columbia University. Email: pietro.ortoleva@columbia.edu.

studied how these preferences should be updated in the presence of new information, documenting the complex interaction between ambiguity aversion and updating. See, amongst many, Gilboa and Schmeidler (1993), Epstein and Le Breton (1993), Epstein and Schneider (2003), Maccheroni, Marinacci, and Rustichini (2006), Hanany and Klibanoff (2007, 2009), Epstein and Schneider (2007), Ghirardato, Maccheroni, and Marinacci (2008), and Siniscalchi (2011). Virtually all of these works, however, have focused primarily on updating rules that extend Bayes' rule to the case of ambiguity aversion, without studying the possibility that subjects may be non-Bayesian in the first place. In particular, almost all of these models reduce to Bayes' rule if the agent is ambiguity neutral, i.e., when she is an Expected Utility maximizer.

On the other hand, motivated by extensive experimental findings, a separate literature has studied the case in which agents may deviate from Bayes' rule. Among the many behavioral models, see for example, Barberis et al. (1998), Daniel et al. (1998), Rabin and Schrag (1999), Kahneman and Frederick (2002), Mullainathan (2002), Rabin (2002), Mullainathan et al. (2008), Gennaioli and Shleifer (2010), and Rabin and Vayanos (2010). Amongst the axiomatic models, see Epstein (2006), Epstein et al. (2008), and Ortoleva (2012). All of these papers, however, have focused on the case in which subjects are Expected Utility maximizers (ambiguity neutral), without considering the issue of ambiguity aversion. In particular, Ortoleva (2012) characterizes axiomatically a model in which agents acts like a standard bayesian ones as long as the information they receive is assigned a likelihood above a certain threshold, but who might react differently if they receive an information to which she assigned a smaller, or zero, probability. This model is called the Hypothesis Testing model. Ortoleva (2012) shows that this updating model can be characterized axiomatically using an axiom called Dynamic Coherence together with standard postulates.

The goal of this paper is to study axiomatically a model in which agents are ambiguity averse both before and after receiving new information, but at the same time follow a non-Bayesian updating rule. We impose axioms that combine standard ones in the literature of ambiguity aversion for preferences before and after receiving new information. In addition, we impose the Dynamic Coherence axiom of Ortoleva (2012). Our main result is to show that these axioms are equivalent to the following model. Both before and after receiving new information, the agent ranks acts following one of the most well-known models of ambiguity averse preferences, the MaxMin Expected Utility model of Gilboa and Schmeidler (1989): agents have not one, but a *set* of priors over the possible states of the world, and evaluate acts using the most pessimist priors for each act. It is well-known that this model can accommodate behavior as in the Ellsberg paradox. The key innovation is how the sets of priors are chosen by the decision maker before and after new information. To do this, our agent has a prior over sets of priors ρ, and a threshold ϵ between 0 and 1. She then chooses her sets of priors as follows:

- Before new information arrives, she chooses the set of priors Π to which the prior over sets of priors ρ assigns the highest likelihood.

Not knowing which set of priors to use, the agent picks the one that she considers the most likely according to her prior over sets of priors – in a way reminiscent of *maximum likelihood*.

As new information i is revealed, our agent tests her set of priors to verify whether she was using the correct one. This is where she uses her threshold ϵ.

- If the probability that *all* priors in her set assign to the new information i is *above* the threshold ϵ, i.e., if $\pi(i) > \epsilon$ for all $\pi \in \Pi$, then the set of priors π is *not* rejected, and the agent uses a set of priors that contains all priors in Π updated using Bayes' rule.

That is, if new information is not "unexpected," the agent follows a procedure similar to Bayesian updating (with sets of priors) – for example, as in the model of Ghirardato et al. (2008).

- If, however, the probability that at least one of the priors assigned to the new information i is *below* the threshold, i.e., if $\pi(i) \leq \epsilon$ for some $\pi \in \Pi$, then the set of priors is *rejected*, and our agent goes back to her prior over sets of priors ρ; updates it with a procedure similar to Bayes' rule for priors over sets of priors; then chooses the set of priors Π' to which the *updated* prior over sets of priors assigns the highest likelihood.

That is, if the set of priors is rejected by the data, then our agent reconsiders which set to use by picking the new maximum likelihood one, which is obtained by looking at her prior over sets of priors after it has been updated using the new information.

It is easy to see that, when $\epsilon = 0$, then our agent follows a prior-by-prior Bayesian updating whenever possible. When we also have that the set of priors Π before the new information is a singleton ($|\Pi| = 1$), then the agent is a standard Expected Utility maximizer both before and after new information.

When $\epsilon > 0$ our agent's behavior instead departs from Bayesian updating: when she receives information to which at least one prior assigned a probability below the threshold ϵ, she does not simply update her set of priors, but rather reconsiders whether she was using the right one in the first place. She might then modify her beliefs more than what prescribed by prior by prior updating, in line with the intuition of the Hypothesis Testing model of Ortoleva (2012), albeit applied to ambiguity averse preferences.

2 The model

2.1 Formal Setup

We adopt a standard dynamic setup of preferences under uncertainty, the same one used in Ortoleva (2012). We have a finite (non-empty) set Ω of states of the world, a σ-algebra Σ over Ω, and a (non-empty) set X of consequences. We assume that X is a convex subset of a vector space. (A typical case is the one in which X is the set of all the lotteries on a set of prizes, which is the classic setting of Anscombe and Aumann 1963.) Denote by $\Delta(\Omega)$ the set of all Borel probability measures (priors) on $\Delta(\Omega)$. By $\mathcal{C}$ denote the set of closed and convex subsets of $\Delta(\Omega)$. Correspondingly, denote by $\Delta(\mathcal{C})$ the set of all probability measures (priors over sets of priors) over $\Delta(\mathcal{C})$. Also, denote by $\mathcal{F}$ the set of all acts, that is, the set of all finite-valued Σ-measurable functions $f : \Omega \to X$. For every state of the world, an act returns a consequence in X.

The primitive of our analysis is a class of non-degenerate preference relations $\{\succeq_A\}_{A\in\Sigma}$, where by $\succeq_A$ we understand the preference of the agent after she receives the information $A \in \Sigma$; we denote by $\succeq = \succeq_\Omega$ the preference at time 0, before the agent receives any information.

With a standard abuse of notation, for any $x \in X$ denote by $x \in \mathcal{F}$ the constant act that yields the consequence x at every state $\omega \in \Omega$. For any $A \in \Sigma$, $f, g \in \mathcal{F}$, denote by $fAg \in \mathcal{F}$ the act that coincides with f in A and with g outside of it, that is, $fAg(\omega) = f(\omega)$ for every $\omega \in A$, and $fAg(\omega) = g(\omega)$ for every $\omega \in \Omega \backslash A$.

2.2 Axiomatic Foundations

We now turn to discuss our axiomatic foundations. For studying ambiguity aversion, we follow the well-known model of Gilboa and Schmeidler (1989) and we will therefore adapt their axiomatic framework. We posit that in every period, and after every information, the agent have preferences that satisfy their axioms: continuity, C-independence, uncertainty aversion, and monotonicity. (We refer to Gilboa and Schmeidler (1989) for a in-depth discussion of these postulates.) In addition, we also posit that the arrival of new information does not affect the ranking of unambiguous act – which implies that, loosely speaking, the arrival of new information affects the beliefs but not the utility. We collect these standard postulate in the following axiom.

Axiom 1 (Well-Behaved Standard Preferences with Ambiguity Aversion (WBP-AA))**.** *For any $A \in \Sigma$, $f, g, h \in \mathcal{F}$:*

1. *(*Continuity*): the sets* $\{f' \in \mathcal{F} : f' \succeq_A f\}$ *and* $\{f' \in \mathcal{F} : f \succeq_A f'\}$ *are closed;*

2. *(*C-Independence*): for any* $\alpha \in (0,1)$, $x \in \Delta(X)$

$$f \succeq_A g \Leftrightarrow \alpha f + (1-\alpha)x \succeq_A \alpha g + (1-\alpha)x;$$

3. *(*Uncertainty Aversion*) for any* $\alpha \in (0,1)$, *if* $f \sim_A g$ *then* $\alpha f + (1-\alpha)g \succeq_A f$.

4. *(*Monotonicity*): if* $f(\omega) \succeq_A g(\omega)$ *for all* $\omega \in \Omega$, *then* $f \succeq_A g$.

5. *(*Constant Preference Invariance*): for any* $B \in \Sigma$, $p, q \in \Delta(X)$, $p \succeq_A q \Leftrightarrow p \succeq_B q$

We next impose that the agent 'believes' in the information she receives: if she is told that the true state lies inside some $A \in \Sigma$, then she is indifferent between two acts that differ only outside of A. This is another very standard postulate, Consequentialism.

Axiom 2 (Consequentialism (C))**.** *For any* $A \in \Sigma$, *and* $f, g \in \mathcal{F}$, *if* $f(\omega) = g(\omega)$ *for all* $\omega \in A$, *then* $f \sim_A g$.

Finally, we impose the Dynamic Coherence axiom introduced in Ortoleva (2012). The basic idea is that want to rule out the possibility that our agent has a *circular* reaction to information. We refer to Ortoleva (2012) for an in-depth discussion.

Axiom 3 (Dynamic Coherence)**.** *For any* $A_1, \ldots, A_n \in \Sigma$, *if* $(\Omega \backslash A_{i+1})$ *is* $\succeq_{A_i}$*-null for* $i = 1, \ldots, (n-1)$, *and* $(\Omega \backslash A_1)$ *is* $\succeq_{A_n}$*-null, then* $\succeq_{A_1} = \succeq_{A_n}$.

The axioms outlined above are the key axioms of our paper. Before we proceed, let us know discuss additional postulates that will allow us to obtain further specifications.

One the most well-known postulates used to study the evolution of preferences with the arrival of information is Dynamic Consistency.

Axiom 4 (Dynamic Consistency (DC))**.** *For any* $A \in \Sigma$, A *not* $\succeq_\Omega$*-null, and for any* $f, g \in \mathcal{F}$, *we have*

$$f \succeq_A g \Leftrightarrow fAg \succeq_\Omega g.$$

It is well known that, when the agent is an Expected Utility maximizer both before and after the arrival of new information, then her preference satisfy Dynamic Consistency if and only if they she updates her prior using Bayes' rule. (See Ghirardato 2002 for an in-depth discussion of this postulate and its implications.) However.

it is well know that when agents are ambiguity averse, especially a' la Gilboa and Schmeidler (1989), Dynamic Consistency might be too strong of a requirement. For example, Epstein and Schneider (2003) show that an agents who updates her set of priors by updating each prior using Bayes' rule may violate Dynamic Consistency.[1] In fact, Ghirardato et al. (2008) show that this latter model can be characterized by a weakening of Dynamic Consistency, in which the same requirements of the axiom apply not to the full preference relation, but only to a *subset* of it, which they call the 'unambiguously preferred' relation: Following Ghirardato et al. (2004), for any $A \in \Sigma$ define the preference relation $\succeq_A^*$ as

$$f \succeq_A^* g \quad \text{if} \quad \lambda f + (1-\lambda)h \succeq_A \lambda g + (1-\lambda)h$$

for all $\lambda \in [0,1]$ and all $h \in \mathcal{F}$. Intuitively, $\succeq_A^*$ is the largest restriction of $\succeq_A$ that satisfy independence (and we clearly have $\succeq_A^* = \succeq_A$ if the latter satisfies independence). This is typically interpreted as the comparisons that the agent feels confident in making.[2]

Correspondingly, we can also define the set of unambiguously non-null events.

Definition 1. *An event $A \in \Sigma$ is $\succeq$-**unambiguously non-null** if for all $p, q \in \Delta(X)$ we have that $\{z \in X : xAy \succeq^* z\} \supset \{z \in X : y \succeq^* z\}$.*

The idea is that an event is unambiguously non-null if betting on A is unambiguously better than getting the loss payoff y for sure. (See Ghirardato et al. (2008) for more discussion.)

Then, following Ghirardato et al. (2008) we impose that the agent has a dynamically consistent behavior on the unambiguous preference and on unambiguously non-null events.

Axiom 5 (Restricted Dynamic Consistency (RDC))**.** *For any $A \in \Sigma$, A $\succeq$-unambiguously non-null, and for any $f, g \in \mathcal{F}$, we have*

$$f \succeq_A^* g \Leftrightarrow fAg \succeq^* g.$$

We refer to Ghirardato et al. (2008) for further discussion. Notice that if every preference satisfies independence, then Restricted Dynamic Consistency (Axiom 5) is clearly equivalent to standard Dynamic Consistency.[3]

[1] In particular, they show that the agent satisfies Dynamic Consistency if and only if the set of priors they use in the first period satisfies a property called *rectangularity*.

[2] Similar notions are used in the case of risk in Cerreia-Vioglio (2009) and Cerreia-Vioglio et al. (2014).

[3] Moreover, if there are no null events, then under WBP-AA and Consequentialism we have that Dynamic Coherence is *weaker* than Restricted Dynamic Consistency. In general, the two axioms are not nested. (Both statements follow as immediate consequence of Theorem 1 below.)

2.3 The Hypothesis Testing Model with Ambiguity Aversion

Before we introduce our representation, let us discuss a few notions of updating priors and sets of priors. First, let us define a notation for Bayesian Updating. For any $\pi \in \Delta(\Omega)$ and $A \in \Sigma$ such that $\pi(A) > 0$, define $\mathrm{BU}(\pi, A) \in \Delta(\Omega)$ (bayesian update of π using A) as

$$\mathrm{BU}(\pi, A)(B) := \frac{\pi(A \cap B)}{\pi(A)} \tag{1}$$

for all $B \in \Sigma$. In a model with ambiguity aversion, however, agents may have not one, but a set of priors. We can then define a natural extension of the notion of Bayesian updating to sets of priors: when all priors in the set are updated following Bayes' rule. For any $\Pi \subseteq \mathcal{C}$ and $A \in \Sigma$ such that $\pi(A) > 0$ for all $\pi \in \Pi$, define $\hat{BU}(\Pi, A)$ as

$$\hat{\mathrm{BU}}(\Pi, A) := \{\mathrm{BU}(\pi, A) : \pi \in \Pi\}.$$

Finally, we define a notion of updating of a prior over sets of priors. Recall the following standard definition: if we have a prior over priors (not over *sets* of priors) $\rho \in \Delta(\Delta(\Omega))$ with finite support, and an event $A \in \Sigma$ such that $\pi(A) > 0$ for some $\pi \in \mathrm{supp}(\rho)$, then the Bayesian updating of this prior over priors is

$$\mathrm{BU}(\rho, A)(\pi) := \frac{\pi(A)\rho(\pi)}{\int_{\Delta(\Omega)} \pi'(A)\rho(\mathrm{d}\pi')} \tag{2}$$

for all $\pi \in \Delta(\Omega)$. We now wish to extend this notion to priors over *sets* of priors. For any $\rho \in \Delta(\mathcal{C})$ and $A \in \Sigma$ such that for some $\Pi \in \mathrm{supp}(\rho)$ we have $\pi(A) > 0$ for all $\pi \in \Pi$, define $\overline{\mathrm{BU}}(\rho, A)$ as

$$\overline{\mathrm{BU}}(\rho, A)(\Pi) := \frac{\min\limits_{\pi \in \Pi} \pi(A)\ \rho(\Pi)}{\int\limits_{\mathcal{C}} \min\limits_{\pi \in \Pi} \pi(A)\ \rho(\mathrm{d}\Pi)}.$$

Intuitively, the agent assign to each set of priors a weight that is depends on the smallest likelihood assigned by any prior π in Π to the new information A. That is, this is a notion of cautious updating, in line with the intuition of the MaxMin Expected Utility model of Gilboa and Schmeidler (1989).

Definition 2. *A class of preferences relations $\{\succeq_A\}_{A \in \Sigma}$ admits an **Hypothesis Testing Representation with Ambiguity Aversion** if there exists a continous function $u : X \to \mathbb{R}$, a prior $\rho \in \Delta(\mathcal{C})$ with finite support, and $\epsilon \in [0, 1]$ such that for any $A \in \Sigma$ there exist $\Pi_A \in \mathcal{C}$ such that:*

1. *for any $f, g \in \mathcal{F}$*

$$f \succeq_A g \Leftrightarrow \min_{\pi \in \Pi_A} \sum_{\omega \in \Omega} \pi(\omega)\mathbb{E}_{f(\omega)}(u) \geq \min_{\pi \in \Pi_A} \sum_{\omega \in \Omega} \pi(\omega)\mathbb{E}_{g(\omega)}(u);$$

2. $\{\Pi_\Omega\} = \arg\max_{\Pi \in \mathcal{C}} \rho(\Pi)$*;*

3.

$$\Pi_A = \begin{cases} \hat{\mathrm{BU}}(\Pi_\Omega, A) & \text{if } \pi(A) > \epsilon \text{ for all } \pi \in \Pi_\Omega \\ \hat{\mathrm{BU}}(\Pi^*_A, A) & \text{otherwise} \end{cases}$$

where $\{\Pi^*_A\} = \arg\max_{\Pi \in \mathcal{C}} \overline{\mathrm{BU}}(\rho, A)(\Pi)$*;*

4. for any $A \in \Sigma$ *there exist* $\Pi \in \mathrm{supp}(\rho)$ *such that* $\pi(A) > 0$ *for all* $\pi \in \Pi$*.*

In a Hypothesis testing representation with Ambiguity Aversion the agent has a utility function u, a prior over sets of priors ρ, and a threshold ϵ. In every period and after any information she ranks act exactly as in the MaxMin Expected Utility (MMEU) model of Gilboa and Schmeidler (1989): she has a *set* of priors over the states of the world, and she evaluate each act using the prior that returns the *lowest* Expected Utility for that act (using utility u). Thus, the model coincides with MMEU when it comes to how to rank acts in each instance. The key innovation is how are the sets of priors used by the agent determined.

First of all, before any new information is revealed, our agent chooses the set of priors Π_Ω to which the prior over sets of priors ρ assigns the highest likelihood, which is *unique*. Thus, she follows an approach similar to maximum likelihood, even though she is using sets of priors.

When she receives some new information A, then our agent follows one of two possible approaches. If the information that was revealed was assigned by *all* priors in her set of priors Π_Ω a probability above the threshold ϵ, i.e., if $\pi(A) > \epsilon$ for all $\pi \in \Pi_\Omega$, then the set of priors is not rejected and our agent "keeps" it. She then adopts a set of priors obtained by updating all priors in Π using Bayes' rule. Notice that this is well-defined, as we must have $\pi(A) > \epsilon \geq 0$. This means that if the information is not unexpected for *any* of the priors in the set, then our agent follows an updating rule that can be seen as a version of Bayes' rule for sets of priors.

If, however, the information received was assigned by *any* of the priors in the set a probability below the threshold ϵ, i.e., if $\pi(A) \leq \epsilon$ for some $\pi \in \Pi_\Omega$, then the set agent reconsiders whether she is using the correct set of priors in the first place. She then updates her prior over sets of priors using a version reminiscent of Bayes' rule, and then she chooses the set of priors to which the *updated* prior over priors assigns the highest likelihood. (Note that she may end up choosing the same set of priors she was using before: this happens when, even after reconsidering, the old set of priors is still the best in the lot.)

Notice that to avoid any indeterminacy, in a Hypothesis Testing model with Ambiguity Aversion the prior over sets of priors ρ is constructed in such a way that the argmax of the updated prior over sets priors is always *unique*: this guarantees that the choice of priors is always well-defined.[4]

Notice also that in any Hypothesis Testing representation, for any event $A \in \Sigma$ there must exist some prior π in the support of ρ such that $\pi(A) > 0$. (Otherwise $\mathrm{BU}(\rho, A)$ would not be defined in part 3 of the representation.) This means that, even if the original belief of the agent assigned probability zero to an event, there must exist *some* belief in the support of ρ that assigns to that event a strictly positive probability.

Just like Dynamic Coherence is neither stronger nor weaker than Restricted Dynamic Consistency, a Hypothesis Testing representation is neither more general nor more restrictive than the model in which all priors in the set of priors are updated using Bayes' rule. On the one hand, when $\epsilon = 0$ the model coincides with that model whenever it is defined, but the agent's beliefs are disciplined also when it does not apply (not-unambiguously null events). Indeed the two models coincide if $\epsilon = 0$ and there are no null events. On the other hand, when $\epsilon > 0$ the model above allows for reactions to non-null events that can be very different from updating all priors using Bayes' rule.

2.4 Representation Theorem

We are now ready to state our representation theorem. To do so, however, it would be convenient to focus on a specific case of our representation: since there could be multiple values of ϵ that represent the same preferences, we focus on the representations with the smallest value of ϵ, what we call a minimal representation.

Definition 3. *An Hypothesis Testing Representation with Ambiguity Aversion* (u, ρ, ϵ)

[4]It is easy to see that one could construct an alternative representation in which this uniqueness is not required, but in which the agent is endowed not only with a prior over sets of priors ρ, but also with a linear order over sets of priors $\gtrdot$, to be used when there are multiple maximizers of the updated ρ. This representation would be equivalent to the one above, which suggests that the only role of the uniqueness requirement above is to guarantee that the behavior is well defined from the representation. We should also emphasize that requiring the uniqueness of the argmax, or adding a linear order to choose between multiple ones, is more than a technical detail: it is an essential condition for the model to have an empirical content. To see why, consider any agent, no matter how her beliefs are formed, and construct ρ as the uniform distribution over all possible sets priors. Then, notice that after any information A, the prior that the agent uses after A must belong to the argmax of the updated ρ, since it must give probability 1 to A: this means that we can represent *any* behavior with a model similar to the one above without the requirement that the argmax is unique (and setting $\epsilon = 1$), or without a fixed rule on how to choose between multiple ones.

*is **minimal** if there is no $\epsilon' \in [0,1)$ such that $\epsilon' < \epsilon$ and (u, ρ, ϵ') is an Hypothesis Testing Representation with Ambiguity Aversion of the same preferences.*

We are now ready to state our main representation theorem.

Theorem 1. *A class of preference relations $\{\succeq_A\}_{A\in\Sigma}$ satisfies WBP-AA, Consequentialism, and Dynamic Coherence if and only if it admits a minimal Hypothesis Testing Representation with Ambiguity Aversion (u, ρ, ϵ).*
Moreover, $\epsilon = 0$ if and only if $\{\succeq_A\}_{A\in\Sigma}$ satisfies also Restricted Dynamic Consistency.

Theorem 1 shows that the axioms described above are necessary and sufficient to characterize the Hypothesis Testing model with Ambiguity Aversion. Moreover, if Restricted Dynamic Consistency is also satisfied, we obtain a representation in which the agent updates the set of prior using Bayes' rule if she faces an event to which every prior in her set assigns positive probability; otherwise, she picks a new set of priors by maximizing the updated prior over sets of priors. Theorem 1 in Ortoleva (2012) to the case of ambiguity aversion, showing that same axiom used int the case of Expected Utility agents, Dynamic Coherence, leads to a very similar form of updating for ambiguity averse agents as well.

Appendix: Proof of Theorem 1

[*Sufficiency of the Axioms*] Our proof follows the one of Theorem 1 in Ortoleva (2012). Given Axiom 1, from Gilboa and Schmeidler (1989) we know that for any $A \in \Sigma$, there exist $u_A : X \to \mathbb{R}$, $\Pi_A \subseteq \Delta(\Omega)$, Π convex and compact, such that for any $f, g \in \mathcal{F}$

$$f \succeq_A g \Leftrightarrow \min_{\pi\in\Pi_A} \sum_{\omega\in A} \pi(\omega)\mathbb{E}_{f(\omega)}(u) \geq \min_{\pi\in\Pi_A} \sum_{\omega\in A} \pi(\omega)\mathbb{E}_{g(\omega)}(u) \tag{A.2}$$

where Π_A is unique and u_A is unique up to a positive affine transformation. It is also standard practice to show that Axiom 1.(5) implies that, for any $A \in \Sigma$, all u_A are positive affine transformations of u_Ω, which means that we can assume $u_\Omega = u_A$ for all $A \in \Sigma$. Define $u : X \to \mathbb{R}$ as $u = u_\Omega$ and $\pi = \pi_\Omega$. Moreover, notice that for any $A, B \in \Sigma$ A is $\succeq_B$-null if and only if $\pi(A) = 0$ for all $\pi \in \Pi_B$.

Notice that Claims 1 of Theorem 1 in Ortoleva (2012) holds true here as well. Moreover, notice the following claim (which parallels Claim 2 in the proof of Theorem 1 in Ortoleva 2012).

Claim 1. *For any $A, B \in \Sigma$, if $\pi_A(B) = 1 = \pi_B(A)$ for all $\pi_A \in \Pi_A$ and $\pi_B \in \Pi_B$, then $\Pi_A = \Pi_B$.*

Proof. Consider any $A, B \in \Sigma$ such that $\pi_A(B) = 1 = \pi_B(A)$ for all $\pi_A \in \Pi_A$ and $\pi_B \in \Pi_B$. Notice that by construction of Π_A and Π_B we must have $\pi_A(A \cap B) = 1 = \pi_B(A \cap B)$ for all $\pi_A \in \Pi_A$ and $\pi_B \in \Pi_B$. Hence $(\Omega \backslash (A \cap B))$ is both $\succeq_A$-null and $\succeq_B$-null But then, by Claim 1 in Ortoleva (2012) we must have $\Pi_A = \Pi_{A\cap B} = \Pi_B$ as sought. □

Define now the set $\mathcal{K}_{AA} \subseteq \Sigma$ as $\mathcal{K}_{\text{AA}} := \{A \in \Sigma : A \text{ is not } \succeq \text{-unambiguously non-null}\} \cup \{A \in \Sigma : \exists f, g \in \mathcal{F} \text{ s.t. } f \succeq_A^* g \text{ and } g \succ^* fAg, \text{ or } f \succ_A^* g \text{ and } g \succeq^* fAg\}$. These are the events after which either Reduced Dynamic Consistency (Axiom 5) does not apply (not unambiguously non-null events), or after which it is violated. Define ϵ as $\epsilon := \max_{A \in \mathcal{K}_{AA}} \max_{\pi \in \Pi_\Omega} \pi(A)$ if $\mathcal{K}_{\text{AA}} \neq \emptyset$, and $\epsilon = 0$ if $\mathcal{K}^*_{\text{AA}} = \emptyset$. (This is well defined since Ω is finite.)

Consider now $A \in \Sigma \backslash \mathcal{K}_{AA}$ (notice that this set includes all $A \in \Sigma$ such that $\pi(A) > \epsilon$ for all $\pi \in \Pi_\Omega$, by construction of ϵ).

Claim 2. *For any $A \in \Sigma \backslash \mathcal{K}_{AA}$, $f, g, h \in \mathcal{F}$, we have*

$$f \succeq_A g \Leftrightarrow fAh \succeq_\Omega gAh$$

Proof. Consider $f, g, h \in \mathcal{F}$, $A \in \Sigma \backslash \mathcal{K}_{AA}$. Notice first of all that by Axiom 2 we have $fAh \sim_A f$ and $gAh \sim_A g$. This implies that we have $f \succeq_A g$ iff $fAh \succeq_A gAh$. Define $f' := fAh$ and $g' := gAh$. Notice that we have $fAh \succeq_A gAh$ iff $f' \succeq_A g'$ iff (since $A \notin \mathcal{K}^*_{\text{AA}}$) $f'Ag' \succeq_\Omega g'$ iff $fAh \succeq_\Omega gAh$, as sought. $\square$

Finally, from Ghirardato, Maccheroni, and Marinacci (2004) we know that for any $A \in \Sigma$, $\succeq_A^*$ satisfies monotonicity, continuity and independence, and it can be represented by

$$f \succeq_A^* g \Leftrightarrow \sum_{\omega \in \Omega} \pi(\omega) \mathbb{E}_{f(\omega)}(u) \geq \sum_{\omega \in \Omega} \pi(\omega) \mathbb{E}_{f(\omega)}(u) \quad \forall \pi \in \Pi_A$$

where Π_A is a compact and convex subset of $\Delta(\Omega)$, it is the same as the one in Equation A.2, and it is unique.[5] This means that for any $A \in \Sigma$ such that $\pi(A) > \epsilon$ for all $\pi \in \Pi_\Omega$ (which also means $\pi(A) > 0$ for all $\pi \in \Pi_\Omega$) we have

$$\begin{aligned}
f \succeq_A^* g \quad &\Leftrightarrow \quad fAh \succeq_\Omega^* gAh \\
&\Leftrightarrow \quad \sum_{\omega \in A} \pi(\omega) \mathbb{E}_{f(\omega)}(u) + \sum_{\omega \in \Omega \backslash A} \pi(\omega) \mathbb{E}_{f(\omega)}(u) \geq \sum_{\omega \in A} \pi(\omega) \mathbb{E}_{g(\omega)}(u) + \sum_{\omega \in \Omega \backslash A} \pi(\omega) \mathbb{E}_{g(\omega)}(u) \quad \forall \pi \in \Pi_\Omega \\
&\Leftrightarrow \quad \sum_{\omega \in A} \pi(\omega) \mathbb{E}_{f(\omega)}(u) \geq \sum_{\omega \in A} \pi(\omega) \mathbb{E}_{g(\omega)}(u) \quad \forall \pi \in \Pi_\Omega \\
&\Leftrightarrow \quad \frac{1}{\pi(A)} \sum_{\omega \in A} \pi(\omega) \mathbb{E}_{f(\omega)}(u) \geq \frac{1}{\pi(A)} \sum_{\omega \in A} \pi(\omega) \mathbb{E}_{g(\omega)}(u) \quad \forall \pi \in \Pi_\Omega
\end{aligned}$$

Since Π_A is unique, this proves that for any $A \in \Sigma$ such that $\pi(A) > \epsilon$ for all $\pi \in \Pi_\Omega$, then $\Pi_A(B) = \hat{BU}(\Pi_\Omega, A)$.

Define now the set $\mathcal{K}^*_{AA}$ as follows.

$$\mathcal{K}^*_{AA} := \{A \in \Sigma : \ \pi(A) \leq \epsilon \text{ for some } \pi \in \Pi\}.$$

Define the sets $H_{AA} := \mathcal{K}^*_{AA} \cup \{\Omega\}$ and $M_{AA} := \{\pi_m \in \Delta(\Omega) : m \in \mathcal{K}^*_{AA}\} \cup \{\pi_\Omega\}$.

[5]See Ghirardato, Maccheroni, and Marinacci (2004). In Section 5.1 they discuss how their Theorem 14 implies that the set of priors found by the representation of $\succeq^*$ using their Theorem 11 must coincide with the one found with a representation of $\succeq$ a la Gilboa and Schmeidler (1989).

We can now proceed replicating exactly the steps in the proof of Theorem 1 in Ortoleva (2012) and prove the claims that parallel Claims 5, 6, and 7, with the following modifications: we use sets of priors Π_m instead of priors π_m; whenever we have $\pi_m(A) = 1$ replace it with $\pi(A) = 1$ for all $\pi \in \Pi_m$; replace the conditions $\pi(A) \leq \epsilon$ with the corresponding condition $\pi(A) \leq \epsilon$ for some $\pi \in \Pi_\Omega$; use the set $\mathcal{K}^*_{AA}, H_{AA}, M_{AA}$ instead of $\mathcal{K}^*, H, M$. In particular, construct the preference $\rhd$ on M_{AA}.

Define $\gamma^*_{AA} := \max\{\min_{\pi\in\Pi_A} \pi(B) : A, B \in \mathcal{K}^*_{AA}, \pi(B) < 1 \text{ for some } \pi \in \Pi_A\}$. (Notice that γ_* is well defined since M is finite.) If $\gamma^* > 0$, define $\delta := \frac{1}{|M|}\frac{1-\gamma_*}{\gamma_*}$; otherwise, if $\gamma^* = 0$, define $\delta = 1$. (Since we must have $\gamma_* \in [0, 1)$, then $\delta > 0$.) Proceed like in the proof of Theorem 1 in Ortoleva (2012) in constructing constructing the transitive closure of $\hat{\rhd}$ of $\rhd$, and the function f and v on M_{AA}, and construct $\rho \in \Delta(\mathcal{C})$ as

$$\rho(\Pi) := \frac{v(\Pi) + \frac{1}{|M|}}{\sum_{m\in M}(v(\Pi) + \frac{1}{|M|})}.$$

for all $\Pi \in M$, $\rho(\Pi) := 0$ otherwise. Finally, we need to prove that M_{AA} and ρ that we just constructed are the ones that we are looking for. But it's easy to replicate the passages in the proof of Claims 8, 9, 10 in the proof of Theorem 1 in Ortoleva (2012) (once again whenever we have $\pi_m(A) = 1$ we need to replace it with $\pi(A) = 1$ for all $\pi \in \Pi_m$, and use M_{AA}, v, and ρ as constructed here). Condition (4) will also be trivially true here as well by construction, since for any event $A \in \mathcal{K}^*_{AA}$ we construct $\Pi \in \text{supp}(\rho)$ such that each $\pi(A) > 0$ for all $\pi \in \Pi$.

[*Necessity of the Axioms*] Axiom 1 and 2 are immediate. We are left with Axiom 3. Consider $A_1, \ldots, A_n \in \Sigma$ such that $\succeq_{A_1} \neq \succeq_{A_n}$, $(\Omega\backslash A_{i+1})$ is $\succeq_{A_i}$-null for $i = 1, \ldots, (n-1)$, and $(\Omega\backslash A_1)$ is $\succeq_{A_n}$-null. Consider first the case in which $\pi(A_i) > \epsilon$ for all $\pi \in \Pi$, for all $i = 1, \ldots, n$. Since $(\Omega\backslash A_{i+1})$ is $\succeq_{A_i}$-null for $i = 1, \ldots, (n-1)$, then it must be that $\pi(A_i\backslash A_{i+1}) = 0$ for all $\pi \in \Pi_{A_i}$ for $i = 1, \ldots, (n-1)$, and $\pi(A_n\backslash A_1) = 0$ for all $\pi \in \Pi_{A_n}$. Notice that by the representation it must be that $\Pi_{A_i} = \text{BU}(\Pi_\Omega, A_i)$ for $i = 1, \ldots, n$. But then, for all $\pi \in \Pi_\Omega$ we have $\pi(A_i\backslash A_{i+1}) = 0$ for $i = 1, \ldots, (n-1)$, and $\pi(A_1\backslash A_n) = 0$. Hence $\pi(\cup_{i=1}^n A_i) = \pi(\cap_{i=1}^n A_i)$ for all $\pi \in \Pi_\Omega$, and so $\text{BU}(\Pi_\Omega, \cap_{i=1}^n A_i) = \text{BU}(\Pi_\Omega, A_j)$ for $j = 1, \ldots, n$. But this implies $\succeq_{A_1} = \succeq_{A_n}$, a contradiction.

Consider now the more general case in which there exist some i such that $\pi(A_i) > \epsilon$ for all $\pi \in \Pi_\Omega$. Say without loss of generality that we have $\pi(A_1) > \epsilon$ for all $\pi \in \Pi_\Omega$. By the representation it must be $\Pi_{A_1} = \text{BU}(\Pi_\Omega, A_1)$. At the same time we have $\pi(A_1\backslash A_2) = 0$ for all $\pi \in \Pi_{A_1}$, and by definition of BU this means that we have $\pi(A_1) = \pi(A_1 \cap A_2)$ for all $\pi \in \Pi_\Omega$. But then $\pi(A_2) \geq \pi(A_1)$, hence $\pi(A_2) > \epsilon$, for all $\pi \in \Pi_\Omega$. Proceed like this to prove that we must have $\pi(A_i) > \epsilon$ for all $\pi \in \Pi_\Omega$ for $i = 1, \ldots, n$. But we have already shown that this leads to a contradiction.

We are left with the case in which for $i = 1, \ldots, n$ we have $\pi(A_i) \leq \epsilon$ for some $\pi \in \Pi$. Since $(\Omega\backslash A_{i+1})$ is $\succeq_{A_i}$-null for $i = 1, \ldots, (n-1)$, then we must have that $\text{BU}(\pi, A_i)(A_{i+1}) = 1$ for all $\pi \in \Pi^*_{A_i}$ for $i = 1, \ldots, (n-1)$. For the same reason, we must have $\text{BU}(\pi, A_n)(A_1) = 1$ for all $\pi \in \Pi^*_{A_n}$. This implies that for $i = 1, \ldots, n-1$ we have that for all $\pi \in \Pi^*_{A_i}$ we have $\pi(A_{i+1}) \geq \pi(A_i)$, and for all $\pi \in \Pi^*_{A_n}$ we have $\pi(A_1) \geq \pi(A_n)$. Therefore, $\min_{\pi\in\Pi^*_{A_i}} \pi(A_{i+1}) \geq \min_{\pi\in\Pi^*_{A_i}} \pi(A_i)$ for $i = 1, \ldots, n-1$ and $\min_{\pi\in\Pi^*_{A_n}} \pi(A_1) \geq \min_{\pi\in\Pi^*_{A_n}} \pi(A_n)$. Now, notice that since $\Pi^*_{A_i}$ is the unique element in $\arg\max_{\Pi\in\mathcal{C}} \bar{\text{BU}}(\rho, A_i)(\Pi)$ for $i = 1, \ldots, n$, then we must have that for all $i, j = 1, \ldots, n$, $\rho(\Pi^*_{A_i}) \min_{\pi\in\Pi^*_{A_i}} \pi(A_i) > \rho(\Pi^*_{A_j}) \min_{\pi\in\Pi^*_{A_j}} \pi(A_i)$ if $\Pi^*_{A_i} \neq \Pi^*_{A_j}$. This, together with the fact that $\min_{\pi\in\Pi^*_{A_i}} \pi(A_{i+1}) \geq \min_{\pi\in\Pi^*_{A_i}} \pi(A_i)$ for $i = 1, \ldots, n-1$ and $\min_{\pi\in\Pi^*_{A_n}} \pi(A_1) \geq \min_{\pi\in\Pi^*_{A_n}} \pi(A_n)$, implies that $\rho(\Pi^*_{A_i}) \geq \rho(\Pi^*_{A_{i+1}})$ for $i = 1, \ldots, (n-1)$, where the inequality is strict if $\Pi^*_{A_i} \neq \Pi^*_{A_{i+1}}$, and $\rho(\Pi^*_{A_n}) \geq \rho(\Pi^*_{A_1})$, where the inequality is strict if $\Pi^*_{A_1} \neq \Pi^*_{A_n}$. But then we have $\rho(\Pi^*_{A_1}) \geq \rho(\Pi^*_{A_2}) \geq \ldots \rho(\Pi^*_{A_n}) \geq \rho(\Pi^*_{A_1})$, and so none of this inequalities can be strict, hence $\Pi^*_{A_1} = \Pi^*_{A_i}$

for $i = 1, \ldots, n$. But then, for all $\pi \in \Pi^*_{A_1}$ we have $\mathrm{BU}(\pi, A_i)(A_{i+1}) = 1$ for $i = 1, \ldots, n-1$, and $\mathrm{BU}(\pi, A_n)(A_1) = 1$, which means $\pi(A_i \backslash A_{i+1}) = 0$ for $i = 1, \ldots, n-1$, and $\pi(A_n \backslash A_1) = 0$. This means that for all $\pi \in \Pi^*_{A_1}$ we have $\pi(\cap_{i=1}^n A_i) = \pi(\cup_{i=1}^n A_i)$. But this implies that $\hat{\mathrm{BU}}(\Pi^*_{A_1}, A_n) = \hat{\mathrm{BU}}(\Pi^*_{A_1}, A_1)$, hence $\Pi_{A_1} = \Pi_{A_n}$ and $\succeq_{A_1} = \succeq_{A_n}$, a contradiction.

[$\epsilon = 0$ *iff Reduced Dynamic Consistency*] Ghirardato, Maccheroni, and Marinacci (2008) show that the agent updates her set of priors using Bayes' rule every time she is told that a $\succeq$-unambiguously non-null event has occurred if and only if she satisfies Reduced Dynamic Consistency. If $\epsilon = 0$, therefore, Reduced Dynamic Consistency applies. Conversely, assume that Reduced Dynamic Consistency is satisfied and let us say, by means of contradiction, that we have a minimal Hypothesis Testing representation with Ambiguity Aversion (u, ρ, ϵ) of $\{\succeq_A\}_{A\in\Sigma}$ in which $\epsilon \neq 0$. Since $\{\succeq_A\}_{A\in\Sigma}$ satisfies Reduced Dynamic Consistency, however, then $(u, \rho, 0)$ must also represent it, contradicting the minimality of (u, ρ, ϵ). *Q.E.D.*

References

Anscombe, F. J. and R. J. Aumann (1963): "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, 34, 199–205.

Barberis, N., A. Shleifer, and R. Vishny (1998): "A model of investor sentiment," *Journal of Financial Economics*, 49, 307–343.

Cerreia-Vioglio, S. (2009): "Maxmin Expected Utility on a Subjective State Space: Convex Preferences under Risk," Mimeo, Bocconi University.

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2014): "Cautious Expected Utility and the Certainty Effect," Mimeo California Institute of Technology.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998): "Investor Psychology and Security Market under- and Overreactions," *The Journal of Finance*, 53, 1839–1885.

Epstein, L. G. (2006): "An axiomatic model of non-Bayesian updating," *Review of Economic Studies*, 73, 413–436.

Epstein, L. G. and M. Le Breton (1993): "Dynamically Consistent Beliefs Must Be Bayesian," *Journal of Economic Theory*, 61, 1–22.

Epstein, L. G., J. Noor, and A. Sandroni (2008): "Non-Bayesian updating: a theoretical framework," *Theoretical Economics*, 3, 193–229.

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

——— (2007): "Learning under ambiguity," *Review of Economic Studies*, 74, 1275–1303.

Gennaioli, N. and A. Shleifer (2010): "What comes to mind," *Quarterly Journal of Economics*, 125, 1399–1433.

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

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): "Differentiating ambiguity and ambiguity attitude," *Journal of Economic Theory*, 118, 133–173.

——— (2008): "Revealed Ambiguity and Its Consequences: Updating," in *Advances in Decision Making under Risk and Uncertainty. Selected Papers from the FUR 2006 conference*, ed. by M. Abdellaoui and J. Hey, Berlin: Springer-Verlag.

Gilboa, I. and M. Marinacci (2011): "Ambiguity and the Bayesian paradigm," Mimeo Bocconi University.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18, 141–153.

——— (1993): "Updating Ambiguous Beliefs," *Journal of Economic Theory*, 59, 34–49.

Hanany, E. and P. Klibanoff (2007): "Updating preferences with multiple priors," *Theoretical Economics*, 2, 261–298.

——— (2009): "Updating ambiguity averse preferences," *The BE Journal of Theoretical Economics*, 9, 37.

Kahneman, D. and S. Frederick (2002): "Representativeness revisited: Attribute substitution in intuitive judgment," *Heuristics and biases: The psychology of intuitive judgment*, 49–81.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Dynamic variational preferences," *Journal of Economic Theory*, 128, 4–44.

Mullainathan, S. (2002): "Thinking through categories," *NBER working paper.*

Mullainathan, S., J. Schwartzstein, and A. Shleifer (2008): "Coarse Thinking and Persuasion," *Quarterly Journal of Economics*, 123, 577–619.

Ortoleva, P. (2012): "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 102, 2410–2436.

Rabin, M. (2002): "Inference By Believers in the Law of Small Numbers*," *Quarterly Journal of Economics*, 117, 775–816.

Rabin, M. and J. L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias*," *Quarterly Journal of Economics*, 114, 37–82.

Rabin, M. and D. Vayanos (2010): "The gambler's and hot-hand fallacies: theory and applications," *Review of Economic Studies*, 77, 730–778.

Siniscalchi, M. (2011): "Dynamic Choice under Ambiguity," *Theoretical Economics*, 6, 379–421.