

Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News

Pietro Ortoleva*

Preliminary Version – May 28, 2010

Abstract

Despite its normative appeal and widespread use, Bayes' rule has two well-known limitations: first, it does not predict how agents should react to an information to which they assigned probability zero; second, a sizable empirical evidence documents how agents systematically deviate from its prescriptions by overreacting to information to which they assigned a positive but *small* probability. We characterize axiomatically the following alternative updating rule. Our agent has a utility function over consequences, a prior over priors ρ , and a threshold. In the first period she chooses the prior that maximizes the prior over priors ρ - a' la maximum likelihood. As new information is revealed: if the chosen prior assigns to this information a probability above the threshold, she follows Bayes' rule and updates it. Otherwise, she goes back to her prior over priors ρ , updates that using Bayes' rule, and then chooses the new prior that maximizes the updated ρ .

JEL: D81, C61

Keywords: Bayes' Rule, Updating, Dynamic Consistency, Ambiguity Aversion.

1. Introduction

1.1 Basic Idea and Motivation

One of the most widespread assumptions in economics is that agents update their beliefs using Bayes' rule. It is easy to see the reasons for this popularity. First, it is extremely intuitive – so intuitive that some people even consider it a feature of rationality. This normative appeal is paired with a positive one: there are many instances in which agent's seem to follow Bayes' rule in revising their beliefs. And finally, its extremely convenient functional form facilitates its application even in complicated economic models.

Despite these advantages, however, Bayes' rule has two well-known limitations. First, it contains no prescription about how should the agent react to an information to which she

*Division of the Humanities and Social Sciences, California Institute of Technology, ortoleva@caltech.edu. I would like to thank Kim Border, Mark Dean, Federico Echenique, Larry Epstein, Paolo Ghirardato, Asen Kochov, Jawwad Noor, Efe Ok, Leonardo Pejsachowicz, Ariel Rubinstein, Gil Riella, Leat Yariv, and the participants at seminars at ASU, Boston University, Caltech, Collegio Carlo Alberto, NYU, USC and SWET 2010 for useful comments and discussions.

assigned probability zero – Bayes’ rule is simply not defined in that case. This limitation has well-known consequences: for example, the notion of Perfect Bayesian Nash Equilibrium is often criticized (and then refined) since it posits no restrictions on the beliefs of the agents out of equilibrium path – which is an immediate consequence of the fact that Bayes’ rule has no prescription about how should these beliefs be formed. Moreover, even in the cases in which Bayes’ rule does apply, a sizable amount of evidence collected in the past two decades has documented that decision makers systematically depart from its prescriptions.¹ While these departures take many forms, one seems to be of particular relevance: people tend to violate Bayes’ rule when they receive news that they did not foresee, i.e. information to which they originally assigned a positive but *small* probability: “in violation of Bayes’ rule most people tend to overreact to unexpected and dramatic news events” ((De Bondt and Thaler, 1985, pg. 804)).

The goal of this paper is to develop axiomatically an alternative updating rule that tries to maintain the elements of appeal of Bayes’ rule, including its simplicity, while reconciling with the two limitations mentioned above. In particular, we model an agent who reacts according to Bayes’ rule when “normal” news is given to her, but who might overreact, and change her prior beyond the bayesian prescriptions, when she receives some information that she “did not expect.” To give an example of the behavior we have in mind, think about an investor who is constructing a portfolio to allocate her wealth. To guide her decision, our investor has a belief over the returns of each possible investment, a belief that could originate from some properly calibrated economic model. As time goes by, our investor receives new information, e.g. from observing the stock market, and revises her belief. In normal times it is reasonable to expect that she revises her belief following Bayes’ rule. But what happens if a big, shocking news is revealed, like a financial crisis? While such crisis might have been considered possible by our investor – in the sense that she might have assigned it a positive probability – it might also have been very *unlikely* for her – she could have assigned to it a very *small* probability. How will she react? The point is, in light of this unexpected event our investor might go beyond updating her prior using Bayes’ rule: she might think that she used the wrong prior to begin with, given that it assigned such a small probability to the realized event, and that she should therefore pick a whole new prior based on the new information. The arrival of unexpected news might therefore lead her to a “change of paradigm” that entails a change of belief beyond Bayes’ rule. For example, if she was using an economic model to form her original belief, and this model assigned a small probability to the information that was later revealed, then she might question the validity of this model and look for a new one. In this paper we aim to study and characterize this behavior.

1.2 Overview of the Results

We consider a standard dynamic model in which we observe the preferences of the agent before and after she receives some information about the state of the world. The main result of this paper is to provide an axiomatic foundation for what we call the *Hypothesis Testing*

¹For experimental evidence of violations of Bayes’ rule in general, and for low probability events in particular, see, among others, Tversky and Kahneman (1974, 1981), Kahneman and Tversky (1982), Griffin and Tversky (1992), Grether (1992) Holt and Smith (2009) and the surveys in Camerer (1995) and Rabin (1998).

representation. According to this representation, the agent has a utility function u over consequences, a prior over priors ρ , and a threshold ϵ between 0 and 1. She then acts as follows.

- In the first period, before any new information arrives, our agent chooses the prior π that maximizes her prior over priors ρ , and then forms her preferences as a standard expected utility maximizer.

That is, not knowing which prior she should use, the agent picks the one that she considers the most likely according to her belief over beliefs. In our example of the investor, we can see ρ as a belief over the possible economic models she can use (each of which entails a certain belief over the state of the world): our agent would then use the model, hence the belief, that she considers the most likely – in a *maximum likelihood* fashion.

As new information i is revealed our agent *tests her prior* to verify whether she was using the correct one – and this is where the threshold ϵ comes into play.

- If the probability that her prior assigned to some new information i is *above* the threshold ϵ , i.e. $\pi(i) > \epsilon$, then the model is *not* rejected and she simply updates her prior π using Bayes’ rule.

That is, if new information is not “unexpected,” our agent acts like a standard one.

- If, however, the probability that her prior assigned to some information i is *below* the threshold, i.e. $\pi(i) \leq \epsilon$, then the model is *rejected* and our agent: goes back to her prior over priors ρ ; updates it using the additional information that she has received; then chooses the prior π' that maximizes her *updated* prior over priors; using this prior π' she forms her preferences maximizing the expected utility.

That is, if the model is rejected by the data, then our agent reconsiders which priors to use by picking the new maximum likelihood model, which is found by looking at the *updated* prior over priors.

It is easy to see that when $\epsilon = 0$ our agent behaves just like a standard Bayesian agent whenever Bayes’ rule applies. But her behavior is here prescribed also when Bayes’ rule does not apply (zero probability events): in that case our agent goes back to her prior over priors and chooses a new prior to use. By contrast, when $\epsilon > 0$ our agent’s behavior departs from the standard one. In accordance to the experimental evidence, when an event with a small but positive probability occurs, our agent might not simply update her prior, but rather reconsider whether she was using the right one in the first place. She might then modify her beliefs more than what prescribed by Bayes’ rule.

One of the features of the Hypothesis Testing model is that the agent chooses to review her prior or not only on the basis of the likelihood it assigns to the new information. Indeed one could think of different behaviors, in which other considerations besides the likelihood are taken into account, but they would not fit the interpretation suggested above:² in this

²For example, consider the state space $[0, 2]$, and an agent who is not “surprised” if she is told that the true state lies in any subset of $[0, 1]$, irrespectively of the likelihood, while she is “surprised” is she is told that the true state lies in $(1, 2]$. Since it could be the case that an event subset of $[0, 1]$ has a smaller likelihood than $(1, 2]$, this behavior might not fit the interpretation of the model suggested above. It turns out, however, that this behavior could also be represented using our model, since if the agent reconsiders her prior, it does not mean that she will change it. See Section 3.1 for more.

sense we are introducing a model in which the element of “shock” comes only from the likelihood.

The contribution of the paper is to characterize axiomatically the Hypothesis Testing model. In particular, we show that requiring that an agent’s behavior could be described using an Hypothesis testing model is equivalent to requiring that it satisfies the same axioms that characterize Bayesian Updating, but in which Dynamic Consistency is replaced with a novel axiom, Dynamic Coherence. The basic idea of this axiom is to impose that the agent has a non-circular reaction to new information. As opposed to Dynamic Consistency, it is based on a comparison of the reactions across possible informations, without going through the original preference. For this reason, it could apply also to those events to which the original preference assigned probability zero, where Dynamic Consistency has no bite. At the same time, on non-null events it is weaker than Dynamic Consistency.

We also extend our analysis to the case in which the agent is not a standard expected utility maximizer, but rather is ambiguity averse. We show that if we posit that both before and after the arrival of information the agent’s preferences satisfy the axioms in Gilboa and Schmeidler (1989), together with the same two axioms that we used before, Dynamic Coherence and Consequentialism, then we obtain a representation with the same intuition, in which the agent has a prior over *sets* of priors.

1.3 *Outline and Related Literature*

In the remainder of the introduction we discuss the related literature. Section 2 presents the formal setup and the axiomatic foundations. Section 3 introduces and discusses the main representation. Section 4 introduces the extensions of the model to the case of ambiguity aversion. Section 5 concludes. In the main text we carry out our analysis assuming that the state space is finite. In Appendix A we study the more general case. The proofs appear in the Appendix B.

Many generalizations of Bayes’ rule have been proposed in the literature. Among the many behavioral models, see for example the Jeffrey’s rule and, more recently, Mullainathan (2002), Rabin (2002), Mullainathan et al. (2008), Gennaioli and Shleifer (2009).³ A common feature of these contributions is that they propose significant generalizations of Bayes’ law; by contrast, we wish to focus only on over-reactions to unexpected news, or to null events, while maintaining the elements of appeal of the standard model for other non-unexpected news. (Moreover, none of these papers have an axiomatic foundation.)

Within the small axiomatic decision theory literature, Epstein (2006) and Epstein et al. (2008) model agents who might be tempted to use a posterior that is different from the Bayesian Updating of their prior: for example, they might be tempted to overweigh or underweight new evidence. (Consequences of this behavior are discussed in Epstein et al. (2010).) Both models differ from ours in several aspects. First, they are not interested in modeling the agent’s reaction to zero-probability events, one of our main goals. Second, they study a setup of preferences over menus, and a central point of their analysis is that the agent is aware of the possible biases she can be subject to in the future. By contrast, we look at a standard dynamic setup, and such awareness is inessential in our case. Furthermore,

³See also Brav and Heaton (2002) and Brandt et al. (2004) for additional references to the behavioral literature.

they present models that allow for significant departures from the standard Bayesian one, while we wish to confine our departure to addressing the two issues mentioned above.⁴

Both the *lexicographic beliefs* of Blume et al. (1991), and the *conditional probability systems* of Myerson (1986a,b) address the issue of beliefs for null events, with important results on the agent’s reaction to them. Neither of these theories, however, seem to extend to non-bayesian reactions to unlikely but non-null events, one of the goals of this paper; furthermore, they are based on preferences that depart from standard expected utility maximization also for the static case – for example in Blume et al. (1991) preferences are not fully Archimedean. By contrast, we wish to study an agent that is completely standard in the static case, but might have a non-standard reaction to news.

In the game theory literature Foster and Young (2003) present a model in which agents “test” whether their current belief about the opponent’s behavior is correct, and perform a paradigm change if this hypothesis is rejected – just like in the Hypothesis Testing model. Their analysis, however, is focused on learning in games, and does not apply to the standard framework. Also, it contains no prescription about how a new belief should be chosen if the current one is rejected, while this is an essential component of our model. Finally, as opposed to what happens in our representation, their agents do not act as standard Bayesian agents if an hypothesis is not rejected. In this sense, their model has a similar spirit in terms of when a paradigm change should take place in a game, but is very different in all other aspects.

Finally, in the literature on ambiguity a large attention is devoted to the issue of updating ambiguous beliefs, and since the Bayesian postulates seems problematic in that framework, many generalizations of them have been proposed.⁵ These generalizations, however, are aimed at reconciling the standard approach with the presence of ambiguity aversion, and most of them reduce to Bayes’ rule in the case in which the agent is an expected utility maximizer (ambiguity neutrality). By contrast, our main goal is to study violations of Bayes’ rule even in this standard case, and even when we allow for ambiguity aversion we obtain a model for updating ambiguous beliefs which does not reduce to Bayes’ rule in the case of ambiguity neutrality.

2. Setup and foundations

2.1 Formal Setup

We adopt a standard dynamic setup. 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.⁶ We also 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).)

⁴In addition, the recent work of Kochov (2009) studies a non-Bayesian reactions in the presence of unforeseen contingencies – agents who fail to properly account for events that will take place in the non-immediate future, a behavior very different from the one we are interested in.

⁵See, among others, Gilboa and Schmeidler (1993), Epstein and Le Breton (1993), Epstein and Schneider (2003), Maccheroni et al. (2006), Siniscalchi (2006), Hanany and Klibanoff (2006), Ghirardato et al. (2008) and Epstein and Schneider (2007). The latter discusses a possible form of hypothesis Testing for updating multiple priors.

⁶In Appendix A we generalize our analysis to the case in which Ω is infinite.

Denote by $\Delta(\Omega)$ the set of all Borel probability measures (priors) on $\Delta(\Omega)$. Correspondingly, denote by $\Delta(\Delta(\Omega))$ the set of all probability measures (priors over priors) over $\Delta(\Omega)$. Also, denote by \mathcal{F} the set of all acts, that is, the set of all finite-valued Σ -measurable functions $f : \Omega \rightarrow X$. Acts are the standard objects on which the preferences are defined: for every state of the world, an act returns a consequence in X .

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 $\omega \in \Omega \setminus A$.

We define null events in the standard way: we say that A is null if the agent is indifferent between any two acts that coincide outside A , independently of what they return if A were to occur. When the agent is a standard expected utility maximizer, as it will be the case here, these are the events to which she assigns probability zero.

Definition 1. For any preference relation \succeq on \mathcal{F} , we say that $B \in \Sigma$ is \succeq -**null** if $fBg \sim g$ for any $f, g \in \mathcal{F}$.

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.

2.2 Axiomatic Foundations

We start by imposing a few standard axioms.

A.1 (Well-Behaved Standard Preferences (WbP)). For any $A \in \Sigma$, $f, g, h \in \mathcal{F}$:

1. (Continuity): the sets $\{\alpha \in [0, 1] : \alpha f + (1 - \alpha)g \succeq h\}$ and $\{\alpha \in [0, 1] : h \succeq \alpha f + (1 - \alpha)g\}$ are closed;
2. (Independence): for any $\alpha \in (0, 1)$

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

3. (Monotonicity): if $f(\omega) \succeq_A g(\omega)$ for all $\omega \in \Omega$, then $f \succeq_A g$.
4. (Constant Preference Invariance): for any $x, y \in X$, $x \succeq_A y \Leftrightarrow x \succeq y$

A.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 $f \sim_A g$.

Axiom WbP (Axiom 1) is a collection standard postulates that guarantee that both before and after the arrival of information our agent acts like a standard expected utility maximizer (part (1), (2) and (3)), and that the arrival of information does not affect the agent's ranking of the consequences in X (part (4)). Consequentialism (Axiom 2) is another standard postulate that guarantees 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 . Both axioms are entirely standard.

We now turn to restrict the way beliefs evolve with information. To this end, the standard assumption made in the literature is Dynamic Consistency.

A.3 (Dynamic Consistency (DC)). *For any $A \in \Sigma$, A not \succeq -null, and for any $f, g \in \mathcal{F}$, we have*

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

The basic idea of Dynamic Consistency is that the arrival of some information $A \in \Sigma$ should not modify the ranking of two acts that coincide outside of A . It is well known that adding Dynamic consistency to WbP and Consequentialism implies that the agent follows Bayes' rule. (We refer to Ghirardato (2002) for an in-depth discussion of this standard postulate and its implications.) While appealing, however, Dynamic Consistency has two important limitations. First of all, it disciplines agent's preferences only if the event A is not \succeq -null, that is only if A is assigned a positive probability to begin with. This implies that any theory that derives from Dynamic Consistency is bound to have no predicting power on the agent's reaction to events to which she assigns probability zero. Second, as we mentioned in the introduction, a sizable experimental evidence documents systematic violations of Bayes' rule in general, and of Dynamic Consistency in particular. This seems to be especially true when the revealed information was assigned a small probability. In other words, despite its normative appeal, on the one hand Dynamic Consistency seems to be too strong to be accepted as a positive axiom, while on the other hand it seems not strong enough since it doesn't restrict the agent's behavior in the case of null events.

To reconcile with the empirical evidence we replace Dynamic Consistency with an axiom that allows the agent to behave in a dynamically-inconsistent manner, but that guarantees that she does so in a regulated way. At the same time, we posit this regularity to apply also to the agent's reaction to null events, so that we can develop a theory that applies to that case as well. The basic idea is that want to rule out the possibility that our agent has a *circular* reaction to information.

We'll present the idea behind this axiom in two steps. First, consider two events A_1 and A_2 , and say that if the agent is told that A_1 occurred, then she is *sure* that also A_2 occurred; but if she is told that A_2 occurred, she is sure that A_1 occurred. That is, for each event she is sure of the other one as well. Then, we would like to say that the *informational content* of the two events is the same – namely, it is $A_1 \cap A_2$. But then, the agent should have the same beliefs after either A_1 or A_2 , since they have the same informational content: we should then have $\succeq_{A_1} = \succeq_{A_2}$. Notice that when A_1, A_2 are non-null, this is clearly implied by Bayes' rule. But this postulate could be meaningful also when A_1 or A_2 are null: how the agent considered these events before the arrival of information should not matter here.

We will now strengthen the idea above to rule out cycles in beliefs. Consider three possible events $A_1, A_2, A_3 \in \Sigma$ such that: after being told that the true state lies in A_1 , then the agent is *sure* that it also lies in A_2 ; if instead she is told that it lies either in A_2 , then she is sure that it lies also in A_3 ; but if she is told that it lies in A_3 , then she is *sure* that it lies in the *first* event A_1 . That is, being told that the true state is in one event, she thinks it is also in the next one; but if she's told that it is in the last one, then she is sure it is also

the first. This would be naturally true if the agent had the *same* preferences after each of these events, i.e. if $\succeq_{A_1} = \succeq_{A_2} = \succeq_{A_3}$: if this were the case, then after each of these events she would be sure that the true state lies in $A_1 \cap A_2 \cap A_3$, giving us the beliefs described above. (In fact, $A_1 \cap A_2 \cap A_3$ would be the informational content of each of these events.) But, what if this is not the case, i.e. $\succeq_{A_1} \neq \succeq_{A_2} \neq \succeq_{A_3}$? Then we would be in a situation in which the agent’s beliefs react to information in a *circular* manner: beliefs keep changing, but they form a loop, since at the end the agent is sure that the true state lies in the *first* event. The idea of the axiom is to rule out this possibility: therefore it will posit that in such situations preferences do not change. (The axiom will be written for cycles of any length and in terms of the primitives: to say that the agent is sure that A_2 has happen after she is told A_1 , we write that $(\Omega \setminus A_2)$ is \succeq_{A_1} -null.)

A.4 (Dynamic Coherence). *For any $A_1, \dots, A_n \in \Sigma$, if $(\Omega \setminus A_{i+1})$ is \succeq_{A_i} -null for $i = 1, \dots, (n - 1)$, and $(\Omega \setminus A_1)$ is \succeq_{A_n} -null, then $\succeq_{A_1} = \succeq_{A_n}$.*

It is not hard to see that any Bayesian agent would satisfy the axiom above as long as all A_i s are non-null. But of course Bayes’ rule is stronger than simply postulating the lack of this circularity – it entails a much stronger form of consistency. At the same time, Dynamic Coherence applies also when some of the A_i s are null. This is possible because of a conceptual difference between this axiom and Dynamic Consistency. The latter regulates the agents reaction to information by comparing the preferences after the information with the original ones: but this implies that it cannot restrict the reaction to null events, since the original preferences are “flat” in that case (by definition of null events). By contrast, Dynamic Coherence imposes a form of consistency of beliefs *across* events, without going through the original belief. This means that it can be imposed for *all* information, regardless of the likelihood assigned by the original belief.

Correspondingly, Dynamic Coherence is neither stronger nor weaker than Dynamic Consistency: while it does allow violations of Dynamic Consistency, albeit regulating them, it also disciplines the reaction to null events, on which Dynamic Consistency has no bite. (Since the two axioms are not nested, then the representation Theorem in Section 3.2 will analyze also the case in which both hold true at the same time.)

3. The Hypothesis Testing Model

3.1 The model

We are now ready to introduce our main representation. For simplicity, let us define a notation for Bayesian Updating. For any $\pi \in \Delta(\Omega)$ and $A \in \Sigma$ such that $\pi(A) \geq 0$, define $\text{BU}(\pi, A) \in \Delta(\Omega)$ (bayesian update of π using A) as

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

for all $B \in \Sigma$. Let us also define the Bayesian updating of a prior over priors. Abusing

notation, for any $\rho \in \Delta(\Delta(\Omega))$ and $A \in \Sigma$ such that $\pi(A) > 0$ for some $\pi \in \text{supp}(\rho)$, define

$$\text{BU}(\rho, A)(\pi) := \frac{\pi(A)\rho(\pi)}{\int_{\Delta(\Omega)} \pi(A)\rho(d\pi)} \quad (2)$$

for all $\pi \in \Delta(\Omega)$. Both definitions are standard.

Definition 2. A class of preferences relations $\{\succeq_A\}_{A \in \Sigma}$ admits an **Hypothesis Testing Representation** if there exists a non-constant affine function $u : X \rightarrow \mathbb{R}$, a prior over priors $\rho \in \Delta(\Delta(\Omega))$ with finite support, and $\epsilon \in [0, 1)$ such that for any $A \in \Sigma$ there exist $\pi_A \in \Delta(\Omega)$ such that:

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

$$f \succeq_A g \Leftrightarrow \sum_{\omega \in \Omega} \pi_A(\omega)u(f(\omega)) \geq \sum_{\omega \in \Omega} \pi_A(\omega)u(g(\omega));$$

2. $\{\pi_\Omega\} = \arg \max_{\pi \in \Delta(\Omega)} \rho(\pi)$;

- 3.

$$\pi_A = \begin{cases} \text{BU}(\pi_\Omega, A) & \text{if } \pi_\Omega(A) > \epsilon \\ \text{BU}(\pi_A^*, A) & \text{otherwise} \end{cases}$$

where $\{\pi_A^*\} = \arg \max_{\pi \in \Delta(\Omega)} \text{BU}(\rho, A)(\pi)$;

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

In an Hypothesis testing representation the agent has a utility function, a prior over priors ρ , and a threshold ϵ . In every period and after every information she maximizes her expected utility using the utility u and a belief π_A (part 1), which is determined as follows. In the first period, before any new information is revealed, she chooses the prior π_Ω that maximizes the prior over priors ρ (part 2). She behaves as if she were choosing which “theory” to use to forecast the states of the world: given a certain belief ρ over the possible theories, she picks the most likely one – in a maximum likelihood fashion.

As new information A is revealed, two things can happen (part 3). If the prior she was using assigned to this event a probability *above* the threshold ($\pi_\Omega(A) > \epsilon$), then the prior “passes the test” and our agent “keeps” it. She then simply follows Bayes’ rule. This means that if the information is not unexpected, then we are in the “business as usual” situation and our agent behaves like a standard Bayesian one. If, however, she is given an information that she did not expect, that is, if the likelihood that her prior assigned to that information is *below* the threshold ($\pi_\Omega(A) \leq \epsilon$), then the prior “fails the test” and our agent revises it. It is as if she thought: “If the prior I am using did not forecast what happened, then, maybe, it is the wrong prior!” From this point of view, our agent acts as if she were “testing” her prior and “rejecting” the hypothesis that it is correct if it falls below the threshold, as if it were a confidence level – hence the name of the model. And how does she choose a new prior? She updates her prior over priors using Bayes’ rule, and then she chooses the prior

that maximizes the *updated* prior over priors. That is, if the “theory” she were using before performed poorly, she updates her beliefs over theories using the new information, and picks the one judged the most likely by this updated belief. (Notice, however, that this might be the same one she was using before: this happens if, after reconsidering, she still believes that the one she was originally using is the best in the lot.)

The decision rule that we have just described in words, however, might leave some room to indeterminacy: what would happen if there is more than one prior that maximizes the updated prior over priors? Which one is chosen? To avoid this indeterminacy, in an Hypothesis Testing model ρ is constructed in such a way that the argmax of the updated prior over priors is always *unique*.⁷

Just like Dynamic Coherence is neither stronger nor weaker than Dynamic Consistency, an Hypothesis Testing representation is neither more general nor more restrictive than the standard Bayesian model. To see why, notice that if $\epsilon = 0$, then our agent behaves exactly like a standard Bayesian agent whenever Bayes’ rule applies; but her behavior is disciplined also when Bayes’ rule does not apply (null events), thus extending the predictive power of the theory to all possible events and making this model a special case of the Bayesian one. On the other hand, when $\epsilon > 0$ the Hypothesis Testing representation allows non-bayesian reactions to non-null events. In this sense, the Hypothesis Testing representation generalizes the Bayesian approach to allow for that over-reaction to unexpected news that has been documented empirically.⁸

As we mentioned in the introduction, in the model described above the agent decides whether she should reconsider her prior or not *only* on the basis of the likelihood assigned to the information. As we argued, there could be cases in which the decision on whether to be “surprised” is affected not only by the likelihood of the new information, but also by other considerations: for example, there could be events which, although unlikely, are not surprising for the agent, while there might be other events that are more likely but also more “surprising.” On the one hand, our representation has an interpretation which would not fit these cases. On the other hand, it turns out that our model might still be able to represent them. The reason is, even when our agent reconsiders her prior, she might very well end up choosing her old one, making her behave as if she did not reconsider her prior at all.⁹

⁷We should emphasize that this condition is more than a technical detail: rather, it is an essential condition for an Hypothesis Testing model to have *any* 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 priors. Then, notice that after every information A , the prior that the agent uses after A must belong to the argmax of the updated ρ , since it must give probability one to A : this means that we can represent any behavior with a model similar to an Hypothesis Testing representation without the requirement that the argmax is unique (and setting $\epsilon = 1$).

⁸More precisely, if $\epsilon = 0$ then the Hypothesis Testing model is a special case of the Bayesian one. The converse is true if there are no null events. This means that if both conditions hold then the two models coincide.

⁹A indirect proof of this claim comes from Theorem 1 below. In fact, it is not hard to see that such agents would still satisfy our axioms (in particular, they would still satisfy Dynamic Coherence). Therefore, we could always find an Hypothesis Testing representation of this behavior. (Being able to identify exactly which behavior can be represented using our model is of course one advantage of having an axiomatic foundation.)

3.2 Representation Theorem

We are now ready to introduce the link between the axioms and the Hypothesis Testing model: the representation theorem. Since there could be multiple values of ϵ that represent the same preferences (see below), we focus on the representations with the smallest possible values of ϵ .

Definition 3. An Hypothesis Testing Representation (u, ρ, ϵ) is **minimal** if there is no $\epsilon' \in [0, 1)$ such that $\epsilon' < \epsilon$ and (u, ρ, ϵ') is an Hypothesis Testing Representation of the same preferences.

Theorem 1. A class of preference relations $\{\succeq_A\}_{A \in \Sigma}$ satisfies *WbP*, *Consequentialism*, and *Dynamic Coherence* if and only if it admits a minimal Hypothesis Testing Representation (u, ρ, ϵ) .

Moreover, $\epsilon = 0$ if and only if $\{\succeq_A\}_{A \in \Sigma}$ satisfies also *Dynamic Consistency*.

Theorem 1 shows that Dynamic Coherence, together with WbP and Consequentialism, is equivalent to the existence of a (minimal) Hypothesis Testing Representation, thus providing a precise characterization of the empirical content of this model. Moreover, it shows that if we further add Dynamic Consistency, these axioms are equivalent to the existence of an Hypothesis Testing model where $\epsilon = 0$: this characterizes the case in which the agent follows Bayes' rule whenever it applies, but reconsider which prior to use whenever she faces an information to which she assigned probability zero.

We now turn to discuss the uniqueness properties of an Hypothesis Testing Representation. It is standard practice to show that the utility function is unique up to a positive affine transformation. The threshold ϵ is unique, but only thanks to our focus on minimal representations: in general there might be a continuum of values of ϵ that would work.¹⁰ As for the prior over priors ρ , it turns out that it is not unique, and that even its support is not unique. There are essentially three reasons why this is the case. (To avoid confusion we address the elements of the support of ρ as “models.”) First, we can always add to the support of ρ one additional model with a likelihood so low that it will never be used. That is, we can always add “redundant” models leaving the behavior unaffected. Second, even if we removed these redundant ones, we are bound to identify models only after the events that trigger the agent to use them, and therefore we have no control over what these models prescribe outside of these events.¹¹ For the same reason, moreover, there might be multiple ways to combine these models, making not unique also the cardinality of the support of ρ .¹²

¹⁰To see why, call a the likelihood of the *least* likely event that does not trigger a violation of Bayes' rule, and call b the likelihood of the *most* likely event that does trigger a violation of Bayes' rule. Then, we must have $b < a$, and any $\epsilon \in [b, a)$ would work. The reason is, our space of events might not be “dense” in this sense, and therefore the threshold ϵ is not uniquely identified.

¹¹Consider a model π used after the arrival of some information $A \in \Sigma$. Since all we observe is $\text{BU}(\pi, A)$, then we can replace π with any π' such that $\text{BU}(\pi, A) = \text{BU}(\pi', A)$ and $\pi(A) = \pi'(A)$.

¹²Consider four events $A, B, C, D \in \Sigma$, all with a likelihood below ϵ , such that $A \supset B$, $C \supset D$ and $A \cap C = \emptyset$. Denote by π_i the priors that are used after event i . Assume that we have $\text{BU}(\pi_A, B) \neq \pi_B$ and $\text{BU}(\pi_C, D) \neq \pi_D$. There are at least three ways to represent this case. First, we can create four distinct models, one for each event, giving to ρ the appropriate weights. Second, we can have only two models, π_1 and π_2 , such that $\text{BU}(\pi_1, A) = \pi_A$, $\text{BU}(\pi_1, C) = \pi_C$, $\text{BU}(\pi_2, B) = \pi_B$, $\text{BU}(\pi_2, D) = \pi_D$. (Model π_1 is used

3.3 Discussion: the Hypothesis Testing model as a form of bounded rationality

One characteristic of the Hypothesis Testing model is that agents forms beliefs in a non-standard way even in the first period, before any information is revealed. In fact, we can think of our agent as if she had in mind a set of “conceivable” priors (the support of ρ), which she can rank in terms of plausibility (forming ρ), of which, however, she uses only the most likely one – a’ la maximum likelihood.¹³ The other priors in the support of ρ , albeit “conceivable,” are in fact not used to make decisions unless some unexpected news is revealed.¹⁴ This is in contrast to the behavior prescribed by the standard Bayesian approach, according to which if the agent has a prior over priors, she should choose as a belief its *expectation*, not its maximizer. That is, she should consider all conceivable priors and weight them appropriately, instead of using only the most likely one. While these two approaches are behaviorally indistinguishable in a static case, since all we can see is the belief that is used and not where it comes from, it is the analysis of the dynamic case that allow us to set them apart. In particular, it is precisely by observing violations of Dynamic Consistency that we can infer how has the agent formed her beliefs in the first place.

While seemingly irrational, we consider the use of one model, or a subset of the possible models, as rather realistic. To see why, let us go back to our original example of the investor. When she chooses a portfolio, our investor could indeed considers *all* possible economic and statistical models that she can come up with, assign to them a relative likelihood, estimate them all, look at their predictions, and finally come up with a belief which is the weighted average of the predictions of all of these models. This is a behavior prescribed by the standard approach. Alternatively, our agent could choose just a subset of these models, the ones that she consider the most likely, estimate only these, and then use as a belief the weighted average of the predictions of only this selected subset of models.¹⁵ She would then consider alternative models only if some unexpected news is revealed. (We consider the latter behavior more realistic.)

This leaves us with the question of why should such behavior be adopted, since it could lead to inaccurate predictions. There are at least four reasons why this could be the case. First, it could be a rational reaction to a form of bounded rationality/costly thinking/cost of considering models. In fact, if estimating a model is costly, as in most cases it is, then our

after A and C , while model π_2 is used after B and D .) Third, we could use two other models, π_3 and π_4 , such that $\text{BU}(\pi_3, A) = \pi_A$, $\text{BU}(\pi_3, D) = \pi_D$, $\text{BU}(\pi_4, B) = \pi_B$, $\text{BU}(\pi_4, C) = \pi_C$. (Model π_3 is used after A and D , while model π_4 is used after B and C .) Unfortunately, this example also shows that we cannot find a unique ρ with “minimal” support.

¹³Of course this chosen prior might itself be the combination of multiple priors with given weight. For example, the many priors in the support of ρ might originate from different weights put on a fixed set of beliefs.

¹⁴This suggests that our agent might have two layers of belief, what is conceivable but not used, and what is actually used. The idea of multiple layers of beliefs is explored also in Blume et al. (1991), who consider the *lexicographic probabilities*. As we mentioned before, however, their model studies the reactions to null-events by allowing non-standard (non-archimedean) preferences in the static case. By contrast, our model has standard preferences on the static case but non-standard dynamics (and moreover allows for non-standard reactions to events that are not null but simply low probability).

¹⁵Once again, notice that in an Hypothesis Testing representation the chosen prior could come from averaging a *set* of models with relative weights – this will be the chosen prior.

agent has an incentive to be parsimonious in the number of models she considers.¹⁶ Such cost could also be psychological: agents might not like to be reminded that they have a limited knowledge of the reality, and that they do not even know which model to use, and might therefore prefer to focus on a single one as if it were true – a tendency highly emphasized in the psychology literature. Second, it could be seen simply as another instance of the standard behavioral bias that leads agents to disregard low probability events.¹⁷ Third, this behavior could stem for a *preference for simplicity*: considering only a few models is simpler, and the agent might prefer it in an Occam’s Razor sense. Finally, considering such simpler theories might actually be *optimal* even if there were no cost of considering models or other additional costs. In fact, Gilboa and Samuelson (2008) show (in a different setting) that considering complex theories might induce people to *overfit* the data and engage in ineffective learning, generating worse predictions. By contrast, simpler theories do not have this problems, and might lead to optimal behavior. We refer to their work for an detailed analysis.

4. The Hypothesis-Testing model with ambiguity aversion

Our analysis thus far was carried out under the assumption that in every period, and after every information, the decision maker behaves like a standard expected utility maximizer with a well-formed (and unique) prior over the states of the world. This is an immediate consequence of axiom WbP (Axiom 1) which includes independence. Since the Ellsberg paradox, however, it is well known that independence is often violated, and agents do not behave like standard expected utility maximizers with a single prior over the states of the world. Rather, they are shown to be ambiguity averse: instead of using expected utility, they dislike betting on outcomes that depend on the realization of unknown states of the world, and have a preference for hedging. This seems all the more problematic in our analysis since the instances in which we should expect a violation of the Bayesian model are also the ones in which we should expect ambiguity aversion. The goal of this section is to show that we can find an appropriate extension of our result also for the case in which the agent is ambiguity averse, by means of the same axiom, Dynamic Coherence.

As we mentioned in the introduction, the issue of updating beliefs under ambiguity has been studied by an extensive literature. Almost all of these works, however, have focused on the complicated interaction between ambiguity aversion and Bayesian updating, without questioning the updating procedure in the first place. In particular, most of these model reduce to the case of Bayes’ rule if the agent is ambiguity neutral (expected utility maximizer). By contrast, in Section 3 we have suggested an alternative to Bayesian updating in the case of ambiguity neutrality, and we now wish to do the same in the case of ambiguity aversion. For simplicity, we carry out this analysis in the same setup as Theorem 1, that is when the state space Ω is finite.

¹⁶In fact, it is not hard to see that we could have re-written our model as one in which the agent chooses the optimal number of models given a cost of considering each of them.

¹⁷That is, the models to which ρ assigns a small probability could be disregarded just like low probability events are often disregarded in choices among gambles.

4.1 Foundations with Ambiguity Aversion

In what follows we model ambiguity aversion using the well-known model of Gilboa and Schmeidler (1989). In particular, we replace standard Independence with their's C-independence and Ambiguity Aversion axioms, and obtain WbP-AA (Axiom 5) to replace WbP (Axiom 1). We refer to Gilboa and Schmeidler (1989) for a more detailed discussion.

A.5 (Well-Behaved Standard Preferences with Ambiguity Aversion (Wbp-AA)).

For any $A \in \Sigma$, $f, g, h \in \mathcal{F}$:

1. (Continuity): the sets $\{\alpha \in [0, 1] : \alpha f + (1 - \alpha)g \succeq h\}$ and $\{\alpha \in [0, 1] : h \succeq \alpha f + (1 - \alpha)g\}$ 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 $B \in \Sigma$, $p, q \in \Delta(X)$, $p \succeq_A q \Leftrightarrow p \succeq_B q$

It is well-known that WbP-AA (Axiom 5) guarantees that both before and after receiving every information the agent has a closed and convex set of prior beliefs over the states of the world (instead of a unique prior), and judges every act by the expected utility computed with the most pessimistic prior for that act. (Again, we refer to Gilboa and Schmeidler (1989) for a discussion on the properties of this representation.)

While we replace WbP (Axiom 1) with WbP-AA above (Axiom 5), we posit that Consequentialism and Dynamic Coherence hold as before – with the same intuition. This is in contrast with what happens with Dynamic Consistency, since 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, consider an agents who updates her set of priors by updating each prior using Bayes' rule: Epstein and Schneider (2003) show that such agent might violate Dynamic Consistency.¹⁸ In fact, Ghirardato et al. (2008) show that this model is equivalent to the agent satisfying a weakening of Dynamic Consistency, in which this axiom applies only to a *subset* of the original preference relation, which they call the “unambiguously preferred” relation. In particular, following the analysis in Ghirardato et al. (2004) and Ghirardato et al. (2008), for any $A \in \Sigma$ define the preference relation \succeq_A^* as follows:

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

for all $\lambda \in [0, 1]$ and all $h \in \mathcal{F}$. (It is not hard to see that \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.) Correspondingly, we can also define the set of unambiguously non-null events.

¹⁸In 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*.

Definition 4. 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.) The idea of the axiom is to impose a dynamically consistent behavior on this restricted preference and on unambiguously non-null events.

A.6 (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 (Axiom 1), then Restricted Dynamic Consistency (Axiom 6) is clearly equivalent to standard Dynamic Consistency (Axiom 3).¹⁹

4.2 The representation with Ambiguity Aversion

In the case of ambiguity aversion modeled a' la Gilboa and Schmeidler (1989) the agent has not one, but a *set* of priors. As she receives new information this set of priors will be modified: a natural extension of the standard Bayesian approach to this case is the model in which the agent reacts to this new information by updating using Bayes' rule each of the priors in her set of priors. Just like in the first part of the paper we started from the Bayesian model and defined the Hypothesis Testing one, in our analysis here we start from this generalization to extend the Hypothesis Testing model to the case of ambiguity aversion. The idea of this generalization is that our agent has a prior over *sets* of priors, and picks the set of priors that maximizes it. Then, she tests this set of prior using a threshold: if it passes the test, she updates every prior in the set using Bayes' rule. Otherwise, she updates her prior over sets of priors, and chooses the new set that maximizes it.

For simplicity, we introduce a notation for the extension of the notion of Bayesian Updating to updating sets of priors and a prior over sets of priors. By \mathcal{C} denote the set of closed and convex subsets of $\Delta(\Omega)$. 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

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

Moreover, for any $\rho \in \Delta(\mathcal{C})$ and $A \in \Sigma$ such that for some $\Pi \in \text{supp}(\rho)$ we have $\pi(A) > 0$ for all $\pi \in \Pi$, define $\overline{BU}(\rho, A)$ as

$$\overline{BU}(\rho, A)(\Pi) := \frac{\min_{\pi \in \Pi} \pi(A) \rho(\Pi)}{\int_{\mathcal{C}} \min_{\pi \in \Pi} \pi(A) \rho(d\Pi)}.$$

¹⁹Moreover, it turns out that if there are no null events, then under WbP and Consequentialism we have that Dynamic Coherence is weaker than Restricted Dynamic Consistency. (This is an immediate consequence of Theorem 2 below.)

Definition 5. A class of preferences relations $\{\succeq_A\}_{A \in \Sigma}$ admits an **Hypothesis Testing Representation with Ambiguity Aversion** if there exists a continuous function $u : X \rightarrow \mathbb{R}$, 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.

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

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

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

Definition 6. An Hypothesis Testing Representation with Ambiguity Aversion (u, ρ, ϵ) 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.

Finally, the representation theorem.

Theorem 2. *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 2 shows that the same postulates that we have imposed for the standard expected utility case, Dynamic Coherence and Consequentialism, gives us the desired representation also for the case in which agents are ambiguity averse a' la Gilboa and Schmeidler (1989). Moreover, when Restricted Dynamic Consistency is satisfied as well, 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.

5. Conclusion

In this paper we have developed axiomatically an alternative to the standard Bayesian model. We study an agent who behaves like a standard Bayesian one when she receives an information that is not “unexpected,” i.e. to which she assigned a probability above a threshold. If this is not the case, however, instead of following Bayes' rule our agent reconsiders which prior to use by updating a prior over priors and selecting the most likely one after the update.

We have also discussed an extensions of the model to the case in which the preferences are ambiguity averse.

Appendix

A. Extension: Infinite State Space

Theorem 1 shows the existence of an Hypothesis Testing representation under the assumption that the state space Ω is finite. We now turn to study the more general case in which Ω could be infinite, and show that the same axioms that we used before are equivalent to the existence of a very similar representation. There are, however, a few differences. The main one is the following. If Ω is finite, in an Hypothesis Testing model when the agent receives an information to which she assigns a probability below the threshold ϵ , she chooses the prior that maximizes the updated prior over priors ρ . As we have discussed in Section 3.1, her behavior is well defined since ρ is constructed in such a way to guarantee that this maximizer is always unique. This might be no longer possible when Ω is infinite, and therefore our agent will have to be endowed with a decision rule that allows her to choose between multiple maximizers. Specifically, when Ω is infinite our agent will also have a linear order \triangleright over the priors such that if there are two priors that maximize the updated ρ , she chooses between them using \triangleright . Moreover, this linear order \triangleright and the likelihood over priors ρ will be coherent: for any $\pi, \pi' \in \Delta(\Omega)$, if $\pi \triangleright \pi'$ then $\rho(\pi) \geq \rho(\pi')$.

We also need to make a few additional changes of a more technical nature. First of all, we obtain a representation with probability charges instead of probability measures, as it usually happens with an infinite state space. (A probability charge is a probability measure that only needs to be finitely additive but not necessarily countably additive.²⁰ For any state space Ω and σ -algebra Σ , denote by $\Delta^F(\Sigma)$ the set of probability charges on Σ .) Second, let us go back to (2), the updating rule for the prior over priors. Notice that when Ω is infinite, then also $\Delta^F(\Sigma)$ is infinite, which means that for any non-simple $\rho \in \Delta(\Delta^F(\Sigma))$ we will have $\rho(\pi) = 0$ for all π : but this makes (2) meaningless, since there is no point in “updating zeros.” In fact, for non-simple priors it is customary in Bayesian statistics to express updating rules using the *density* of the probability measure. We will therefore modify (2) along these lines. There are, however, two things we need to notice. First, it is not obvious what do we mean by a density of a prior over priors, since there is no obvious reference measure on $\Delta^F(\Sigma)$ that we should use to define it.²¹ Second, we might be in a situation in which there is no non-zero density that can express the relative likelihood in

²⁰Whether countable additivity is a desirable property in the context of subjective probabilities is an issue that dates back to de Finetti and Savage. (See de Finetti (1931, 1970), Savage (1954).) Since countably additivity could be considered too restrictive, because for example it rules out the possibility that the agent has a uniform prior over the natural numbers, here we focus only on probability charges. At the same time, if we wish to obtain probability measures instead of charges, we only need to add the requirement that every preference in the class $\{\succeq_A\}_{A \in \Sigma}$ satisfies Arrow’s Monotone Continuity. (See Chateauneuf et al. (2005) for more.)

²¹That is, while the density of some $\pi \in \Delta(\mathbb{R})$ is of course the Radon-Nikodym derivative w.r.t. the Lebesgue measure ℓ , there is no obvious reference measure to use to define densities on, for example, $\Delta(\Delta(\mathbb{R}))$. (See (Ghosh and Ramamoorthi, 2003, Ch. 2) for some discussion.)

the mind of the agent. (For example, consider the case of an agent who consider *equally likely* every prior in $\Delta(\mathbb{R})$: whatever reference measure we use on $\Delta(\mathbb{R})$, there is no non-zero density on $\Delta(\mathbb{R})$ that can express this belief.) For this reasons, to express the relative likelihood of events we will simply use a function $\rho : \Delta^F(\Sigma) \rightarrow [0, 1]$ such that $\rho(\pi) > 0$ for some $\pi \in \Delta^F(\Sigma)$. This ρ will play the role of a density in our analysis by representing the relative likelihood:²² before receiving any information our agent chooses the prior that maximizes ρ ; and if information A was assigned a probability below the threshold, she would pick the prior that maximizes the “updated” ρ , i.e. $\rho(\pi)\pi(A)$.

For simplicity let us add the following notation. For any linear order \triangleright on a set Y , and for any $A \subseteq Y$, define

$$\max_{\triangleright} A := \{x \in A : x \triangleright y \ \forall y \in A\}.$$

This is the set of elements of A that maximize the linear order \triangleright . Finally, when Ω is infinite we focus on preferences defined on *simple* acts, i.e. the set of finite-valued Σ -measurable functions from Ω to X .

Definition 7. A class of preferences relations $\{\succeq_A\}_{A \in \Sigma}$ admits an **Extended Hypothesis Testing Representation** if there exists a continuous function $u : X \rightarrow \mathbb{R}$, a function $\rho : \Delta^F(\Sigma) \rightarrow [0, 1]$, a linear order \triangleright over $\Delta^F(\Sigma)$, and $\epsilon \in [0, 1]$, such that for any $A \in \Sigma$ there exist $\pi_A \in \Delta^F(\Sigma)$ such that:

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

$$f \succeq_A g \Leftrightarrow \int_{\Omega} \mathbb{E}_f(u) d\pi_A \geq \int_{\Omega} \mathbb{E}_g(u) d\pi_A;$$

2. $\{\pi_{\Omega}\} = \max_{\triangleright} \arg \max_{\pi \in \Delta^F(\Sigma)} \rho(\pi)$;

- 3.

$$\pi_A = \begin{cases} \text{BU}(\pi_{\Omega}, A) & \text{if } \pi_{\Omega}(A) > \epsilon \\ \text{BU}(\pi_A^*, A) & \text{otherwise} \end{cases}$$

where $\{\pi_A^*\} = \max_{\triangleright} \arg \max_{\pi \in \Delta^F(\Sigma)} \rho(\pi)\pi(A)$;

4. for any $A \in \Sigma$ there exist $\pi \in \Delta^F(\Sigma)$ such that $\rho(\pi) > 0$ and $\pi(A) > 0$;
5. for any $\pi, \pi' \in \Delta^F(\Sigma)$, if $\pi \triangleright \pi'$ then $\rho(\pi) \geq \rho(\pi')$;

Definition 8. An Extended Hypothesis Testing Representation $(u, \rho, \triangleright, \epsilon)$ is **minimal** if there is no $\epsilon' \in [0, 1)$ such that $\epsilon' < \epsilon$ and $(u, \rho, \triangleright, \epsilon')$ is an Hypothesis Testing Representation of the same preferences.

Finally, the representation theorem.²³

²²This means that if a density exists, then we could simply use it instead of ρ and our approach would coincide with the standard Bayesian updating of densities.

²³We should point out that the proof of Theorem 3 relies on the Szpilrajn Extension Theorem and therefore implicitly requires the Axiom of Choice.

Theorem 3. For any state space Ω , a class of preference relations $\{\succeq_A\}_{A \in \Sigma}$ satisfies *WbP*, *Consequentialism*, and *Dynamic Coherence* if and only if it admits a minimal *Extended Hypothesis Testing Representation* $(u, \rho, \triangleright, \epsilon)$.
Moreover, $\epsilon = 0$ if and only if $\{\succeq_A\}_{A \in \Sigma}$ satisfies also *Dynamic Consistency*.

Theorem 3 shows that the same axioms that we used when Ω is finite are necessary and sufficient for the existence of an *Extended Hypothesis Testing* representation for any Ω .

B. Proofs

B.1 Proof of Theorem 1

[*Sufficiency of the Axioms*] Given Axiom 1, it is standard practice to show that for any $A \in \Sigma$, there exist a non-constant affine function $u_A : X \rightarrow \mathbb{R}$, $\pi_A \in \Delta(\Omega)$ such that for any $f, g \in \mathcal{F}$

$$f \succeq_A g \Leftrightarrow \sum_{\omega \in \Omega} \pi_A(\omega) u_A(f(\omega)) \geq \sum_{\omega \in \Omega} \pi_A(\omega) u_A(g(\omega)); \quad (\text{A.1})$$

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.(4) 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 $u : X \rightarrow \mathbb{R}$ as $u = u_\Omega$ and $\pi = \pi_\Omega$.

Claim 1. For any $A, B \in \Sigma$, $A \supseteq B$, if $(\Omega \setminus B)$ is \succeq_A -null, then $\succeq_A = \succeq_B$.

Proof. Consider $A, B \in \Sigma$, $A \supseteq B$ such that $(\Omega \setminus B)$ is \succeq_A -null. Notice that since $A = A \cup B$, then $(\Omega \setminus B)$ is $\succeq_{A \cup B}$ -null. At the same time, since $(\Omega \setminus A)$ is $\succeq_{A \cup B}$ -null by Axiom 2. But then, Axiom 4 implies $\succeq_A = \succeq_B$ as sought. \square

Claim 2. For any $A, B \in \Sigma$, if $\pi_A(B) = 1 = \pi_B(A)$, then $\pi_A = \pi_B$.

Proof. Consider any $A, B \in \Sigma$ such that $\pi_A(B) = 1 = \pi_B(A)$. Notice that by construction of π_A and π_B we must have $\pi_A(A \cap B) = 1 = \pi_B(A \cap B)$. But then, by Claim 1 we must have $\pi_A = \pi_{A \cap B} = \pi_B$ as sought. \square

Define now the set $\mathcal{K} \subseteq \Sigma$ as follows:

$$\mathcal{K} := \{A \in \Sigma : A \text{ is } \succeq \text{-null}\} \cup \{A \in \Sigma : \exists f, g \in \mathcal{F} \text{ s.t. } f \succeq_A g \text{ and } g \succ fAg, \text{ or } f \succ_A g \text{ and } g \succeq fAg\}.$$

These are the events after which either *Dynamic Consistency* does not apply (null-events), or after which it is violated. Define now ϵ as $\epsilon := \max\{\pi(B) : B \in \mathcal{K}\}$ if $\mathcal{K} \neq \emptyset$, $\epsilon = 0$ if $\mathcal{K} = \emptyset$. Notice also that, by construction of \mathcal{K} , this implies that $\epsilon \geq \pi(A)$ for all $A \in \mathcal{K}$. Also, by construction we must have $\epsilon \in [0, 1)$. To see why, consider any $A \in \Sigma$ such that $\pi(A) = 1$. This implies that $(\Omega \setminus A)$ is \succeq -null. But then, $\succeq_A = \succeq_\Omega$ by Claim 1, hence $A \notin \mathcal{K}$. This implies that $\max\{\pi(B) : B \in \mathcal{K}\} < 1$, hence $\epsilon \in [0, 1)$.

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

Claim 3. For any $f, g, h \in \mathcal{F}$, $A \in \Sigma \setminus \mathcal{K}$ we have

$$f \succeq_{I_t, A} g \Leftrightarrow fAh \succeq_{I_t} gAh$$

Proof. Consider $f, g, h \in \mathcal{F}$, and $A \in \Sigma \setminus \mathcal{K}$. By construction of \mathcal{K} , for any $r, s \in \mathcal{F}$ we have $rAs \succeq s$ iff $r \succeq_A s$. Notice now that by Axiom 2 we have $fAh \sim_A f$ and $gAh \sim_A g$. This implies $f \succeq_A g$ 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 $f'Ag' \succeq_{I_t} g'$ iff $fAh \succeq_{I_t} gAh$ (where the last passages use the fact that $\pi(A) > \epsilon$). \square

Claim 4. For any $A \in \Sigma \setminus \mathcal{K}$, $\pi_A(B) = \frac{\pi(A \cup B)}{\pi(A)} = \text{BU}(\pi, A)$.

Proof. Now first of all that for any $A \in \Sigma \setminus \mathcal{K}$ we must have $\pi(A) > 0$, because any A which is \succeq -null belongs to \mathcal{K} . We then have that for any $f, g \in \mathcal{F}$

$$\begin{aligned} f \succeq_A g &\Leftrightarrow fAh \succeq gAh \\ &\Leftrightarrow \sum_{\omega \in A} \pi(\omega)u(f(\omega)) + \sum_{\omega \in \Omega \setminus A} \pi(\omega)u(f(\omega)) \geq \sum_{\omega \in A} \pi(\omega)u(g(\omega)) + \sum_{\omega \in \Omega \setminus A} \pi(\omega)u(g(\omega)) \\ &\Leftrightarrow \sum_{\omega \in A} \pi(\omega)u(f(\omega)) \geq \sum_{\omega \in A} \pi(\omega)u(g(\omega)) \\ &\Leftrightarrow \frac{1}{\pi(A)} \sum_{\omega \in A} \pi(\omega)u(f(\omega)) \geq \frac{1}{\pi(A)} \sum_{\omega \in A} \pi(\omega)u(g(\omega)) \end{aligned}$$

Since π_A is unique, this proves that for any $A \in \Sigma \setminus \mathcal{K}$, $\pi_A(B) = \frac{\pi(A \cup B)}{\pi(A)}$ as sought. In particular, this is also true for any $A \in \Sigma$ such that $\pi(A) > \epsilon$. \square

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

$$\mathcal{K}^* := \{A \in \Sigma : \pi_\Omega(B) \geq \pi_\Omega(A) \text{ for some } B \in \mathcal{K}\}.$$

Notice that we must have $\mathcal{K}^* = \{A \in \Sigma : \pi(A) \leq \epsilon\}$ by construction of ϵ . This is the set of events such that there exist an event in \mathcal{K} that is more likely than some of them. Define now the sets $H := \mathcal{K}^* \cup \{\Omega\}$ and $M := \{\pi_m \in \Delta(\Omega) : m \in \mathcal{K}^*\} \cup \{\pi_\Omega\}$.

Claim 5. The following holds for no $A, B, C, D \in \mathcal{K}^* \cup \{\Omega\}$: $\pi_A = \pi_B$, $\pi_C = \pi_D$, $\pi_C(A) = 1$, $\pi_A(C) < 1$, $\pi_B(D) = 1$ and $\pi_D(B) < 1$.

Proof. Say by means of contradiction that such $A, B, C, D \in M$ and notice that $\pi_C = \pi_D$ and $\pi_C(A) = 1$ imply $\pi_D(A) = 1$. Similarly, $\pi_A = \pi_B$ and $\pi_B(D) = 1$ imply $\pi_A(D) = 1$. But then, Claim 2 implies $\pi_A = \pi_D$, which means $\pi_A = \pi_C$. But by construction of π_C this contradicts $\pi_A(C) < 1$. \square

Define now the binary relation \triangleright on M as

$$\pi_m \triangleright \pi_{m'} \Leftrightarrow \pi_{m'}(m) = 1 \text{ and } \pi_m(m') < 1.$$

Notice that \triangleright is well defined by Claim 5, and it is also irreflexive for the same reason.

Claim 6. Consider $\pi_1, \dots, \pi_n \in M$ such that $\pi_1 \triangleright \dots \triangleright \pi_n$. Then, there exist $A_1, \dots, A_n \in \mathcal{K}^*$ such that $\pi_i = \pi_{A_i}$, $\pi_{A_{i+1}}(A_i) = 1$ and $\pi_{A_i}(A_{i+1}) < 1$ for $i = 1, \dots, (n-1)$.

Proof. For simplicity focus on the case in which $n = 3$: it is trivial to show that the proof extends to the general case. Consider $\pi_1, \pi_2, \pi_3 \in M$ such that $\pi_1 \triangleright \pi_2 \triangleright \pi_3$. By construction of \triangleright we know that there exist $A_1, A_2, A'_2, A_3 \in \mathcal{K}^*$ such that $\pi_1 = \pi_{A_1}$, $\pi_2 = \pi_{A_2} = \pi_{A'_2}$, $\pi_3 = \pi_{A_3}$ and $\pi_{A_2}(A_1) = 1$, $\pi_{A_1}(A_2) < 1$, $\pi_{A_3}(A'_2) = 1$, $\pi_{A'_2}(A_3) < 1$. Since $\pi_{A_2} = \pi_{A'_2}$, we must also have $\pi_{A'_2}(A_1) = 1$. If we can prove that we also have $\pi_{A_1}(A'_2) < 1$, then we are done. Say, by contradiction, that $\pi_{A_1}(A'_2) = 1$. But then, since $\pi_{A'_2}(A_1) = 1$, by Claim 2 we have $\pi_{A_1} = \pi_{A'_2}$, hence $\pi_1 = \pi_2$, which contradicts the fact that \triangleright is irreflexive. \square

Claim 7. \triangleright is acyclic.

Proof. By means of contradiction consider $\pi_1, \dots, \pi_n \in M$ such that $\pi_1 \triangleright \dots \triangleright \pi_n \triangleright \pi_1$. By Claim 6 there exist $A_1, \dots, A_n \in \mathcal{K}^*$ such that $\pi_i = \pi_{A_i}$, $\pi_{A_{i+1}}(A_i) = 1$ and $\pi_{A_i}(A_{i+1}) < 1$ for $i = 1, \dots, (n-1)$, and $\pi_{A_1}(A_n) = 1$ and $\pi_{A_n}(A_1) < 1$. Since \triangleright is irreflexive notice that we must have that $\pi_{A_i} \neq \pi_{A_{i+1}}$ for $i = 1, \dots, (n-1)$, and $\pi_{A_1} \neq \pi_{A_n}$. But this contradicts Axiom 4. \square

Define $\gamma_* := \max\{\pi_A(B) : A, B \in \mathcal{K}^*, \pi_A(B) < 1\}$ if $\{\pi_A(B) : A, B \in \mathcal{K}^*, \pi_A(B) < 1\} \neq \emptyset$, $\gamma_* = 0$ otherwise. (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 $\gamma_* \in [0, 1)$, we must have $\delta > 0$.) Notice also that since M is finite and \triangleright is acyclic. Consider now the transitive closure of \triangleright and call it $\hat{\triangleright}$. Since \triangleright is irreflexive and antisymmetric, so must $\hat{\triangleright}$. To see why, say, by contradiction, that we have $A, B \in M$ such that $A \hat{\triangleright} B \hat{\triangleright} A$. Then, by definition of transitive closure, there must exist $m_1, \dots, m_{n+m} \in M$ such that $A \triangleright m_1 \triangleright \dots \triangleright m_n \triangleright B$ and $B \triangleright m_{n+1} \triangleright \dots \triangleright m_{m+n} \triangleright A$. But this violates the acyclicity of \triangleright .

Since M is finite, enumerate it and construct the function f as follows. Set $f(m_1) = 0$. Consider m_n . Assign to $f(m_n)$ any value such that: for all $i < n$, $f(m_n) \neq f(m_i)$; $f(m_n) > f(m_i)$ if $m_n \hat{\triangleright} m_i$; and $f(m_n) < f(m_i)$ if $m_i \hat{\triangleright} m_n$. To see why this is always possible, notice that for all m_n, m_i, m_j , with $n \geq i, j$, if we have $m_n \hat{\triangleright} m_i$ and $m_j \hat{\triangleright} m_n$, then we must also have that $m_j \hat{\triangleright} m_i$ since $\hat{\triangleright}$ is transitive, which implies that we must also have $f(m_j) > f(m_i)$. Thus f is well defined. Normalize now the function f so that it has a range in $(0, \delta)$ and call it v . Notice that we must have that $\pi \triangleright \pi'$ implies $v(\pi) > v(\pi')$, and such that $v(\pi) \neq v(\pi')$ for all $\pi, \pi' \in M$.

Construct now $\rho \in \Delta(\Delta(\Omega))$ 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.

We now turn to show that M and ρ that we just constructed are the ones that we are looking for.

Claim 8. $\{\pi_\Omega\} = \arg \max_{\pi \in \Delta(\Omega)} \rho(\pi)$.

Proof. Say, by means of contradiction that there exist $\pi \in \Delta(\Omega)$ such that $\pi \neq \pi_\Omega$ and $\rho(\pi) \geq \rho(\pi_\Omega)$. For this to be possible we must have $\rho(\pi) > 0$, which in turns implies that $\pi = \pi_m$ for some $m \in H$. Say first that we have that $\pi_\Omega(m) = 1$ for all $m \in H$ such that $\pi_m = \pi$. Then, since we also have that $\pi_m(\Omega) = 1$ for all $m \in H$ such that $\pi_m = \pi$, then Claim 2 would imply $\pi = \pi_\Omega$, a contradiction. This means that we must have $\pi_\Omega(m) < 1$ for some $m \in H$ such that $\pi_m = \pi$. Also, notice that we must have $\pi_m(\Omega) = 1$ (by definition), which implies that we have $\pi_\Omega \triangleright \pi_m = \pi$. But then we must have $v(\pi_\Omega) > v(\pi)$ by construction of v , which in turns implies that we must have $\rho(\pi_\Omega) > \rho(\pi)$, a contradiction. \square

Notice now that for any $A \in \Sigma$, given the definition of ϵ , the agent will behave as prescribed in the representation if $\pi(A) > \epsilon$. We now turn to analyze the events with probability below the threshold.

Claim 9. For any $A \in H$, if $\pi \in \arg \max_{m \in M} \text{BU}(\rho, A)$, then $\pi(A) = 1$.

Proof. Consider $A \in H$ and say by means of contradiction that there exist $\pi \in \arg \max_{m \in M} \text{BU}(\rho|A)$ such that $\pi(A) < 1$. This means that we must have $\pi(A)\rho(\pi) \geq \pi_A(A)\rho(\pi_A)$. Notice that since $A \in H$, by construction we must have $\pi_A(A) = 1$, which implies that we must have $\pi(A)\rho(\pi) \geq \rho(\pi_A)$. By construction of ρ this is possible only if $\pi \in M$ and

$$(v(\pi) + \frac{1}{|M|})\pi_A \geq v(\pi_A) + \frac{1}{|M|}.$$

Notice that by construction $v(\pi_A) > 0$. Moreover, since $\pi(A)\rho(\pi) \geq \pi_A(A)\rho(\pi_A)$, it must be that $\pi(A) > 0$, hence $\gamma^* > 0$. Therefore, by definition of $\gamma^* > 0$, we have $\pi_A \leq \gamma^*$. This implies

$$(v(\pi) + \frac{1}{|M|})\gamma^* > \frac{1}{|M|} \Rightarrow v(\pi) > \frac{1}{|M|} \frac{1 - \gamma^*}{\gamma^*}.$$

But since v has range $(0, \frac{1}{|M|} \frac{1 - \gamma^*}{\gamma^*})$, this is a contradiction. \square

Claim 10. For any $A \in H$ we have $\{\pi_A\} = \arg \max_{m \in M} \text{BU}(\rho, A)$.

Proof. Consider $A \in H$ and say, by means of contradiction, that we have $\pi \in \arg \max_{m \in M} \text{BU}(\rho|A)$ for some $\pi \neq \pi_A$. This means that we have $\pi(A)\rho(\pi) \geq \pi_A(A)\rho(\pi_A)$. By Claim 9 we know that we must have $\pi(A) = 1$ and since $\pi_A(A) = 1$, then this means that we must have $\rho(\pi) \geq \rho(\pi_A)$. For this to be possible we must have $\pi \in M$, which implies that there exists $B \in H \cup \{\Omega\}$ such that $\pi_B = \pi$. Now, if we have $\pi_A(B) = 1$, then since $\pi_B(A) = \pi(A) = 1$, by Claim 2 we have $\pi = \pi_A$, a contradiction. Therefore, we must have $\pi_A(B) < 1$. But then, we must have $\pi_A \triangleright \pi_B = \pi$, which implies $v(\pi_A) > v(\pi_B)$, hence $\rho(\pi_A) > \rho(\pi)$, a contradiction. \square

Claim 10 proves the representation. Notice also that point (4) of the representation is trivially true by our construction of ρ . We have therefore found an Hypothesis Testing Representation (u, ρ, ϵ) . Notice that by the definition of ϵ , u, ρ, ϵ is also a minimal representation.

[*Necessity of the Axioms*] The proof the necessity of Axiom 1 (WbP) and Axiom is standard practice. Axiom 2 is immediate from the representation. We are left with Axiom 4. Say by means of contradiction that there exist $A_1, \dots, A_n \in \Sigma$ such that $\succeq_{A_1} \neq \succeq_{A_n}$, $(\Omega \setminus A_{i+1})$ is \succeq_{A_i} -null for $i = 1, \dots, (n-1)$, and $(\Omega \setminus A_1)$ is \succeq_{A_n} -null. Consider first the case in which $\pi(A_i) > \epsilon$ for $i = 1, \dots, n$. By the representation, this means that we must have $\pi_{A_i} = \text{BU}(\pi, A_i)$ for $i = 1, \dots, n$. But since $(\Omega \setminus A_{i+1})$ is \succeq_{A_i} -null for $i = 1, \dots, (n-1)$, hence $\pi_{A_i}(A_i \setminus A_{i+1}) = 0$, then by construction of BU we must have $\pi(A_i \setminus A_{i+1}) = 0$ for $i = 1, \dots, n$. For the same reason we must also have $\pi(A_n \setminus A_1) = 0$. But then we must have that $\pi(\bigcap_{i=1}^n A_i) = \pi(\bigcup_{i=1}^n A_i)$, and so, by definition of BU, we must have $\pi_{A_1} = \pi_{A_n}$, hence $\succeq_{A_1} = \succeq_{A_n}$, a contradiction.

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

We are left with the case in which $\pi(A_i) \leq \epsilon$ for $i = 1, \dots, n$. By construction we must have that $\text{BU}(\pi_{A_i}^*, A_i)(A_{i+1}) = 1$ for $i = 1, \dots, n-1$ and $\text{BU}(\pi_{A_n}^*, A_n)(A_1) = 1$. This implies $\pi_{A_i}^*(A_{i+1}) \geq \pi_{A_i}^*(A_i)$ for $i = 1, \dots, n-1$ and $\pi_{A_n}^*(A_1) \geq \pi_{A_n}^*(A_n)$. Now, notice that since $\pi_{A_i}^*$ is the unique element in $\arg \max_{\pi \in \Delta(\Omega)} \text{BU}(\rho, A_i)(\pi)$ for $i = 1, \dots, n$, then we must have that for all $i, j = 1, \dots, n$, $\rho(\pi_{A_i}^*)\pi_{A_i}^*(A_i) > \rho(\pi_{A_j}^*)\pi_{A_j}^*(A_i)$ if $\pi_{A_i}^* \neq \pi_{A_j}^*$. This, together with the fact that $\pi_{A_i}^*(A_{i+1}) \geq \pi_{A_i}^*(A_i)$ for $i = 1, \dots, n-1$ and $\pi_{A_n}^*(A_1) \geq \pi_{A_n}^*(A_n)$, implies that $\rho(\pi_{A_i}^*)\pi_{A_{i+1}}^*(A_{i+1}) \geq \rho(\pi_{A_{i+1}}^*)\pi_{A_{i+1}}^*(A_{i+1})$, where the inequality is strict if $\pi_{A_i}^* \neq \pi_{A_{i+1}}^*$. Hence for $i = 1, \dots, (n-1)$, $\rho(\pi_{A_i}^*) \geq \rho(\pi_{A_{i+1}}^*)$, 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 $\pi_{A_n}^* \neq \pi_{A_1}^*$. But then we have $\rho(\pi_{A_1}^*) \geq \rho(\pi_{A_2}^*) \geq \dots \geq \rho(\pi_{A_n}^*) \geq \rho(\pi_{A_1}^*)$, and so none of these inequalities can be strict, hence $\pi_{A_1}^* = \pi_{A_i}^*$ for $i = 1, \dots, n$. But then, $\text{BU}(\pi_{A_1}^*, A_i)(A_{i+1}) = 1$ for $i = 1, \dots, n-1$, and $\text{BU}(\pi_{A_1}^*, A_n)(A_1) = 1$, which means $\pi_{A_1}^*(A_i \setminus A_{i+1}) = 0$ for $i = 1, \dots, n-1$, and $\pi_{A_1}^*(A_n \setminus A_1) = 0$. This means $\pi_{A_1}^*(\bigcap_{i=1}^n A_i) = \pi_{A_1}^*(\bigcup_{i=1}^n A_i)$. But this implies that $\text{BU}(\pi_{A_1}^*, A_n) = \text{BU}(\pi_{A_1}^*, A_1)$, hence $\pi_{A_1} = \pi_{A_n}$ and $\succeq_{A_1} = \succeq_{A_n}$, a contradiction.

[$\epsilon = 0$ iff *Dynamic Consistency*] Notice first of all that if $\epsilon = 0$, then the agent updates her prior using Bayes' rule every time she is told that a non-null event has occurred, which it is well known to imply that Dynamic Consistency is satisfied. Consider now the case in which the $\{\succeq_A\}_{A \in \Sigma}$ satisfies Dynamic Consistency. Let us say, by means of contradiction, that we have a minimal Hypothesis Testing representation (u, ρ, ϵ) of $\{\succeq_A\}_{A \in \Sigma}$ in which $\epsilon \neq 0$. Since $\{\succeq_A\}_{A \in \Sigma}$ satisfies Dynamic Consistency, however, then $(u, \rho, 0)$ must also represent it, contradicting the minimality of (u, ρ, ϵ) .

B.2 Proof of Theorem 2

[*Sufficiency of the Axioms*] We proceed in a similar way to how we proceeded for the proof of Theorem 1. Given Axioms 5, from Gilboa and Schmeidler (1989) we know that for any $A \in \Sigma$, there exist $u_A : X \rightarrow \mathbb{R}$, $\Pi_A \subseteq \Delta(\Omega)$, Π convex and compact, such that for any

$f, g \in \mathcal{F}$

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

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 5.(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 $u : X \rightarrow \mathbb{R}$ 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 Claim 1 holds true here as well. Moreover, notice the following claim (which parallels Claim 2 in the proof of Theorem 1).

Claim 11. 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 $(\Omega \setminus (A \cap B))$ is both \succeq_A -null and \succeq_B -null. But then, by Claim 1 we must have $\Pi_A = \Pi_{A \cap B} = \Pi_B$ as sought. \square

Define now the set $\mathcal{K}_{AA} \subseteq \Sigma$ as $\mathcal{K}_{AA} := \{A \in \Sigma : A \text{ is not } \succeq\text{-unambiguously non-null}\} \cup \{A \in \Sigma : \exists f, g \in \mathcal{F} \text{ s.t. } f \succeq_A^* g \text{ and } g \succ^* fAg, \text{ or } f \succ_A^* g \text{ and } g \succeq^* fAg\}$. These are the events after which either Reduced Dynamic Consistency (Axiom 6) 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 $\mathcal{K}_{AA} \neq \emptyset$, $\epsilon = 0$ if $\mathcal{K}_{AA}^* = \emptyset$. (This is well defined since Ω is finite.)

Consider now $A \in \Sigma \setminus \mathcal{K}_{AA}$ (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 12. For any $A \in \Sigma \setminus \mathcal{K}_{AA}$, $f, g, h \in \mathcal{F}$, we have

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

Proof. Consider $f, g, h \in \mathcal{F}$, $A \in \Sigma \setminus \mathcal{K}_{AA}$. 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 $A \notin \mathcal{K}_{AA}^*$) $f'Ag' \succeq_\Omega g'$ 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

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

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.²⁴ 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

²⁴See 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). (For completeness, let us point out that our continuity (Axiom ??1) implies the Archimedean axiom that they impose.)

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

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 $\Pi_A(B) = \hat{B}U(\Pi_\Omega, A)$.

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

We can now proceed replicating exactly the steps in the proof of Theorem 1 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 \triangleright on M_{AA} .

Define $\gamma_{AA}^* := \max_{\pi \in \Pi_A} \{\min \pi(B) : A, B \in \mathcal{K}_{AA}^*, \pi(B) < 1 \text{ for some } \pi \in \Pi_A\}$. (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 constructing constructing the transitive closure of $\hat{\triangleright}$ of \triangleright , 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 (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 Axiom 2 are immediate. We are left with Axiom 4. Consider $A_1, \dots, A_n \in \Sigma$ such that $\succeq_{A_1} \neq \succeq_{A_n}$, $(\Omega \setminus A_{i+1})$ is \succeq_{A_i} -null for $i = 1, \dots, (n-1)$, and $(\Omega \setminus A_1)$ is \succeq_{A_n} -null. Consider first the case in which $\pi(A_i) > \epsilon$ for all $\pi \in \Pi$, for all $i = 1, \dots, n$. Since $(\Omega \setminus A_{i+1})$ is \succeq_{A_i} -null for $i = 1, \dots, (n-1)$, then it must be that $\pi(A_i \setminus A_{i+1}) = 0$ for all $\pi \in \Pi_{A_i}$ for $i = 1, \dots, (n-1)$, and $\pi(A_n \setminus A_1) = 0$ 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 $i = 1, \dots, n$. But

then, for all $\pi \in \Pi_\Omega$ we have $\pi(A_i \setminus A_{i+1}) = 0$ for $i = 1, \dots, (n-1)$, and $\pi(A_1 \setminus A_n) = 0$. 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 $j = 1, \dots, n$. 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 $\pi(A_1 \setminus A_2) = 0$ 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 $i = 1, \dots, n$. But we have already shown that this leads to a contradiction.

We are left with the case in which for $i = 1, \dots, n$ we have $\pi(A_i) \leq \epsilon$ for some $\pi \in \Pi$. Since $(\Omega \setminus A_{i+1})$ is \succeq_{A_i} -null for $i = 1, \dots, (n-1)$, then we must have that $\text{BU}(\pi, A_i)(A_{i+1}) = 1$ for all $\pi \in \Pi_{A_i}^*$ for $i = 1, \dots, (n-1)$. 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 $i = 1, \dots, n-1$ 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 $i = 1, \dots, n-1$ 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 $\arg \max_{\Pi \in \mathcal{C}} \text{BU}(\rho, A_i)(\Pi)$ for $i = 1, \dots, n$, then we must have that for all $i, j = 1, \dots, n$, $\rho(\Pi_{A_i}^*) \min_{\pi \in \Pi_{A_i}^*} \pi(A_i) > \rho(\Pi_{A_j}^*) \min_{\pi \in \Pi_{A_j}^*} \pi(A_i)$ 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 $i = 1, \dots, n-1$ 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 $i = 1, \dots, (n-1)$, 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 $\rho(\Pi_{A_1}^*) \geq \rho(\Pi_{A_2}^*) \geq \dots \geq \rho(\Pi_{A_n}^*) \geq \rho(\Pi_{A_1}^*)$, and so none of this inequalities can be strict, hence $\Pi_{A_i}^* = \Pi_{A_j}^*$ for $i = 1, \dots, n$. But then, for all $\pi \in \Pi_{A_1}^*$ we have $\text{BU}(\pi, A_i)(A_{i+1}) = 1$ for $i = 1, \dots, n-1$, and $\text{BU}(\pi, A_n)(A_1) = 1$, which means $\pi(A_i \setminus A_{i+1}) = 0$ for $i = 1, \dots, n-1$, and $\pi(A_n \setminus A_1) = 0$. 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 $\hat{\text{BU}}(\Pi_{A_1}^*, A_n) = \hat{\text{BU}}(\Pi_{A_1}^*, A_1)$, 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.

B.3 Proof of Theorem 3

To prove the sufficiency of the axioms we follow similar steps as in the proof of the sufficiency of the axioms in Theorem 1. (For brevity we highlight only the difference between the two proofs.) In particular, define u, π_A for all $A \in \Sigma, \mathcal{K}$ and \mathcal{K}^* . (Notice that for each $A \in \Sigma$, here π_A will belong to $\Delta^F(\Sigma)$ and not necessarily to $\Delta(\Omega)$.) Define now $\bar{\mathcal{K}}^*$ as follows.

$$\bar{\mathcal{K}}^* := \{A \in \Sigma : \exists (A_n) \in (\mathcal{K}^*)^\infty \text{ such that } \pi(A_n) \rightarrow \pi(A)\}.$$

Define $\epsilon := \sup_{A \in \mathcal{K}^*} \pi(A)$ if $\mathcal{K} \neq \emptyset$ (hence $\mathcal{K}^* \neq \emptyset$), $\epsilon = 0$ if $\mathcal{K} = \emptyset$. Proceed following the steps of the proof of Theorem 1 simply using \bar{K}^* and $\Delta^F(\Sigma)$ whenever we used \mathcal{K}^* and $\Delta(\Omega)$, therefore defining $H := \bar{\mathcal{K}}^* \cup \{\Omega\}$ and $M := \{\pi_m \in \Delta^F(\Sigma) : m \in \bar{\mathcal{K}}^*\} \cup \{\pi_\Omega\}$, until the point in which we define $\hat{\triangleright}$ as the transitive closure of the binary relation \triangleright . Notice that here as well we would have that the acyclicity of \triangleright implies that $\hat{\triangleright}$ must be antisymmetric, which implies that $\hat{\triangleright}$ is a strict partial order. By means of the Szpilrajn Extension Theorem we know that there exist a linear order on M that completes $\hat{\triangleright}$ and hence \triangleright . (See, for example, (Rothmaler, 2000, p. 88).) Call this linear order \triangleright' . Extend now \triangleright' to a linear order $\bar{\triangleright}$ on $\Delta^F(\Sigma)$ in any way that guarantees that for any $\pi, \pi' \in \Delta^F(\Sigma)$, if $\pi \in M$ and $\pi' \notin M$, we have $\pi \bar{\triangleright} \pi'$. (The existence of this extension is trivial.) Define now $\rho : \Delta^F(\Sigma) \rightarrow \mathbb{R}$ as $\rho(\pi) = 1$ for all $\pi \in M$, and $\rho(\pi) = 0$ otherwise.

We now prove that $(u, \rho, \bar{\triangleright}, \epsilon)$ is an Extended Hypothesis Testing representation. Notice first of all that we must have that $\pi_A \in \arg \max_{\pi \in \Delta(\Omega)} \rho(\pi)\pi(A)$: this is trivially true since ρ is uniform and $\pi_A(A) = 1$. For the same reason we also have that if $\pi \in \arg \max_{\pi \in \Delta(\Omega)} \rho(\pi)\pi(A)$, then $\pi(A) = 1$. Finally, condition (5) is trivially true since ρ is uniform. We are left with the following claim.

Claim 13. For all $A \in \Sigma$, $\{\pi_A\} = \max_{\substack{\triangleright \\ \pi \in \Delta^F(\Sigma)}} \arg \max \rho(\pi)\pi(A)$.

Proof. Consider any $\pi \in \arg \max_{\pi \in \Delta^F(\Sigma)} \rho(\pi)\pi(A)$. Since $\rho(\pi_A) = 1$, and $\rho(\pi) = 1$ iff $\pi \in M$, then we must have that $\pi = \pi_B$ for some $B \in \bar{\mathcal{K}}^* \cup \{\Omega\}$. At the same time, we must also have that $\pi_B(A) = 1$. If we also had $\pi_A(B) = 1$, then by Claim 2 in the proof of Theorem 1 (which applies here as well) we would have $\pi_A = \pi_B$. If, instead, $\pi_B(A) < 1$, then we have $\pi_A \triangleright \pi_B$, hence $\pi_A \bar{\triangleright} \pi_B$. This proves the claim. \square

The minimality follows trivially. Finally, the proof of the necessity of the axioms is then identical to the one of the proof of Theorem 1, and the same holds true for the proof that $\epsilon = 0$ if and only if the preference satisfies also Dynamic Consistency.

References

- ANSCOMBE, F. J. AND R. J. AUMANN (1963): “A Definition of Subjective Probability,” *The Annals of Mathematical Statistics*, 34, 199–205.
- BLUME, L., A. BRANDENBURGER, AND E. DEKEL (1991): “Lexicographic probabilities and choice under uncertainty,” *Econometrica: Journal of the Econometric Society*, 61–79.
- BRANDT, M., Q. ZENG, AND L. ZHANG (2004): “Equilibrium stock return dynamics under alternative rules of learning about hidden states,” *Journal of Economic Dynamics and Control*, 28, 1925–1954.
- BRAV, A. AND J. HEATON (2002): “Competing theories of financial anomalies,” *Review of Financial Studies*, 15, 575.
- CAMERER, C. (1995): “Individual decision making,” in *Handbook of Experimental Economics*, ed. by A. R. J. Kagel, Princeton University Press, Princeton, vol. 2.

- CHATEAUNEUF, A., F. MACCHERONI, M. MARINACCI, AND J. TALLON (2005): “Monotone continuous multiple priors,” *Economic Theory*, 26, 973–982.
- DE BONDT, W. AND R. THALER (1985): “Does the stock market overreact?” *The Journal of Finance*, 40, 793–805.
- DE FINETTI, B. (1931): “Sul significato soggettivo della probabilita,” *Fundamenta mathematicae*, 17, 298–329.
- (1970): *Teoria delle probabilita*, Einaudi Torino.
- EPSTEIN, L. (2006): “An axiomatic model of non-Bayesian updating,” *Review of Economic Studies*, 73, 413–436.
- EPSTEIN, L. AND M. LE BRETON (1993): “Dynamically Consistent Beliefs Must Be Bayesian,” *Journal of Economic Theory*, 61, 1–22.
- EPSTEIN, L., J. NOOR, AND A. SANDRONI (2008): “Non-Bayesian updating: a theoretical framework,” *Theoretical Economics*, 3, 193–229.
- (2010): “Non-Bayesian Learning,” *The B.E. Journal of Theoretical Economics (Advances)*, 10.
- EPSTEIN, L. 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.
- FOSTER, D. AND H. YOUNG (2003): “Learning, hypothesis testing, and Nash equilibrium,” *Games and Economic Behavior*, 45, 73–96.
- GENNAIOLI, N. AND A. SHLEIFER (2009): “What comes to mind,” *NBER Working Paper*.
- 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.
- GHOSH, J. AND R. RAMAMOORTHY (2003): *Bayesian nonparametrics*, Springer Verlag.
- GILBOA, I. AND L. SAMUELSON (2008): “Preferring Simplicity,” Mimeo Yale University.
- GILBOA, I. AND D. SCHMEIDLER (1989): “Maxmin expected utility with a non-unique prior,” *Journal of Mathematical Economics*, 18, 141–153.
- (1993): “Updating Ambiguous Beliefs,” *Journal of Economic Theory*, 59, 34–49.

- GREETHER, D. (1992): “Testing Bayes rule and the representativeness heuristic: Some experimental evidence,” *Journal of Economic Behavior & Organization*, 17, 31–57.
- GRIFFIN, D. AND A. TVERSKY (1992): “The weighing of evidence and the determinants of confidence,” *Cognitive Psychology*, 24, 411–435.
- HANANY, E. AND P. KLIBANOFF (2006): “Updating Preferences with Multiple Priors,” M, Northwestern University.
- HOLT, C. AND A. SMITH (2009): “An update on Bayesian updating,” *Journal of Economic Behavior and Organization*, 69, 125–134.
- KAHNEMAN, D. AND A. TVERSKY (1982): *Intuitive Prediction: Biases and Corrective Procedures*, London: Cambridge University Press.
- KOCHOV, A. (2009): “A Model of Limited Foresight,” Miemo, University of Rochester.
- 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.
- MYERSON, R. (1986a): “Axiomatic foundations of bayesian decision theory,” *Center for Mathematical Studies in Economics and Management Science, Northwestern University Working Paper*, 671.
- (1986b): “Multistage games with communication,” *Econometrica*, 54, 323–358.
- RABIN, M. (1998): “Psychology and economics,” *Journal of Economic Literature*, 11–46.
- (2002): “Inference By Believers in the Law of Small Numbers*,” *Quarterly Journal of Economics*, 117, 775–816.
- ROTHMALER, P. (2000): *Introduction to model theory*, CRC.
- SAVAGE, L. (1954): *The foundations of statistics*, Wiley Publications in Statistics.
- SINISCALCHI, M. (2006): “Dynamic Choice under Ambiguity,” Mimeo, Northwestern University.
- TVERSKY, A. AND D. KAHNEMAN (1974): “Judgment under uncertainty: Heuristics and biases,” *Science*, 185, 1124–1131.
- (1981): “The framing of decisions and the psychology of choice,” *Science*, 211, 453–458.