American Economic Review 2012, 102(6): 2410–2436
http://dx.doi.org/10.1257/aer.102.6.2410

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

By Pietro Ortoleva*

Bayes' rule has two well-known limitations: 1) it does not model the reaction to zero-probability events; 2) a sizable empirical evidence documents systematic violations of it. We characterize axiomatically an alternative updating rule, the Hypothesis Testing model. According to it, the agent follows Bayes' rule if she receives information to which she assigned a probability above a threshold. Otherwise, she looks at a prior over priors, updates it using Bayes' rule for second-order priors, and chooses the prior to which the updated prior over priors assigns the highest likelihood. We also present an application to equilibrium refinement in game theory. (*JEL* D11, D81, D83)

One of the most common assumptions in economics is that agents update their beliefs using Bayes' rule. It is easy to see the reasons for this widespread use: a very strong normative appeal—so strong that it is sometimes considered a feature of rationality—is coupled with a very convenient functional form, which facilitates its application even in complicated economic models.

Despite these advantages, however, Bayes' rule has two well-known limitations. First of all, it contains no prescription on how the agent should react to information to which she assigned probability *zero*—Bayes' rule is simply not defined in that case. This limitation has relevant consequences in economic modeling. For example, in game theory the notion of Perfect Bayesian Nash Equilibrium is often criticized, and then refined, because it posits no restrictions on the beliefs out of the equilibrium path. This is an immediate consequence of the fact that Bayes' rule, on which the equilibrium notion is defined, does not specify how these beliefs should be formed.

Even in the cases in which Bayes' rule does apply, however, it might not provide an accurate description of behavior. Indeed, it is now an established finding in psychology and behavioral economics that decision makers tend to systematically

* Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd., MC 228-77, Pasadena, CA 91125 (e-mail: ortoleva@caltech.edu). I thank two anonymous referees, Kim Border, Mark Dean, Federico Echenique, Larry Epstein, Paolo Ghirardato, Andrea Mattozzi, Asen Kochov, Jawwad Noor, Efe Ok, Leonardo Pejsachowicz, Ariel Rubinstein, Gil Riella, Marciano Siniscalchi, Leeat Yariv, the participants at seminars at ASU, Boston University, Caltech, Collegio Carlo Alberto, New York University, USC, and at ESSET 2010, RUD 2010, SWET 2010, and AEA Meetings 2011 for useful comments and discussions.

† To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.102.6.2410.

deviate from its prescriptions.[1] Among the documented violations, particularly interesting is the evidence of non-Bayesian reactions to "unexpected" news: subjects tend to violate Bayes' rule when they are given information to which they originally assigned a positive but *small* probability, often overreacting to the new information.[2] This very issue has also been the central topic of a large number of empirical studies in finance. In two influential papers, De Bondt and Thaler (1985, 1987) suggest that stock market returns show evidence of non-Bayesian reactions to unexpected information. Similar patterns are argued in Chopra, Lakonishok, and Ritter (1992) for stocks, Larson and Madura (2001) for exchange rates, and Poteshman (2001) for options markets. Many studies have also analyzed the importance of "surprise" in earning announcements for stock behavior,[3] while De Bondt and Thaler (1990) argue that patterns of non-Bayesian reactions could be observed also in the behavior of security analysts. Finally, and importantly, also introspection suggests that decision makers could have non-Bayesian reactions to unexpected news when beliefs are *subjective*, i.e., when there is no objectively known distribution of the outcomes, and agents need to form their own subjective belief.

To illustrate this intuition and the behavior that we are interested in, consider the case of an investor who wants to forecast the future return of a certain stock. Since there is no objectively known distribution about the likelihood of these returns, our decision maker will have to form a subjective belief about them. As she receives new information, she will update this belief. If this new information is not too "unexpected," e.g., in the "business-as-usual" situations, it is reasonable to expect the investor to revise her beliefs following Bayes' rule. However, what would happen if some really unlikely event is realized? For example, our agent might observe the announcement of extremely high, or low, earnings; or witness the market enter an unexpected financial crisis. If this event was assigned probability *zero*, then Bayes' rule is not defined, and our agent will have to somehow choose a new prior given the new evidence. To study this case, we would like to have a model that governs how this new prior is chosen, possibly imposing that this is done "rationally."

[1] For experimental evidence see, among others, Kahneman and Tversky (1972, 1973, 1982, 1983); Tversky and Kahneman (1974, 1981); Camerer (1987, 1989); Griffin and Tversky (1992); Grether (1980, 1992); Holt and Smith (2009); and the surveys in Camerer (1995), Rabin (1998), and Camerer and Loewenstein (2004). In all these experiments subjects are given an objective prior probability, which renders straightforward checking for violations of Bayes' rule. More recently, a few papers document violations of Bayes' rule in environments in which probabilities are subjective: among them, Asparouhova, Hertzel, and Lemmon (2009), Kremer, Moritz, and Siemsen (2011), and Dominiak, Dürsch, and Lefort (2009).

[2] For example, Kahneman and Tversky (1973) argues that subjects tend to overweight "salient" new evidence with respect to prior probability, to "predict the outcomes that appear most representative of the evidence" disregarding the base rate and "erroneously predict rare values and events if these happen to be representative" (Kahneman and Tversky 1973, p. 237). Holt and Smith (2009) present an experiment (with objective probabilities) in which subjects tend to follow Bayes' rule after events in an intermediate range, but who significantly overweight new evidence when there was a lower probability of receiving it. (These same subjects seem also to underreact to very high probability events.) The data collected in Grether (1992) show higher deviations from Bayes' rule for rare events. Finally, Griffin and Tversky (1992) present an experiment in which subjects "focus more on strength and extremeness of evidence [...] with insufficient regard for its weight" (Griffin and Tversky 1992, p. 411), in violation of Bayes' rule; these results are later confirmed with an incentive-compatible mechanism in Antoniou et al. (2010). Notice that, since in all of these experiments probabilities are objective, it is easy to separate the case of zero-probability events, in which Bayes' rule would hold vacuously, from that of small but positive probability events.

[3] See Bernard and Thomas (1989); Chan, Jegadeesh, and Lakonishok (1996); Vuolteenaho (2002); and La Porta et al. (1997). For surveys of the behavioral finance literature, see Shleifer (2000); Hirshleifer (2001); and Barberis and Thaler (2003).

Even if the realized event was assigned a positive probability, however, our investor might still not follow Bayes' rule. This would happen if the realized event was assigned a positive but also *small* probability—if it was considered "possible but also unlikely." To see why, recall that our investor is using a *subjective* prior on the states of the world: it is a prior that she has chosen herself, and that does not constitute an objective description of the true likelihoods. But then, if the prior used by the agent assigned a small probability to the realized event, she might ask herself if she was using the *wrong prior to begin with*: it is as if she told herself "if the prior I am using assigned such a low probability to what actually happened, then maybe it's the wrong prior!" In this case she could then decide to "change her prior" instead of simply updating it—violating Bayes' rule. For example, our investor could have used an economic model to form her original belief over stock returns; but if she saw that this model assigned a really small probability to the information that was later revealed, she might question its validity and look for an alternative one.

The goal of this article is then to study an alternative updating rule that (i) is defined also for zero-probability events, after which the agent "rationally" chooses a new prior, and (ii) allows for non-Bayesian reactions to unexpected news. To this end, we take a specific methodological approach, the axiomatic one: instead of discussing the model directly, we suggest a novel behavioral property (axiom), called Dynamic Coherence, and we show that, together with other standard assumptions, it is necessary and sufficient to guarantee that the behavior of the agent can be represented using a model with the two features above. To illustrate a possible application, we then use this model to construct, in a simple example, a refinement of Perfect Bayesian Nash Equilibrium in which some of the unappealing equilibria are ruled out.

I. Overview of the Results and of the Literature

A. *Overview of the Results*

We consider a standard model in which we observe the preferences of the agent before and after she receives some information about the state of the world. The first contribution of the paper is to suggest a novel axiom on how beliefs evolve after different pieces of information. To illustrate, let us look again at our investor and consider the case in which, if she receives the information that there is a "High Unemployment" rate in the United States, then she is *sure* that there also is a "Weak Housing Market." (That is, after she receives the information "High Unemployment," she has a belief that assigns to the event "Weak Housing Market" probability one). At the same time, let us also say that if the same investor were instead told that there is a "Weak Housing Market," then she would be *sure* that there is also "High Unemployment." That is, after either piece of information, "High Unemployment" and "Weak Housing Market," she is sure of the other one as well. If this is the case, we would like to say that both pieces of information *carry the same informational content* for this agent. But then, the updated belief of the agent should be the same after either piece of information.

The behavioral property that we propose, Dynamic Coherence, is a strengthening of the intuition above. Consider some other investor such that, if she is told that

there is "High Unemployment," then she is sure that there is also a "Weak Housing Market"; if she is told that there is a "Weak Housing Market," then she is sure that there is "Low Inflation"; but if she is told that there is "Low Inflation," then she is sure that there is also "High Unemployment." Notice that this agent has "circular" beliefs in this case: after each piece of information, she is *sure* also of the next one; but after the last one, she is *sure* of the first one as well. Then, just as before, we would like to posit that each of these pieces of information has the same informational content for the agent, and that each of them leads her to the *same beliefs*. We call this postulate Dynamic Coherence.

It is not hard to see that, as long as each of the events involved was assigned a strictly positive probability, then any Bayesian agent must satisfy Dynamic Coherence. But this is no longer true when zero-probability events are involved. In this case Dynamic Coherence would still apply and regulate the agent's beliefs—the intuition behind it is independent of the beliefs of the agent before the arrival of new information.[4] And since Dynamic Coherence applies after zero-probability events, while Bayes' rule does not, then Dynamic Coherence is *not* implied by the standard Bayesian postulates. Rather, as we shall see, it could be imposed together with them precisely to regulate the agent's reaction to zero-probability events. On the other hand, Dynamic Coherence alone is not enough to guarantee the full Bayesian behavior after non–zero probability events—it still allows for violations of Bayes' rule.

The main result of this paper is to show that Dynamic Coherence, together with other standard postulates, is equivalent to the existence of a *Hypothesis Testing representation* of the agent's behavior. According to this representation, the agent has a utility function u over consequences; a prior over priors ρ, i.e., a belief over which belief she should use; 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 to which the prior over priors ρ assigns the highest likelihood; then, she 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 prior over priors. In our example of the investor, we can see ρ as a prior over the possible economic models that the agent can use, each of which entails a certain prior over the state of the world: she would then use the model, hence, the prior, 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. This is where the threshold ϵ comes into play.

[4] In fact, if the agent originally assigned probability zero to "High Unemployment" or to "Weak Housing Market," Bayes' rule would not be defined, but the argument above would still hold. The key point here is that Dynamic Coherence imposes a coherence between beliefs *across* different pieces of information, independently of the belief before information. By contrast, the axiom underlying Bayes' rule, Dynamic Consistency, imposes a coherence between the agent's beliefs before and after information, which is ineffective if the information was assigned probability zero. (See Section IIB.)

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

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

- If, however, the probability that her prior assigned to the new information i is *below* the threshold, i.e., $\pi(i) \leq \epsilon$, then the prior is *rejected*, and our agent goes back to her prior over priors ρ ; updates it with Bayes' rule using the new information; then chooses the prior π' to which the *updated* prior over priors assigns the highest likelihood; using this prior π' she forms her preferences maximizing the expected utility.

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

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.

When $\epsilon > 0$ our agent's behavior departs from the Bayesian one: when she receives information to which she assigned a probability below the threshold ϵ, our agent does not simply update her prior but rather reconsiders whether she was using the right one in the first place. She might then modify her beliefs more than what was prescribed by Bayes' rule.

We also discuss the uniqueness properties of the representation, and we present a simple application of the Hypothesis Testing model to game theory: focusing on the well-known Beer/Quiche game, we use it to construct an alternative to the Perfect Bayesian Nash Equilibrium in which off-equilibrium beliefs are restricted. We show how, in this example, this alternative notion coincides with the Intuitive Criterion. As opposed to the latter, however, off-equilibrium behavior is not interpreted in light of a specific equilibrium, but rather it leads players to put in question their beliefs about what others are playing. Similar results could be obtained in the signaling model of Spence (1973), in which the only separating equilibrium that is not ruled out is the efficient one.

B. *Outline and Related Literature*

In the remainder of this section we discuss the related literature. Section II presents the formal setup and the axiomatic foundations. Section III introduces and discusses the main representation, the Hypothesis Testing model. Section IV develops the simple application to game theory. Section V concludes. The proofs appear in the Appendix.

Many generalizations of Bayes' rule have been proposed in the literature. Among the many behavioral models see, for example, Barberis, Shleifer, and Vishny (1998); Daniel, Hirshleifer, and Subrahmanyam (1998); Rabin and Schrag (1999); Kahneman and Frederick (2002); Mullainathan (2002); Rabin (2002);

Mullainathan, Schwartzstein, and Shleifer (2008); Gennaioli and Shleifer (2010); and Rabin and Vayanos (2010). A common feature of these contributions is that they propose significant generalizations of Bayes' law; by contrast, we wish to suggest a rule that coincides with Bayes' law for "normal" events and departs from it only for "unexpected" or zero-probability ones. Moreover, almost none of the papers above address how the agent should react to events to which she assigned probability zero—one of our main goals. Finally, none of these models is derived from testable behavioral rules. In behavioral finance, Hong, Stein, and Yu (2007) suggest a model in which an investor can use one of two explicitly oversimplified and incorrect models of the world to make predictions on the behavior of a stock. The investor chooses one of them and uses it until she sees it performing poorly. When this happens, she switches to the other one, with a paradigm shift similar to that of our representation, although applied to a very specific case and random process, and with the assumption that the agent cannot ever use the correct model.

Within the axiomatic decision theory literature, Epstein (2006) and Epstein, Noor, and Sandroni (2008) model agents who might be tempted to use a posterior that is different from the Bayesian update of their prior: for example, they might be tempted to overweigh or underweight new evidence. (Consequences of this behavior are discussed in Epstein, Noor, and Sandroni (2010).) Both models differ from ours in several aspects. First, they are not concerned with agent's reaction to zero-probability events, one of our main goals. Moreover, both Epstein (2006) and Epstein, Noor, and Sandroni (2008) study a setup of preferences over menus, and the latter develops a full dynamic model with an infinite horizon. By contrast, we look at the simple setup of preferences conditional on information, in which time plays no role.[5] Weinstein (2011) shows that subjects who violate Bayes' rule and have paradigm shifts can be subject to Dutch-book arguments, but only using bets that depend on events subsequent to a paradigm shift; for all other bets their behavior is indistinguishable from that of a Bayesian agent. Shmaya and Yariv (2011) consider a model in which agents form beliefs on a state space after observing a sequence of signals, in a setup in which the link between the states and the signals is unknown. They provide conditions on how beliefs evolve with information that guarantee that there exists a connection between signals and states that renders these beliefs compatible with Bayes' rule. That is, they provide conditions that guarantee that a general behavior could be considered Bayesian; by contrast, we focus on a behavior which is in clear violation of Bayes' rule, and we provide an alternative representation for it.

Both the *lexicographic beliefs* of Blume, Brandenburger, and Dekel (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, seems to extend to non-Bayesian reactions to non-null events, one of the goals of this article; furthermore, they are based on preferences that depart from standard expected utility maximization also for the static case—for example in Blume, Brandenburger, and Dekel (1991) preferences are not

[5] In addition, Kochov (2009) studies non-Bayesian reactions of agents who fail to properly account for events that will take place in the nonimmediate future, a behavior very different from the one we are interested in.

fully continuous (Archimedean). By contrast, we study an agent that is completely standard in the static case, but might have a nonstandard 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 as 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 our model, their agents need not follow Bayes' rule if a 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 great deal of attention is devoted to the issue of updating ambiguous beliefs, and since the Bayesian postulates seem problematic in that framework, many generalizations of them have been proposed.[6] 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 neutral). By contrast, our main goal is to study violations of Bayes' rule even in this standard case.[7]

II. Setup and Foundations

A. *Formal Setup*

We adopt a standard setup used to study the role of information on preferences. We have a finite (nonempty) set Ω of states of the world, and a (nonempty) set X of *consequences*, which is the set of "objects" that the decision maker could receive when one of the states of the world is realized. Following the literature, we assume that X is a convex subset of a vector space. For example, X could be a set of possible dollar amounts that the agent could receive (e.g., $X = \mathbb{R}_+$); or X could be the set of all the lotteries on a set of prizes Y (this would correspond to the classic setting of Anscombe and Aumann (1963)). We denote by Σ the set of subsets of Ω, which we refer to as *events*, and we denote by $\Delta(\Omega)$ the set of all beliefs (probability measures) on Ω.[8] Correspondingly, denote by $\Delta(\Delta(\Omega))$ the set of all beliefs over beliefs, i.e., the set of all (Borel) probability measures over $\Delta(\Omega)$.

We study the preferences of an agent over functions from Ω to X, which are referred to as *acts*. The typical interpretation is that an act is a "bet" which returns, in every state of the world, a specific consequence in X, e.g., an amount of money.

[6] See, among others, Gilboa and Schmeidler (1993); Epstein and Le Breton (1993); Epstein and Schneider (2003); Maccheroni, Marinacci, and Rustichini (2006); Siniscalchi (2011); Hanany and Klibanoff (2007); Ghirardato, Maccheroni, and Marinacci (2008); Hanany and Klibanoff (2009); and Epstein and Schneider (2007). The latter discusses a form of hypothesis testing for updating multiple priors.

[7] It turns out that our results extend also to the case of ambiguity aversion. The short note Ortoleva (2010a) shows that a very similar representation could be obtained also when agents are ambiguity averse à la Gilboa and Schmeidler (1989) by means of the same axiom that we use here, Dynamic Coherence.

[8] The analysis could be trivially extended to the case in which Σ is a generic algebra on Ω. Moreover, our analysis can also be generalized to an infinite state space in a rather straightforward manner: see the short note Ortoleva (2010b) for more.

Denote by $\mathcal{F}$ the set of all acts, i.e., $\mathcal{F} := X^{\Omega}$. Our analysis studies the preferences of an agent over the acts in $\mathcal{F}$ *conditional on the information* that she receives about the state of the world: we take as a primitive a class of (complete, nondegenerate) preference relations $\{\succeq_A\}_{A\in\Sigma}$ over $\mathcal{F}$.[9] For every event $A \in \Sigma$, by $\succeq_A$ we understand the preferences of the agent after she receives the information $A \in \Sigma$, i.e., after she is told that the true state lies in A. Correspondingly, $\succeq_\Omega$ denotes the preferences of the agent when she receives no information, or before she receives it. To avoid confusion, we should emphasize that what we study is not a proper dynamic environment, since time plays no role in our analysis. Rather, it is a model of how agents condition their preferences on the information that they receive.

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

Finally, we call *null* an event that the decision maker disregards when forming her preferences: we identify it when we observe that the agent is indifferent between any two acts that coincide outside one event, independently of what these acts return if this event were to occur. When the agent is a standard expected utility maximizer, as will be the case here, null events are the ones that have probability *zero*.

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

B. *Axiomatic Foundations*

We start by imposing a few standard axioms.

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

(*i*) (Continuity): *the sets $\{\alpha \in [0,1] : \alpha f + (1-\alpha) g \succeq_A h\}$ and $\{\alpha \in [0,1] : h \succeq_A \alpha f + (1-a) g\}$ are closed.*

(*ii*) (Independence): *for any $\alpha \in (0,1)$*

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

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

(*iv*) (Constant Preference Invariance): *for any $x, y \in X$, $x \succeq_A y \Leftrightarrow x \succeq_\Omega y$.*

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

[9] We say that a preference relation $\succeq$ on $\mathcal{F}$ is *complete* if either $f \succeq g$ or $g \succeq f$ (or both) for all $f, g \in \mathcal{F}$; it is *nondegenerate* if there exists some $f, g \in \mathcal{F}$ such that $f \succeq g$ and we don't have $g \succeq f$. Moreover, for any preference relation $\succeq$ on $\mathcal{F}$, we denote by $\sim$ and $\succ$ its symmetric and asymmetric components.

Axiom WbP (Axiom 1) is a collection of classical postulates. Parts (i), (ii), and (iii) are the assumptions of continuity, independence, and monotonicity, which are known to be equivalent to requiring that our agent is a standard expected utility maximizer (see Anscombe and Aumann 1963). In turn, part (iv) guarantees that the arrival of new information does not affect the agent's ranking over the consequences in X. This follows the rationale that the arrival of new information should affect the agent's beliefs over the states of the world, but not her utility over the consequences that she could receive—for example, it should not affect her risk aversion.

Consequentialism (Axiom 2) is another standard axiom that guarantees that the agent "believes" in the information that 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.

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

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

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

The 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. In particular, if the agent prefers what f returns in A to what g returns in A before she receives new information ($fAg \succeq_\Omega g$), then if she is told that the true state lies in fact in A, she should prefer f to g ($f \succeq_A g$). (We refer to Ghirardato 2002 for an in-depth discussion of this postulate and its implications.)

While normatively appealing, however, our analysis will depart from Dynamic Consistency for two reasons. First of all, Dynamic Consistency disciplines the agent's preferences only for events that are not $\succeq_\Omega$-null, that is, only for events that are assigned a *strictly positive* probability by the agent before the arrival of new information. 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 assigned probability zero. Second, it is well known that adding Dynamic Consistency to WbP and Consequentialism implies that the agent follows Bayes' rule. (See Ghirardato 2002). And since we wish to allow for non-Bayesian behavior, then we would need to relax Dynamic Consistency. In other words, for our purposes, on the one hand Dynamic Consistency is too strong, since it implies Bayes' rule, while on the other hand it is not strong enough, since it does not restrict the agent's behavior in the case of null events.

We replace Dynamic Consistency with an axiom that imposes that our agent has a coherent reaction to different pieces of information, even if they were assigned probability zero. We'll present the idea behind this behavioral postulate in two steps. (These steps are simply a more formal treatment of the examples of the investor in the introduction.) 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 and A_2 are non-null, this is clearly implied by Bayes' rule. But this postulate is 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 "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 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 in the first. Just as before, we would like to say that the agent finds the same informational content in each of the A_is, since they induce beliefs that are "circular." But if the informational content is the same, then her beliefs, and her preferences, should be the same after each of these pieces of information: we should have $\succeq_{A_1} = \succeq_{A_2} = \succeq_{A_3}$.[10] This is what we posit with the axiom below, Dynamic Coherence. (The axiom will be written for cycles of any finite length. Moreover, it will be expressed in terms of the primitives: to say that the agent is sure that A_2 has happened after she is told A_1, we write that $(\Omega \backslash A_2)$ is $\succeq_{A_1}$-null.)

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

It is not hard to see that any Bayesian agent satisfies Dynamic Coherence as long as all A_is are non-null. But of course Bayes' rule is stronger than simply postulating this coherence—it entails a much stronger form of consistency. At the same time, Dynamic Coherence applies also when some of the A_is are null: the intuition behind this axiom is independent of the beliefs that the agent had before the arrival of new information. In fact, there is a conceptual difference between this axiom and Dynamic Consistency. The latter regulates the agent's behavior by comparing the preferences before and after the arrival of new information: but this implies that it cannot restrict the agent's reaction to null events, since the preferences before information are "flat" in that case (by definition of null events). By contrast, Dynamic Coherence imposes a form of consistency of beliefs *across* different events, without going through the belief before the arrival of new information. This means that it can be imposed for *all* possible pieces of information, regardless of whether the original belief assigned probability zero to them or not.

Correspondingly, Dynamic Coherence is neither stronger nor weaker than Dynamic Consistency: while it is does allow for 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, in the representation Theorem in Section IIIB we analyze also the case in which both hold true.)

[10] If this were not the case, i.e., if $\succeq_{A_1} \neq \succeq_{A_2} \neq \succeq_{A_3}$, then the agent's beliefs would react to information in a *circular* manner: beliefs keep changing with these events, but they form a loop, since at the end the agent is sure that the true state lies in the *first* event. The axiom below is meant to rule out this possibility.

III. The Hypothesis Testing Model

A. *The Model*

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

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

for all $B \in \Sigma$. Next, we define the Bayesian update of a prior over priors. Abusing notation, for any $\rho \in \Delta(\Delta(\Omega))$ with finite support, and for any $A \in \Sigma$ such that $\pi(A) > 0$ for some $\pi \in \mathrm{supp}(\rho)$, define

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

for all $\pi \in \Delta(\Omega)$. Both definitions are standard. Notice that $\mathrm{BU}(\rho, A)$ is *not* defined when $\pi(A) = 0$ for all π in the support of ρ.

DEFINITION 2: *A class of preferences relations* $\{\succeq_A\}_{A \in \Sigma}$ *admits a* ***Hypothesis Testing Representation*** *if there exists a nonconstant affine function* $u : X \to \mathbb{R}$, *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*:

(*i*) *for any* $f, g \in \mathcal{F}$

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

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

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

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

In a Hypothesis Testing representation the agent has a utility function u, a prior over priors ρ, and a threshold ϵ. In every period and after any information she maximizes her expected utility using the utility u and a belief π_A (part i), which is determined as follows. Before any new information is revealed, our agent chooses the prior π_Ω to which the prior over priors ρ assigns the highest likelihood, which is unique (part ii). 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 iii). If the prior she was using assigned to the realized event A 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 a piece of 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 π_Ω "fails the test," and our agent revises it. It is as if she decided to reconsider whether she was using the wrong prior to begin with, given that it did a poor job in forecasting what happened. 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 the latter were a confidence level. How is the new prior chosen? The agent updates her prior over priors using Bayes' rule (for second-order priors), and then she chooses the prior to which the *updated* prior over priors assigns the highest likelihood. That is, if the "theory" she was 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 over theories. (Of course, she might choose the same prior she was using before: this happens when, even after reconsidering, the old prior is still the best in the lot.)

The decision rule that we have just described in words, however, might leave some room for indeterminacy: what happens if there is more than one prior to which the updated prior over priors assigns the highest likelihood? Which one is chosen? To avoid this indeterminacy, in a Hypothesis Testing model ρ is constructed in such a way that the arg max of the updated prior over priors is always *unique* (see part iii of the representation). Alternatively, we could have had a representation in which this uniqueness is not required, but in which the agent is endowed not only with a prior over priors ρ, but also with a linear order over priors $\succcurlyeq$. This linear order $\succcurlyeq$ would then be used by the agent to choose in case there were multiple maximizers of the updated ρ. This representation would be equivalent to a Hypothesis Testing representation, which suggests that the only role of the uniqueness requirement above is to guarantee that the behavior is well defined from the representation.[11]

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

Just as Dynamic Coherence is neither stronger nor weaker than Dynamic Consistency, a Hypothesis Testing representation is neither more general nor more restrictive than the standard Bayesian model. To see why, notice that when $\epsilon = 0$ the model coincides with Bayes' rule whenever it is defined, but the agent's beliefs

[11] We should emphasize that requiring the uniqueness of the arg max, or adding a linear order to choose between multiple ones, is more than a technical detail: rather, it is an essential condition for a Hypothesis Testing model to have an empirical content. To see why, consider any agent, no matter how her beliefs are formed, and construct ρ as the uniform distribution over all possible priors. Then, notice that after any information A, the prior that the agent uses after A must belong to the arg max of the updated ρ, since it must give probability 1 to A: this means that we can represent any behavior with a model similar to a Hypothesis Testing representation without the requirement that the arg max is unique (and setting $\epsilon = 1$), or without a fixed rule on how to choose between multiple ones.

are disciplined also when Bayes' rule does not apply (null events). (The two models clearly coincide if $\epsilon = 0$ and there are no null events.) On the other hand, when $\epsilon > 0$ the Hypothesis Testing model allows non-Bayesian reactions to non-null events, generalizing the Bayesian approach. Notice that there are no constraints on ϵ to be "small": we allow agents to review their prior also when they receive a piece of information to which they assigned a relatively large probability. This will be the case for those agents that are more prone to reviewing their beliefs. (As we shall see in Section IIIC, those will also be the agents who are more prone to violate Bayes' rule and Dynamic Consistency.)

Finally, we should emphasize one last feature of the Hypothesis Testing model. When a standard Bayesian agent has a prior over priors, she uses as a belief its expectation, and therefore considers *all* the priors in the support of the prior over priors, weighting them appropriately. By contrast, in a Hypothesis Testing model the agent uses only the prior to which the prior over priors assigns the highest likelihood—the other priors in the support are not used unless some unexpected news is revealed. This aspect of the representation is seemingly irrational, because by restricting her attention to a specific prior the agent could end up with inaccurate predictions. At the same time, however, this could be exactly the behavior followed by a boundedly rational agent who has some "cost of considering models." For example, consider again the case of an investor who is trying to forecast the future return of a certain stock. To form her belief she could indeed use *all* possible existing models in economics and finance, weighting them appropriately. But considering so many models is likely to be very costly for her. Instead, she could choose to focus on a particular one, the one that she believes is the most likely to be correct; she would then use this model until it is "rejected" by the data, at which point she will look for a new one—which is exactly the behavior described in a Hypothesis Testing representation.

B. *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 the next section), we focus on the representations with the smallest value of ϵ.

DEFINITION 3: *A 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 a 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* $\{\succeq_A\}_{A \in \Sigma}$ *also satisfies Dynamic Consistency.*

Theorem 1 provides a foundation for the Hypothesis Testing model by showing its equivalence with Dynamic Coherence, together with WbP and Consequentialism. The second part of the theorem shows that, if we further add Dynamic Consistency,

we obtain a Hypothesis Testing model with $\epsilon = 0$. This provides a foundation for the special case of the Bayesian model in which the agent follows Bayes' rule whenever it applies, but "rationally" reconsiders which prior to use whenever she faces zero-probability events.

C. *Uniqueness*

We now turn to discuss the uniqueness properties of a 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, although this is thanks to our focus on minimal representations: in general, there might be a continuum of values of ϵ that would work.[12] At the same time, we could give a "behavioral" foundation for the value of ϵ: in any minimal Hypothesis Testing representation, the value of ϵ must coincide with the probability assigned to the most likely event after which the agent violates Dynamic Consistency.[13] In turn, this implies that if two agents have the same initial belief, but the first agent violates Dynamic Consistency not only when the second one does, but more often, then in any minimal Hypothesis Testing representation the first agent should have a *higher* threshold ϵ. In this sense, once we fix the initial belief, we can consider ϵ a measure of "how much" the agent violates Dynamic Consistency.

Finally, in general the prior over priors ρ might not be unique, and even its support might not be unique. (To avoid confusion we address the elements of the support of ρ as "models.") To see why, notice first that 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. Moreover, 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. This also means that there might be multiple ways to combine these models, making not unique also the cardinality of the support of ρ.

In order to obtain a form of uniqueness we need to restrict our attention to "focused" representations, i.e., representations in which the support of ρ contains only the beliefs that are *actually used* by the agent. More precisely, for every Hypothesis Testing representation (u, ρ, ϵ), define $\pi_\rho := \underset{\pi \in \Delta(\Omega)}{\arg\max}\ \rho(\pi)$ and

$$\Pi_{\rho,\epsilon} := \big\{\pi \in \Delta(\Omega) \ : \ \exists E \in \Sigma \ \text{ s.t. } \ \pi_\rho(E) \ \leq \ \epsilon \quad \text{and}$$

$$\pi \ = \ \mathrm{BU}\big(\underset{\pi \in \Delta(\Omega)}{\arg\max}\ \mathrm{BU}\big(\rho, E\big)(\pi), E\big)\big\}.$$

[12] To see why, call a the likelihood of the *most* likely event that triggers a violation of Dynamic Consistency, and call b the likelihood of the *least* likely event that has a likelihood above a. We must have $a < b$, and any $\epsilon \in [a, b)$ would work. That is, our space of events might not be "dense" in this sense, and, therefore, the threshold ϵ is not uniquely identified.

[13] To see why, call A the most likely event after which the agent violates Dynamic Consistency, and notice that we cannot have $\epsilon < \pi_\Omega(A)$, otherwise the representation would prescribe that the agent follows Bayes' rule after A, therefore satisfying Dynamic Consistency. Conversely, ϵ wouldn't need to be larger, since for the events with a likelihood above $\pi_\Omega(A)$ the agent does satisfy Dynamic Consistency and, hence, Bayes' rule.

It is easy to see that π_ρ is the belief used by the agent before new information arrives, while $\Pi_{\rho,\epsilon}$ is the set of all beliefs that are used after the arrival of some information A such that $\pi_\rho(A) \leq \epsilon$. We then define as "focused" a representation in which ρ contains only these beliefs in its support—only the beliefs that the agent could end up using (for some information).

DEFINITION 4 (Focused Hypothesis Testing Representation (FHT)): *A Hypothesis Testing Representation* (u, ρ, ϵ) *is* ***focused*** *if* $\text{supp}(\rho) := \Pi_{\rho,\epsilon} \cup \{\pi_\rho\}$.

It turns out that looking at focused representations is without loss of generality, as the following corollary shows.

COROLLARY 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 Focused Hypothesis Testing representation* (*MFHT*) (u, ρ, ϵ). *Moreover*, $\epsilon = 0$ *if and only if* $\{\succeq_A\}_{A\in\Sigma}$ *satisfies also Dynamic Consistency.*

By restricting our attention to focused representation we obtain the following uniqueness result.

PROPOSITION 2: *For every class of preference relations* $\{\succeq_A\}_{A\in\Sigma}$ *there exists a strict partial order* $\rhd$ *over* $\Delta(\Omega)$ *such that if* (u, ϵ, ρ) *is a MFHT of* $\{\succeq_A\}_{A\in\Sigma}$, *then for every MFHT representation* (u', ϵ', ρ') *the following are equivalent:*

(*i*) (u', ϵ', ρ') *is a MFHT of* $\{\succeq_A\}_{A\in\Sigma}$;

(*ii*) (a) u' *is a positive affine transformation of* u;

(b) $\epsilon' = \epsilon$;

(c) $\text{supp}(\rho) = \text{supp}(\rho')$;

(d) $\pi \rhd \pi' \Rightarrow (\rho(\pi) > \rho(\pi')$ and $\rho'(\pi) > \rho'(\pi'))$.

Moreover, if $|\text{supp}(\rho)| > 1$, *then* $\rhd$ *is not empty*, *and for every* $\pi \in \text{supp}(\rho)$ *there exists some* $\pi' \in \text{supp}(\rho)$ *such that either* $\pi \rhd \pi'$ *or* $\pi' \rhd \pi$.

While part (a) is standard, and parts (b) and (c) derive, respectively, from requiring a minimal and a focused representation, the content of Proposition 2 lies in the connection between i) and part (d). In fact, the proposition shows that there exists a partial order $\rhd$ over priors such that the prior over priors ρ of *any* MFHT representation must agree with this partial order: this implies that the ρs of all possible MFHT representations must be ordinally similar. Moreover, Proposition 2 shows that the converse is true as well: if ρ' has the same support as ρ and agrees with the partial order $\rhd$, *then* it is an MFHT representation of the same preferences. That is, respecting the ranking of $\rhd$ (together with conditions (a)–(c)) is not only necessary, but also sufficient, for a prior over priors to be part of a representation of the same preferences.

This latter point suggests that the strict partial order $\rhd$ is, in fact, what is uniquely identified in this representation. The last part of Proposition 2 then shows that, while $\rhd$ need not be complete, it is at least "partially" complete: excluding the trivial case in which the support of ρ is a singleton, we know that $\rhd$ is not empty, and also that every prior in the support of ρ is compared by $\rhd$ to at least one other prior in the support. (That is, no prior in the support of ρ is left "uncompared.")

IV. A (Simple) Application to Equilibrium Refinements

In this section we briefly discuss an application of the Hypothesis Testing model to game theory. Focusing on a simple example, we develop an alternative to the Perfect Bayesian Nash Equilibrium (PBNE) that is built on the Hypothesis Testing model instead of Bayes' rule. This alternative notion exploits the fact that the former is also defined for zero-probability events, while the latter is not, allowing us to restrict also the beliefs off the equilibrium path.[14] In the example that we consider, we obtain a refinement which coincides with the Intuitive Criterion of Cho and Kreps (1987). As opposed to it, however, off-equilibrium behavior is not interpreted in light of the equilibrium that is currently played. Rather it leads players to reevaluate their beliefs on the strategy followed by the other players.

Since our goal at present is simply to illustrate one possible application of the model, we focus on a specific game: the well-known Beer/Quiche game of Cho and Kreps (1987), depicted in Figure 1.[15] There are two players, 1 and 2, who are in a bar. Player 1 can be of two types, *Wimp* or *Surly*, chosen by Nature, who picks *Surly* with probability 0.9. Player 1 moves first, and decides whether to have *Beer* or *Quiche* for breakfast, observed by Player 2. While the *Surly* type would rather have *Beer*, the *Wimp* type prefers *Quiche*. After observing Player 1's choice, Player 2 decides whether to *Duel* or not. If she does, Player 1's utility is reduced no matter what her type is, while Player 2's utility depends on Player 1's type: she'd rather *Duel* only if Player 1 is *Wimp*. Moreover, she would rather *Not Duel* if her belief is identical to the prior belief (the probability of *Surly* being 0.9). We denote by μ_2 the beliefs of Player 2—for example $\mu_2(Surly \mid Beer)$ denotes the belief that Player 2 assigns to Player 1 being *Surly* after observing *Beer*.

It is easy to see that this game has two PBNE in pure strategies. In the first one, Player 1 plays *Beer* no matter what her type is, and Player 2 does *Not Duel* if she sees *Beer* and *Duel*s if she sees *Quiche*. This is supported by a belief such that $\mu_2(Surly \mid Beer) = 0.9$ and $\mu_2(Surly \mid Quiche) \leq 0.5$. In the second equilibrium, Player 1 plays *Quiche* no matter what her type is, and Player 2 does *Not Duel* if she sees *Quiche* and *Duel*s if she sees *Beer*. This is supported by a belief such that $\mu_2(Surly \mid Quiche) = 0.9$ and $\mu_2(Surly \mid Beer) \leq 0.5$. As pointed out in Cho and Kreps (1987), however, this second equilibrium can be seen as problematic: a belief such that $\mu_2(Surly \mid Beer) < 0.9$ seems incompatible with the observation that only

[14] In this application we use only one of the two features of the Hypothesis Testing model: the fact that it is defined after zero-probability events. We do not use the other feature of the model, non-Bayesian reactions to unexpected news: we will impose $\epsilon = 0$.

[15] An equivalent notion could naturally be constructed for any game. For instance, we will later argue that similar results would be obtained in the signaling game of Spence (1973).

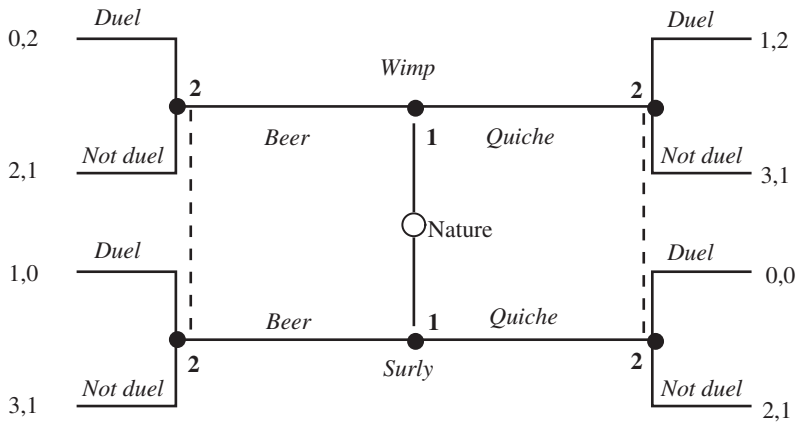

FIGURE 1. THE BEER/QUICHE GAME

the *Surly* type is better off with *Beer* instead of *Quiche*. But since the PBNE imposes no restrictions on off-equilibrium beliefs, this equilibrium is not ruled out.

We now construct an alternative notion of equilibrium based on the Hypothesis Testing model instead of Bayes' rule. The idea of this alternative notion is that Player 2 forms her beliefs over the appropriate state space, $\{Surly, Wimp\} \times \{Beer, Quiche\}$, using a Hypothesis Testing model with a prior over priors ρ and a threshold $\epsilon = 0$.[16] Moreover, we impose that ρ satisfies two additional requirements. First, the priors in the support of ρ must be compatible with the initial information of the game: the unconditional probability of *Surly* must be 0.9. Second, they must assign zero-weight to actions that are not best responses to at least *some* strategy of Player 2. That is, we admit only priors that can be, in weak sense, "rationalized" by some possible strategy of Player 2. Formally, we say that a prior over priors ρ is *consistent* if for all π in the support of ρ we have: (i) $\pi(Surly \times \{Beer, Quiche\}) = 0.9$; and (ii) there exists some strategy s_2 of Player 2 such that, for all $s_1 \in \{Beer, Quiche\}$ and $\theta_1 \in \{Surly, Wimp\}$, $\pi(\{s_1\} \times \{\theta_1\}) = 0$ if s_1 is not a best response to s_2 for type θ_1.

We can now define our notion of equilibrium. A strategy profile σ and a belief system μ_2 over $\{Surly, Wimp\}$ is a Hypothesis-Testing Equilibrium (HTE) if: (i) the strategies are sequentially rational given the beliefs—just as in a PBNE; (ii) there exists a Hypothesis Testing model $(\rho, 0)$ for Player 2 such that ρ is consistent, and it induces beliefs over $\{Surly, Wimp\}$ that coincide with $\mu_2(\cdot \mid s_1)$ for all $s_1 \in S_1$. That is, we require that the belief system could be derived from *some* consistent Hypothesis Testing model for Player 2, instead of Bayes' rule as in the PBNE.

We now look at the HTE of our game (in pure strategies). Notice first that any HTE must also be a PBNE by construction, which means that we have only two candidates. It is also easy to see that the PBNE in which Player 1 plays *Beer* must also be an HTE. In fact, a belief such that $\mu_2(Surly \mid Beer) = 0.9$ and $\mu_2(Surly \mid Quiche) = 0$ can be derived by a Hypothesis Testing model in which the support of ρ contains

[16] We again emphasize that by setting $\epsilon = 0$ we do not allow for non-Bayesian reactions for unexpected news: our agent is a standard Bayesian agent whenever Bayes' rule applies, but her behavior is defined also after zero-probability events.

only π and π' such that: $\rho(\pi) > \rho(\pi')$, π assigns weight zero to *Quiche* being chosen by any type, while π' assigns probability zero to *Surly* choosing *Quiche* and *Wimp* choosing *Beer*.[17] In the Hypothesis Testing model, the player starts with belief π and keeps it if she observes *Beer* (on the equilibrium path), while she moves to belief π' if she observes *Quiche* (off the equilibrium path).

On the other hand, the PBNE in which Player 1 plays *Quiche* is not an HTE. To see why, notice that if it were, then there would exist a consistent Hypothesis Testing model that selected, after the event *Beer*, a prior $\hat{\pi}$ such that $\hat{\pi}(Surly \mid Beer) \leq 0.5$. This, however, implies $\hat{\pi}(\{Wimp\} \times \{Beer\}) > 0$, which in turn implies that any s_2 of Player 2 that "rationalizes" $\hat{\pi}$ must be such that *Beer* is a best response for the *Wimp* type. But if *Beer* is a best response for the *Wimp* type, then it is the *unique* best response for the *Surly* one. This means $\hat{\pi}(\{Surly\} \times \{Quiche\}) = 0$, hence $\hat{\pi}(Surly \mid Beer) \geq 0.9$, contradicting $\hat{\pi}(Surly \mid Beer) \leq 0.5$. Therefore, this is not an HTE.

Almost identical arguments apply to the well-known signaling game of Spence (1973), in which workers use schooling to signal their ability. It is not hard to see that, of the many separating PBNE of this game, only the "efficient" separating one is an HTE.[18]

In this example, the HTE coincides with the selection of the Intuitive Criterion of Cho and Kreps (1987). There is, however, an important difference between the two approaches. The Intuitive Criterion evaluates deviations *in light of the specific equilibrium that is being played*: disequilibrium behavior is interpreted as a signal, the meaning of which depends on the starting point.[19] By contrast, in an HTE the players' reaction to a deviation from equilibrium does not depend on the current equilibrium. Rather, the current equilibrium is discarded, and the agent reevaluates what to expect under the restriction that the other player must be at least minimally rational—she must be best responding to some belief on what the first player is playing. The idea of questioning one's original belief, and "rationally" choosing a new one, is in fact one of the essential ideas of the Hypothesis Testing model.

V. Conclusion

In this article we discuss, and justify axiomatically, a novel updating rule according to which the agent follows Bayes' rule when she receives information that is not "unexpected," i.e., to which she assigned a probability above a threshold. Otherwise, instead of following Bayes' rule, our agent reconsiders which prior to use by updating a prior over priors and selecting the prior with the highest likelihood after the update. The main result of the paper shows that the existence of this representation is equivalent to a novel behavioral axiom, Dynamic Coherence, together with other

[17] Notice also that π is compatible with Player 1 best responding to Player 2 dueling only if she sees *Quiche* (the equilibrium strategy), while π' is compatible with Player 1 best responding to Player 2 never dueling.

[18] Recall that, in the model in Spence (1973), any PBNE in which the high ability worker puts an effort above the minimal separating one (e^*) is supported by an (off-equilibrium) belief of the firm that assigns a strictly positive probability to the worker being of low ability after observing e^*. However, this belief cannot be generated in any HTE, since choosing any effort above e^* is a *dominated* strategy for a low ability worker.

[19] Cho and Kreps (1987) point out that this feature might not always be appealing. See Mailath, Okuno-Fujiwara, and Postlewaite (1993) for more discussion.

standard postulates. We also characterize the case in which the agent follows Bayes' rule whenever it is defined, while she "rationally" chooses a new prior when it is not.

We conclude by mentioning possible applications and extensions. Among the former, besides the one discussed in Section IV, the Hypothesis Testing model could be applied to study portfolio choice: investors would have a standard behavior in the "business as usual" situations but might strongly change their beliefs in case of unexpected news. This could generate behaviors in line with some of the empirical evidence on overreactions to unexpected news in financial markets, like "surprise" in earning announcements or financial crisis.[20]

Natural extensions of the Hypothesis Testing model examine the case in which the state space is infinite, and the one in which the agent is not an expected utility maximizer but is ambiguity averse. Very similar representations could be found in both cases by means of the same axiom, Dynamic Coherence: see Ortoleva (2010b) for the extension to an infinite state space, and Ortoleva (2010a) for the corresponding one with ambiguity averse agents.

Appendix

PROOF OF THEOREM 1 AND COROLLARY 1:

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

$$\text{(A1)} \qquad f \succeq_A g \Leftrightarrow \sum_{\omega \in \Omega} \pi_A(\omega)\, u_A\big(f(\omega)\big) \geq \sum_{\omega \in \Omega} \pi_A(\omega)\, u_A\big(g(\omega)\big),$$

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.(iv) implies that, for any $A \in \Sigma$, all u_A are positive affine transformations of u_Ω, which means that we can assume $u_\Omega = u_A$ for all $A \in \Sigma$. Define $u : X \to \mathbb{R}$ as $u = u_\Omega$ and $\pi = \pi_\Omega$.

CLAIM 1: *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 $\pi_A(B) = 1 = \pi_B(A)$. By construction of π_A and π_B we must then have $\pi_A(A \cap B) = 1 = \pi_B(A \cap B)$. In turns, this means that $(\Omega \backslash (A \cap B))$ is $\succeq_A$-null and $\succeq_B$-null. At the same time, by Consequentialism (Axiom 2) we have that both $(\Omega \backslash A)$ and $(\Omega \backslash B)$ are $\succeq_{A \cap B}$-null. Dynamic Coherence (Axiom 4) then implies $\succeq_A = \succeq_{A \cap B}$, and $\succeq_B = \succeq_{A \cap B}$. This means $\succeq_A = \succeq_B$ and $\pi_A = \pi_B$, as sought.

[20] The arrival of a financial crisis might induce not only a "change of paradigm," but also, at the same time, a sharp increase in ambiguity. (See Caballero and Krishnamurthy 2009; Caballero and Simsek 2009; Guidolin and Rinaldi 2010; and Routledge and Zin 2009.) Both features should then be included in models that study these phenomena.

Define now $\mathcal{K}$ as the set of events after which the agent violates Dynamic Consistency, i.e., $\mathcal{K}_{\succeq} := \{A \in \Sigma : \exists f, g \in \mathcal{F}$ such that $f \succeq_A g$ and $g \succ fAg$, or $f \succ_A g$ and $g \succeq fAg\}$, and define 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 $\pi(A) = 1$. By Claim 1 this implies $\succeq_A = \succeq_\Omega$, hence, $A \notin \mathcal{K}$. This means that $\max\{\pi(B) : B \in \mathcal{K}\} < 1$, hence, $\epsilon \in [0,1)$. Define now the set $\overline{\mathcal{K}} := \mathcal{K} \cup \{A \in \Sigma : A$ is $\succeq$-null$\}$, and notice that, by construction of ϵ, if $\pi(A) > \epsilon$ then $A \in \Sigma \backslash \overline{\mathcal{K}}$.

We now turn to prove that for any $f, g, h \in \mathcal{F}, A \in \Sigma \backslash \overline{\mathcal{K}}$ we have $f \succeq_A g$ if, and only if, $fAh \succeq gAh$. To see why, consider $f, g, h \in \mathcal{F}$, and $A \in \Sigma \backslash \overline{\mathcal{K}}$. By construction of $\overline{\mathcal{K}}$, for any $r, s \in \mathcal{F}$ we have $rAs \succeq s$ if and only if $r \succeq_A s$. Notice now that by Axiom 2 we have $fAh \sim_A f$ and $gAh \sim_A g$. This implies $f \succeq_A g$ if and only if $fAh \succeq_A gAh$. Define $f' := fAh$ and $g' := gAh$. Notice that we have $fAh \succeq_A gAh$ if and only if $f' \succeq_A g'$ if and only if $f'Ag' \succeq g'$ if and only if $fAh \succeq gAh$ (where the last passages use the fact that $\pi(A) > \epsilon$).

We now prove that $\pi_A(B) = \pi(A \cup B)/\pi(A) = \mathrm{BU}(\pi, A)$ for any $A \in \Sigma \backslash \overline{\mathcal{K}}$. To see why, notice that for any $A \in \Sigma \backslash \overline{\mathcal{K}}$ we must have $\pi(A) > 0$, because any A which is $\succeq$-null belongs to $\overline{\mathcal{K}}$. We then have that for any $f, g \in \mathcal{F}$

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

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

Define now the set $\mathcal{K}^* := \{A \in \Sigma : \pi_\Omega(B) \geq \pi_\Omega(A)$ for some $B \in \overline{\mathcal{K}}\}$ and 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 exists an event in $\mathcal{K}$ that is more likely than some of them. Define now the sets $H := \mathcal{K}^* \cup \{\Omega\}$ and $M := \{\pi_C : C \in \mathcal{K}^*\} \cup \{\pi_\Omega\}$.

CLAIM 2: *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 1 implies $\pi_A = \pi_D$, which means $\pi_A = \pi_C$. But by construction of π_C this contradicts $\pi_A(C) < 1$.

Define now the binary relation $\rhd$ on M as

$$\pi_C \rhd \pi_{C'} \Leftrightarrow \pi_{C'}(C) = 1 \quad \text{and} \quad \pi_C(C') < 1.$$

Notice that $\rhd$ is well defined by Claim 2, and it is also irreflexive for the same reason.

CLAIM 3: *Consider* $\pi_1, \ldots, \pi_n \in M$ *such that* $\pi_1 \rhd \cdots \rhd \pi_n$. *Then, there exist* $A_1, \ldots, A_n \in \mathcal{K}^*$ *such that* $\pi_i = \pi_{A_i}$, $\pi_{A_{i+1}}(A_i) = 1$, *and* $\pi_{A_i}(A_{i+1}) < 1$ *for* $i = 1, \ldots, (n-1)$.

PROOF:

For simplicity we 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 $\pi_1 \rhd \pi_2 \rhd \pi_3$. By construction of $\rhd$ 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 1 we have $\pi_{A_1} = \pi_{A_2'}$, hence, $\pi_1 = \pi_2$, which contradicts the fact that $\rhd$ is irreflexive.

CLAIM 4: $\rhd$ *is acyclic.*

PROOF:

By means of contradiction consider $\pi_1, \ldots, \pi_n \in M$ such that $\pi_1 \rhd \ldots \pi_n \rhd \pi_1$. By Claim 3 there exist $A_1, \ldots, A_n \in \mathcal{K}^*$ such that $\pi_i = \pi_{A_i}$, $\pi_{A_{i+1}}(A_i) = 1$, $\pi_{A_i}(A_{i+1}) < 1$ for $i = 1, \ldots, (n-1)$, $\pi_{A_1}(A_n) = 1$, and $\pi_{A_n}(A_1) < 1$. Since $\rhd$ is irreflexive notice that we must have that $\pi_{A_i} \neq \pi_{A_{i+1}}$ for $i = 1, \ldots, (n-1)$, and $\pi_{A_1} \neq \pi_{A_n}$. But this contradicts Axiom 4.

Define $\gamma_* := \max\{\pi_A(B) : A, B \in \mathcal{K}^*, \pi_A(B) < 1\}$ if $\{\pi_A(B) : A, B \in \mathcal{K}^*, \pi_A(B) < 1\} \neq 0/$, $\gamma_* = 0$ 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 $\gamma_* \in [0, 1)$ and M is finite, we must have $\delta > 0$). Consider now the transitive closure of $\rhd$ and call it $\hat{\rhd}$. Since $\rhd$ is irreflexive and antisymmetric, so must be $\hat{\rhd}$. To see why, say, by contradiction, that we have $\pi, \pi' \in M$ such that $\pi \hat{\rhd} \pi' \hat{\rhd} \pi$. Then, by definition of transitive closure, there must exist $\pi_1, \ldots, \pi_{n+m} \in M$ such that $\pi \rhd \pi_1 \rhd \cdots \rhd \pi_n \rhd \pi'$ and $\pi' \rhd \pi_{n+1} \rhd \cdots \rhd \pi_{m+n} \rhd \pi$. But this violates the acyclicity of $\rhd$.

Since M is finite, enumerate it and construct the function $f : M \to \mathbb{R}$ as follows. Set $f(\pi_1) = 0$. Consider π_n. Assign to $f(\pi_n)$ any value such that, for all $i < n$: $f(\pi_n) \neq f(\pi_i)$; $f(\pi_n) > f(\pi_i)$ if $\pi_n \hat{\rhd} \pi_i$; and $f(\pi_n) < f(\pi_i)$ if $\pi_i \hat{\rhd} \pi_n$. To see why this is always possible, notice that for all π_n, π_i, π_j, with $n \geq i, j$, if we have $\pi_n \hat{\rhd} \pi_i$ and $\pi_j \hat{\rhd} \pi_n$, then we must also have that $\pi_j \hat{\rhd} \pi_i$ since $\hat{\rhd}$ is transitive, which implies that we must also have $f(\pi_j) > f(\pi_i)$. 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

$\pi \rhd \pi'$ implies $v(\pi) > v(\pi')$, and that $v(\pi) \neq v(\pi')$ for all $\pi, \pi' \in M$. Construct now $\rho \in \Delta(\Delta(\Omega))$ as

$$\rho(\pi) := \frac{v(\pi) + \frac{1}{|M|}}{\sum_{m \in M}\left(v(\pi) + \frac{1}{|M|}\right)}$$

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.

We will first of all show that $\{\pi_\Omega\} = \underset{\pi \in \Delta(\Omega)}{\arg\max}\ \rho(\pi)$. To prove it, 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 turn, implies that $\pi = \pi_C$ for some $C \in H$. Say first that we have that $\pi_\Omega(C) = 1$ for all $C \in H$ such that $\pi_C = \pi$. Then, since we also have that $\pi_C(\Omega) = 1$ for all $C \in H$ such that $\pi_C = \pi$, Claim 1 would imply $\pi = \pi_\Omega$, a contradiction. This means that we must have $\pi_\Omega(C) < 1$ for some $C \in H$ such that $\pi_C = \pi$. Also, notice that we must have $\pi_C(\Omega) = 1$ (by definition), which implies that we have $\pi_\Omega \rhd \pi_C = \pi$. But then we must have $v(\pi_\Omega) > v(\pi)$ by construction of v, which, in turn, implies that we must have $\rho(\pi_\Omega) > \rho(\pi)$, a contradiction.

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 5: *For any $A \in H$, if $\pi \in \arg\max \mathrm{BU}(\rho, A)(\pi)$, then $\pi(A) = 1$.*

PROOF:

Consider $A \in H$ and say by means of contradiction that there exist $\pi \in \arg\max \mathrm{BU}(\rho, A)(\pi)$ such that $\pi(A) < 1$. This means $\pi(A)\rho(\pi) \geq \pi'(A)\rho(\pi')$ for all $\pi' \in M$, and, therefore, also $\pi(A)\rho(\pi) \geq \pi_A(A)\rho(\pi_A)$. Also, since $A \in H$, then by construction we must have $\pi_A(A) = 1$, which implies $\pi(A)\rho(\pi) \geq \rho(\pi_A)$. By construction of ρ this is possible only if $\pi \in M$ and

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

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

$$\left(v(\pi) + \frac{1}{|M|}\right)\gamma_* > \frac{1}{|M|} \Rightarrow v(\pi) > \frac{1}{|M|}\frac{1 - \gamma_*}{\gamma_*}.$$

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

Our final step to prove the existence of the desired representation is to prove that for any $A \in H$ we have $\{\pi_A\} = \arg\max \mathrm{BU}(\rho, A)$. Consider $A \in H$ and say, by means of contradiction, that we have $\pi \in \arg\max \mathrm{BU}(\rho, A)$ for some $\pi \neq \pi_A$. This means that we have $\pi(A)\rho(\pi) \geq \pi_A(A)\rho(\pi_A)$. By Claim 5 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 $B \in H$ such that $\pi_B = \pi$. Now, if we have $\pi_A(B) = 1$, then since $\pi_B(A) = \pi(A) = 1$, by Claim 1 we have $\pi = \pi_A$, a contradiction. Therefore, we must have $\pi_A(B) < 1$. But then, we must have $\pi_A \rhd \pi_B = \pi$, which implies $v(\pi_A) > v(\pi_B)$, hence $\rho(\pi_A) > \rho(\pi)$, a contradiction.

The steps above prove the existence of a Hypothesis Testing Representation (u, ρ, ϵ). Notice that by the definition of ϵ, (u, ρ, ϵ) is a representation both minimal and focused.

[*Necessity of the Axioms*] The proof of the necessity of Axiom 1 (WbP) is standard practice. Axiom 2 is immediate from the representation. We are left with Axiom 4. Say by means of contradiction that there exist $A_1, \ldots, A_n \in \Sigma$ such that $\succeq_{A_1} \neq \succeq_{A_n}$, $(\Omega \backslash A_{i+1})$ is $\succeq_{A_i}$-null for $i = 1, \ldots, (n-1)$, and $(\Omega \backslash A_1)$ is $\succeq_{A_n}$-null. Consider first the case in which $\pi(A_i) > \epsilon$ for $i = 1, \ldots, n$. By the representation, this means that we must have $\pi_{A_i} = \mathrm{BU}(\pi, A_i)$ for $i = 1, \ldots, n$. But since $(\Omega \backslash A_{i+1})$ is $\succeq_{A_i}$-null for $i = 1, \ldots, (n-1)$, hence $\pi_{A_i}(A_i \backslash A_{i+1}) = 0$, then by construction of BU we must have $\pi(A_i \backslash A_{i+1}) = 0$ for $i = 1, \ldots, n$. For the same reason we must also have $\pi(A_n \backslash A_1) = 0$. But then we must have that $\pi(\cap_{i=1}^n A_i) = \pi(\cup_{i=1}^n A_i)$, 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 exists 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 $\pi_{A_1} = \mathrm{BU}(\pi, A_1)$. At the same time we have $\pi_{A_1}(A_1 \backslash A_2) = 0$, 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 $i = 1, \ldots, n$. 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 $i = 1, \ldots, n$. By construction we must have that $\mathrm{BU}(\pi^*_{A_i}, A_i)(A_{i+1}) = 1$ for $i = 1, \ldots, n-1$ and $\mathrm{BU}(\pi^*_{A_n}, A_n)(A_1) = 1$. This implies $\pi^*_{A_i}(A_{i+1}) \geq \pi^*_{A_i}(A_i)$ for $i = 1, \ldots, n-1$ and $\pi^*_{A_n}(A_1) \geq \pi^*_{A_n}(A_n)$. Now, notice that since $\pi^*_{A_i}$ is the unique element in $\arg\max_{\pi \in \Delta(\Omega)} \mathrm{BU}(\rho, A_i)(\pi)$ for $i = 1, \ldots, n$, then we must have that for all $i, j = 1, \ldots, n$, $\rho(\pi^*_{A_i})\pi^*_{A_i}(A_i) > \rho(\pi^*_{A_j})\pi^*_{A_j}(A_i)$ 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 $i = 1, \ldots, n-1$ 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 $i = 1, \ldots, (n-1)$, $\rho(\pi^*_{A_i}) \geq \rho(\pi^*_{A_{i+1}})$, 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 $\pi^*_{A_1} \neq \pi^*_{A_n}$. But then we have $\rho(\pi^*_{A_1}) \geq \rho(\pi^*_{A_2}) \geq \ldots \rho(\pi^*_{A_n}) \geq \rho(\pi^*_{A_1})$, and so none of these inequalities can be strict, hence, $\pi^*_{A_1} = \pi^*_{A_i}$ for $i = 1, \ldots, n$. But then, $\mathrm{BU}(\pi^*_{A_1}, A_i)(A_{i+1}) = 1$ for $i = 1, \ldots, n-1$, and $\mathrm{BU}(\pi^*_{A_1}, A_n)(A_1) = 1$, which means $\pi^*_{A_1}(A_i \backslash A_{i+1}) = 0$ for $i = 1, \ldots, n-1$, and $\pi^*_{A_1}(A_n \backslash A_1) = 0$. This means $\pi^*_{A_1}(\cap_{i=1}^n A_i) = \pi^*_{A_1}(\cup_{i=1}^n A_i)$. But this implies that $\mathrm{BU}(\pi^*_{A_1}, A_n) = \mathrm{BU}(\pi^*_{A_1}, A_1)$, hence $\pi_{A_1} = \pi_{A_n}$ and $\succeq_{A_1} = \succeq_{A_n}$, a contradiction.

[$\epsilon = 0$ *If and Only If 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 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).

PROOF OF PROPOSITION 2:

For every class of preference relations $\{\succeq_A\}_{A\in\Sigma}$, define $M := \{E \in \Sigma : \pi_\rho(E) \leq \epsilon\} \cup \{\Omega\}$, and for all $E \in M$ define $\pi_E = \mathrm{BU}(\arg\max_{\pi\in\Delta(\Omega)} \mathrm{BU}(\rho, E)(\pi), E)\}$, and define $\pi_\Omega = \pi_\rho$: then, define $\overline{\rhd}$ on $\Pi_{\rho,\epsilon} \cup \{\pi_\rho\}$ as

$$\pi_C \;\overline{\rhd}\; \pi_{C'} \Leftrightarrow \pi_{C'}(C) = 1 \quad \text{and} \quad \pi_C(C') < 1.$$

Claims 2 and 4 in the proof of Theorem 1 show that it is well defined, acyclic, irreflexive, and asymmetric.[21] Moreover, notice that, as long as $\mathrm{supp}(\rho)$ is not empty, we must have $\pi_\Omega \overline{\rhd} \pi_E$ for all $\pi_E \in \Pi_{\rho,\epsilon}$, which implies that $\overline{\rhd}$ is not empty and that for every $\pi \in \mathrm{supp}(\rho)$ there exists some $\pi' \in \mathrm{supp}(\rho)$ such that either $\pi \overline{\rhd} \pi'$ or $\pi' \overline{\rhd} \pi$. Now define $\rhd$ as the transitive closure of $\overline{\rhd}$, and notice that it is a strict partial order.

[(i) $\Rightarrow$ (ii)] The only nontrivial part is to show that, for all $\pi, \pi \in \Delta(\Omega)$, if $\pi \rhd \pi'$ then $\rho'(\pi) > \rho'(\pi')$. Notice that we can have $\pi \rhd \pi'$ either if $\pi \overline{\rhd} \pi'$, or if $\pi \overline{\rhd} \pi_1 \overline{\rhd} \ldots \overline{\rhd} \pi_n \overline{\rhd} \pi'$ for some $\pi_1, \ldots, \pi_n$. Consider some $\pi, \pi' \in \mathrm{supp}(\rho)$ such that $\pi \overline{\rhd} \pi'$, and notice that by construction of $\overline{\rhd}$, it means that there exist $E, E' \in M$ such that $\pi = \pi_E$, $\pi' = \pi_{E'}$, and $\pi_\rho(E), \pi_\rho(E') < \epsilon$. Moreover, since $\pi \overline{\rhd} \pi'$, then $\pi_E \overline{\rhd} \pi_{E'}$, hence $\pi_E(E') < 1$ and $\pi_{E'}(E) = 1$. Say, by means of contradiction, that $\rho'(\pi_{E'}) \geq \rho'(\pi_E)$. Then notice that we must have $\rho'(\pi_{E'}) = \pi_{E'}(E)\rho'(\pi_{E'}) > \pi_E(E')\rho'(\pi_E)$, which implies $\mathrm{BU}(\rho', E)(\pi_{E'}) > \mathrm{BU}(\rho', E)(\pi_E)$. But then, notice that, since $\pi_{E'}(E) = 1$, then $\mathrm{BU}(\pi_{E'}, E) = \pi_{E'}$. But since $\pi_E \neq \pi'_E$, this contradicts the fact that (u', ρ', ϵ') represents $\{\succeq_A\}_{A\in\Sigma}$. Consider now the case in which $\pi \rhd \pi'$ and $\pi \overline{\rhd} \pi_1 \overline{\rhd} \ldots \overline{\rhd} \pi_n \overline{\rhd} \pi'$ for some $\pi_1, \ldots, \pi_n$. We have just showed that this implies $\rho'(\pi) > \rho'(\pi_2) > \cdots > \rho'(\pi_n) > \rho'(\pi')$. Hence, $\rho'(\pi) > \rho'(\pi')$.

[(ii) $\Rightarrow$ (i)] The only nontrivial part is to show that for all $E \in M$ we have $\pi_E(E)\rho'(\pi_E) > \pi(E)\rho'(\pi)$ for all $\pi \in \Pi_{\rho,\epsilon} \cup \{\pi_\rho\}$. Now, define $\hat{\succ}$ on M as $A \hat{\succ} E$ if and only if $\rho'(\pi_A)\pi_A(E) > \rho'(\pi_E)$. Notice that $\hat{\succ}$ is antisymmetric. Moreover, it is also acyclic. To see why, notice that if $A_1 \hat{\succ} A_2 \hat{\succ} \ldots \hat{\succ} A_n$, then $\rho'(\pi_{A_1})\pi_{A_1}(A_2) > \rho'(\pi_{A_2})$, and $\rho'(\pi_{A_2})\pi_{A_2}(A_3) > \rho'(\pi_{A_3}), \ldots, \rho'(\pi_{A_{n-1}})\pi_{A_{n-1}}(A_n) > \rho'(\pi_{A_n})$. This means that we must have $\rho'(\pi_{A_1}) > \rho'(\pi_{A_n})$, which, in turn, means that we cannot have $A_n \hat{\succ} A_1$. So $\hat{\succ}$ is acyclic.

Now, since $\mathrm{supp}(\rho) = \mathrm{supp}(\rho')$, then we must have that $\Pi_{\rho,\epsilon} = \Pi_{\rho',\epsilon'}$, since both are focused Hypothesis Testing representations. But this implies that for all $E \in M$, we must have $\pi_E \in \Pi_{\rho',\epsilon'}$. Consider now some $E \in M$ such that $A \hat{\succ} E$ for some $A \in M$ and $B \hat{\succ} E$ for no $B \in M$. We now argue that such E cannot exist. To see why, notice that π_E cannot ever be chosen in the HT representation (u', ρ', ϵ'), because it is not chosen after E (since $A \hat{\succ} E$), and not after any other event B, since there is no $B \in M$ such that $B \hat{\succ} E$. But then, we must have that $\pi_E \notin \Pi_{\rho',\epsilon'}$. At the same time, we argued before that for all $E \in M$ we have $\pi_E \in \Pi_{\rho',\epsilon'}$; therefore, that there does exist some $E \in M$ such that $A \hat{\succ} E$ for some $A \in M$, and $B \hat{\succ} E$ for no $B \in M$. But since $\hat{\succ}$ is acyclic (and Σ is finite), this could happen only if $\hat{\succ}$ is empty. But then, it means that for all $E \in M$ we have $\pi_E(E)\rho'(\pi_E) > \pi_A(E)\rho'(\pi_A)$ for all $A \in M$; hence, $\mathrm{BU}(\rho', E)(\pi_E) > \mathrm{BU}(\rho', E)(\pi)$ for all $\pi \in \mathrm{supp}(\rho')$ as sought.

[21] Notice that $\{\succeq_A\}_{A\in\Sigma}$ must satisfy the axioms in Theorem 1 because it admits an Hypothesis Testing representation (u, ϵ, ρ). Therefore we know that Claim 2 and 4 apply in this case.

REFERENCES

Anscombe, Frank J., and Robert J. Aumann. 1963. "A Definition of Subjective Probability." *The Annals of Mathematical Statistics* 34 (1): 199–205.

Antoniou, Constantinos, Glenn W. Harrison, Morten Lau, and Daniel Read. 2010. "Subjective Bayesian Beliefs." Unpublished.

Asparouhova, Elena, Micheal Hertzel, and Micheal Lemmon. 2009. "Inference from Streaks in Random Outcomes: Experimental Evidence on Beliefs in Regime-Shifting and the Law of Small Numbers." *Management Science* 55 (11): 1766–82.

Barberis, Nicholas, and Richard Thaler. 2003. "A Survey of Behavioral Finance." In *Handbook of the Economics of Finance: Financial Markets and Asset Pricing Volume 1B*, edited by George M. Constantinides, Milton Harris, and Rene Stulz, 1053–123. Amsterdam: Elsevier, North Holland.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics* 49 (3): 307–43.

Bernard, Victor L., and Jacob K. Thomas. 1989. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research* 27 (Supplement): 1–36.

Blume, Lawrence, Adam Brandenburger, and Eddie Dekel. 1991. "Lexicographic Probabilities and Choice under Uncertainty." *Econometrica* 59 (1): 61–79.

Caballero, Ricardo J., and Arvind Krishnamurthy. 2009. "Global Imbalances and Financial Fragility." *American Economic Review* 99 (2): 584–88.

Caballero, Ricardo J., and Alp Simsek. 2009. "Complexity and Financial Panics." National Bureau of Economic Research Working Paper 14997.

Camerer, Colin F. 1987. "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence." *American Economic Review* 77 (5): 981–97.

Camerer, Colin F. 1989. "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review* 79 (5): 1257–61.

Camerer, Colin. 1995. "Individual Decision Making." In *The Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth, 587–703. Princeton, NJ: Princeton University Press.

Camerer, Colin F., and George Loewenstein. 2004. "Behavioral Economics: Past, Present, Future." In *Advances in Behavioral Economics*, edited by Colin F. Camerer, George Loewenstein, and Matthew Rabin, 3–51. Princeton, NJ: Princeton University Press.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok. 1996. "Momentum Strategies." *Journal of Finance* 51 (5): 1681–713.

Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102 (2): 179–221.

Chopra, Navin, Josef Lakonishok, and Jay R. Ritter. 1992. "Measuring Abnormal Performance: Do Stocks Overreact?" *Journal of Financial Economics* 31 (2): 235–68.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance* 53 (6): 1839–85.

De Bondt, Werner F. M., and Richard Thaler. 1985. "Does the Stock Market Overreact?" *Journal of Finance* 40 (3): 793–805.

De Bondt, Werner F. M., and Richard H. Thaler. 1987. "Further Evidence on Investor Overreaction and Stock Market Seasonality." *Journal of Finance* 42 (3): 557–81.

De Bondt, Werner F. M., and Richard H. Thaler. 1990. "Do Security Analysts Overreact?" *American Economic Review* 80 (2): 52–57.

Dominiak, Adam, Peter Dürsch, and Jean-Philippe Lefort. 2012. "A Dynamic Ellsberg Urn Experiment." *Games and Economic Behavior* 75 (2): 625–38.

Epstein, Larry G. 2006. "An Axiomatic Model of Non-Bayesian Updating." *Review of Economic Studies* 73 (2): 413–36.

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

Epstein, Larry G., and Martin Schneider. 2003. "Recursive Multiple-Priors." *Journal of Economic Theory* 113 (1): 1–31.

Epstein, Larry G., and Martin Schneider. 2007. "Learning under Ambiguity." *Review of Economic Studies* 74 (4): 1275–303.

Epstein, Larry G., Jawwad Noor, and Alvaro Sandroni. 2008. "Non-Bayesian Updating: A Theoretical Framework." *Theoretical Economics* 3 (2): 193–229.

Epstein, Larry G., Jawwad Noor, and Alvaro Sandroni. 2010. "Non-Bayesian Learning." *B. E. Journal of Theoretical Economics: Advances in Theoretical Economics* 10 (1).

Foster, Dean P., and H. Peyton Young. 2003. "Learning, Hypothesis Testing, and Nash Equilibrium." *Games and Economic Behavior* 45 (1): 73–96.

Gennaioli, Nicola, and Andrei Shleifer. 2010. "What Comes to Mind." *Quarterly Journal of Economics* 125 (4): 1399–433.

Ghirardato, Paolo. 2002. "Revisiting Savage in a Conditional World." *Economic Theory* 20 (1): 83–92.

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci. 2008. "Revealed Ambiguity and Its Consequences: Updating." In *Advances in Decision Making Under Risk and Uncertainty*, edited by Mohammed Abdellaoui and John D. Hey, 3–18. Berlin: Springer.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with Non-unique Prior." *Journal of Mathematical Economics* 18 (2): 141–53.

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

Grether, David M. 1980. "Bayes Rule as a Descriptive Model: The Representativeness Heuristic." *Quarterly Journal of Economics* 95 (3): 537–57.

Grether, David M. 1992. "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence." *Journal of Economic Behavior and Organization* 17 (1): 31–57.

Griffin, Dale, and Amos Tversky. 1992. "The Weighing of Evidence and the Determinants of Confidence." *Cognitive Psychology* 24 (3): 411–35.

Guidolin, Massimo, and Francesca Rinaldi. 2010. "A Simple Model of Trading and Pricing Risky Assets under Ambiguity: Any Lessons for Policy-Makers?" *Applied Financial Economics* 20 (1-3): 105–35.

Hanany, Eran, and Peter Klibanoff. 2007. "Updating Preferences with Multiple Priors." *Theoretical Economics* 2 (3): 261–98.

Hanany, Eran, and Peter Klibanoff. 2009. "Updating Ambiguity Averse Preferences." *B E Journal of Theoretical Economics: Advances in Theoretical Economics* 9 (1): 37.

Hirshleifer, David. 2001. "Investor Psychology and Asset Pricing." *Journal of Finance* 56 (4): 1533–97.

Holt, Charles A., and Angela M. Smith. 2009. "An Update on Bayesian Updating." *Journal of Economic Behavior and Organization* 69 (2): 125–34.

Hong, Harrison, Jeremy C. Stein, and Jialin Yu. 2007. "Simple Forecasts and Paradigm Shifts." *Journal of Finance* 62 (3): 1207–42.

Kahneman, Daniel, and Shane Frederick. 2002. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Griffin, and D. Kahneman, 49–81. New York: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective probability: A Judgment of Representativeness." *Cognitive Psychology* 3(3): 430–54.

Kahneman, Daniel, and Amos Tversky. 1973. "On the Psychology of Prediction." *Psychological Review* 80 (4): 237–51.

Kahneman, Daniel, and Amos Tversky. 1982. "Intuitive Prediction: Biases and Corrective Procedures." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic and Amos Tversky, 414–21. New York: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky. 1983. "Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90 (4): 293–315.

Kochov, Asen. 2009. "A Model of Limited Foresight." Unpublished.

Kremer, Mirko, Brent Moritz, and Enno Siemsen. 2011. "Demand Forecasting Behavior: System Neglect and Change Detection." *Management Science* 57 (10): 1827–43.

La Porta, Rafael, Josef Lakonishok, Andrei Schleifer, and Robert Vishny. 1997. "Good News for Value Stocks: Further Evidence on Market Efficiency." *Journal of Finance* 52 (2): 859–74.

Larson, Stephen J., and Jeff Madura. 2001. "Overreaction and Underreaction in the Foreign Exchange Market." *Global Finance Journal* 12 (2): 153–77.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006. "Dynamic Variational Preferences." *Journal of Economic Theory* 128 (1): 4–44.

Mailath, George J., Masahiro Okuno-Fujiwara, and Andrew Postlewaite. 1993. "Belief-Based Refinements in Signalling Games." *Journal of Economic Theory* 60 (2): 241–76.

Mullainathan, Sendhil. 2002. "Thinking through Categories." Unpublished.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. "Coarse Thinking and Persuasion." *Quarterly Journal of Economics* 123 (2): 577–619.

Myerson, Roger B. 1986a. "Axiomatic Foundations of Bayesian Decision Theory." Northwestern University, Center for Mathematical Studies in Economics and Management Science Discussion Paper 671.

Myerson, Roger B. 1986b. "Multistage Games with Communication." *Econometrica* 54 (2): 323–58.

Ortoleva, Pietro. 2010a. "Hypothesis Testing and Ambiguity Aversion." Unpublished.

Ortoleva, Pietro. 2010b. "Hypothesis-Testing Model with an Infinite State Space." Unpublished.

Poteshman, Allen M. 2001. "Underreaction, Overreaction, and Increasing Misreaction to Information in the Options Market." *Journal of Finance* 56 (3): 851–76.

Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature* 36 (1): 11–46.

Rabin, Matthew. 2002. "Inference by Believers in the Law of Small Numbers." *Quarterly Journal of Economics* 117 (3): 775–816.

Rabin, Matthew, and Joel L. Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics* 114(1): 37–82.

Rabin, Matthew, and Dimitri Vayanos. 2010. "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies* 77 (2): 730–78.

Routledge, Bryan R., and Stanley E. Zin. 2009. "Model Uncertainty and Liquidity." *Review of Economic Dynamics* 12 (4): 543–66.

Shleifer, Andrei. 2000. *Inefficient Markets: An Introduction to Behavioral Finance*. New York: Oxford University Press.

Shmaya, Eran, and Leeat Yariv. 2011. "Experiments on Decisions Under Uncertainty: A Theoretical Framework." Unpublished.

Siniscalchi, Marciano. 2011. "Dynamic Choice under Ambiguity." *Theoretical Economics* 6 (3): 379–421.

Spence, A. Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87 (3): 355–74.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58.

Vuolteenaho, Tuomo. 2002. "What Drives Firm-Level Stock Returns?" *Journal of Finance* 57 (1): 233–64.

Weinstein, Jonathan. 2011. "Provisional Probabilities and Paradigm Shifts." Unpublished.