Do Bayesians learn their way out of ambiguity?

Alexander Zimper

September 2011

Do Bayesians learn their way out of ambiguity?*

Alexander Zimper[†]

September 11, 2011

Abstract

In standard models of Bayesian learning agents reduce their uncertainty about an event's true probability because their consistent estimator concentrates almost surely around this probability's true value as the number of observations becomes large. This paper takes the empirically observed violations of Savage's (1954) sure thing principle seriously and asks whether Bayesian learners with ambiguity attitudes will reduce their ambiguity when sample information becomes large. To address this question, I develop closed-form models of Bayesian learning in which beliefs are described as Choquet estimators with respect to neo-additive capacities (Chateauneuf, Eichberger, and Grant 2007). Under the optimistic, the pessimistic, and the full Bayesian update rule, a Bayesian learner's ambiguity will increase rather than decrease to the effect that these agents will express ambiguity attitudes regardless of whether they have access to large sample information or not. While consistent Bayesian learning occurs under the Sarin-Wakker update rule, this result comes with the descriptive drawback that it does not apply to agents who still express ambiguity attitudes after one round of updating.

Keywords: Non-additive Probability Measures, Bayesian Learning, Choquet Expected Utility Theory

JEL Classification Numbers: C11, D81, D83

1 Introduction

Consider the situation of an agent who is uncertain about the probability with which a specific event (e.g., a coin comes up heads) occurs. If this agent can run a statistical experiment such that the event in question occurs independently and with identical probability across different trials, we would typically expect that he will learn the true probability as the number of trials becomes large. More specifically, in the standard framework of Bayesian learning it is assumed that the agent holds a subjective additive probability measure defined on an event space that combines possible true parameter values—here the event's true probability—with all observations that could possibly be made when running these trials. Bayesian learning refers to the situation in which the agent updates his expectation about the event's true probability in accordance with Bayes' rule; that is, the agent's Bayesian estimator is his expectation of the event's true probability with respect to a conditional additive probability measure that takes into account the observed sample information. Well-known consistency results for Bayesian estimators (Doob 1949; Breiman, LeCam, and Schwartz 1964; Lijoi, Pruenster, and Walker 2004) establish that Bayesian learning of an event's true probability converges towards this true probability whenever the agent considers the value of this true probability as possible. In particular, Doob's (1949) consistency theorem demonstrates that for almost all parameter values in the support of the agent's prior probability distribution the conditional additive probability measure will concentrate around the true parameter value almost surely if the agent observes sample information drawn from an independently and identically distributed (=i.i.d.) process. To sum up: In the standard framework of Bayesian learning an agent will reduce his ex ante uncertainty about the probability of an event whereby he almost surely learns the true probability in the long run.

This powerful concept of consistent Bayesian learning is based on the assumption that there exists some additive probability measure defined on the joint space of parameter-values and i.i.d. sample observations. According to Savage's (1954) seminal contribution, the existence of a unique subjective additive probability measure is guaranteed for agents whose preferences over Savage acts (loosely speaking: bets) obey several behavioral and technical axioms. However, starting with Ellsberg (1961), subsequent empirical studies have collected overwhelming evidence showing that one of Savage's key axioms—the sure thing principle—is consistently violated by real-life decision makers. As a reaction to the descriptive shortcomings of Savage's subjective expected utility theory, models of decision making under ambiguity have been developed which relax the sure thing principle to the effect that an agent no longer resolves his uncertainty in terms of a unique additive probability measure but, for instance, in terms of a unique non-additive

probability measure (Choquet expected utility models) or in terms of sets of additive probability measures (multiple priors models).

The present paper takes the violation of the sure thing principle seriously and asks what happens to Bayesian learning if the agent's belief over the joint space of parameter-values and i.i.d. sample observations is not expressed in terms of an additive, but rather in terms of a non-additive, probability measure. In particular, I will address the following question:

Q1: Does the Bayesian estimator of an agent—whose ambiguity is expressed in terms of a non-additive probability measure—converge to the event's true probability as the number of sample observations becomes large?

Non-additive probability measures arise in Choquet expected utility (CEU) theory (Schmeidler 1986, 1989; Gilboa 1987), which generalizes standard subjective expected utility (EU) theory to allow for ambiguity attitudes as expressed in the violation of the sure thing principle. My formal analysis restricts attention to so-called neo-additive capacities in the sense of Chateauneuf, Eichberger and Grant (2007) according to which an agent's non-additive belief about the probability of an event is a weighted average of an ambiguity part and an additive part. In particular, I develop different closed-form models of Bayesian learning—characterized by different Bayesian update rules for non-additive probability measures—such that the agent holds a neo-additive capacity defined on the joint space of parameter-values and sample observations resulting from i.i.d. Bernoulli trials. We can think of this stylized set-up as a situation under ambiguity in which the agent may have some (possibly small) lack of confidence—as expressed by a positive ambiguity parameter of a neo-additive capacity—in the accuracy of a subjective additive probability measure which would otherwise obtain in a Savage framework without ambiguity.

Working with neo-additive capacities is attractive for two reasons. Firstly, neo-additive capacities reduce the potential complexity of non-additive probability measures in a very tractable way such that important empirical features (e.g., inversely S-shaped probability transformation functions) are portrayed (cf. Chapter 11 in Wakker 2010). Secondly, Bayesian learning with respect to neo-additive capacities is closed in the following sense: The neo-additive capacity defined on the joint parameter and i.i.d. sample space gives rise to neo-additive priors and posteriors over possible parameter values, which give in turn rise to a Bayesian estimator that is itself a neo-additive capacity. Applied to a coin-flipping example, this means that the agent expresses his uncertainty about the probability that a coin-flip results in heads (or tails, for that matter) in terms of a neo-additive capacity. As it turns out, the additive part of this neo-additive estimator will converge through Bayesian learning (almost surely) to the event's true additive

probability so that question Q1 can be conveniently reformulated within the neo-additive framework as follow:

Q2: Does the ambiguity part of the neo-additive Bayesian estimator converge towards zero as the number of sample observations becomes large?

An intuitive answer to the above question is not straightforward. Since we are used to the fact that the agent of the standard—additive—Bayesian learning model reduces his uncertainty towards zero as the number of observation becomes large, we might expect a similar result for the reduction of ambiguity. After all, an agent's confidence in the additive part of his estimator should grow if he runs a statistical experiment with more and more trials. Or should it not? Namely, observing more and more observations also means observing events with smaller and smaller likelihood to the effect that a Bayesian agent who observes a lot of data updates his beliefs based on information that looks very unlikely from his ex ante perspective. Updating beliefs in the face of unlikely information, however, should rather increase than decrease the agent's ambiguity.

To investigate question Q2 in an analytically rigorous way, I consider four popular Bayesian update rules for non-additive probability measures which give rise to four different closed-form models of a neo-additive Bayesian estimator. In particular, I analyze neo-additive Bayesian learning with respect to the so-called *full Bayesian* (Pires 2002; Eichberger, Grant, and Kelsey 2006; Siniscalchi 2010) as well as the *optimistic*, the *pessimistic* (Gilboa and Schmeidler 1993) and the *Sarin-Wakker* (Sarin and Wakker 1998a) update rules. As my main formal finding I derive the following answer to the questions Q1 and Q2, respectively:

Answer (part a) If the agent either applies the full Bayesian, the optimistic, or the pessimistic update rule, then his ambiguity will increase rather than decrease as the number of sample observations becomes large. As a consequence, such a neo-additive Bayesian estimator will—in general—stay bounded away from the event's true probability as the number of sample observations becomes large.

Answer (part b) If the agent applies the Sarin-Wakker update rule, then his ambiguity drops immediately to zero after one observation. As a consequence, the agent's neo-additive Bayesian estimator reduces to the standard additive Bayesian estimator after one observation so that it converges to the event's true probability as the number of sample observations becomes large.

On the one hand, part a) gives a negative answer to questions Q1 and Q2, respectively. Namely, if ambiguity is introduced into a framework of Bayesian learning, then

there exist perfectly reasonable update rules such that Bayesian agents will never learn their way out of ambiguity. On the other hand, part b) shows that there exists at least one update rule for non-additive beliefs such that the agent's ambiguity will completely vanish when he observes new information. The problem with this seemingly "positive" result, however, is its apparent lack of descriptive relevance. Namely, whenever agents with neo-additive beliefs still exhibit some ambiguity attitudes after having made one observation, their updating behavior is not consistent with the Sarin-Wakker update rule. Under the realistic assumption that Bayesian learners still express ambiguity attitudes after one observation, the Sarin-Wakker rule does not, in my opinion, offer a viable answer to the question of how Bayesian agents might learn their way out of ambiguity.

It is well known in the decision theoretic literature that the two dynamic principles of *consequentialism* and *dynamic consistency* together imply the sure thing principle for the reductionist Bayesian framework which is relevant to this paper (Sarin and Wakker 1998b; Ghirardato 2002). That is, any dynamic decision theoretic model that satisfies consequentialism as well as dynamic consistency within a Bayesian framework is unable to express ambiguity attitudes. Loosely speaking, consequentialism refers to a situation where the agent "never looks back", i.e., where he does not care about how he has arrived at a specific decision node but where he only looks forward to the future consequences of his actions. Dynamic consistency refers to a situation in which all ex post realizations of an agent will stick to the plan of actions that is optimal from the agent's ex ante perspective. Interestingly, the full Bayesian, the optimistic, and the pessimistic update rule satisfy consequentialism but violate dynamic consistency whereas the Sarin-Wakker update rule satisfies dynamic consistency but violates consequentialism. This finding suggests a close relationship between two seemingly unrelated concepts, namely, consistent Bayesian learning, on the one hand, and dynamic consistency, on the other hand. While the formal analysis of the present paper is restricted to the most popular Bayesian update rules for non-additive probability measures used in the literature, I plan to address this apparent relationship at a more general level in future research.

The remainder of the analysis is structured as follows. Section 2 discusses the relationship between this paper and existing literature. Section 3 recalls Choquet decision theory and neo-additive capacities. Bayesian updating of neo-additive capacities is introduced in Section 4. In Section 5 the probability space and the benchmark model of Bayesian learning with respect to an additive probability measure are constructed. Section 6 presents the main results for Bayesian learning with neo-additive capacities. Finally, Section 7 concludes.

2 Related literature and further motivation

Prospect theory and statistical information. CEU theory is formally equivalent to (cumulative) prospect theory (=PT 1992) (Tversky and Kahneman 1992; Wakker and Tversky 1993) whenever PT 1992 is restricted to gains (for a comprehensive treatment of this equivalence see Wakker 2010). PT 1992, in turn, extends the celebrated concept of original prospect theory (=OPT 1979) by Kahneman and Tversky (1979) to the case of several possible gain values in a way that satisfies first-order stochastic dominance. In a recent review on applications of OPT 1979 and PT 1992 (in their terminology, PT and CPT, respectively) in management science, Holmes, Bromiley, Devers, Holcomb, and McGuire (2011) write:

> "According to Web of Science, scholars have cited the Kahneman and Tversky (1979) article that introduced PT more than 6,100 times. In addition, scholars have cited Tversky and Kahneman's (1992) article that introduced CPT more than 1,300 times. More than 550 articles citing either of these two studies (or both) have appeared in periodicals consistently recognized as top academic journals for management research [...]." (p. 1079)

Holmes et al. (2011) proceed to list applications of OPT 1979 and PT 1992 to management science, including executive compensation, negotiations, affect and motivation, human resource management, organizational risk and return, and firm risk-taking behavior. Despite the huge popularity of OPT 1979 and PT 1992 as alternatives to EU theory in applications, the present paper is, to the best of my knowledge, the first mathematically rigorous approach for addressing the question of whether the non-additive decision weights of OPT 1979 and PT 1992 (in my context: the non-additive probability measures of CEU theory) will reduce through Bayesian learning to the additive probabilities of EU theory or not. Since my formal findings answer this question in the negative, this paper's learning model establishes that meaningful applications of OPT 1979 and PT 1992 are not restricted to decision situations in which a Bayesian learner lacks statistical data but that these concepts may be also relevant to situations in which the agent takes into account statistical observations.

Bayesian learning in economic applications. Standard models of consistent Bayesian learning have been applied to several topics in economics and management science. For instance, in early contributions Cyert, DeGroot, and Holt (1978) as well as Tonks (1983) apply a Bayesian learning model—formulated within a normal distribution framework— to a firm's decision problem to invest in different technological processes. Viscusi (1979)

and Viscusi and O'Connor (1984) apply a Bayesian learning model—formulated within a Beta-distribution framework—to the risk-learning behavior of workers in injury-prone industries whereas Viscusi (1990, 1991) applies the same learning model to the question of how far smokers underestimate their health risk. The present paper's learning model is closely related to Viscusi's formal approach in that his additive Beta-distribution model obtains as the special case of my model whenever there is no ambiguity. In case the agent expresses ambiguity attitudes, however, the neo-additive Bayesian estimator does not converge to the underlying true probability and it stands to hope that my approach may explain real-life learning that is not compatible with standard models of Bayesian learning. For example, we argue in Ludwig and Zimper (2011a) that models of consistent Bayesian learning cannot explain the age-dynamics of mortality expectations as expressed in the *health retirement study* according to which objective survival probabilities and subjective survival beliefs diverge rather than converge when more statistical information becomes available.

Rational expectations in financial economics. While Savage's approach establishes conditions for the existence of subjective additive probability measures, Muth's (1961) *rational expectations hypothesis* stipulates that subjective estimators correctly reflect average objective probabilities which drive economic fundamentals. The rational expectations hypothesis is thereby justified by the assumption that agents are standard Bayesian learners whose subjective probabilities have already converged to their objective counterparts. Especially in financial economics, however, the rational expectations hypothesis has come under heavy scrutiny. In particular, there exist several different proposals for relaxing the rational expectations assumption to address so-called *asset pricing puzzles* (cf., e.g., Mehra and Prescott 1985; Barberis and Thaler 2003). Along this line, Cecchetti, Lam and Mark (2000) consider the implications of rules of thumb with respect to estimates of the consumption growth process and Abel (2002) studies the effects of pessimism and doubt on asset returns. Related to this literature but with less of an *ad hoc* flavor are robust control applications to asset pricing puzzles where pessimism results from an agent's caution in responding to concerns about model misspecification (Hansen, Sargent and Tallarini 1999; Hansen and Sargent 2001; Maenhout 2004; Hansen and Sargent 2007). An apparent drawback of both approaches is, however, their incompatibility with standard learning models with Bayesian features or overtones by which subjective beliefs converge to their objective counterparts in the long-run, cf. Barsky and DeLong (1993), Timmermann (1993), Brav and Heaton (2002), Cogley and Sargent (2008) and Adam, Marcet and Nicolini (2008). As a potential contribution to this financial economics literature, the non-converging Bayesian learning model of this

paper provides a consistent theoretical foundation for why a subjective estimator may remain biased even in the long-run. Similar to the learning models considered in this paper, we develop and calibrate—within a normal distribution framework—in Ludwig and Zimper (2011b) a model of Choquet Bayesian learning of the consumption growth rate that results in a biased subjective estimator for the stochastic discount factor. As a consequence, our approach of non-additive Bayesian learning may, at least partially, explain the so-called *risk-free rate puzzle* according to which a realistically calibrated standard consumption based asset pricing model yields a return on risk-free assets of about $5\% - 6\%$ compared to a real-world actual risk-free rate in the range of $1 - 2\%$.

Alternative models of learning under ambiguity. Marinacci (1999, 2002) studies a non-Bayesian learning environment with non-additive beliefs whereby he considers a decision maker who observes an experiment where the outcomes in each trial are identically and independently distributed with respect to the decision-maker's non-additive belief. In this setup, Marinacci derives for (basically convex) capacities laws of large numbers as counterparts to the additive case thereby admitting for the possibility that ambiguity does not vanish in the long-run. While Marinacci's approach may thus be regarded as a frequentist approach towards non-additive probabilities, our approach is a Bayesian one according to which an agent has a subjective prior belief over the whole event space while he uses sample information from an objective process to update his subjective belief.

Epstein and Schneider (2007) consider a model of learning under ambiguity for max min expected utility (MMEU) preferences of Gilboa and Schmeidler (1989) which shares with our learning model the feature that ambiguity does not necessarily vanish in the long run. Their learning model is based on the *recursive multiple priors* approach of Epstein and Schneider (2003) which goes back to Sarin and Wakker (1998b, Theorem 2.1) who consider fixed decision trees and show that updating for the information structure of a given tree satisfies consequentialism and dynamic consistency whenever the set of multiple priors is given as the *reduced family* of probability measures (=*rectangular priors* in the terminology of Epstein and Schmeidler). Similar to the fixed decision trees of Sarin and Wakker (1998b), Epstein and Schneider (2003) fix some information structure and assume that Savage's sure thing principle holds exactly at all observable events. Or in the words of Siniscalchi (2010):

> "Epstein and Schneider's dynamic consistency requirement [...] implies that prior preferences must satisfy Savage's Postulate P2 relative to every conditioning event in the filtration under consideration." (p. 35)

Linking ambiguity attitudes to a given information structure imposes strong restrictions on the ambiguity attitudes that an agent may express. For example, an agent who will be able to observe in the future every possible event cannot exhibit any ambiguity attitudes within the Epstein and Schneider framework since the sure thing principle must hold—as in Savage (1954)—at all possible events.

As one main difference to Epstein and Schneider (2007), the agents of this paper's learning model may express ambiguity attitudes which are independent from their information partition. As another difference, Epstein and Schneider establish long-run ambiguity through the assumption that the agent permanently receives ambiguous signals formally described as a multitude of different likelihood functions at each information stage. While this introduction of multiple likelihoods is rather ad hoc, this paper's axiomatically founded approach offers a straightforward explanation for the possible existence of a long-run bias under Bayesian learning under ambiguity. Namely, the bias of the neo-additive Bayesian estimator is the greater the more surprised the agent is by the information he receives.

3 Choquet decision theory and neo-additive capacities

This section briefly recalls basic elements of Choquet expected utility theory and neo-additive capacities. CEU theory was first axiomatized by Schmeidler (1986, 1989) within the Anscombe and Aumann (1963) framework, which assumes preferences over objective probability distributions. Subsequently, Gilboa (1987) as well as Sarin and Wakker (1992) have presented CEU axiomizations within the Savage (1954) framework, assuming a purely subjective notion of likelihood. Moreover, as a representation of preferences over lotteries, CEU theory coincides with *rank dependent utility theory* as introduced by Quiggin (1981, 1982). Within the context of CEU theory, properties of such capacities are used in the literature for formal definitions of, e.g., *ambiguity* and *uncertainty attitudes* (Schmeidler 1989; Epstein 1999; Ghirardato and Marinacci 2002), *pessimism* and *optimism* (Eichberger and Kelsey 1999; Wakker 2001), as well as *sensitivity to changes in likelihood* (Wakker 2004).

Consider a measurable space $(\Omega, \mathcal{F})$ with $\mathcal{F}$ denoting a σ-algebra on the state space Ω and a non-additive probability measure $(=capacity)$ $\nu : \mathcal{F} \rightarrow [0, 1]$ satisfying

(i) $\nu(\emptyset) = 0$, $\nu(\Omega) = 1$

(ii) $A \subset B \Rightarrow \nu(A) \leq \nu(B)$ for all $A, B \in \mathcal{F}$.

The Choquet integral of a bounded function $f : \Omega \to \mathbb{R}$ with respect to capacity ν is defined as the following Riemann integral extended to domain Ω (Schmeidler 1986):

$$E[f, \nu(d\omega)] = \int_{-\infty}^{0} (\nu(\{\omega \in \Omega \mid f(\omega) \geq z\}) - 1) \, dz + \int_{0}^{+\infty} \nu(\{\omega \in \Omega \mid f(\omega) \geq z\}) \, dz \tag{1}$$

whereby I simply write $E[f, \nu]$ for $E[f, \nu(d\omega)]$. For example, assume that f takes on m different values such that $A_1, ..., A_m$ is the unique partition of Ω with $f(\omega_1) > ... > f(\omega_m)$ for $\omega_i \in A_i$. Then the Choquet expectation (1) becomes

$$E[f, \nu] = \sum_{i=1}^{m} f(\omega_i) \cdot [\nu(A_1 \cup ... \cup A_i) - \nu(A_1 \cup ... \cup A_{i-1})]. \tag{2}$$

This paper focuses on non-additive probability measures that are defined as *neo-additive capacities* in the sense of Chateauneuf, Eichberger and Grant (2007). Recall that the set of *null events*, denoted $\mathcal{N}$, collects all events that the decision maker deems impossible.

Definition. *Fix some set of null-events $\mathcal{N} \subset \mathcal{F}$ for the measurable space $(\Omega, \mathcal{F})$. The neo-additive capacity, ν, is defined, for some $\delta, \lambda \in [0, 1]$ by*

$$\nu(A) = \delta \cdot \nu_\lambda(A) + (1 - \delta) \cdot \mu(A) \tag{3}$$

for all $A \in \mathcal{F}$ such that μ is some additive probability measure satisfying

$$\mu(A) = \begin{cases} 0 & \text{if } A \in \mathcal{N} \\ 1 & \text{if } \Omega \backslash A \in \mathcal{N} \end{cases} \tag{4}$$

and the non-additive probability measure ν_λ is defined as follows

$$\nu_\lambda(A) = \begin{cases} 0 & \text{iff } A \in \mathcal{N} \\ \lambda & \text{else} \\ 1 & \text{iff } \Omega \backslash A \in \mathcal{N}. \end{cases} \tag{5}$$

Throughout this paper I restrict attention to sets of null-events $\mathcal{N}$ such that $A \in \mathcal{N}$ if and only if $\mu(A) = 0$, which implies $\nu(A) = 0$ (resp. $\nu(A) = 1$) if and only if $\mu(A) = 0$ (resp. $\mu(A) = 1$). Call A an *essential* event whenever $A \notin \mathcal{N}$ and $\Omega \backslash A \notin \mathcal{N}$ (i.e., neither A nor its complement is null) and observe that, for any essential A, the neo-additive capacity ν in (3) simplifies to

$$\nu(A) = \delta \cdot \lambda + (1 - \delta) \cdot \mu(A) \tag{6}$$

with $0 < \mu(A) < 1$. Neo-additive capacities can thus be interpreted as non-additive beliefs that stand for deviations from additive beliefs such that a parameter δ (*degree of ambiguity*) measures the lack of confidence the agent has in some subjective additive probability measure μ. The following observation—formally proved in the appendix— extends a result (Lemma 3.1) of Chateauneuf, Eichberger and Grant (2007) for finite random variables to the more general case of random variables with a bounded range.

Observation 1. *Let $f : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable function with bounded range. The Choquet expected value (1) of f with respect to a neo-additive capacity (3) is then given by*

$$E[f, \nu] = \delta\left(\lambda \sup f + (1 - \lambda) \inf f\right) + (1 - \delta) E[f, \mu]. \tag{7}$$

According to Observation 1, the Choquet expected value of a random variable f with respect to a neo-additive capacity is a convex combination of the expected value of f with respect to some additive probability measure π and an ambiguity part. If there is no ambiguity, i.e., $\delta = 0$, then the Choquet expected value (7) reduces to the standard expected value of a random variable with respect to an additive probability measure. In case there is some ambiguity, however, the second parameter λ measures how much weight the decision maker puts on the least upper bound of the range of f. Conversely, $(1 - \lambda)$ is the weight he puts on the greatest lower bound. For high (resp. low) values of λ, the Choquet expected value $E[f, \nu]$ is greater (resp. less) than the corresponding additive expected values $E[f, \mu]$ whenever the agent expresses some ambiguity, i.e., $\delta > 0$. This formal feature gives rise to the possibility of over-estimation (resp. under-estimation) with respect to the rational benchmark case.

4 Bayesian updating of neo-additive capacities

The Bayesian framework that I consider in this paper satisfies the so-called *reduction principle* (cf., e.g., Siniscalchi 2010), that is, it reduces dynamic decision situations in the standard way to a static Savage framework such that ex ante and ex post preferences over Savage acts are well-defined. Whenever the reduction principle is satisfied, the dynamic principles of consequentialism and dynamic consistency imply Savage's sure thing principle (Sarin and Wakker 1998b; Ghirardato 2002). As a consequence, any Bayesian update rule that allows the expression of ambiguity attitudes cannot simultaneously satisfy both principles.

At first, I explain in some detail the relationship between consequentialism and dynamic consistency, on the one hand, and the sure thing principle, on the other hand. In a next step, I am going to present the four most popular (in terms of appearances in the literature) Bayesian update rules for non-additive probability measures whereby the optimistic, the pessimistic as well as the full Bayesian update rule satisfy consequentialism but violate dynamic consistency whereas the converse statement is true for the Sarin-Wakker update rule. As this section's main formal result I derive a proposition that applies all considered update rules to neo-additive capacities to obtain formal expressions for conditional neo-additive capacities.

4.1 Consequentialism, dynamic consistency, and the sure thing principle

As in the previous section, consider some state space Ω and some σ-algebra $\mathcal{F}$ on Ω. Recall that a Savage act f is an $\mathcal{F}$-measurable function that maps the state space Ω into an arbitrary set of consequences X, i.e., $f : \Omega \to X$. The event space $\mathcal{F}$ is supposed to be rich enough to cover all aspects of uncertainty relevant to the decision maker. Whenever $\omega \in \Omega$ is the (unique) true state of the world and the agent has chosen Savage act f, he ends up with the consequence $f(\omega)$ after the uncertainty is resolved. Given two Savage acts f, g, let $B, \neg B \in \mathcal{F}$ denote two complementary events and define the Savage act $f_B g : \Omega \to X$ as follows:

$$f_B g (\omega) = \left\{ \begin{array}{ll} f(\omega) & \text{for } \omega \in B \\ g(\omega) & \text{for } \omega \in \neg B. \end{array} \right. \tag{8}$$

Ex ante preferences over Savage acts, denoted $\succeq$, are interpreted as the decision maker's preferences before he receives any information. In contrast, ex post preferences over Savage acts, denoted $\succeq_B$, are interpreted as preferences conditional on B, i.e., after the decision maker has observed the occurrence of some essential event $B \in \mathcal{F}$.

Definition: Consequentialism. *For all Savage acts f, g, h, h', h'', h''' and all essential events $B \in \mathcal{F}$, the following condition holds for ex post preferences:*

$$f_B h \succeq_B g_B h' \quad \Leftrightarrow \quad f_B h'' \succeq_A g_B h'''. \tag{9}$$

The notion of consequentialism in the context of decision trees under risk goes back to Hammond (1989) and Machina (1989) whereas the above definition is closest to Machina's interpretation once the reduction principle is applied to decision trees:

"The philosophy behind this approach [=consequentialism; the author] is that the uncertainty that was involved in the rest of the tree, as represented by the probabilities at the snipped-off chance nodes and the planned choices at the snipped-off choice nodes, is now irrelevant and should be treated as if it never existed." (p. 1641)

In other words, (9) states that it should not matter for ex post preferences conditional on observation B whatever consequences, governed either by $h, h', h'', $ or h''', might have happened outside of B.

Definition: Dynamic consistency. *For all Savage acts f, g, h and all essential events $B \in \mathcal{F}$, the following condition holds for the relationship between ex ante and ex post preferences:*

$$f_B h \succeq g_B h \quad \Leftrightarrow \quad f_B h \succeq_B g_B h. \tag{10}$$

In a nutshell, dynamic consistency means that a decision maker who ex ante prefers for event B the consequences of act f to the consequences of act g will not change his mind after the observation of B and vice versa. Finally, recall the formal definition of Savage's sure thing principle

Definition: Sure thing principle. *For all Savage acts f, g, h, h' and all events $B \in \mathcal{F}$, the following condition holds for ex ante preferences:*

$$f_B h \succeq g_B h \quad \Rightarrow \quad f_B h' \succeq g_B h'. \tag{11}$$

The following finding restates—for our slightly different definitions of (9) and (10)—a formal result which is already implied by Theorem 1 in Ghirardato (2002).

Observation 2. *Consequentialism (9) combined with dynamic consistency (10) implies the sure thing principle (11).*

Proof: Consider any essential $B \in \mathcal{F}$. Observe that

$$
\begin{aligned}
f_B h \quad &\succeq \quad g_B h \\
&\Rightarrow \quad f_B h \succeq_B g_B h \text{ by (10)} \\
&\Rightarrow \quad f_B h' \succeq_B g_B h' \text{ by (9)} \\
&\Rightarrow \quad f_B h' \succeq g_B h' \text{ by (10)},
\end{aligned}
$$

which proves the claim.$\square$

4.2 Bayesian update rules

A Bayesian update rule specifies how the ex ante preference ordering $\succeq$ determines, for all essential $B \in \mathcal{F}$, the ex post preference orderings $\succeq_B$. At first, I consider Bayesian update rules such that for every essential $B \in \mathcal{F}$ and every pair of Savage acts f, g there is some Savage act h satisfying

$$f_B h \succeq g_B h \text{ implies } f_B h' \succeq_B g_B h'' \text{ for all } h', h''. \tag{12}$$

Such update rules obviously satisfy consequentialism since the ex post preferences are independent of consequences that happened outside of observation B.

Gilboa and Schmeidler (1993) introduce the class of so-called h-*Bayesian update rules* such that, for all non-null B and some fixed act $E \in \mathcal{F}$, h in (12) is given as

$$h = (x^*, E; x_*, \neg E) \tag{13}$$

whereby x^* denotes the best and x_* denotes the worst consequence possible, which exist by assumption. (Note that my framework uses sup and inf rather than max and min, which is possible by Observation 1.) The different possible specifications of E in (13) can result in a multitude of different h-Bayesian update rules. For example, for the so-called *optimistic* (=naive) update rule, h is the constant act where $E = \emptyset$. That is, under the optimistic update rule the null-event becomes associated with the worst consequence possible. Gilboa and Schmeidler (1993, p.41) offer the following psychological motivation for this update rule:

> "[...] when comparing two actions given a certain event B, the decision maker implicitly assumes that had B not occurred, the worst possible outcome [...] would have resulted. In other words, the behavior given B [...] exhibits 'happiness' that B has occurred; the decisions are made as if we are always in 'the best of all possible worlds'."

The corresponding **optimistic Bayesian update rule** for conditional beliefs of CEU decision makers is given by

$$\nu^{opt}(A \mid B) = \frac{\nu(A \cap B)}{\nu(B)}. \tag{14}$$

For the *pessimistic* (=Dempster-Shafer or maximum likelihood) update rule h is the constant act where $E = \Omega$, associating with the null-event the best consequence possible. Gilboa and Schmeidler (1993, p.41):

14

"[...] we consider a 'pessimistic' decision maker, whose choices reveal the hidden assumption that all the impossible worlds are the best conceivable ones."

The corresponding **pessimistic Bayesian update rule** for CEU decision makers is

$$\nu^{pess}\left(A \mid B\right) = \frac{\nu\left(A \cup \neg B\right) - \nu\left(\neg B\right)}{1 - \nu\left(\neg B\right)}. \tag{15}$$

A further updating rule, which does not belong to the class of h-Bayesian update rules but satisfies (12) and therefore consequentialism, is the so-called *full* (or *generalized*) Bayesian update rule. The full Bayesian update rule is popular in the literature because it avoids the extreme updating behavior of the optimistic and pessimistic update rules. More specifically, suppose that, for all acts f, g,

$$f \succeq_B g \text{ if and only if } f_B h \succeq g_B h \tag{16}$$

such that h is the *conditional certainty equivalent* of g, i.e., h is the constant act such that $g \sim_B h$. The corresponding **full Bayesian update rule** for non-additive probability measures is then given as follows (Eichberger, Grant, and Kelsey 2006):

$$\nu\left(A \mid B\right) = \frac{\nu\left(A \cap B\right)}{\nu\left(A \cap B\right) + 1 - \nu\left(A \cup \neg B\right)} \tag{17}$$

for $A, B \in \mathcal{F}$ such that $B \neq \emptyset$.

Finally, I consider the **Sarin-Wakker update rule** (Sarin and Wakker 1998a), which violates consequentialism but satisfies dynamic consistency. Let $D_f\left[A\right] \in \mathcal{F}$ denote an event *that dominates event A with respect to f* in the sense that $D_f\left[A\right] \cap A = \varnothing$ and $f\left(\omega\right) \succeq f\left(\omega'\right) \succeq f\left(\omega''\right)$ for all $\omega \in D_f\left[A\right]$, $\omega' \in A$, and $\omega'' \in \neg\left(D_f\left[A\right] \cup A\right)$. The Choquet expected value (2) can then be rewritten as

$$E\left[f, \nu\left(d\omega\right)\right] = \sum_{\omega \in \Omega} f\left(\omega\right) \cdot \tilde{\nu}\left(A_i\right) \tag{18}$$

where the decision weight $\tilde{\nu}$ is given by

$$\tilde{\nu}\left(A_i\right) = \nu\left(A_i \cup D_f\left(A_i\right)\right) - \nu\left(D_f\left(A_i\right)\right). \tag{19}$$

Sarin and Wakker (1998a) propose the following act-dependent update rule in terms of conditional decision weights

$$\tilde{\nu}_f\left(A \mid B\right) = \frac{\nu\left(\left(A \cap B\right) \cup D_f\left[A \cap B\right]\right) - \nu\left(D_f\left[A \cap B\right]\right)}{\nu\left(B \cup D_f\left[B\right]\right) - \nu\left(D_f\left[B\right]\right)} \tag{20}$$

for essential $A, B \in \mathcal{F}$, whereby

$$E\left[f, \tilde{\nu}_f\left(d\omega \mid B\right)\right] = \sum_{\omega \in \Omega} f\left(\omega\right) \cdot \tilde{\nu}_f\left(A \mid B\right). \tag{21}$$

The following proposition is proved in the appendix.

Proposition 1: *Let $\delta, \lambda \in [0,1]$ and consider essential events A, B.*

   (i) *An application of the full Bayesian update rule (17) to a prior belief (6) results in the posterior belief*

$$\nu^{FB}\left(A \mid B\right) = \delta_B^{FB} \cdot \lambda + \left(1 - \delta_B^{FB}\right) \cdot \mu\left(A \mid B\right) \tag{22}$$

   such that

$$\delta_B^{FB} = \frac{\delta}{\delta + \left(1 - \delta\right) \cdot \mu\left(B\right)}. \tag{23}$$

   (ii) *An application of the optimistic update rule (14) to a prior belief (6) results in the conditional belief*

$$\nu^{opt}\left(A \mid B\right) = \delta_B^{opt} + \left(1 - \delta_B^{opt}\right) \cdot \mu\left(A \mid B\right) \tag{24}$$

   with

$$\delta_B^{opt} = \frac{\delta \cdot \lambda}{\delta \cdot \lambda + \left(1 - \delta\right) \cdot \mu\left(B\right)}. \tag{25}$$

   (iii) *An application of the pessimistic update rule (15) to a prior belief (6) results in the conditional belief*

$$\nu^{pess}\left(A \mid B\right) = \left(1 - \delta_B^{pess}\right) \cdot \mu\left(A \mid B\right) \tag{26}$$

   with

$$\delta_B^{pess} = \frac{\delta \cdot \left(1 - \lambda\right)}{\delta \cdot \left(1 - \lambda\right) + \left(1 - \delta\right) \cdot \mu\left(B\right)}. \tag{27}$$

   (iv) *An application of the Sarin-Wakker update rule (15) to a prior belief (6) results in the conditional decision weight*

$$\tilde{\nu}_f^{SW}\left(A \mid B\right) = \mu\left(A \mid B\right). \tag{28}$$

Remark 1. While the question "Which Bayesian updating rules are actually used by real-life individuals?" is ultimately an empirical one, I know (thanks to an anonymous referee) of only one experimental investigation which directly addresses this question.

Namely, Cohen, Gilboa, Jaffray, and Schmeidler (2000) investigate the question whether the pessimistic—in their terms: the maximum likelihood (=ML)—update rule or the full Bayesian (=FB) update rule is consistent with ambiguity averse subjects' choice behavior. Their experimental findings establish an approximate ratio of 2:1 in favor for the full Bayesian update rule. More specifically, Cohen et al. (2000) write:

> "Thus, out of the remaining 33 subjects who must have multiple priors in sample I, 12 chose in accordance with MLU prediction (36, 36%), whereas 19 of them made choices consistent with FBU (57, 58%), and 2 were inconsistent with either theory (6, 06%). Results in sample II are similar: out of 43 subjects, MLU accounts for the choices of 10 of them (23, 25%), FBU for the choices of 23 others (53, 5%), whereas the choices of the remaining 10 (23, 25%) are unexplained." (p. 130)

Remark 2. Recall that the MMEU theory of Gilboa and Schmeidler (1989) formally coincides with CEU theory restricted to convex capacities (e.g., neo-additive capacities for which the λ parameter equals zero) whenever the set of priors is given as the core of the convex capacity. Hanany and Klibanoff (2007, 2009) establish the existence of dynamically consistent—but consequentialism-violating—update rules for the MMEU framework by proving the existence of a dynamically consistent update rule which uniquely maximizes ambiguity in terms of the *more-ambiguous-than* relation as introduced by Ghirardato and Marinacci (2002). An interesting question for future research is whether this ambiguity maximizing update rule for multiple priors is formally equivalent to the Sarin-Wakker update rule for non-additive probability measures or not.

5   Bayesian learning: Information structure and additive beliefs

This section derives in some mathematical detail a closed-form learning model with additive beliefs as introduced to the economics literature by Viscusi and O'Connor (1984) and Viscusi (1985). The reader who wants to skip the mathematical details may immediately jump to the—intuitive—additive Bayesian estimator (37).

Consider the situation of an agent who is uncertain about the probability with which an event, say H (e.g., a coin comes up heads), occurs but who observes a statistical experiment with n independent trials where H occurs in every trial with identical probability. Formally, let us consider a probability space $(\mu, \Omega, \mathcal{F})$ such that μ denotes a

17

subjective additive probability measure defined on the events in $\mathcal{F}$. In what follows, I describe the construction of the state space, of the event space and of the additive probability measure. The event space is thereby closely related to the information structure, which I also construct in some detail. Finally, I establish consistency of the additive estimator for this closed-form model of Bayesian learning.

Construction of the state space Ω. Denote by $\Pi = (0,1)$ the parameter-space that collects all possible values of the "true" probability of H in any given trial and denote by $S^\infty = \times_{i=1}^\infty S_i$ with $S_i = \{H, T\}$, $i = 1, 2, ...,$ the sample-space that collects all possible infinite sequences of outcomes. Furthermore, define the state space

$$\Omega = (0,1) \times S^\infty,$$

with generic element $\omega = (\pi, s^\infty)$. Observe that this definition excludes the possibility that the true probability is degenerate, i.e., either zero or one.

Construction of the information structure. The information structure of the learning model is described by an infinite sequence of ever *finer* information partitions $\mathcal{P}_1, \mathcal{P}_2, ...$ such that, for $n = 1, 2, ...,$ $\mathcal{P}_n$ consists of all information cells of the following form

$$(0,1) \times \{x_1\} \times ... \times \{x_n\} \times S_{n+1} \times ...$$

with $x_i \in S_i$, $i = 1, 2, .., n$. According to this information structure, the period n agent does not make any observation about the true parameter value but he observes the exact sample information up till period n. The distinctive feature of any model of Bayesian learning of a parameter value is that learning exclusively follows from (indirect) likelihood considerations but never from (direct) observation of the parameter value. Loosely speaking, this feature gives rise to the possibility that the agent of our model may never learn the true parameter value when his likelihood considerations are described by a non-additive probability measure.

Construction of the event space $\mathcal{F}$. Endow $(0,1)$ with the Euclidean metric and denote by $\mathcal{B}$ the Borel σ-algebra in $(0,1)$, i.e., the smallest σ-algebra containing all open subsets of the open Euclidean unit interval. Similarly, endow S^∞ with the discrete metric and denote by $\mathcal{S}^\infty$ the Borel σ-algebra in S^∞, i.e., $\mathcal{S}^\infty$ coincides with the power-set of S^∞. The event space $\mathcal{F}$ is then defined as the σ-algebra generated by the members in $\mathcal{B} \otimes \mathcal{S}^\infty$, where $\mathcal{B} \otimes \mathcal{S}^\infty$ denotes the usual product algebra of $\mathcal{B}$ and $\mathcal{S}^\infty$.

In a next step, define by Σ_n the σ-algebra generated by $\mathcal{P}_n$, for $n = 1, 2, ...$ That is, Σ_n is the smallest collection of subsets of Ω that is a σ-algebra which contains all

information cells in $\mathcal{P}_n$. Observe that $\Sigma_1 \subset \Sigma_2 \subset ... \subset \mathcal{F}$ so that the sequence of σ-algebras $(\Sigma_1, \Sigma_2, ..., \mathcal{F})$ constitutes a *filtration*.

Construction of the additive probability measure μ. Define $\tilde{\pi} : \Omega \to (0,1)$ such that $\tilde{\pi}(\pi, s^\infty) = \pi$ the $\mathcal{F}$-measurable random variable that assigns to every state of the world the true probability of outcome H. For our purpose it is convenient to denote by $\boldsymbol{\pi}$ the event in $\mathcal{F}$ such that the value $\pi \in \Pi$ is the true probability, i.e.,

$$\boldsymbol{\pi} = \{\omega \in \Omega \mid \tilde{\pi}(\omega) = \pi\}. \tag{29}$$

I assume that the agent's prior over $\tilde{\pi}$ is given as a Beta distribution with parameters $\alpha, \beta > 0$ so that

$$\mu(\boldsymbol{\pi}) = K_{\alpha,\beta} \pi^{\alpha-1}(1-\pi)^{\beta-1} \tag{30}$$

for $\pi \in (0,1)$ such that the normalizing constant is given as

$$K_{\alpha,\beta} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

with the gamma function defined as $\Gamma(y) = \int\limits_0^\infty x^{y-1}e^{-x}dx$ for $y > 0$. Define by $X_n : \Omega \to \{0,1\}$, with $n = 1, 2, ...$, the Σ_n-measurable random variable that takes on value one if a head occurs in the n-th trial and zero otherwise. I assume that, conditional on the parameter-value $\pi \in (0,1)$, each X_n is Bernoulli distributed with probabilities

$$\mu(\{\omega \in \Omega \mid X_n(\omega) = x\} \mid \boldsymbol{\pi}) = \pi^x(1-\pi)^{1-x} \text{ for } x \in \{0,1\}. \tag{31}$$

Furthermore, define by $I_n : \Omega \to \{0, ..., n\}$ the Σ_n-measurable random variable counting the number of heads occurring in n trials, i.e., $I_n = \sum_{k=1}^n X_k$. For convenience, I denote by $\mathbf{I}_n^k$ the event in $\mathcal{F}$ such that outcome H has occurred k-times in the n first trials, i.e.,

$$\mathbf{I}_n^k = \{\omega \in \Omega \mid I_n(\omega) = k\}. \tag{32}$$

Since the X_n are independently and identically Bernoulli distributed, each I_n is, conditional on the parameter-value $\pi \in (0,1)$, binomially distributed with probabilities

$$\mu(\mathbf{I}_n^k \mid \boldsymbol{\pi}) = \binom{n}{k} \pi^k(1-\pi)^{n-k} \text{ for } k \in \{0, ..., n\}. \tag{33}$$

By Bayes' rule we obtain the following posterior probability that π is the true value conditional on information $\mathbf{I}_n^k$

$$\mu(\boldsymbol{\pi} \mid \mathbf{I}_n^k) = \frac{\mu(\boldsymbol{\pi} \cap \mathbf{I}_n^k)}{\mu(\mathbf{I}_n^k)} \tag{34}$$

$$= \frac{\mu(\mathbf{I}_n^k \mid \boldsymbol{\pi})\mu(\boldsymbol{\pi})}{\int_{[0,1]} \mu(\mathbf{I}_n^k \mid \boldsymbol{\pi})\mu(\boldsymbol{\pi})\,d\pi} \tag{35}$$

$$= K_{\alpha+k,\beta+n-k} \pi^{\alpha+k-1}(1-\pi)^{\beta+n-k-1}. \tag{36}$$

The Bayesian estimator. The agent's prior estimator of the true probability of H is defined as the expected value of $\tilde{\pi}$ with respect to the prior distribution, i.e., $E[\tilde{\pi}, \mu]$. Accordingly, the agent's posterior estimator of π conditional on information $\mathbf{I}_n^k$ is defined as the expected value of $\tilde{\pi}$ with respect to the resulting posterior distribution, i.e., $E[\tilde{\pi}, \mu(\pi \mid \mathbf{I}_n^k)]$. In the case of a Beta prior we therefore obtain as the prior estimator the expected value of the Beta distribution, implying $E[\tilde{\pi}, \mu] = \frac{\alpha}{\alpha+\beta}$. Furthermore, since the agent's posterior $\mu(\pi \mid \mathbf{I}_n^k)$ is itself a Beta distribution with parameters $\alpha+k, \beta+n-k$, we have $E[\tilde{\pi}, \mu(\pi \mid \mathbf{I}_n^k)] = \frac{\alpha+k}{\alpha+\beta+n}$ as the agent's posterior estimator of π conditional on information $\mathbf{I}_n^k$. Or equivalently,

$$E[\tilde{\pi}, \mu(\boldsymbol{\pi} \mid \mathbf{I}_n^k)] = \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right) E[\tilde{\pi}, \mu] + \left(\frac{n}{\alpha+\beta+n}\right) \frac{k}{n} \tag{37}$$

where $\frac{k}{n}$ is the sample mean. That is, the agent's posterior estimator of the probability of H is a weighted average of his prior estimator and the sample mean whereby the weight attached to the sample mean increases in the number of trials. If the number of trials approaches infinity, i.e., $n \to \infty$, the sample mean information $\mathbf{I}_n^k$ converges, by the strong law of large numbers, with probability one to the sample information $\mathbf{I}^*$ according to which outcome H has occurred with relative frequency π, i.e., the true probability of outcome H. As a consequence, we obtain the following consistency result for this closed-form model of Bayesian learning with additive beliefs.

Observation 3: *The posterior estimator $E[\tilde{\pi}, \mu(\boldsymbol{\pi} \mid \mathbf{I}_n^k)]$ of the probability of H converges with probability one to the true probability π as n approaches infinity.*

Remark. Observe that the convergence result of Observation 3 holds for all true parameter values in $(0, 1)$ so that it is somewhat stronger than Doob's (1949) consistency theorem which establishes convergence of Bayesian estimators only for almost all true parameter values. Related to Doob's consistency theorem is Blackwell and Dubins' (1962) convergence theorem for different additive probability measures within the frequentist framework. In particular, Diaconis and Freedman (1986, Theorem 3) establish a formal link between Doob's consistency theorem and Blackwell and Dubins' convergence theorem by basically showing that the Bayesian estimate is consistent if and only if any corresponding conditional probability measures merge in the *weak topology*.

6 Bayesian learning with neo-additive capacities

In this section I formally combine the updating of neo-additive capacities, as derived in Proposition 1, with the additive learning model (37) to derive closed form models of Bayesian learning under ambiguity. Formally, I now consider the neo-additive probability space $(\nu, \Omega, \mathcal{F})$. As a generalization of the Bayesian learning model discussed in the previous section, we have now a neo-additive prior about the unknown parameter π such that

$$\nu(\boldsymbol{\pi}) = \delta\lambda + (1-\delta) \cdot K_{\alpha,\beta}\pi^{\alpha-1}(1-\pi)^{\beta-1}, \tag{38}$$

for $\pi \in (0,1)$, i.e., the additive part of this prior is the Beta distribution with parameters α and β. Accordingly, the agent's prior estimator of the true value of π is now given as the Choquet expected value of $\tilde{\pi}$ with respect to his neo-additive prior, i.e.,

$$E[\tilde{\pi}, \nu] = \int_\Omega \tilde{\pi}(\omega) \, d\nu(\boldsymbol{\pi}) \tag{39}$$

$$= \delta\lambda + (1-\delta) E[\tilde{\pi}, \mu] \tag{40}$$

by Proposition 1 and the fact that $\inf \tilde{\pi} = 0$ and $\sup \tilde{\pi} = 1$. Observe that this prior estimator is itself a neo-additive capacity. Combining the results of Proposition 1 with the posterior estimator

$$E\left[\tilde{\pi}, \nu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right] = \int_\Omega \tilde{\pi}(\omega) \, d\nu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right) \tag{41}$$

gives the following proposition, which constitutes this paper's main formal result.

Proposition 2. *Suppose that the agent receives sample information* $\mathbf{I}_n^k$. *Contingent on the applied update rule we obtain the following conditional neo-additive beliefs and posterior estimators of parameter* $\pi \in (0,1)$ *whereby* $E\left[\tilde{\pi}, \mu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right]$ *is given by (37).*

(i) *Full Bayesian updating.*

$$\nu^{FB}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right) = \delta_{\mathbf{I}_n^k}^{FB}\lambda + \left(1 - \delta_{\mathbf{I}_n^k}^{FB}\right) \cdot K_{\alpha+k,\beta+n-k}\pi^{\alpha+k-1}(1-\pi)^{\beta+n-k-1} \tag{42}$$

with

$$\delta_{\mathbf{I}_n^k}^{FB} = \frac{\delta}{\delta + (1-\delta) \cdot \mu\left(\mathbf{I}_n^k\right)} \tag{43}$$

so that

$$E\left[\tilde{\pi}, \nu^{FB}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right] = \delta_{\mathbf{I}_n^k}^{FB}\lambda + \left(1 - \delta_{\mathbf{I}_n^k}^{FB}\right) \cdot E\left[\tilde{\pi}, \mu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right]. \tag{44}$$

(ii) *Optimistic updating.*

$$\nu^{opt}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right) = \delta_{\mathbf{I}_n^k}^{opt} + \left(1 - \delta_{\mathbf{I}_n^k}^{opt}\right) \cdot K_{\alpha+k,\beta+n-k}\pi^{\alpha+k-1}\left(1-\pi\right)^{\beta+n-k-1} \quad (45)$$

with

$$\delta_{\mathbf{I}_n^k}^{opt} = \frac{\delta \cdot \lambda}{\delta \cdot \lambda + (1-\delta) \cdot \mu\left(\mathbf{I}_n^k\right)} \quad (46)$$

so that

$$E\left[\tilde{\pi}, \nu^{opt}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right] = \delta_{\mathbf{I}_n^k}^{opt} + \left(1 - \delta_{\mathbf{I}_n^k}^{opt}\right) \cdot E\left[\tilde{\pi}, \mu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right]. \quad (47)$$

(iii) *Pessimistic updating.*

$$\nu^{pess}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right) = \left(1 - \delta_{\mathbf{I}_n^k}^{pess}\right) \cdot K_{\alpha+k,\beta+n-k}\pi^{\alpha+k-1}\left(1-\pi\right)^{\beta+n-k-1} \quad (48)$$

with

$$\delta_{\mathbf{I}_n^k}^{pess} = \frac{\delta \cdot (1-\lambda)}{\delta \cdot (1-\lambda) + (1-\delta) \cdot \mu\left(\mathbf{I}_n^k\right)} \quad (49)$$

so that

$$E\left[\tilde{\pi}, \nu^{pess}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right] = \left(1 - \delta_{\mathbf{I}_n^k}^{pess}\right) \cdot E\left[\tilde{\pi}, \mu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right]. \quad (50)$$

(iv) *Sarin-Wakker updating.*

$$\tilde{\nu}_{\tilde{\pi}}^{SW}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right) = K_{\alpha+k,\beta+n-k}\pi^{\alpha+k-1}\left(1-\pi\right)^{\beta+n-k-1} \quad (51)$$

so that

$$E\left[\tilde{\pi}, \tilde{\nu}_{\tilde{\pi}}^{SW}\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right] = E\left[\tilde{\pi}, \mu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right]. \quad (52)$$

Except for Sarin-Wakker updating, where the agent's estimator coincides with the additive estimator, the above neo-additive estimators are given as a weighted average between the additive estimator $E\left[\tilde{\pi}, \mu\left(\boldsymbol{\pi} \mid \mathbf{I}_n^k\right)\right]$ and the numbers λ (for full Bayesian learning), 1 (for optimistic Bayesian learning), and 0 (for pessimistic Bayesian learning), respectively. Thus, Bayesian learning in accordance with the update rules (i) - (iii) does not—in general—coincide with the additive estimator $E\left[\tilde{\pi}, \mu\left(\pi \mid \mathbf{I}_n^k\right)\right]$ whenever there is some positive initial degree of ambiguity, i.e., $\delta > 0$. More specifically, observe that for the update rules (i) - (iii) the limit estimator's weight on the true parameter value decreases in the prior additive probability attached to the limit information so that, in general, the estimate $E\left[\tilde{\pi}, \nu\left(\pi \mid \mathbf{I}_n^k\right)\right]$ is the more biased the smaller $\mu\left(\mathbf{I}_n^k\right)$ is. As an intuitive explanation of this formal relationship one could say that that agent's learning behavior puts the less weight on the sample information the more he is surprised to actually receive this information.

Under this paper's distributional assumptions we obtain the following formal expression for the probability $\mu\left(\mathbf{I}_n^k\right)$.

Observation 4: *The prior additive probability of receiving information $\mathbf{I}_n^k$ is given by*

$$\mu\left(\mathbf{I}_n^k\right) = \binom{n}{k} \frac{(\alpha + k - 1) \cdot \ldots \cdot \alpha \cdot (\beta + n - k - 1) \cdot \ldots \cdot \beta}{(\alpha + \beta + n - 1) \cdot \ldots \cdot (\alpha + \beta)}. \tag{53}$$

In the limit of a Bayesian learning process, the agent's posterior estimator of π will then converge with probability one to $E\left[\tilde{\pi}, \nu\left(\boldsymbol{\pi} \mid \mathbf{I}^*\right)\right]$ where $\mathbf{I}^*$ denotes the information according to which the agent has observed the occurrence of outcome H with relative frequency $\pi \in (0, 1)$. Suppose, for example, that $\alpha = \beta = 1$ so that the agent's beta prior is given as the uniform distribution. Substituting $\alpha = \beta = 1$ in (53) gives

$$\mu\left(\mathbf{I}_n^k\right) = \binom{n}{k} \frac{k!\,(n-k)!}{(n+1) \cdot n!} \tag{54}$$

$$= \frac{1}{n+1}. \tag{55}$$

Thus, if the agent's prior is given by the uniform distribution, ex ante he regards every possible observation regarding the sample mean as equally likely. Taking the limit of (55) shows that $\mu\left(\mathbf{I}^*\right) = 0$ for all $\mathbf{I}^*$ which gives us results (i)-(iii) of the following corollary. Result (iv) of the corollary follows from Proposition 2 (iv) combined with Observation 3.

Corollary 1. *Consider the case that $\alpha = \beta = 1$, i.e., the agent's prior over $\tilde{\pi}$ is given as the uniform distribution. Contingent on the applied update rule, the agent's estimator of the true probability of outcome H converges with probability one to the following values as n becomes large.*

(i) *Full Bayesian updating.*

$$E\left[\tilde{\pi}, \nu^{FB}\left(\boldsymbol{\pi} \mid \mathbf{I}^*\right)\right] = \lambda. \tag{56}$$

(ii) *Optimistic updating.*

$$E\left[\tilde{\pi}, \nu^{opt}\left(\boldsymbol{\pi} \mid \mathbf{I}^*\right)\right] = 1. \tag{57}$$

(iii) *Pessimistic updating.*

$$E\left[\tilde{\pi}, \nu^{pess}\left(\boldsymbol{\pi} \mid \mathbf{I}^*\right)\right] = 0. \tag{58}$$

(iv) *Sarin-Wakker updating.*

$$E\left[\tilde{\pi}, \tilde{\nu}_{\tilde{\pi}}^{SW}\left(\boldsymbol{\pi} \mid \mathbf{I}^*\right)\right] = \pi. \tag{59}$$

7 Concluding remarks

In this paper I have developed closed-form models of Bayesian learning for agents that express ambiguity attitudes by violating Savage's sure thing principle. The paper demonstrates that—for a neo-additive capacity framework—Bayesian learning with respect to either the full Bayesian, the optimistic, or the pessimistic update rule does not, in general, converge towards an event's true probability but results in a neo-additive capacity that stays bounded away from this probability. For example, Corollary 1 describes the extreme situation in which the agent completely ignores the statistical information in the limit of his learning process because he regards this information as highly "unlikely". In that case, Bayesian learning only re-enforces the agent's ambiguity attitudes expressed by his neo-additive beliefs and his update rule.

The paper also demonstrates that Bayesian learning with respect to the Sarin-Wakker update rule implies convergence towards the event's true probability. The problem with this finding, however, is its lack of descriptive appeal. Namely, Sarin-Wakker updating implies in our framework the arguably unrealistic feature that agents get completely rid of their ambiguity attitudes after only one round of updating.

To sum-up: The analysis of this paper suggests that Bayesian learners do not, in general, reduce their ambiguity through the observation of more and more statistical information. To the contrary, the formal findings about full Bayesian, optimistic and pessimistic updating demonstrate that for perfectly reasonable update rules an agent's ambiguity will increase rather than decrease through Bayesian learning.

Appendix: Formal proofs

Proof of Observation 1: By an argument in Schmeidler (1986), it suffices to restrict attention to a non-negative function f so that

$$E\left[f,\nu\right] = \int_0^{+\infty} \nu\left(\{\omega \in \Omega \mid f\left(\omega\right) \geq z\}\right) dz,$$

which is equivalent to

$$E\left[f,\nu\right] = \int_{\inf f}^{\sup f} \nu\left(\{\omega \in \Omega \mid f\left(\omega\right) \geq z\}\right) dz \tag{60}$$

since f bounded. We consider a partition P_n, $n = 1, 2, ...$, of Ω with members

$$A_n^k = \{\omega \in \Omega \mid a_{k,n} < f\left(\omega\right) \leq b_{k,n}\} \text{ for } k = 1, ..., 2^n$$

such that

$$a_{k,n} = \left[\sup f - \inf f\right] \cdot \frac{(k-1)}{2^n} + \inf f$$

$$b_{k,n} = \left[\sup f - \inf f\right] \cdot \frac{k}{2^n} + \inf f.$$

Define the step functions $a_n : \Omega \to \mathbb{R}$ and $b_n : \Omega \to \mathbb{R}$ such that, for $\omega \in A_n^k$, $k = 1, ..., 2^n$,

$$a_n\left(\omega\right) = a_{k,n}$$

$$b_n\left(\omega\right) = b_{k,n}.$$

Obviously,

$$E\left[a_n,\nu\right] \leq E\left[f,\nu\right] \leq E\left[b_n,\nu\right]$$

for all n and

$$\lim_{n \to \infty} E\left[b_n,\nu\right] - E\left[a_n,\nu\right] = 0.$$

That is, $E\left[a_n,\nu\right]$ and $E\left[b_n,\nu\right]$ converge to $E\left[f,\nu\right]$ for $n \to \infty$. Furthermore, observe that

$$\inf a_n = \inf f \text{ for all } n, \text{ and}$$

$$\sup b_n = \sup f \text{ for all } n.$$

Since $\lim_{n \to \infty} \inf b_n = \lim_{n \to \infty} \inf a_n$ and $E\left[b_n,\mu\right]$ is continuous in n, we have

$$\lim_{n \to \infty} E\left[b_n,\nu\right] = \delta\left(\lambda \lim_{n \to \infty} \sup b_n + (1-\lambda) \lim_{n \to \infty} \inf b_n\right) + (1-\delta) \lim_{n \to \infty} E\left[b_n,\mu\right]$$

$$= \delta\left(\lambda \sup f + (1-\lambda) \inf f\right) + (1-\delta) E\left[f,\mu\right].$$

In order to prove Observation 1, it therefore remains to be shown that, for all n,

$$E[b_n, \nu] = \delta \left(\lambda \sup b_n + (1 - \lambda) \inf b_n \right) + (1 - \delta) E[b_n, \mu].$$

Since b_n is a step function, (60) becomes

$$
\begin{aligned}
E[b_n, \nu] &= \sum_{A_n^k \in P_n} \nu \left(A_n^{2^n} \cup ... \cup A_n^k \right) \cdot (b_{k,n} - b_{k-1,n}) \\
&= \sum_{A_n^k \in P_n} b_{k,n} \cdot \left[\nu \left(A_n^{2^n} \cup ... \cup A_n^k \right) - \nu \left(A_n^{2^n} \cup ... \cup A_n^{k-1} \right) \right],
\end{aligned}
$$

implying for a neo-additive capacity

$$
\begin{aligned}
E[b_n, \nu] &= \sup b_n \left[\delta \lambda + (1 - \delta) \mu \left(A_n^{2^n} \right) \right] + \sum_{k=2}^{2^n - 1} b_{k,n} (1 - \delta) \mu \left(A_n^k \right) \\
&\quad + \inf b_n \left[1 - \delta \lambda - (1 - \delta) \sum_{k=2}^{2^n} \mu \left(A_n^k \right) \right] \\
&= \delta \lambda \sup b_n + (1 - \delta) \sum_{k=1}^{2^n} b_{k,n} \mu \left(A_n^k \right) + \inf b_n \left[\delta - \delta \lambda \right] \\
&= \delta \left(\lambda \sup b_n + (1 - \lambda) \inf b_n \right) + (1 - \delta) E[b_n, \mu].
\end{aligned}
$$

$\square$

Proof of Proposition 1.

Part (i). An application of the full Bayesian update rule to a neo-additive capacity gives

$$
\begin{aligned}
\nu^{FB}(A \mid B) &= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap B)}{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap B) + 1 - (\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cup \neg B))} \\
&= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap B)}{1 + (1 - \delta) \cdot (\mu(A \cap B) - \mu(A \cup \neg B))} \\
&= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap B)}{1 + (1 - \delta) \cdot (\mu(A \cap B) - \mu(A) - \mu(\neg B) + \mu(A \cap \neg B))} \\
&= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap B)}{1 + (1 - \delta) \cdot (-\mu(\neg B))} \\
&= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap B)}{\delta + (1 - \delta) \cdot \mu(B)} \\
&= \frac{\delta \cdot \lambda}{\delta + (1 - \delta) \cdot \mu(B)} + \frac{(1 - \delta) \cdot \mu(B)}{\delta + (1 - \delta) \cdot \mu(B)} \mu(A \mid B) \\
&= \delta_B^{FB} \cdot \lambda + \left(1 - \delta_B^{FB} \right) \cdot \mu(A \mid B)
\end{aligned}
$$

26

such that

$$\delta_B^{FB} = \frac{\delta}{\delta + (1 - \delta) \cdot \mu(B)}.$$

☐

Part(ii). An application of the optimistic Bayesian update rule to a neo-additive capacity gives

$$
\begin{aligned}
\nu^{opt}(A \mid B) &= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap B)}{\delta \cdot \lambda + (1 - \delta) \cdot \mu(B)} \\
&= \frac{\delta \cdot \lambda}{\delta \cdot \lambda + (1 - \delta) \cdot \mu(B)} + \frac{(1 - \delta) \cdot \mu(B)}{\delta \cdot \lambda + (1 - \delta) \cdot \mu(B)} \cdot \mu(A \mid B) \\
&= \delta_B^{opt} + \left(1 - \delta_B^{opt}\right) \cdot \mu(A \mid B)
\end{aligned}
$$

such that

$$\delta_B^{opt} = \frac{\delta \cdot \lambda}{\delta \cdot \lambda + (1 - \delta) \cdot \mu(B)}.$$

☐

Part (iii). An application of the pessimistic Bayesian update rule to a neo-additive capacity gives

$$
\begin{aligned}
\nu^{pess}(A \mid B) &= \frac{\nu(A \cup \neg B) - \nu(\neg B)}{1 - \nu(\neg B)} \\
&= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cup \neg B) - \delta \cdot \lambda - (1 - \delta) \cdot \mu(\neg B)}{1 - \delta \cdot \lambda - (1 - \delta) \cdot \mu(\neg B)} \\
&= \frac{(1 - \delta) \cdot \mu(\neg(\neg A \cap B)) - (1 - \delta) \cdot \mu(\neg B)}{1 - \delta \cdot \lambda - (1 - \delta) \cdot \mu(\neg B)} \\
&= \frac{(1 - \delta) \cdot (1 - \mu(\neg A \cap B)) - (1 - \delta) \cdot (1 - \mu(B))}{1 - \delta \cdot \lambda - (1 - \delta) \cdot \mu(\neg B)} \\
&= \frac{(1 - \delta) \cdot (\mu(B) - \mu(\neg A \cap B))}{1 - \delta \cdot \lambda - (1 - \delta) \cdot (\mu(\neg B))} \\
&= \frac{(1 - \delta) \cdot (\mu(B) - \mu(B)\mu(\neg A \mid B))}{1 - \delta \cdot \lambda - (1 - \delta) \cdot (\mu(\neg B))} \\
&= \frac{(1 - \delta) \cdot (\mu(B) - \mu(B)(1 - \mu(A \mid B)))}{1 - \delta \cdot \lambda - (1 - \delta) \cdot \mu(\neg B)} \\
&= \left(1 - \delta_B^{pess}\right) \cdot \mu(A \mid B)
\end{aligned}
$$

such that

$$\delta_B^{pess} = \frac{\delta(1 - \lambda)}{\delta(1 - \lambda) + (1 - \delta) \cdot \mu(B)}.$$

☐

Part (iv). An application of the Sarin Wakker update rule to a neo-additive capacity gives

$$
\begin{aligned}
\tilde{\nu}_f \left(A \mid B \right) &= \frac{\delta \cdot \lambda + (1 - \delta) \, \mu \left((A \cap B) \cup D_f \left[A \cap B \right] \right) - \left[\delta \cdot \lambda + (1 - \delta) \, \mu \left(D_f \left[A \cap B \right] \right) \right]}{\delta \cdot \lambda + (1 - \delta) \, \mu \left(B \cup D_f \left[B \right] \right) - \left[\delta \cdot \lambda + (1 - \delta) \, \mu \left(D_f \left[B \right] \right) \right]} \\
&= \frac{\mu \left((A \cap B) \cup D_f \left[A \cap B \right] \right) - \mu \left(D_f \left[A \cap B \right] \right)}{\mu \left(B \cup D_f \left[B \right] \right) - \mu \left(D_f \left[B \right] \right)} \\
&= \frac{\mu \left(A \cap B \right)}{\mu \left(B \right)},
\end{aligned}
$$

since E and $D_f \left[E \right]$ are, by construction, disjoint events for all E.

Proof of Observation 4: An application of Bayes' rule gives

$$
\begin{aligned}
\mu \left(\mathbf{I}_n^k \right) &= \frac{\mu \left(\mathbf{I}_n^k \mid \boldsymbol{\pi} \right) \mu \left(\boldsymbol{\pi} \right)}{\mu \left(\boldsymbol{\pi} \mid \mathbf{I}_n^k \right)} \\
&= \frac{\binom{n}{k} \pi^k \left(1 - \pi \right)^{n-k} \mu \left(\boldsymbol{\pi} \right) \cdot K_{\alpha,\beta} \pi^{\alpha-1} \left(1 - \pi \right)^{\beta-1}}{K_{\alpha+k,\beta+n-k} \pi^{\alpha+k-1} \left(1 - \pi \right)^{\beta+n-k-1}} \\
&= \binom{n}{k} \frac{K_{\alpha,\beta}}{K_{\alpha+k,\beta+n-k}} \\
&= \binom{n}{k} \frac{\Gamma \left(\alpha + \beta \right)}{\Gamma \left(\alpha \right) \Gamma \left(\beta \right)} \frac{\Gamma \left(\alpha + k \right) \Gamma \left(\beta + n - k \right)}{\Gamma \left(\alpha + \beta + n \right)} \\
&= \binom{n}{k} \frac{\Gamma \left(\alpha + \beta \right)}{\Gamma \left(\alpha + \beta + n \right)} \cdot \frac{\Gamma \left(\alpha + k \right)}{\Gamma \left(\alpha \right)} \cdot \frac{\Gamma \left(\beta + n - k \right)}{\Gamma \left(\beta \right)} \\
&= \binom{n}{k} \frac{\Gamma \left(\alpha + \beta \right)}{\left(\alpha + \beta + n - 1 \right) \cdot \ldots \cdot \left(\alpha + \beta \right) \cdot \Gamma \left(\alpha + \beta \right)} \\
&\quad \cdot \frac{\left(\alpha + k - 1 \right) \cdot \ldots \cdot \alpha \cdot \Gamma \left(\alpha \right)}{\Gamma \left(\alpha \right)} \\
&\quad \cdot \frac{\left(\beta + n - k - 1 \right) \cdot \ldots \cdot \beta \cdot \Gamma \left(\beta \right)}{\Gamma \left(\beta \right)} \\
&= \binom{n}{k} \frac{\left(\alpha + k - 1 \right) \cdot \ldots \cdot \alpha \cdot \left(\beta + n - k - 1 \right) \cdot \ldots \cdot \beta}{\left(\alpha + \beta + n - 1 \right) \cdot \ldots \cdot \left(\alpha + \beta \right)}
\end{aligned}
$$

whereby the last equality readily follows from the fact that $\Gamma \left(x \right) = \left(x - 1 \right) \cdot \Gamma \left(x - 1 \right)$ for $x > 1$ (cf. Theorem 8.18 in Rudin 1976).$\square$

References

Abel, A. B. 2002. An exploration of the effects of pessimism and doubt on asset returns. *Journal of Economic Dynamics and Control* **26**(7-8) 1075-1092.

Adam, K., A. Marcet, J. P. Nicolini. 2008. Stock market volatility and learning. ECB Working Paper **862** European Central Bank.

Anscombe, F. J., R. J. Aumann. 1963. A definition of subjective probability. *Annals of American Statistics* **34**(1) 199-205.

Barberis, N., R. Thaler. 2003. A survey of behavioral finance. In: G. M. Constantinides, M. Harris, R. M. Stulz [eds.] *Handbook of the Economics and Finance* pp. 1053-1123. Elsevier, Amsterdam.

Barsky, R. B., J. B. DeLong. 1993. Why does the stock market fluctuate? *Quarterly Journal of Economics* **108**(2) 291-311.

Blackwell, D., L. Dubins. 1962. Merging of opinions with increasing information. *The Annals of Mathematical Statistics* **33**(3) 882-886.

Brav, A., J. B. Heaton. 2002. Competing theories of financial anomalies. *Review of Financial Studies* **15**(2) 575-606.

Breiman, D., L. LeCam, L. Schwartz. 1964. Consistent estimates and zero-one sets. *The Annals of Mathematical Statistics* **35**(1) 157-161.

Cecchetti, S. G., P. Lam, C. M. Nelson. 2000. Asset pricing with distorted beliefs: Are equity returns too good to be true? *American Economic Review* **90**(4) 787-805.

Chateauneuf, A., J. Eichberger, S. Grant. 2007. Choice under uncertainty with the best and worst in mind: Neo-additive capacities. *Journal of Economic Theory* **137**(1) 538-567.

Cogley, T., T. J. Sargent. 2008. The market price of risk and the equity premium: A legacy of the Great Depression? *Journal of Monetary Economics* **55**(3) 454-476.

Cohen, M., I. Gilboa, J. Y. Jaffray, D. Schmeidler. 2000. An experimental study of updating ambiguous beliefs. *Risk, Decision, and Policy* **5**(2) 123–133.

Cyert, R. M., M. H. DeGroot, C. A. Holt. 1978. Sequential investment decisions with Bayesian learning. *Management Science* **24**(7) 712-718.

Diaconis, P., D. Freedman. 1986. On the consistency of Bayes estimates. *The Annals of Statistics* **14**(1) 1-26.

Doob, J. L. 1949. Application of the theory of martingales. In: *Le Calcul des Probabilites et ses Applications*, Colloques Internationaux du Centre National de la Recherche Scientifique 13, pp. 23–27. Paris: CNRS.

Eichberger, J., D. Kelsey. 1999. E-capacities and the Ellsberg paradox. *Theory and Decision* **46**(2) 107-140.

Eichberger, J., S. Grant, D. Kelsey. 2006. Updating Choquet beliefs. *Journal of Mathematical Economics* **43**(7-8) 888-899.

Ellsberg, D. 1961. Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics* **75**(4) 643-669.

Epstein, L. G. 1999. A definition of uncertainty aversion. *The Review of Economic Studies* **66**(3) 579-608.

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

Epstein, L. G., M. Schneider. 2007. Learning under ambiguity. *Review of Economic Studies* **74**(4) 1275-1303.

Freedman, D. A. 1963. On the asymptotic behavior of Bayes' estimates in the discrete case. *The Annals of Mathematical Statistics* **33**(4) 1386-1403.

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

Ghirardato, P., M. Marinacci. 2002. Ambiguity made precise: A comparative foundation. *Journal of Economic Theory* **102**(2) 251–289.

Gilboa, I. 1987. Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics* **16**(1) 65-88.

Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with non-unique priors. *Journal of Mathematical Economics* **18**(2) 141-153.

Gilboa, I., D. Schmeidler. 1993. Updating ambiguous beliefs. *Journal of Economic Theory* **59**(1) 33-49.

Hammond, P. J. 1988. Consequentialist foundation for expected utility. *Theory and Decision* **25**(1) 25-78.

Hanany E., P. Klibanoff. 2007. Updating preferences with multiple priors. *Theoretical Economics* **2**(3) 261–298.

Hanany E., P. Klibanoff. 2009. Updating ambiguity averse preferences. *The B.E. Journal of Theoretical Economics* **9** (Advances) Article 37.

Hansen, L. P., T. J. Sargent. 2001. Robust control and model uncertainty. *American Economy Review Papers and Proceedings* **91**(2) 60-66.

Hansen, L. P., T. J. Sargent. 2007. *Robustness.* Princeton University Press, Princeton.

Hansen, L. P., T. J. Sargent, T. D. Tallarini. 1999. Robust permanent income and pricing. *The Review of Economic Studies* **66**(4) 873-907.

Holmes, Jr, R. M., P. Bromiley, C. E. Devers, T. R. Holcomb, J. B. McGuire. 2011. Management theory applications of prospect theory: Accomplishments, challenges, and opportunities. *Journal of Management* **37** (4) 1069-1107.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263-291.

Lijoi, A., I. Pruenster, S. O. Walker. 2004. Extending Doob's consistency theorem to nonparametric densities. *Bernoulli* **10**(4) 651-663.

Ludwig, A., A. Zimper. 2011a. A parsimonious Choquet model of subjective life-expectancy. mimeo.

Ludwig, A., A. Zimper. 2011b. Choquet Bayesian learning with an application to the risk-free rate puzzle. mimeo.

Machina, M. 1989. Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature* **27**(4) 1622-1668.

Maenhout, P. J. 2004. Robust portfolio rules and asset pricing. *The Review of Financial Studies* **17**(4) 951-983.

Marinacci, M. 1999. Limit laws for non-additive probabilities and their frequentist interpretation. *Journal of Economic Theory* **84**(2) 145-195.

Marinacci, M. 2002. Learning from ambiguous urns. *Statistical Papers* **43**(1) 145-151.

Mehra, R., E. C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* **15**(2) 145-161.

Muth, J. F. 1961. Rational expectations and the theory of price movements. *Econometrica* **29**(3) 315-335.

Pires, C. P. 2002. A rule for updating ambiguous beliefs. *Theory and Decision* **53**(2) 137-152.

Quiggin, J. P. 1981. Risk perception and risk aversion among Australian farmers. *Australian Journal of Agricultural Economics* **25**(2) 160-169.

Quiggin, J. P. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3**(4) 323-343.

Rudin, W. 1976. *Principles of Mathematical Analysis*. McGraw-Hill, New York.

Sarin, R., P. P. Wakker. 1992. A simple axiomatization of nonadditive expected utility. *Econometrica* **60**(6) 1255-1272.

Sarin, R., P. P. Wakker. 1998a. Revealed likelihood and Knightian uncertainty. *Journal of Risk and Uncertainty* **16**(3) 223-250.

Sarin, R., P. P. Wakker. 1998b. Dynamic choice and nonexpected utility. *Journal of Risk and Uncertainty* **17**(2) 87-119.

Savage, L. J. 1954. *The Foundations of Statistics*. John Wiley and & Sons, Inc., New York.

Schmeidler, D. 1986. Integral representation without additivity. *Proceedings of the American Mathematical Society* **97**(2) 255-261.

Schmeidler, D. 1989. Subjective probability and expected utility without additivity. *Econometrica* **57**(3) 571-587.

Siniscalchi, M. 2010. Dynamic choice under ambiguity. *Theoretical Economics* (forthcoming).

Timmermann, A. 1993. How learning in financial markets generates excess volatility and predictability in stock prices. *Quarterly Journal of Economics* **108**(4) 1135-1145.

Tonks, I. 1983. Bayesian learning and the optimal investment decision of the firm. *The Economic Journal* **93** (Supplement: Conference papers) 87-98.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty* **5**(4) 297-323.

Tversky, A., P. P. Wakker. 1995. Risk attitudes and decision weights. *Econometrica* **63**(6) 1255-1280.

Viscusi, W. K. 1979. *Employment hazards: An investigation of market performance.* Harvard University Press, Cambridge, Mass.

Viscusi, W. K. 1985. A Bayesian perspective on biases in risk perception. *Economics Letters* **17**(1-2) 59-62.

Viscusi, W. K. 1990. Do smokers underestimate risks? *The Journal of Political Economy* **98**(6) 1253-1269.

Viscusi, W. K. 1991. Age variations in risk perceptions and smoking decisions. *The Review of Economics and Statistics* **73**(4) 577-588.

Viscusi, W. K., C. J. O'Connor. 1984. Adaptive responses to chemical labeling: Are workers Bayesian decision makers? *The American Economic Review* **74**(5) 942-956.

Wakker, P. P. 2001. Testing and characterizing properties of nonadditive measures through violations of the sure-thing principle. *Econometrica* **69**(4) 1039-1059.

Wakker, P. P. 2004. On the composition of risk preference and belief. *Psychological Review* **111**(1) 236-241.

Wakker, P. P. 2010. *Prospect Theory: For Risk and Ambiguity.* Cambridge University Press, Cambridge, UK.

Wakker, P. P., A. Tversky. 1993. An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty* **7**(2) 147-176.