Bayesian consistent belief selection

Christopher P. Chambers*and Takashi Hayashi[†‡§]

August 2005

Abstract

A subjective expected utility agent is given information about the state of the world in the form of a set of possible priors. She is assumed to form her beliefs given this information. A set of priors may be updated according to Bayes' rule, prior-by-prior, upon learning that some state of the world has not obtained. In a model in which information is completely summarized by this set of priors, we show that there exists no decision maker who obeys Bayes' rule, conditions her prior only on the available

*Division of the Humanities and Social Sciences, Mail Code 228-77, California Institute of Technology, Pasadena, CA 91125. Email: chambers@hss.caltech.edu. Phone: (626) 395-3559.

[†]Department of Economics, University of Texas at Austin, BRB 1.116, Austin, TX 78712. Email: th925@eco.utexas.edu. Phone: (512) 475-8543.

[‡]Corresponding author.

[§]We would like to thank Kim Border, Federico Echenique, Larry Epstein, Bart Lipman, Max Stinchcombe, John Quiggin, and Bill Zame for helpful discussions and comments. The associate editor and two anonymous referees also provided comments which were very helpful. All errors are our own.

information (by selecting a belief in the announced set), and who updates the information prior-by-prior using Bayes' rule.

1 Introduction

The classical theories of Savage [18] and Anscombe and Aumann [2] provide behavioral foundations in environments of subjective uncertainty justifying the hypothesis of subjective expected utility. The theories are elegant and the behavioral conditions posited are intuitive. However, a significant gap in these theories is that they do not specify how such a probability measure should be formed.

Recently, several models have been proposed which attempt to fill this gap. One example is the case-based decision theory of Gilboa and Schmeidler [13]. In this model, past experiences of the decision maker are explicitly modelled and incorporated in the formation of a prior. Other models assume the decision maker is given some information about the true state of the world, in the form of a *set of possible objective probabilities* (the literature is initiated by Damiano [3], see also Ahn [1], Gajdos et al. [7, 8], Stinchcombe [20], and Wang [21]). The decision-maker is given the objective information that the "true" objective probability lies in some set. She then forms a subjective probability which depends on this information. This is the approach we follow.

While our study is not motivated by ambiguity, it helps to consider the classical setup of the Ellsberg paradox [4]. An urn contains balls of many colors. The decision maker is told that the composition (relative proportions of colors) of the balls in the urn lies within some set, but is told nothing else. A ball is drawn, and the state of the world is the color of ball drawn. Mathematically, a composition of balls in an urn is identical to a probability measure over the colors. The decision maker's subjective probability over the states of the world is formed as a function of this information (the set of possible probabilities) she is given. She forms a rule which specifies, for every possible set of probability measures, a subjective probability over the states of the world. The assumption here is that *information available to the decision maker in forming a subjective belief is completely summarized by a set of probability measures.*

Our result demonstrates a logical conflict that revelation of information about the state of the world produces in such a model. When it is revealed that the ball drawn is not green, for example, we imagine our decision maker updates her information in a natural sense–she updates each probability measure in the set according to Bayes' rule. We assume that the *only* informational content of this information revelation is contained in the set of updated probability measures. The tension results when we also require our decision maker to use Bayes' rule in updating *her subjective probability.* Our result demonstrates that when she uses Bayes' rule, her posterior belief is typically different than the one she would have chosen had she used her rule and the objective information at hand, *no matter what rule she uses.*

The result is thus a logical incompatibility between the idea that a decision maker only has information available in the form of a set of probability measures, and the idea that a decision maker should use Bayes' rule.

Under our maintained assumption, updating the set of probability measures according to Bayes' rule prior-by-prior is the result of an *objective* or *logical deduction process*. Imagine that there are just two possible probabilities: either p or q. Suppose that the decision maker learns that the event E is true. If p were true, the true conditional probability would be $p(\cdot|E)$. If q were true, the true conditional probability would be $q(\cdot|E)$. Since only p or q could be true, we may conclude that the true conditional probability is either $p(\cdot|E)$ or $q(\cdot|E)$. However, under the hypothesis that all information available is summarized completely by the set of probability measures, there is no additional information gained by the revelation of E as to *which* of these conditional probabilities is correct. The informational content of learning E is the set of updated measures. This stands in stark contrast to models which feature a set of subjective priors, where different subjective updating rules have been proposed (see, for example, Gilboa and Schmeidler [12]).

The subjective expected utility hypothesis is pervasive in economic modelling, and it is with this motivation that we study Bayesian decision makers. To understand why we believe our approach holds some interest, consider the following thought experiment. A decision maker might be placed in one of two situations–one in which he faces an urn containing red and blue balls, and is told nothing about the composition; and another in which an urn contains red, blue, and green balls, again is told nothing about the composition, but is told that the ball drawn will not be green. In forming a belief over the states "red" and "blue," while the two situations are not informationally *identical*, they appear at least to be informationally *symmetric*. The information that the decision maker has about the likelihood of drawing a red or blue ball in either situation is identical. As green balls cannot be drawn, the only *seemingly* relevant information is the information about the composition of red and blue balls. Hence, the presence of the undrawn green balls in the second situation appears superfluous, and it is difficult to think of a logical reason why any "rational" individual would distinguish between the two situations. Our result shows that this reasoning is logically unsound for a Bayesian; the presence of the seemingly irrelevant green balls cannot be superfluous.

Ideas similar to ours are found in the statistics literature on probability aggregation. Here, a collection of experts each form probabilistic beliefs over some collection of events. The interest is in aggregating these probability measures into a group probability measure. A classical reference to this literature is Genest and Zidek [10]. Genest [9] studies those aggregation procedures which commute with respect to the application of likelihood functions (which are nonzero-valued). Using this axiom, Genest characterizes the "log-opinion pool." While this axiom is conceptually similar to ours, it results in a possibility while ours results in an impossibility.

Section 2 discusses and proves the main theorem. Section 3 discusses the set-valued selection case. Section 4 concludes. An Appendix discusses an environment in which the all events which can be realized belong to a filtration.

2 The model and result

Let Ω be a finite set of states of the world. For any nonempty subset $E \subset \Omega$, the set of probability measures over E is denoted $\Delta(E)$. The set of all nonempty, convex, and compact subsets of $\Delta(E)$ consisting only of measures having full support on E is denoted $\mathcal{K}^{fs}(E)$.

A **belief selection problem** consists of a nonempty subset of Ω, say, E, and a set $P \in \mathcal{K}^{fs}(E)$. The domain of all belief selection problems is denoted $\mathcal{X}^{fs}$. For all nonempty $E \subset \Omega$, and for all nonempty $F \subset E$, for all $P \subset \mathcal{K}^{fs}(E)$, the set $P^F \subset \mathcal{K}^{fs}(F)$ is given by

$$P^F \equiv \left\{ \frac{p}{p(F)}|_{2^F} : p \in P \right\}.$$

Here, 2^F refers to the subsets of F. Therefore, $\left(\frac{p}{p(F)}\right)|_{2^F}$ is the restriction of $\frac{p}{p(F)}$ to subsets of F. Hence, P^F is simply that set of probabilities that results from updating P prior-by-prior, restricted so that their support lies in F. To save on notation, we will often drop the "$|_{2^F}$" notation when its requirement is obvious.

A **belief selection rule** is a function $\psi : \mathcal{X}^{fs} \to \bigcup_{E \in 2^\Omega \setminus \varnothing} \Delta(E)$, such that for all $(E, P) \in \mathcal{X}^{fs}$, $\psi(E, P) \in P$.

While our model is non-behavioral, it is intended to be understood as a product of a model which explicitly describes behavior. Specifically, we can imagine a set of outcomes X, and a corresponding set of Anscombe-Aumann acts, $\mathcal{F} \equiv \Delta(X)^\Omega$. Suppose now that for each $(E, P) \in \mathcal{X}^{fs}(E)$, there is a preference relation over $\mathcal{F}|_E$, denoted by $\succsim_{(E,P)}$. The well-known Anscombe-Aumann axioms [2], applied to each $\succsim_{(E,P)}$, deliver a pair $\left(u_{(E,P)}, p_{(E,P)}\right)$, where $u_{(E,P)}$ is a cardinally unique expected utility functional $u_{(E,P)} : \Delta(X) \to \mathbb{R}$, and $p_{(E,P)}$ is a unique probability measure on 2^E. By postulating that for all $x, y \in \Delta(X)$ (here, we abuse notation in a standard way by identifying constant acts with the value the constant act takes) and for all $(E,P),(E',P') \in \mathcal{X}^{fs}$, $x \succsim_{(E,P)} y \iff x \succsim_{(E',P')} y$, we can guarantee that for all $(E,P),(E',P') \in \mathcal{X}^{fs}$, $u_{(E,P)} = u_{(E',P')}$. That is, all such representations are unique up to affine transformation, and can be thus chosen to be identical as there are no choices across belief selection problems. This is a normalization and has no real content–it is similar to the "state-independence" normalization of Anscombe and Aumann. See Fishburn [6] or Karni [15] for more on this point. To tie subjective belief to objective probability, require that for all $E \subset \Omega$, all $F \subset E$, all $p \in \Delta(E)$, and all $x, y \in \Delta(X)$, $xFy \sim_{(E,\{p\})} p(F)x + (1 - p(F))y$.[1] This implies that for all $p \in \Delta(E)$, $p_{(E,\{p\})} = p$. Lastly, we assert a simple dominance condition: for all (E,P), and all $f, g \in \mathcal{F}|_E$, if $f \succsim_{(E,\{p\})} g$ for all $p \in P$, then $f \succsim_{(E,P)} g$. The dominance requirement implies by a simple separation argument that for all (E,P), $p_{(E,P)} \in P$. This primitive model implies a model of belief selection whereby $\psi(E,P) = p_{(E,P)}$.[2]

[1] Here, $xFy(\omega) = x$ if $\omega \in F$, and y otherwise. $p(F)x + (1 - p(F))y$ is the lottery placing probability $p(F)$ on x and $1 - p(F)$ on y.

[2] Typically, models incorporating objective information differ in two respects. First, preference is usually non-parametric, and the set of states under consideration is always assumed to be Ω (as opposed to E, which can happen in our model). Our work formally incorporates conditional preference in order to talk about information revelation. Second, such models usually study act-information pairs; whereby comparisons of the sort $(f,P) \succsim (g,Q)$ are made. Such

In choosing to model belief selection in the way that we have, we are assuming the following:

Assumption 1: The totality of objective information available to the decision maker in forming a belief is summarized by a set of probability measures.

This assumption is implicit in the specification of our domain. To remove the assumption, we would have to enlarge the domain of a belief selection rule to include other types of information. Importantly, in a dynamic setup, after an event is revealed, the decision maker is not allowed to use information about the previous belief selection problem being faced.

Our main interest is in studying the prior-by-prior updating rule for sets of priors, and in understanding when a belief selection rule commutes with respect to Bayesian updating. Under Assumption 1, the objective informational content of learning an event has obtained must be contained solely in some set of probability measures. Here, we make explicit that this set is the set of prior-by-prior updates of the set of probability measures.

Assumption 2: The only informational content of learning an event has obtained is contained in the set of updated probability measures.

A belief selection rule is **Bayesian consistent** if for all $(E, P) \in \mathcal{X}^{fs}$, and for all nonempty $F \subset E$ for which $\psi(E,P)(F) \neq \varnothing$, $\psi\left(F, P^F\right) = \frac{\psi(E,P)}{\psi(E,P)(F)}$. Bayesian consistency thus ties together two methods of forming beliefs and

a preference induces an obvious conditional preference on the set of acts, conditional on the set of probabilities under consideration.

requires their consistency. One method is described in Assumption 1 and Assumption 2. The other method is simply Bayes' rule. Bayesian consistency can be viewed as an analogue of Savage's [18] sure thing principle in the behavioral context. Specifically, it states that for all $(E, P) \in \mathcal{X}^{fs}$, all $f, f', g \in \mathcal{F}|_E$, and all $F \subset E$, $fFg \succsim_{(E,P)} f'Fg \iff f|_F \succsim_{(F,P^F)} f'|_F$. The decision maker facing (E, P) and a comparison between fFg and $f'Fg$ should "look ahead," and evaluate these acts based upon what she would do conditional on the event F.

Assumption 3: Upon the revelation of information, the decision maker must update her subjective probability according to Bayes' rule.

Bayesian consistency thus postulates the conjunction of Assumptions 1, 2, and 3. It requires that a decision maker using Bayes' rule does not come into conflict with her pre-specified selection rule upon learning an event has obtained, when the only informational content of learning is the set of updated probability measures.

Our primary result establishes that there is no rule satisfying Bayesian consistency. The conjunction of the three Assumptions are mutually incompatible.

Theorem 1: Suppose that $|\Omega| \geq 3$. Then there exists no Bayesian consistent belief selection rule.

Proof. Let $E \subset \Omega$ such that $|E| = 3$. Without loss of generality, label $E \equiv \{a, b, c\}$. We will construct four elements of $\mathcal{K}^{fs}(E)$. These four sets are illustrated in Figure 1. To do so, we need to define some preliminary elements of $\Delta(E)$. Define $\{p^i\}_{i=1}^6$ as follows:

	a	b	c
p^1	$3/13$	$9/13$	$1/13$
p^2	$3/7$	$3/7$	$1/7$
p^3	$1/5$	$3/5$	$1/5$
p^4	$1/3$	$1/3$	$1/3$
p^5	$3/5$	$1/5$	$1/5$
p^6	$3/7$	$1/7$	$3/7$

.

We define the following four elements of $\mathcal{K}^{fs}(E)$. Here, conv denotes the convex hull.

$$\begin{aligned} P_1 &\equiv \text{conv}\{p^1, p^2, p^3\}, \\ P_2 &\equiv \text{conv}\{p^2, p^3, p^4\}, \\ P_3 &\equiv \text{conv}\{p^2, p^4, p^5\}, \\ P_4 &\equiv \text{conv}\{p^4, p^5, p^6\}. \end{aligned}$$

Consider the problems $(E, P_1), (E, P_2), (E, P_3), (E, P_4) \subset \mathcal{X}^{fs}$. We claim that $P_1 \cap P_4 = \varnothing$. This is obvious; for all $p \in P_1$, $p(b) \geq 3p(c)$. However, for all $p \in P_4$, $p(c) \geq p(b)$.

We will now establish that $\psi(E, P_1) = \psi(E, P_4)$, which is a contradiction.

The argument is very geometric, so we will illustrate the first step in Figure 2.

Let $p^* = \psi(E, P_1)$. We claim that $P_1^{\{a,b\}} =$

$\operatorname{conv}\left\{\frac{p^1}{p^1(\{a,b\})},\frac{p^2}{p^2(\{a,b\})},\frac{p^3}{p^3(\{a,b\})}\right\}$. To see this, note that it follows by definition that $\operatorname{conv}\left\{\frac{p^1}{p^1(\{a,b\})},\frac{p^2}{p^2(\{a,b\})},\frac{p^3}{p^3(\{a,b\})}\right\}\subset P_1^{\{a,b\}}$. Conversely, let $p\in P_1^{\{a,b\}}$. Then $p=\lambda p^1+\mu p^2+\rho p^3$, where $\lambda+\mu+\rho=1$ and $\lambda,\mu,\rho\geq 0$. Consequently,

$$\begin{aligned}p\left(\cdot\right|\{a,b\}) &= \frac{\lambda p^1(\{a,b\})}{\lambda p^1(\{a,b\})+\mu p^2(\{a,b\})+\rho p^3(\{a,b\})}\left(\frac{p^1}{p^1(\{a,b\})}\right)\\ &\quad+\frac{\mu p^2(\{a,b\})}{\lambda p^1(\{a,b\})+\mu p^2(\{a,b\})+\rho p^3(\{a,b\})}\left(\frac{p^2}{p^2(\{a,b\})}\right)\\ &\quad+\frac{\rho p^3(\{a,b\})}{\lambda p^1(\{a,b\})+\mu p^2(\{a,b\})+\rho p^3(\{a,b\})}\left(\frac{p^3}{p^3(\{a,b\})}\right),\end{aligned}$$

so that $\frac{p}{p(\{a,b\})}\in\operatorname{conv}\left\{\frac{p^1}{p^1(\{a,b\})},\frac{p^2}{p^2(\{a,b\})},\frac{p^3}{p^3(\{a,b\})}\right\}$. Similarly, $P_1^{\{a,c\}}=\operatorname{conv}\left\{\frac{p^1}{p^1(\{a,c\})},\frac{p^2}{p^2(\{a,c\})},\frac{p^3}{p^3(\{a,c\})}\right\}$. By Bayesian consistency, $\psi\left(\{a,b\},P_1^{\{a,b\}}\right)=\frac{p^*}{p^*(\{a,b\})}$ and $\psi\left(\{a,c\},P_1^{\{a,c\}}\right)=\frac{p^*}{p^*(\{a,c\})}$.

The shaded lines on the faces of the simplex on Figure 2 are $P_1^{\{a,b\}}$ and $P_1^{\{a,c\}}$. In particular, this illustrates geometrically how to find the Bayesian update of an information set. It is a projection of the information set onto the corresponding face of the simplex from the opposite vertex.

Note that $P_2^{\{a,b\}}=P_1^{\{a,b\}}$ and that $P_2^{\{a,c\}}=P_1^{\{a,c\}}$. (This is also clearly visible from Figure 2). Therefore, $\psi\left(\{a,b\},P_2^{\{a,b\}}\right)=\frac{p^*}{p^*(\{a,b\})}$ and $\psi\left(\{a,c\},P_2^{\{a,c\}}\right)=\frac{p^*}{p^*(\{a,c\})}$. Let $p^{**}=\psi\left(E,P_2\right)$. By Bayesian consistency, $\frac{p^{**}}{p^{**}(\{a,b\})}=\frac{p^*}{p^*(\{a,b\})}$ and $\frac{p^{**}}{p^{**}(\{a,c\})}=\frac{p^*}{p^*(\{a,c\})}$. This is only possible if $p^{**}=p^*$. Hence, we conclude that $\psi\left(E,P_2\right)=\psi\left(E,P_1\right)$.

The remainder of the proof lies in establishing that $\psi\left(E,P_3\right)=\psi\left(E,P_2\right)$ and that $\psi\left(E,P_4\right)=\psi\left(E,P_3\right)$. That $\psi\left(E,P_3\right)=\psi\left(E,P_2\right)$ follows from the fact that $P_3^{\{a,c\}}=P_2^{\{a,c\}}$ and $P_3^{\{b,c\}}=P_2^{\{b,c\}}$, and an identical argument using

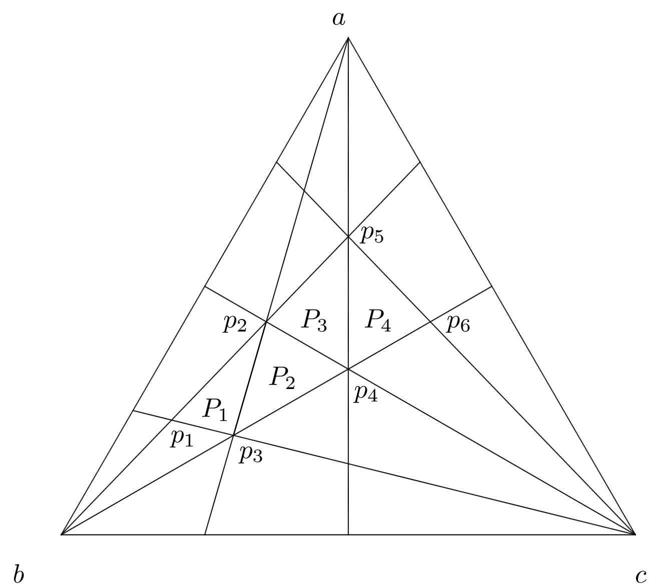

Figure 1: Information sets

Bayesian consistency.

That $\psi(E, P_4) = \psi(E, P_3)$ follows from the fact that $P_4^{\{a,b\}} = P_3^{\{a,b\}}$ and $P_4^{\{a,c\}} = P_3^{\{a,c\}}$, and an identical argument using Bayesian consistency. Hence, $\psi(E, P_4) = \psi(E, P_1)$, which demonstrates the existence of $p^* \in P_1 \cap P_4$, a contradiction. ■

We remark that the theorem is significantly robust; it can be generalized to apply to many restricted domains (including restrictions on the possible conditioning events or restrictions on the possible objective information sets). It can also be generalized to set-valued selections. However, the main intuition and message are clear enough from the simple case presented above.

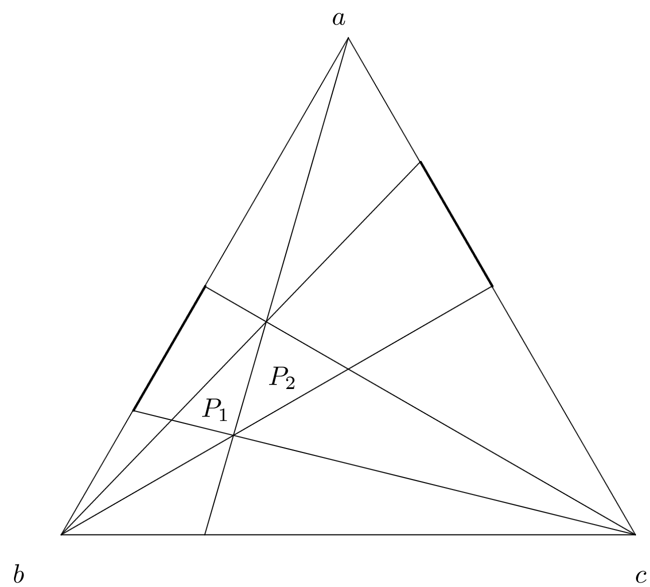

Figure 2: Constructing Bayesian updates of information sets

3 Discussion and conclusion

Our result demonstrates that any Bayesian decision maker who forms a belief when given imprecise information must use something more than just the set of probability measures available at hand. The only time a decision maker can truly use *only this information* is at time zero–when no events have been realized. Otherwise, a decision maker comes into conflict with Bayes' rule. A consequence of this result is the fact that the timing of the resolution of subjective uncertainty is relevant for understanding the beliefs of a decision maker. In forming beliefs, it is not only relevant what "objective" information is available to them, but also how they arrived at that information. An outsider observing two decision makers with seemingly identical objective information should not conclude that their subjective beliefs are the same. This may have implications for the common prior doctrine for example (which maintains that agents who have the same objective information should hold the same prior belief).

We close by connecting our work to the literature on dynamic consistency with non-separable preferences. A maxmin expected utility (as in Gilboa and Schmeidler [11]) decision maker is characterized by a set of endogenously defined subjective probability measures and a utility index. Such a decision maker evaluates the utility of an act by finding the minimum expected utility of the act across all of the measures in the set. As these preferences are inherently non-separable across mutually exclusive events, preference over acts conditional on any given event necessarily depends on the original decision problem faced. A good reference for understanding this phenomenon (in the case of objective uncertainty) is Machina [16]. Different models have been proposed to discuss this phenomenon. Recent works include Epstein and Schneider [5], who restrict the set of events which can be conditioned upon, and Hanany and Klibanoff [14], who discuss update rules which depend on the original problem under consideration. Our set of objective probabilities bears formal resemblance to the set of subjective priors endogenously derived in these models, and thus it is reasonable to investigate the connection between the two. It is true that in posulating a model of conditional preferences in the maxmin expected utility framework, the preference which is generated by the prior-by-prior update rule is not in general dynamically consistent. This, however, does not imply, nor is implied by, our result. A maxmin expected utility maximizer is endowed with a *single* set of probability measures–lack of dynamic consistency is equivalent to the absence of additive separability of the support function of this set. Our result, in contrast, requires *at least four* sets of probability measures to be considered–at least in our proof. Our result relies on the impossibility of identifying a set of priors uniquely by its set of Bayesian updates on two events. Put differently, if the domain were restricted to include only one objective information set and its updates, Bayesian consistency would be trivial to satisfy. It is also conceivable that different "objective" updating rules might allow for Bayesian consistency to be satisfied; but the prior-by-prior update is the only one following from a logical deduction process.

References

[1] D.S. Ahn, Ambiguity without a state space, *Review of Economic Studies* 75 (2008), 3-28.

[2] F.J. Anscombe and R. Aumann, A definition of subjective probability, *Annals of Mathematical Statistics* 34 (1963), 199-205.

[3] E. Damiano, Choice under limited uncertainty, *Advances in Theoretical Economics* 6 (2006), article 5.

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

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

[6] P.C. Fishburn, A mixture set axiomatization of conditional expected utility, *Econometrica* 41 (1973), 1-25.

[7] T. Gajdos, J.-M. Tallon, and J.-C. Vergnaud, Decision making with imprecise probabilistic information, *Journal of Mathematical Economics* 40 (2004), 647-681.

[8] T. Gajdos, T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud, Attitude toward imprecise information, *Journal of Economic Theory* 140 (2008), 27-65.

[9] C. Genest, A characterization theorem for externally Bayesian groups, *Ann. Stat.* 12 (1984), 1100-1105.

[10] C. Genest and J.V. Zidek, Combining probability distributions: a critique and an annotated bibliography, *Stat. Sci.* 1 (1986), 114-135.

[11] I. Gilboa and D. Schmeidler, Maxmin expected utility with nonunique prior, *Journal of Mathematical Economics* 18 (1989), 141-153.

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

[13] I. Gilboa and D. Schmeidler, Case-based decision theory, *Quarterly Journal of Economics* 110 (1995), 605-639.

[14] E. Hanany and P. Klibanoff, Updating preferences with multiple priors, *Theoretical Economics* 2 (2007), 261-298.

[15] E. Karni, A definition of subjective probabilities with state-dependent preferences, *Econometrica* 61 (1993), 187-198.

[16] M.J. Machina, Dynamic consistency and non-expected utility models of choice under uncertainty, *Journal of Economic Literature* 28 (1989), 1622-1688.

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

[18] L.J. Savage, "The Foundations of Statistics," 2nd edition, Dover Publications, Minneola, NY 1972.

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

[20] M. Stinchcombe, Choice and games with ambiguity as sets of probabilities, Working Paper, University of Texas, Austin, 2003.

[21] T. Wang, A class of multi prior preferences, manuscript, 2003.