JOURNAL OF ECONOMIC THEORY 59, 33–49 (1993)

Updating Ambiguous Beliefs*

ITZHAK GILBOA

Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, Evanston, Illinois 60208

AND

DAVID SCHMEIDLER

Department of Economics, Ohio State University, Columbus, Ohio 43210, and Department of Statistics and Department of Economics, Tel Aviv University, Tel Aviv, Israel.

Received April 17, 1991; revised July 3, 1991.

We present and axiomatize several update rules for probabilities (and preferences) where there is no unique additive prior. In the context of non-additive probabilities we define and axiomatize Bayesian update rules; in the context of multiple (additive) priors we define maximum likelihood rules. It turns out that for decision makers which can be described by both theories, the two approaches coincide. Thus, we suggest an axiomatically based ambiguous beliefs update rule, which is needed for applications in many economic theory models. *Journal of Economic Literature* classification numbers: D80, D81, C11, C71.

1. INTRODUCTION

The Bayesian approach to decision making under uncertainty prescribes that a decision maker have a unique prior probability and a utility function such that decisions are made so as to maximize the expected utility. In particular, in a statistical inference problem the decision maker is assumed to have a probability distribution over all possible distributions which may govern a certain random process.

This paradigm was justified by axiomatic treatments, most notably that of Savage [26], and it enjoys unrivaled popularity in economic theory, game theory, and so forth.

However, this theory is challenged by two classes of evidence: on the one hand, there are experiments and thought experiments (such as Ellsberg

* We thank James Dow, Joe Halpern, Jean-Yves Jaffray, Bart Lipman, Klaus Nehring, Sergiu Werlang, and an anonymous referee for stimulating discussions and comments. NSF grants SES-9113108 and SES-9111873 are gratefully acknowledged.

0022-0531/93 $5.00

[12] and many others) which seem to show that individuals tend to violate the consistency conditions underlying (and implied by) the Bayesian approach. On the other hand, people seem to have difficulties with specification of a prior for actual statistical inference problems. Thus, classical—rather than Bayesian—methods are used for practical purposes, although they are theoretically less satisfactory.

The last decade has witnessed—among numerous and various generalizations of von Neumann and Morgenstern's [25] expected utility theory—generalizations of the Bayesian paradigm as well. We will not attempt to provide a survey of them here. Instead, we only mention the models which are relevant to the sequel.

1. *Non-additive Probabilities.* First introduced by Schmeidler [27, 29, 30] and also axiomatized in Gilboa [16], Fishburn [14], and Wakker [35], non-additive probabilities are monotone set-functions which do not have to satisfy additivity. Using the Choquet integral (Choquet [7]) one may define expected utility, and the works cited before axiomatize preference relations which are representable by expected utility in this sense.

2. *Multiple Priors.* As axiomatized by Gilboa and Schmeidler [18], this model assumes that the decision maker has a *set* of priors, and each alternative is assessed according to its minimal expected utility, where the minimum is taken over all priors in the set. (This idea is also related to Bewley [3–5], who suggests a partial order over alternatives, such that one alternative is preferred to another only if its expected utility is higher according to *all* priors in the set.)

Of particular interest to this study is the intersection of the two models: it turns out that if a non-additive measure exhibits uncertainty aversion (technically, if it is convex in the sense

$$v(A \cup B) + v(A \cap B) \geqslant v(A) + v(B)),$$

then the Choquet integral of a real-valued function with respect to v equals the minimum of all its integrals with respect to additive priors taken from the core of v. (The core is defined as in cooperative game theory, i.e., p is in the core of v if $p(A) \geqslant v(A)$ for every event A with equality for the whole sample space. Convex non-additive measures have nonempty cores.)

While these models—as many others—suggested generalizations of the Bayesian approach for a one-shot decision problem, they shed very little light on the problem of dynamically updating probabilities as new information is gathered. We find this problem to be of paramount importance for several interrelated reasons:

1. The theoretical validity of any model of decision making under uncertainty is quite dubious if it cannot cope successfully with the dynamic aspect.

2. The updating problem is at the heart of statistical theory. In fact, it may be viewed as the problem statistical inference is trying to solve. Some of the works in the statistical literature which pertain to this study are Agnew [1], Genest and Schervish [15], and Lindley, Tversky, and Brown [23].

3. Applications of these models to economic and game theory models require some assumptions on how economic agents change their beliefs over time. The question naturally arises, then: What are reasonable ways to update such beliefs?

4. The theory of artificial intelligence, which in general seems to have much in common with the foundations of economic, decision, and game theory, also tries to cope with this problem. See, for instance, Fagin and Halpern [13], Halpern and Fagin [19], and Halpern and Tuttle [20].

In this study we try to deal with the problem axiomatically and suggest plausible update rules which satisfy some basic requirements. We present a family of pseudo-Bayesian rules, each of which may be considered a generalization of Bayes' rule for a unique additive prior. We also present a family of "classical" update rules, each of which starts out with a given set of priors, possibly rules some of them out in the face of new information, and continues with the (Bayesian) updates of the remaining ones.

In particular, a maximum-likelihood update rule would be the following: consider only those priors which ascribe the maximal probability to the event that is known to have occurred, update each of them according to Bayes' rule, and continue in this fashion.

It turns out that if the set of priors one starts out with can also be represented by a non-additive probability, the results of this rule are independent of the order in which information is gathered.

Furthermore, for those preferences which can be simultaneously represented by a non-additive measure and by multiple priors, the maximum likelihood update rule coincides with one of the more intuitive Bayesian rules, and they boil down to the Dempster–Shafer rule (see Dempster [8, 9], Shafer [31], and Smets [34]). For recent work on belief functions and their updating, see Jaffray [21], Chateauneuf and Jaffray [6], and especially Jaffray [22].

Thus, we find that an axiomatically based generalization of the Bayesian approach can accommodate multiple priors (which are used in classical statistics). Moreover, the maximum likelihood principle, which is at the heart of statistical inference (and implicit in the techniques of confidence sets and hypothesis testing) coincides with the generalized Bayesian updating.

Due to the prominence of this rule, it may be a source of insight to study it in a simple example. Consider Ellsberg's example in which an urn with 90 balls is given, out of which 30 are red, and 60 are either blue or yellow. For simplicity of exposition, let us model this situation in a somewhat extreme fashion, allowing for *all* distributions of blue and yellow balls. Maxmin expected utility with respect to this set of priors is equivalent to the maximization of the Choquet integral of utility w.r.t. to a non-additive measure v defined as

$$v(R)=\tfrac{1}{3}, \qquad v(B)=v(Y)=0$$

$$v(R\cup B)=v(R\cup Y)=\tfrac{1}{3}, \qquad v(B\cup Y)=\tfrac{2}{3}$$

$$v(R\cup B\cup Y)=1,$$

where R, B, and Y denote the events of a red, blue, or yellow ball being drawn, respectively.

Assume now that it is known that a ball (which, say, has already been drawn) is not red. Conditioning on the event $B\cup Y$, *all* priors in the set ascribe probability of $\frac{2}{3}$ to it. Thus, they are all left in the set and updated according to Bayes' rule. This captures our intuition that no ambiguity was resolved, and our complete ignorance regarding the event $B\cup Y$ has not changed. (Actually, it is now highlighted by the fact that this event, about which we know the least, is now known to have occurred.)

Consider, on the other hand, the same update rule in the case that $R\cup B$ is known. The priors we started out with ascribe to this event probabilities ranging from $\frac{1}{3}$ to $\frac{2}{3}$. According to the maximum likelihood principle, only one of them is chosen—namely, the p which satisfies

$$p(R)=\tfrac{1}{3} \qquad p(B)=\tfrac{2}{3} \qquad p(Y)=0.$$

In this particular case, the set of priors shrinks to a singleton and, equivalently, the updated measure v is additive (and equals p itself). Ambiguity is thus reduced (in the case, eliminated) by the generalized Bayesian learning embodied in the exclusion of some priors.

In the context of such examples it is sometimes argued that the maximum-likelihood rule is too extreme, and that priors which, say, only ε-maximize the likelihood function should not be ruled out. Indeed, classical statistics techniques such as hypothesis testing do allow for ranges of the likelihood function.

At present we are not aware of a nice axiomatization of such rules. We point out, however, that the other extreme rule, i.e., updating all priors without excluding any of them (see, for instance, Fagin and Halpern [13], and Jaffray [22]), does not appear to be any less "extreme" in general, nor does it seem to be implied by more compelling axioms.

We believe that our theory can be applied to a variety of economic models, explaining phenomena which are incompatible with the Bayesian theory, and possibly providing better predictions. As a matter of fact, this belief may be updated given new evidence: Dow and Werlang [10] and Simonsen and Werlang [32] have already applied the multiple prior theory to portfolio selection problems. These studies have shown that a decision maker having ambiguous beliefs will have a (non-trivial) range of prices at which he/she will neither buy or sell an uncertain asset, exhibiting inertia in portfolio selection. Applying our new results regarding ambiguous beliefs update, one may study the conditions under which these price ranges will shrink in the face of new information.

Dow, Madrigal, and Werlang [11] studied trade among agents, at least one of whom has ambiguous beliefs. They show that the celebrated no-trade result of Milgrom and Stokey [24] fails to hold in this context. In this study, the Dempster–Shafer rule for updating non-additive measures was used, a rule which is justified by the current paper. Casting the trade set-up into a dynamic context raises the question of an asymptotic no-trade theorem: Under what conditions will additional information reduce the volume or probability of trade?

In another recent study, Yoo [36] addressed the question of why stock prices tend to fall after the initial public offering and rise at a later stage. Although Yoo uses ambiguous beliefs mostly as in Bewley's [3] model, his results can also be obtained using the models mentioned above. It seems that the update rule justified by our study may explain the price dynamics.

These various models seem to point at a basic problem: given a convex non-additive measure (or, equivalently, a set of priors which is the core of such a measure), under what conditions will the Dempster–Shafer rule yield convergence of beliefs to a single additive prior? Obviously, the answer cannot be "always." Consider a "large" measurable space with all possible priors (equivalently, with the "unanimity game" as a non-additive measure). In this set-up of "complete ignorance," no conclusions about the future may be drawn from past observations—that is, the updated beliefs still include all possible priors. However, with some initial information (say, finitely many extreme points of the set of priors) convergence is possible. Conditions that will guarantee such convergence call for further study.

The rest of this paper is organized as follows. Section 2 presents the framework and quotes some results. Section 3 defines the update rules and states the theorems. Finally, Section 4 includes proofs, related analysis, and some remarks regarding possible generalizations.

2. Framework and Preliminaries

Let X be a set of *consequences* endowed with a weak order $\geqslant$. Let (S, Σ) be a measurable space of *states of the world*, where Σ is the algebra of *events*. A function $f: S \to X$ is *Σ-measurable* if for every $x \in X$

$$\{s \mid f(s) > x\}, \qquad \{s \mid f(s) \geqslant x\} \in \Sigma.$$

Let $F = \{f\colon S \to X \mid f \text{ is } \Sigma\text{-measurable}\}$ be the set of *acts*. Let $F_0 = \{f \in F \mid |\text{range}(f)| < \infty\}$ be the set of *simple acts*. A function $u\colon X \to \mathbb{R}$, which represents $\geqslant$, i.e.,

$$u(x) \geqslant u(y) \Leftrightarrow x \geqslant y, \qquad \forall x, y \in X$$

is called a *utility* function.

A function $v\colon \Sigma \to [0, 1]$ satisfying

(i) $v(\varnothing) = 0$; $v(S) = 1$;

(ii) $A \subseteq B \Rightarrow v(A) \leqslant v(B)$

is a *non-additive measure*. It is *convex* if

$$v(A \cup B) + v(A \cap B) \geqslant v(A) + v(B)$$

for all $A, B \in \Sigma$. It is *additive*, or simply a *measure*, if the above inequality is always satisfied as an equality.

A real-valued function is Σ-measurable if for every $t \in \mathbb{R}$

$$\{s \mid w(s) \geqslant t\}, \; \{s \mid w(s) > t\} \in \Sigma.$$

Given such a function w and a non-additive measure v, the (Choquet) integral of w w.r.t. (with respect to) v on S is

$$\int w\, dv = \int_S w\, dv = \int_0^{\infty} v(\{s \mid w(s) \geqslant t\})\, dt + \int_{-\infty}^{0} [v(\{s \mid w(s) \geqslant t\}) - 1]\, dt.$$

For a non-additive measure v we define the core as for a cooperative game, i.e.,

$$\text{Core}(v) = \{p \mid p \text{ is a measure s.t. } p(A) \geqslant v(A) \forall A \in \Sigma\}.$$

Recall that a convex v has a nonempty core (see Shapley [33]).

We are now about to define two classes of binary relations on F: those represented by maximization of expected utility with non-additive measures (NA), and those represented by maxmin of expected utility with multiple priors (MP).

Denote by NA_0 ($=\mathrm{NA}_0(X, \geqslant, S, \Sigma)$) the set of binary relations $\geqslant$ on F such that there are a utility u, unique up to p.l.t. (positive linear transformation), and a unique non-additive measure v satisfying:

(i) for every $f \in F$, $u \circ f$ is Σ-measurable;

(ii) for every $f, g \in F$

$$f \geqslant g \Leftrightarrow \int u \circ f \, dv \geqslant \int u \circ g \, dv.$$

Note that in general the measurability of f does not guarantee that of $u \circ f$, and that (ii) implies that $\geqslant$ on F, when restricted to constant functions, extends $\geqslant$ on X. Hence we use this convenient abuse of notation. Similarly, we will not distinguish between $x \in X$ and the constant act which equals x on S.

Characterizations of NA_0 were given by Schmeidler [29, 30] for the Anscombe–Aumann [2] framework, where X is a mixture space and u is assumed affine; by Gilboa [16] in the Savage [26] framework, where X is arbitrary but $\Sigma = 2^S$ and v is nonatomic; and by Wakker [35] for the case where X is a connected topological space. Fishburn [14] extended the characterization to non-transitive relations.

Let MP_0 ($=\mathrm{MP}_0(X, \geqslant, S, \Sigma)$) denote the set of binary relations $\geqslant$ of F such that there are a utility u unique up to a p.l.t., and a unique nonempty, closed (in the weak* topology), and convex set C of (finitely additive) measures on Σ such that:

(i) for every $f \in F$, $u \circ f$ is Σ-measurable;

(ii) for every $f, g \in F$

$$f \geqslant g \Leftrightarrow \min_{p \in C} \int u \circ f \, dp \geqslant \min_{p \in C} \int u \circ g \, dp.$$

A characterization of MP_0 in the Anscombe–Aumann framework was given in Gilboa and Schmeidler [18]. To the best of our knowledge, there is no such axiomatization in the framework of Savage. However, the set $\mathrm{NA}_0 \cap \mathrm{MP}_0$, which will play an important role in the sequel, may be characterized by strengthening the axioms in Gilboa [16].

It will be convenient to include the trivial weak order $\geqslant^* = F \times F$ in both NA and MP. Hence, we define $\mathrm{NA} = \mathrm{NA}_0 \cup \{\geqslant^*\}$ and $\mathrm{MP} = \mathrm{MP}_0 \cup \{\geqslant^*\}$.

For simplicity we assume that X has $\geqslant$-maximal and -minimal elements. More specifically, let $x^*, x_* \in X$ satisfy $x_* \leqslant x \leqslant x^*$ for all $x \in X$. W.l.o.g. (without loss of generality), assume that x_* and x^* are unique. Since for both NA_0 and MP_0 the utility function is unique only up to a p.l.t. we will assume w.l.o.g. that $u(x_*)=0$ and $u(x^*)=1$ for all utilities henceforth considered.

When X is a mixture space we define NA′ and MP′ to be the subsets of NA and MP, respectively, where the utility u is also required to be affine. For such spaces X we recall the following results.

PROPOSITION 2.1. *Suppose that* $\geqslant \in$ NA′ *and let* v *be the associated non-additive measure. Then* $\geqslant \in$ MP′ iff v *is convex.*

PROPOSITION 2.2. *Suppose that* $\geqslant \in$ MP′ *and let* C *be the associated set of measures. Define*

$$v(A)=\min_{p \in C} p(A) \textit{ for } A \in \Sigma.$$

Then v *is a non-additive measure and* $\geqslant \in$ NA′ iff v *is convex and* $C=\text{Core}(v)$.

The proofs of these appear, explicitly or implicitly, in Schmeidler [28–30]. Note that the axiomatization of NA′ (Schmeidler [30]) uses comonotonic independence, and given this property the convexity of v is equivalent to uncertainty aversion. The axiomatization of MP′ (Gilboa and Schmeidler [18]) uses a weaker independence notion—termed C-independence—and uncertainty aversion. Given these, the convexity of v and the equality $C=\text{Core}(v)$ (where v is defined as in Proposition 2.2) is equivalent to comonotonic independence.

We now define update rules. We need the following definitions. Given a measurable partition $\Pi=\{A_i\}_{i=1}^n$ of S and $\{f_i\}_{i=1}^n \subseteq F$, let $(f_1, A_1; \ldots; f_n, A_n)$ denote the act $g \in F$ satisfying $g(s)=f_i(s)$ for all $s \in A_i$ and all $1 \leqslant i \leqslant n$. Given a binary relation $\geqslant$ on F, an event $A \in \Sigma$ is $\geqslant$*-null* iff the following holds: for every $f, g, h_1, h_2 \in F$,

$$f \geqslant g \qquad \text{iff } (f, A^c; h_1, A) \geqslant (g, A^c; h_2, A).$$

Let $\bar{\mathscr{B}}$ denote the set of all binary relations on F. Given $\mathscr{B} \subseteq \bar{\mathscr{B}}$, an *update rule for* $\mathscr{B}$ is a collection of functions, $U=\{U_A\}_{A \in \Sigma}$, where $U_A : \mathscr{B} \to \bar{\mathscr{B}}$ such that for all $\geqslant \in \mathscr{B}$ and $A \in \Sigma$, A^c is $U_A(\geqslant)$-null and $U_S(\geqslant)=\geqslant$. $U_A(\geqslant)$ should be thought of as the preference relation once A is known to have occurred.

Given $\mathscr{B}$ and an update rule for it, $U=\{U_A\}_{A \in \Sigma}$, U is said to be *commutative w.r.t.* $\geqslant$ or $\geqslant$-commutative if for every $A, B \in \Sigma$ we have $U_A(\geqslant) \in \mathscr{B}$ and

$$U_B(U_A(\geqslant))=U_{A \cap B}(\geqslant).$$

It is *commutative* if it is commutative w.r.t. $\geqslant$ for all $\geqslant \in \mathscr{B}$. (Note that this condition is stronger than strict commutativity, i.e., $U_A \circ U_B = U_B \circ U_A$. However, "commutativity" seems to be a suggestive name which is not overburdened with other meanings.)

3. Bayesian and Classical Rules

Given a set $\mathscr{B}$ of binary relations of F, every $f \in F$ suggests a natural update rule for $\mathscr{B}$: define $\mathrm{BU}^f = \{\mathrm{BU}_A^f\}_{A \in \Sigma}$ by

$$g\, \mathrm{BU}_A^f(\geqslant)\, h \Leftrightarrow (g, A; f, A^c) \geqslant (h, A; f, A^c) \qquad \text{for all } g, h \in F.$$

It is obvious that for every f, BU^f is an update rule, i.e., that A^c is $\mathrm{BU}_A^f(\geqslant)$-null for all $\geqslant \in \mathscr{B}$ and $A \in \Sigma$. We will refer to it as the *f-Bayesian update rule* and $\{\mathrm{BU}^f\}_{f \in F}$ will be called the set of Bayesian update rules.

Note that for $\geqslant \in \mathrm{NA}$ with an additive v, all the Bayesian update rules coincide with Bayes' rule, hence the definition of the Bayesian update rules may be considered a formulation and axiomatization of Bayes' rule in the case of (a unique) additive prior.

Proposition 3.1. *For every $\geqslant \in \bar{\mathscr{B}}$ and $f \in F$, BU^f is $\geqslant$-commutative.*

Theorem 3.2. *Let $f \in F$ and assume that $|\Sigma| > 4$. Then the following are equivalent*:

(i) $\mathrm{BU}_A^f(\mathrm{NA}') \subseteq \mathrm{NA}'$ *for all $A \in \Sigma$;*

(ii) $f = (x^*, T; x_*, T^c)$ *for some $T \in \Sigma$.*

Of particular interest are the Bayesian update rules corresponding to $f = x^*$ and $f = x_*$ (i.e., $T = S$ or $T = \varnothing$ in (ii) above). For the latter (x_*) there is an "optimistic" interpretation: when comparing two actions given a certain event A, the decision maker implicitly assumes that had A not occurred, the worst possible outcome (x_*) would have resulted. In other words, the behavior given A—$\mathrm{BU}_A^f(\geqslant)$—exhibits "happiness" that A has occurred; the decisions are made as if we are always in "the best of all possible worlds."

Note that the corresponding non-additive measure is

$$v_A(B) = v(B \cap A)/v(A).$$

On the other hand, for $f = x^*$, we consider a "pessimistic" decision maker, whose choices reveal the hidden assumption that all the impossible worlds are the best conceivable ones. This rule defines the non-additive function by

$$v_A(B) = [v((B \cap A) \cup A^c) - v(A^c)]/(1 - v(A^c)),$$

which is identical to the Dempster–Shafer rule for updating probabilities.

It should not surprise us that this "pessimistic" rule is going to play a major role in relation to MP—i.e., to uncertainty-averse decision makers who follow a maxmin (expected utility) decision rule. In a similar way one may develop a "dual" theory of "optimism" in which uncertainty-seeking will replace uncertainty-aversion, concavity of v will replace convexity, and maxmax will supercede maxmin. For this "dual" theory, the update rule

$$v_A(B) = v(B \cap A)/v(A)$$

would be the "approriate" one (in a sense that will be clear shortly). Note that this rule was used—without axiomatization—as a definition of probability update in Gilboa [17].

Taking a classical statistics point of view, it is natural to start out with a set of priors. Hence we only define classical update rules for $\mathscr{B} = \mathrm{MP}'$. A natural procedure in the classical updating process is to rule out some of the given priors, and update the rest according to Bayes' rule. Thus, we get a family of update rules, which differ in the way the priors are selected.

Formally, a *classical update rule* is characterized by a function $R : (C, A) \mapsto C'$ such that $C' \subseteq C$ is a closed and convex set of measures for every such C and every $A \in \Sigma$, with $R(C, S) = C$. The associated update rule will be denoted $CU^R = \{CU_A^R\}_{A \in \Sigma}$. (If $R(C, A) = \varnothing$ we define $CU_A^R(\geqslant) = \geqslant^*$.) Note that these are indeed update rules, i.e., for every $\geqslant \in \mathrm{MP}'$, every R and every $A \in \Sigma$, A^c is $CU_A^R(\geqslant)$-null. Furthermore, for $\geqslant \in \mathrm{MP}'$ with an associated set C, $CU_A^R(\geqslant) \in \mathrm{MP}'$ provided that $\inf\{p(A) \mid R(C, A)\} > 0$ for all $A \in \Sigma$.

Of particular interest will be the classical update rule called maximum likelihood and defined by

$$R^0(C\, A) = \{p \in C \mid p(A) = \max_{q \in C} q(A) > 0\}.$$

THEOREM 3.3. *CU^{R^0} is cummutative on $\mathrm{NA}' \cap \mathrm{MP}'$. Furthermore, for $\geqslant \in \mathrm{NA}' \cap \mathrm{MP}'$,*

$$\mathrm{BU}_A^{(x^*, S)}(\geqslant) = \mathrm{CU}_A^{R^0}(\geqslant) \in \mathrm{NA}' \cap \mathrm{MP}'.$$

I.e., the Bayesian update rule with $f = (x^*, S)$ coincides with the maximum-likelihood classical update rule. Moreover, they are also equivalent to the

Dempster–Shafer update rule for belief functions. (Note that every belief function (see Shafer [31]) is convex, though the converse is false. Yet one may apply the Dempster–Shafer rule for every non-additive measure v.)

4. Proofs and Related Analysis

4.1. *Proof of Proposition* 3.1. It only requires to note that for every $f, g, \in F$, $A, B \in \Sigma$

$$((g, A; f, A^c), B; f, B^c) = (g, A \cap B; f, (A \cap B)^c).$$

4.2. *Proof of Theorem* 3.2. First assume (ii). Let there be given $\geqslant \in \mathrm{NA}'$ with associated u and v. Define for $B \in \Sigma$ a non-additive measure v_B by

$$v_B(A) = [v((A \cap B) \cup (T \cap B^c)) - v(T \cap B^c)]/[v(B \cup T) - v(T \cap B^c)]$$

if the denominator is positive. (Otherwise the result is trivial.) For every $g \in F$ we have

$$\begin{aligned}\int_S u \circ (g, B; f, B^c)\, dv &= \int_0^1 v(\{s \mid u \circ (g, B; f, B^c)(s) \geqslant t\})\, dt \\ &= \int_0^1 v((T \cap B^c) \cup (\{s \mid u \circ g(s) \geqslant t\} \cap B))\, dt \\ &= \int_0^1 [v(T \cap B^c) \\ &\qquad + [v(B \cup T) - v(T \cap B^c)]\, v_B(\{s \mid u \circ g(s) \geqslant t\})]\, dt \\ &= v(T \cap B^c) + [v(B \cup T) - v(T \cap B^c)] \int u \circ g\, dv_B,\end{aligned}$$

where v_B and u represent BU^f_B, which implies that the latter is in NA'.

Conversely, assume (i) holds. Assume, to the contrary, that $f(s) \sim x$ for $s \in D$ where $D \in \Sigma$, $D \neq S$ and $x_* < x < x^*$ (where $\sim$ denotes $\geqslant$-equivalence). Let $E, F \in \Sigma$ satisfy $E \cap F = F \cap D = D \cap F = \varnothing$. Denote $\alpha = u(x)$ (where $0 < \alpha < 1$). Choose $m \in (\alpha, 1)$ and a non-additive v such that

$$v(E) = v(F) = v(D) = m$$

$$v(E \cup F) = v(E \cup D) = v(F \cup D) = m$$

and $v(T) = v(T \cap (E \cup F \cup D))$ for all $T \in \Sigma$. Next define $\geqslant \in \mathrm{NA}'$ by v and (the unique) u.

Choose g_1, g_2 such that

$$\begin{aligned} u \circ g_1(s) &= u \circ g_2(s) = \alpha & s \in D \\ u \circ g_1(s) &= 1,\ u \circ g_2(s) = \alpha + (1 - \alpha/m) & s \in E \\ u \circ g_1(s) &= 0,\ u \circ g_2(s) = \alpha + (1 - \alpha/m) & s \in F. \end{aligned}$$

Let $\geqslant'$ be $\mathrm{BU}^f_{E \cup F}(\geqslant)$. By assumption it belongs to NA'; hence, there correspond to it $u' = u$ and v'. Note that v' is unique as $\geqslant'$ is nontrivial, and that $v'(T) = v'(T \cap (E \cup F))$ for all $T \in \Sigma$.

As $\int u \circ g_1\, dv = \int u \circ g_2\, dv$, $g_1 \sim g_2$, whence $g_1 \sim' g_2$. Hence, $\int u \circ g_1\, dv' = \int u \circ g_2\, dv'$, i.e., $v'(E) = \alpha + (1 - \alpha/m)$.

Next chose $\beta \in (0, \alpha)$ and choose an act $g_3 \in F$ such that

$$u \circ g_3(s) = \begin{cases} \alpha & s \in D \\ \beta & s \in E \\ 0 & s \in F. \end{cases}$$

For every $\gamma \in (0, \alpha)$ choose $g_\gamma \in F$ such that

$$u \circ g_\gamma(s) = \begin{cases} \alpha & s \in D \\ \gamma & s \in E \cup F. \end{cases}$$

Then $\int u \circ g_3\, dv = \alpha m$ and $\int u \circ g_\gamma\, dv = \alpha m + \gamma(1 - m)$. Hence, $g_\gamma > g_3$ and $g_\gamma >' g_3$ for all $\gamma > 0$. However, $\int u \circ g_\gamma\, dv' = \gamma$ and $\int u \circ g_3\, dv' = \beta v'(E)$, where $v'(E) = 0$, a contradiction.

Remark 4.3. In the case of no extreme outcomes, i.e., when X has no $\geqslant$-maximal or no $\geqslant$-minimal elements, and in particular when the utility is not bounded, there are no update rules BU^f which map NA' into itself. However, one may choose for $g, h \in F$ $x^*, x_* \in X$ such that $x^* \geqslant g(s)$, $h(s) \geqslant x_*$, $\forall s \in S$, and for every $T \in \Sigma$ define $\mathrm{BU}^f(\geqslant) = \{\mathrm{BU}^f_A\}_{A \in \Sigma}$ between g and h by $f = (x^*, T; x_*, T^c)$. If $\geqslant \in \mathrm{NA}$, this definition is independent of the choice of x^* and x_*. The resulting update rule will be commutative for any (fixed) $T \in \Sigma$.

4.4. *Proof of Theorem* 3.3. Let $\geqslant \in \mathrm{MP}'$ be given, and let C denote its associated set of additive measures. Define $v(\cdot) = \min_{p \in C} p(\cdot)$.

Assume that v is convex and $C = \mathrm{Core}(v)$. For $A \in \Sigma$ with $q(A) > 0$ for some $q \in C$, we have

$$R^0(C, A) = \{p \in C \mid p(A) = \max_{q \in C} q(A)\} = \{p \in C \mid p(A^c) = v(A^c)\}.$$

(Note that if $v(A^c) = 1$, $\mathrm{CU}^{R^0}_A(\mathrm{CU}^{R^0}_B(\geqslant)) = \mathrm{CU}^{R^0}_B(\mathrm{CU}^{R^0}_A(\geqslant)) = \geqslant^*$.)

As was shown in Schmeidler [28], v is convex iff for every chain $\varnothing = E_0 \subseteq E_1 \subset \cdots \subset E_n = S$ there is an additive measure p in $\text{Core}(v) = C$ such that $p(E_i) = v(E_i)$, $0 \leqslant i \leqslant n$. Furthermore, this requirement for $n = 3$ is also equivalent to convexity.

Next define

$$v'_A(T) = \min\{p(T \cap A) \mid p \in R^0(C, A)\}.$$

Claim. $v'_A(T) = v((T \cap A) \cup A^c) - v(A^c)$.

Proof. For $p \in R^0(C, A)$ we have

$$\begin{aligned} p(T \cap A) &= p((T \cap A) \cup A^c) - p(A^c) \\ &= p((T \cap A) \cup A^c) - v(A^c) \\ &\geqslant v((T \cap A) \cup A^c) - v(A^c) \end{aligned}$$

whence

$$v'_A(T) \geqslant v((T \cap A) \cup A^c) - v(A^c).$$

To show the converse inequality, consider the chain $\varnothing \subseteq A^c \subseteq A^c \cup (A \cap T) \subseteq S$. By convexity there is $p \in \text{Core}(v) = C$ satisfying $p(A^c) = v(A^c)$ and $p(A^c \cup (T \cap A)) = v(A^c \cup (T \cap A))$ which also implies $p \in R^0(C, A)$. Then

$$\begin{aligned} v'_A(T) \leqslant p(T \cap A) &= p((T \cap A) \cup A^c) - p(A^c) \\ &= v((T \cap A) \cup A^c) - v(A^c). \quad \blacksquare \end{aligned}$$

Consider $\text{CU}_A^{R^0}(\geqslant)$. If it is not equal to $\geqslant^*$, it has to be the case that $v(A^c) < 1$, and then it is defined by the set of additive measures

$$C_A = \{p_A \mid p \in R^0(C, A)\}$$

where

$$p_A(T) = p(T \cap A)/p(A) = p(T \cap A)/(1 - v(A^c)).$$

Note that C_A is nonempty, closed, and convex. Define

$$v_A(T) = \min\{p(T) \mid p \in C_A\},$$

and observe tht $v_A(T) = v'_A(T)/(1 - v(A^c))$, i.e.,

$$v_A(T) = [v((T \cap A) \cup A^c) - v(A^c)]/[1 - v(A^c)]. \tag{*}$$

Hence, v_A is also convex. We have to show that $C_A = \text{Core}(v_A)$.

To see this, let $p \in \mathrm{Core}(v_A)$. We will show that $p = q_A$ for some $q \in R^0(C, A)$. Take any $q' \in \mathrm{Core}(v)$ and define

$$q(T) = p(T \cap A)[1 - v(A^c)] + q'(T \cap A^c).$$

Note that

$$\begin{aligned} q(T \cap A) &= p(T \cap A)[1 - v(A^c)] \geqslant v_A(T \cap A)[1 - v(A^c)] \\ &= v((T \cap A) \cup A^c) - v(A^c). \end{aligned}$$

(As $p \in \mathrm{Core}(v_A)$ and by definition of the latter.) Next, since $q' \in \mathrm{Core}(v)$,

$$q(T \cap A^c) = q'(T \cap A^c) \geqslant v(T \cap A^c).$$

Hence,

$$\begin{aligned} q(T) &= q(T \cap A) + q(T \cap A^c) \\ &\geqslant v((T \cap A) \cup A^c) - v(A^c) + v(T \cap A^c) \\ &= v(T \cup A^c) - v(A^c) + v(T \cap A^c) \geqslant v(T), \end{aligned}$$

where the last inequality follows from the convexity of v. Finally,

$$q(S) = q(A) + q(A^c) = p(A)[1 - v(A^c)] + v(A^c) = 1.$$

Hence, $q \in \mathrm{Core}(v)$. Furthermore, $q \in R^0(C, A)$. Obviously, $p = q_A$.

Thus we establish $\mathrm{CU}_A^{R^0}(\geqslant) \in \mathrm{NA}'$. Furthermore, $\mathrm{CU}_A^{R^0}(\geqslant) = \mathrm{BU}_A^{(x^*, S)}(\geqslant)$ and the non-additive probability update rule (*) coincides with the Dempster–Shafer rule. Any of these two facts, combined with the observation $\mathrm{CU}_A^{R^0}(\geqslant) \in \mathrm{NA}'$, implies that CU^{R^0} is cummutative. ▮

Remark 4.5. It is not difficult to see that the maximum-likelihood update rule is not commutative in general. In fact, one may ask whether the coverse of Theorem 3.3 is true, i.e., whether a relation $\geqslant \in \mathrm{MP}'$ with respect to which CU^{R_0} is commutative has to define a set C which is a core of a non-additive measure. The negative answer is given by the following example: $S = \{1, 2, 3, 4\}$, $\Sigma = 2^S$, $C = \mathrm{conv}\{p_1, p_2\}$ defined by

	1	2	3	4
p_1	.7	.1	.1	.1
p_2	.1	.3	.3	.3

It is easily verifiable that the maximum-likelihood update rule is commutative w.r.t. the induced $\geqslant \in \mathrm{MP}'$, though C is not the core of any v.

Remark 4.6. It seems that the maximum-likelihood update rule is not commutative in general, because it lacks some "look-ahead" property. One is tempted to define an update rule that will retain all the priors which may, at some point in the future, turn out to be likelihood maximizers. Thus, we are led to the "semi-generalized maximum likelihood":

$$R^1(C, A) = \mathrm{cl}\,\mathrm{conv}\,\{p \in C \mid p(E) = \max_{q \in C} q(E) > 0 \text{ for some measurable } E \subseteq A\}$$

(where cl means closure in the weak* topology). Note that the resulting set of measures may include $p \in C$ such that $p(A) = 0$. In this case define $\mathrm{CU}_A^{R^1}(\geqslant) = \geqslant^*$.

However, the following example shows that this update rule also fails to be commutative in general.

Consider $S = \{1, 2, 3, 4, 5\}$, $\Sigma = 2^S$, and let C be $\mathrm{conv}\{p_1, p_2, p_3, p_4\}$ defined by the following table:

	1	2	3	4	5
p_1	.2	.2	.01	.09	.5
p_2	0	0	.4	.4	.2
p_3	.27	0	.03	0	.7
p_4	0	.27	.03	0	.7

Taking $A = \{1, 2, 3, 4\}$ and $B = \{1, 2, 3\}$, one may verify that

$$R^1(R^1(C, A), B) = \{p_2, p_3, p_4\}$$

and

$$R^1(C, B) = \{p_1, p_2, p_3, p_4\}$$

and that p_{1_B} is not in the convex hull of $\{p_{2_B}, p_{3_B}, p_{4_B}\}$.

We may try an even more generalized version of the maximum likelihood criterion: retain all priors according to which the integral of some nonnegative simple function is maximized. I.e., define

$$R^2(C, A) = \mathrm{cl}\,\mathrm{conv}\left\{p \in C \mid \int u \circ f\,dp = \max\left\{\int u \circ f\,dq \mid q \in C\right\} > 0 \text{ for some } f \in F_0\right\}.$$

The maximization of $\int u \circ f\,dp$ for some f may be viewed as maximization of some convex combination of the likelihood function at several points of time.

However, the same example shows that CU^{R^2} is not commutative in general. ∎

Remark 4.7. Although our results are formulated for NA′ and MP′, they may be generalized easily. First, one should note that none of the results actually requires that X be a mixture space. All that is needed is that the utility on X be uniquely defined (up to a p.l.t.) and that its range will contain an open interval. In particular, connected topological spaces with a continuous utility function will do.

Moreover, most of the results do not even require such richness of the utility's range. In fact, this richness was only used in the proof of (i) ⇒ (iii) in Theorem 3.2.

REFERENCES

1. C. E. AGNEW, Multiple probability assessments by dependent experts, *J. Amer. Statist. Assoc.* **80** (1985), 343–347.
2. F. J. ANSCOMBE AND R. J. AUMANN, A definition of subjective probability, *Ann. Math. Statist.* **34** (1963), 199–205.
3. T. BEWLEY, "Knightian Decision Theory: Part I," Cowles Foundation Discussion paper No. 807, Yale University, 1986.
4. T. BEWLEY, "Knightian Decision Theory: Part II: Intertemporal Problems," Cowles Foundation Discussion paper No. 835, Yale University, 1987.
5. T. BEWLEY, "Knightian Decision Theory and Econometric Inference," Cowles Foundation Discussion paper No. 868, Yale University, 1988.
6. A. CHATEAUNEUF AND J.-Y. JAFFRAY, Some characterizations of lower probabilities and other monotone capacities through the use of Moebius inversion, *Mathematical Social Sciences* **17** (1989), 263–283.
7. G. CHOQUET, Theory of capacities, *Ann. l'Institut Fourier* **5** (1953–1954), 131–295.
8. A. P. DEMPSTER, Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Statist.* **38** (1967), 325–339.
9. A. P. DEMPSTER, A generalization of Bayesian inference, *J. Roy. Statist. Soc. Series B* **30** (1968), 205–247.
10. J. DOW AND S. R. C. WERLANG, "Risk Aversion, Uncertainty Aversion and the Optimal Choice of Portfolio," London Business School Working Paper, 1990.
11. J. DOW, V. MADRIGAL, AND S. R. C. WERLANG, "Preferences, Common Knowledge and Speculative Trade," Fundacao Getulio Vargas Working paper, 1989.
12. D. ELLSBERG, Risk, ambiguity and the Savage axioms, *Quart. J. Econ.* **75** (1961), 643-669.
13. R. FAGIN AND J. Y. HALPERN, A new approach to updating beliefs, mimeo (1989).
14. P. C. FISHBURN, Uncertainty aversion and Separated Effects in decision making under uncertainty, *in* "Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making," (J. Kacprzyk and M. Fedrizzi, Eds.), pp. 10–25, Springer-Verlag, New York/Berlin, 1988.
15. C. GENEST AND M. J. SCHERVISH, Modeling expert judgments for Bayesian updating, *Ann. Statist.* **13** (1985), 1198–1212.
16. I. GILBOA, Expected utility theory with purely subjective non-additive probabilities, *J. Math. Econ.* **16** (1987), 65–88.
17. I. GILBOA, Additivizations of non-additive measures, *Math. Operations Res.* **14** (1989), 1–17.
18. I. GILBOA AND D. SCHMEIDLER, Maxmin expected utility with non-unique prior, *J. Math. Econ.* **18** (1989), 141–153.

19. J. Y. HALPERN AND R. FAGIN, Two views of belief: Belief as generalized probability and belief as evidence, mimeo (1989).
20. J. Y. HALPERN AND M. R. TUTTLE, Knowledge, probability and adversaries, mimeo (1989).
21. J.-Y. JAFFRAY, Linear utility theory for belief functions, *Operations Res. Lett.* **8** (1989), 107–112.
22. J.-Y. JAFFRAY, Bayesian updating and belief functions, mimeo (1990).
23. D. V. LINDLEY, A. TVERSKY, AND R. V. BROWN, On the reconciliation of probability assessments, *J. Roy. Statist. Soc. Series A* **142** (1979), 146–180.
24. P. MILGROM AND N. STOKEY, Information, trade and common knowledge, *J. Econ. Theory* **26** (1982), 17–27.
25. J. VON NEUMANN AND O. MORGENSTERN, "Theory of Games and Economic Behavior," 2nd ed., Princeton Univ. Press, Princeton, NJ, 1947.
26. L. J. SAVAGE, "The Foundations of Statistics," Wiley, New York, 1954.
27. D. SCHMEIDLER, "Subjective Probability Without Additivity," (temporary title), Working paper, Foerder Institute for Economic Research, Tel Aviv University (1982).
28. D. SCHMEIDLER, Nonadditive probabilities and convex games, mimeo (1984).
29. D. SCHMEIDLER, Integral representation without additivity, *Proc. Amer. Math. Soc.* **97** (1986), 253–261.
30. D. SCHMEIDLER, Subjective probability and expected utility without additivity, *Econometrica* **57** (1989), 571–587.
31. G. SHAFER, "A Mathematical Theory of Evidence," Princeton Univ. Press, Princeton, NJ, 1976.
32. M. H. SIMONSEN AND S. R. C. WERLANG, Subadditive probabilities and portfolio inertia, mimeo (1990).
33. L. S. SHAPLEY, "Notes on n-person Games VII: Cores of Convex Games," The Rand Corporation R. M. (1965); also as Cores of convex games, *Int. J. Game Theory* **1** (1971), 12–26.
34. PH. SMETS, "Combining Non-distinct Evidences," Technical report ULB-IRIDIA-86/3, 1986.
35. P. WAKKER, Continuous subjective expected utility with non-additive probabilities, *J. Math. Econ.* **18** (1989), 1–17.
36. K. R. YOO, A theory of the underpricing of initial public offerings, mimeo (1990).