Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Economic Theory 108 (2003) 286–321
JOURNAL OF Economic Theory
http://www.elsevier.com/locate/jet

Conditional preferences and updating

Tan Wang*

Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, BC, Canada V6T 1Z2

Received 26 March 2000; final version received 6 December 2001

Abstract

This paper axiomatizes updating rules for preferences that are not necessarily in the expected utility class. Two sets of results are presented. The first is the axiomatization of conditional preferences. The second consists of the axiomatization of three updating rules: the traditional Bayes rule, the Dempster–Shafer rule, and the generalized Bayes rule. The last rule can be regarded as the updating rule for the multi-prior expected utility (Gilboa and Schmeidler, J. Math. Econom. 18 (1989) 141). Operationally, it is equivalent to updating each prior by the traditional Bayes rule.

JEL classification: D80; D81; D83

Keywords: Updating rules; Learning; Uncertainty; Ambiguity

1. Introduction

The traditional rule for updating is the Bayes rule. The decision-theoretic foundation of this rule is laid by the axiomatization of Savage [35], which shows that, in situations of uncertainty, if an individual's preference satisfies certain axioms, his preference can be represented by an expected utility with respect to a subjective probability measure. This probability measure can be viewed as the individual's belief about the likelihoods of uncertainty events. Moreover, in light of new information, the individual updates his belief according to the Bayes rule.

Over the past 50 years, this Savage paradigm has been the foundation for much of the economic theories under uncertainty. At the same time, however, the Savage

*Fax: 604-822-4695.
E-mail address: tan.wang@commerce.ubc.ca.

doi:10.1016/S0022-0531(03)00028-0

paradigm has been challenged by behavior exhibited in the Ellsberg paradox [14], which questions the very notion of representing an individual's belief by a probability measure and consequently the validity of the Bayes rule. The purpose of this paper is to study how an individual updates when his preference does not necessarily fall into the expected utility class.

We axiomatize three updating rules: the traditional Bayes rule, the Dempster–Shafer rule, and the generalized Bayes rule. The last rule can be viewed as the updating rule for the multi-prior expected utility preferences [4,20]. Operationally, it is equivalent to updating each prior by the traditional Bayes rule.

The rest of the paper is organized as follows: Section 2 contains a brief review of the related literature and further motivation of this paper. Section 3 introduces the set of multi-period consumption–information profiles. Section 4 axiomatizes a class of conditional preferences which serves as the basis for our study of updating rules. The results on updating rules appear in Section 5. Finally, Section 6 concludes with some applications. Proofs and supporting technical details are collected in the appendix.

2. Motivation and related literature

Motivated by well-documented empirical facts such as the Ellsberg paradox, Savage's expected utility theory has been significantly extended over the past decade or so. In this regard, two classes of preferences stand out as the most prominent. One class is the Choquet expected utility. It was axiomatized first in the Anscombe–Aumann framework by Schmeidler [37], and later in the Savage setting by Gilboa [19], Nakamura [32], Sarin and Wakker [33], and Chew and Karni [7]. The other class, the multi-prior expected utility, was axiomatized by Gilboa and Schmeidler [20] in the Anscombe–Aumann framework. Recently, Casadesus-Masanell et al. [4] succeeded in axiomatizing it in the Savage setting. For a comprehensive survey of this literature, the reader is referred to Camerer and Weber [3] and Sarin and Wakker [34].

These new theories have significantly improved our understanding of individual decision making under uncertainty. In particular, they have for the first time modelled analytically preferences whose belief component cannot be represented by a probability measure. The natural question is then: how do these individuals update their beliefs? In the economic literature, Epstein and LeBreton [16] show that if all axioms of Savage, except the sure-thing principle, plus dynamic consistency are to be maintained, then the updating rule must be Bayesian. Gilboa and Schmeidler [21] show that for Choquet expected utility, if the updated preference is again representable by a Choquet expected utility, the admissible f-Bayesian updating rules must be those which assume that the unrealized event yields either the highest or the lowest possible payoff.

In the statistics literature, extensions of the Bayes rule have been concentrated mainly on two updating rules. The first rule, called the Dempster–Shafer rule, is for convex capacities. It dates back to Dempster [11,12]. The more recent work includes

Shafer [38,39]. The second updating rule, called the generalized Bayes rule, is for multi-prior likelihood functions. It has been studied by Walley [44], Wasserman and Kadane [46], Seidenfeld and Wasserman [40], and Herron et al. [25].[1]

All these updating rules were developed axiomatically. However, the primitives of the axiomatizations differ. In statistics literature, for example Walley [44], the primitives are essentially likelihood functions. In economics literature, the main interest is individuals' choice behavior. Thus, any updating rule has to be integrated consistently into individuals' preferences. In this paper, taking individuals' preferences as the primitives, we axiomatize the Dempster–Shafer rule and the generalized Bayes rule.

The motivation for our study of updating rules comes also from the progress on the application front. Epstein and Wang [17,18] developed an intertemporal asset pricing model under Knightian uncertainty. In that model, the agent's preference is represented by a multi-period version of the multi-prior expected utility developed by Gilboa and Schmeidler [20]. The evolution of the agent's belief is modelled by a transition belief kernel that maps a state to a set of (conditional) probability measures, rather than to a single (conditional) probability measure as in the Savage paradigm. Recently, Hansen et al. [24] and Anderson et al. [1] introduced preference for robustness into an otherwise standard intertemporal asset pricing model. The issue of robustness arises from the agent's concern over misspecification of the economic model and his preference for his decision rule to be robust to the misspecification. In both of these models, learning/updating was not considered explicitly. One potential justification for it is that the Knightian uncertainty or the potential error in model specification is taken by the agent as the state of affairs. In other words, these models are the reduced form of models with earning.[2] In principle, the ultimate support for the justification requires the specification of an appropriate rule of updating to connect the reduced form to a model with learning. Our study can be viewed as a step in that direction. In particular, in [17], the agent can be viewed as using the generalized Bayes rule, axiomatized in this paper, to update his belief. Thus, this paper provides the foundation in the learning aspect for the preference used in that paper (see Section 6).

3. Consumption–information profiles

As the first step to axiomatizing updating rules, we introduce in this section the objects of choice for the conditional preferences we will introduce in the next section. Following the approach started by Skiadas [42] we model them by multi-period consumption–information profiles. Each profile has as its components a consumption process and an information filtration.

[1] See also [21] and the references therein.

[2] Seidenfeld and Wasserman [40], Dow and Werlang [13], Herron et al. [25], and Marinacci [31] show that learning and updating will not necessarily reduce Knightian uncertainty to a case of probability. Thus learning and updating need not necessarily eliminate Knightian uncertainty. See also the example in Section 5.3.

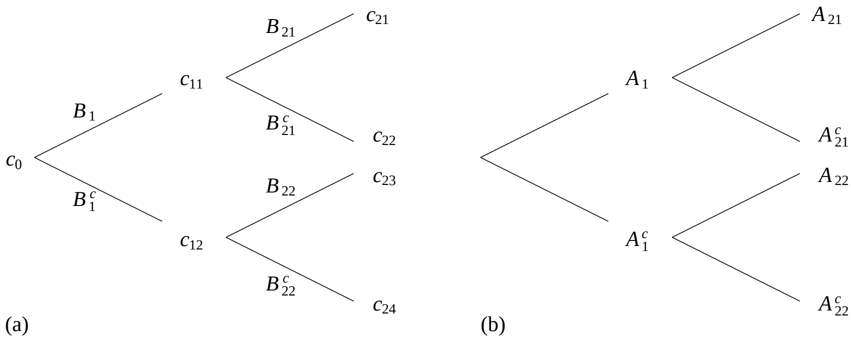

Fig. 1. Two-period consumption–information profile.

The general structure of a domain of multi-period choice objects is well understood in the literature [6,15,29]. To focus on the main issue of this paper, we will illustrate our domain, denoted by $R_+ \times D$, by an example and relegate the details of its construction to the appendix.

We begin with some notations. Let Ω be a finite set, which is taken as the state space for a generic period in the infinite time horizon. The full state space is Ω^∞. For a generic space Y, let $\mathscr{B}(Y)$ denote the space of bounded functions $\tilde{x} : \Omega \to Y$. Elements of $\mathscr{B}(R_+)$ are viewed as the state-contingent consumption in a period. Denote by $\mathbf{F}$ the following set of mappings:

$$\mathbf{F} = \left\{ \mathscr{F} \;\middle|\; \begin{array}{l} \mathscr{F} \text{ is a mapping from } \Omega \text{ to the subsets of } \Omega, 2^\Omega, \\ \text{such that } \mathscr{F}(\omega) = F_i \text{ if } |\omega \in F_i,\ i = 1, \ldots, n \\ \text{and } \{F_1, \ldots, F_n\} \text{ is a partition of } \Omega \end{array} \right\}.$$

Clearly, there is a one-to-one correspondence between elements of $\mathbf{F}$ and partitions. To simplify notation, we will use $\mathscr{F} = \{F_1, \ldots, F_n\}$ to denote both the partition and the corresponding element of $\mathbf{F}$. Members of $\mathbf{F}$ can be viewed as the information revealed in a period.

The example is a two-period profile. The profile is described diagrammatically by the two trees in Fig. 1. Fig. 1(a) describes the consumption process. The interpretation is that at time 1, if event B_1 realizes, the consumption is c_{11}; if further at time 2 event B_{21} realizes, the consumption at time 2 is c_{21}. Fig. 1(b) describes the information filtration. The interpretation is that at time 1, the individual learns one of the two possible events, A_1 or A_1^c; at time 2, he learns further one of the four possible events, A_{21}, A_{21}^c, A_{22}, A_{22}^c.

In a setting where learning is not the major concern, the object of choice is typically described by one tree as in Fig. 1(a). Here in this paper, however, our main interest is in how individual makes his choice over time in light of new information. Thus how the information evolves over time is crucial for the choice decision of the individual. Questions may arise at this point regarding whether there should be any relation between the information implied by the trees in Figs. 1(a) and (b). We will postpone the discussion until we finish the example and provide the formal definition of a consumption–information profile.

Algebraically, the profile depicted diagrammatically above is described by an element $(c_0, d_1) \in R_+ \times D$ where

$$d_1 : \omega_1 \to (c_1(\omega_1), \mathscr{F}_1(\omega_1), d_2(\omega_1)) \quad \text{is such that } \mathscr{F}_1 \in \mathbf{F},$$

$$c_1(\omega_1) = \begin{cases} c_{11} & \text{if } \omega_1 \in B_1, \\ c_{12} & \text{if } \omega_1 \in B_1^c, \end{cases} \qquad \mathscr{F}_1(\omega_1) = \begin{cases} A_1 & \text{if } \omega_1 \in A_1, \\ A_1^c & \text{if } \omega_1 \in A_1^c \end{cases}$$

and for each ω_1,

$$d_2(\omega_1) : \omega_2 \to (c_2(\omega_1; \omega_2), \mathscr{F}_2(\omega_1; \omega_2), 0)$$

is such that for each $\omega_1, \mathscr{F}_2(\omega_1) \in \mathbf{F}$ and

$$c_2(\omega_1; \omega_2) = \begin{cases} c_{21} & \text{if } (\omega_1, \omega_2) \in B_1 \times B_{21}, \\ c_{22} & \text{if } (\omega_1, \omega_2) \in B_1 \times B_{21}^c, \\ c_{23} & \text{if } (\omega_1, \omega_2) \in B_1^c \times B_{22}, \\ c_{24} & \text{if } (\omega_1, \omega_2) \in B_1^c \times B_{22}^c, \end{cases}$$

$$\mathscr{F}_2(\omega_1; \omega_2) = \begin{cases} A_{21} & \text{if } (\omega_1, \omega_2) \in A_1 \times A_{21}, \\ A_{21}^c & \text{if } (\omega_1, \omega_2) \in A_1 \times A_{21}^c, \\ A_{22} & \text{if } (\omega_1, \omega_2) \in A_1^c \times A_{22}, \\ A_{22}^c & \text{if } (\omega_1, \omega_2) \in A_1^c \times A_{22}^c. \end{cases}$$

The 0 in mapping d_2 denotes the residual profile whose consumption process is zero and whose information partition at any time is the trivial partition $\mathscr{F}^0 = \{\Omega\}$. Note that $\mathscr{F}_1$ is a partition of Ω and $\mathscr{F}_2(\omega_1)$ is also a partition of Ω given any $\omega_1 \in \Omega$. Furthermore, $\mathscr{F}_2$ is a partition of Ω^2. Thus $\{\mathscr{F}_1, \mathscr{F}_2\}$ induces a filtration on Ω^∞, which is the information filtration component of the profile.

In the algebraic representation above, the consumption–information profile is described in a recursive fashion. Such recursive description can be applied to any finite-horizon consumption–information profiles. In an infinite horizon setting, the recursion takes the form of a homeomorphism, as stated in the following theorem. This homeomorphism is crucial for our interpretation of elements of $R_+ \times D$ as trees.

Theorem 3.1. *Let $R_+ \times D$ denote the domain of consumption–information profiles. Then D is homeomorphic to $\mathscr{B}(R_+ \times \mathbf{F} \times D)$ where, for notational convenience, $\mathscr{B}(R_+ \times \mathbf{F} \times D)$ denotes the subset of elements $(c_1, \mathscr{F}_1, d_2)$ in $\mathscr{B}(R_+ \times 2^\Omega \times D)$ such that $\mathscr{F}_1 \in \mathbf{F}$.*

Remark. Despite the fact that the consumption–information profiles in $R_+ \times D$ are all infinite-horizon, some elements of D can be naturally identified with their finite-horizon counterparts.[3] For instance, the profile described in the example above corresponds naturally to a two-period consumption–information profile. In the rest of the paper, for expositional and notational simplicity, we will simply call such profiles two-period consumption–information profiles and suppress the description of the profiles beyond period two. Similar treatment is given to one-period profiles.

In modelling consumption–information profiles, we have distinguished two filtrations. One is the filtration implied by the consumption process, i.e., $\mathscr{G}_t = \sigma(c_1, \ldots, c_t)$, and the other is the information filtration induced by $\{\mathscr{F}_1, \mathscr{F}_2\}$ as in the example. One may ask why we do not require that the consumption process be adapted to the information filtration. The answer lies in the difference between the uncertainty about the objective states and the uncertainty about the probability law governing the realization of the states. To elaborate, consider a situation where there are two state variables. One state variable, following an iid binomial process, determines the payoff of the aggregate endowment of the economy. The other state variable is the signal about the likelihood of the state in which the endowment has high payoff. In this situation, while consumptions can be contingent on both state variables, only the signal is relevant for updating the knowledge about the probability law. In this example, the information filtration is that generated by the signals, and the consumption process need not be adapted to the information filtration.

It is well understood that there is a conceptual difference between one's knowledge about the objective states of the world, which is often expressed in terms of a probability law, and his/her knowledge about the probability law itself [14,27].[4] Because of the difference, at least at the conceptual level, updating of the knowledge about the probability law should be treated differently from updating of the probability due to an observation of the realization of the objective states. In the example above, the signal is relevant for the updating of knowledge of the probability law, while the first state variable is relevant for the calculation of conditional probability *given* a probability law. In a Bayesian world, such difference is irrelevant at the operational level because updating of the two different types of knowledge is done according to the same Bayes rule. Updating amounts to calculating conditional probabilities whether the updating is due to an unknown parameter of the probability distribution which one wants to learn about in order to know the true objective probability law, or due to realization of the current state even when the probability law is known precisely. In a Knightian world, uncertainty about the probability law itself is treated differently from the uncertainty about the objective states of the world. In particular, the uncertainty about the probability law

[3] See the appendix for definition of finite-horizon consumption–information profiles and their natural infinite-horizon counterparts.

[4] There can be several interpretation of Knightian uncertainty. One example is that uncertainty is purely subjective. The view we take in this paper corresponds to the setting in the Ellsberg experiment. That is, there is a true probability law out there that governs the realization of states, but the individual decision maker does not have precisely knowledge of it.

itself may not be described by another probability law. As a result, while the updating of a probability when it is *given* can be according to the Bayes rule, the updating of the knowledge of the probability law itself need not be so.

The information filtration in our consumption–information profile is to capture such difference. The example above then suggests that the adaptedness requirement may be overly restrictive for that purpose. To provide a slightly different perspective and to relate to the more familiar setting where the consumption process is required to be adapted to the information filtration, we introduce a third filtration. Let $\mathscr{G}_t = \sigma(c_s : s \leqslant t)$ be the filtration implied by the consumption process. Define the third filtration by $\mathscr{H}_t = \sigma(\mathscr{G}_t, \mathscr{F}_t)$. This filtration represents the information in $\mathscr{F}_t$ and $\mathscr{G}_t$ combined. The consumption process c_t is naturally adapted to this filtration $\mathscr{H}_t$, but $\mathscr{F}_t$ is meant to be only that part of $\mathscr{H}_t$ that is relevant for updating the individual's imprecise knowledge about the probability law.

We close this section with a comparison of our domain D with those in the literature. In [6,15,29], the space D consists of multi-period lotteries, i.e., trees with a probability attached to each of its branches. Modelling consumption profiles as multi-period lotteries implicitly assumes that the probabilities associated with various events have already been evaluated. In a world with Knightian uncertainty, probabilities are not given. To allow for the derivation of subjective beliefs, we model the space D at a more primitive level by removing the assumption of exogenously given probabilities. Wang [45] also models consumption processes as multi-period trees without probabilities. However, information filtrations are not modelled there. Our consumption–information profiles are closest to the acts in Skiadas [42], where both the consumption process and the information filtration are explicitly modelled. However, Skiadas [42] does not exploit the recursive structure in D as described in Theorem 3.1. He does not treat $\mathscr{F}_t$ and $\mathscr{H}_t$ separately either.

4. Conditional preferences

The objective of this section is to axiomatize a class of conditional preferences that will serve as the primitives for our later study of updating rules.

To reduce the clustering of notation, denote, at any time and state, a generic consumption–information profile by (c_0, d_1). Here c_0 indicates the current consumption and d_1 the remaining profile starting next period. That is, if the profile (c_0, d_1) is evaluated at time t given the past history of events $h_t = F_1 \times \cdots \times F_t$, then c_0 is the consumption at time t in that event and d_1 is the remaining profile from time $t+1$ onward. Similarly, we will also denote d_1 as $d_1 : \omega \to (c_1(\omega), \mathscr{F}_1(\omega), d_2(\omega))$, where $c_1(\omega)$ and $\mathscr{F}_1(\omega)$ are the realized consumption and information one period ahead, while $d_2(\omega)$ is the remaining profile from time $t+2$ onward.

Let $t \geqslant 1$ and $h_t \subset \Omega^t$ be a history of past events. A conditional preference given h_t, denoted by $\succcurlyeq_{h_t}$, is a complete ordering on $R_+ \times D$. A family of conditional preferences is a collection of conditional preferences indexed by all possible evolution of past events, i.e., $\{\succcurlyeq_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$.

Axiom 1 (Continuity).[5] *For all* h_t *and all sequences* $\{(c_{0,n}, d_{1,n})\}$ *and* $\{(c'_{0,n}, d'_{1,n})\}$ *in* $R_+ \times D$ *with* $(c_{0,n}, d_{1,n}) \to (c_0, d_1)$ *and* $(c'_{0,n}, d'_{1,n}) \to (c'_0, d'_1)$, *if* $(c_{0,n}, d_{1,n}) \succsim_{h_t} (c'_{0,n}, d'_{1,n})$ *for all* n, *then* $(c_0, d_1) \succsim_{h_t} (c'_0, d'_1)$. *Moreover, for each* $(c_0, d_1) \in R_+ \times D$, *there exists a* $\hat{c}_0 \in R_+$ *such that* $(\hat{c}_0, 0) \sim_{h_t} (c_0, d_1)$.

Axiom 2 (Uncertainty separability). *For all* h_t, $(c_0, c'_0) \in R^2_+$ *and* $(d_1, d'_1) \in D^2$, $(c_0, d_1) \succsim_{h_t} (c_0, d'_1)$ *if and only if* $(c'_0, d_1) \succsim_{h_t} (c'_0, d'_1)$. *Moreover,* $(c_0, d_1) \succ (c'_0, d_1)$ *whenever* $c_0 > c'_1$.

Axiom 3 (Stationarity). *For all* (c_0, d_1), $(c'_0, d'_1) \in R_+ \times D$ *and* h_t, $(c_0, d_1) \succsim_{h_t} (c'_0, d'_1)$ *if and only if* $(c_0, d_1) \succsim_{h_t \times \Omega} (c'_0, d'_1)$.

A consumption–information profile is called deterministic if its consumption process is deterministic and its information filtration consists of the trivial partition $\mathscr{F}^0 = \{\Omega\}$. We will sometimes denote these profiles by $(c_0, c_1, \ldots)$.

Axiom 4 (Deterministic information independence). *For all* h_t, h'_t, $(c_0, c_1, \ldots)$ *and* $(c'_0, c'_1, \ldots)$, $(c_0, c_1, \ldots) \succsim_{h_t} (c'_0, c'_1, \ldots)$ *if and only if* $(c_0, c_1, \ldots) \succsim_{h'_t} (c'_0, c'_1, \ldots)$. *Moreover, for any* $\mathscr{F}$ *and* $\mathscr{G} \in \mathbf{F}$, $(c_0, d_1) \sim_{h_t} (c_0, d'_1)$ *whenever* $d_1 : \omega \to (c_1(\omega), \mathscr{F}(\omega), d_2(\omega))$ *and* $d'_1 : \omega \to (c_1(\omega), \mathscr{G}(\omega), d_2(\omega))$ *are such that* $d_2(\omega)$ *is a deterministic profile for all* $\omega \in \Omega$.

These four axioms are straightforward to interpret. Continuity is standard. Uncertainty separability says that if two consumption–information profiles have identical current consumption c_0, or identical future consumption–information flow d_1, then the ranking of these two profiles should be independent of the common component. Stationarity requires that if there is no information revealed over the period, the conditional preference remains unchanged. Finally, the first part of Deterministic information independence requires that all conditional preferences rank deterministic consumption–information profiles the same way. The second part says if all uncertainty in the consumption process is resolved in the next period, then any updating of preference due to the information received in the next period is irrelevant.

To state the next axiom we need the following definition. An event $A \subset \Omega$ is said to be null given the history h_t if for all $c_0 \in R_+$ and d_1, d'_1 and $d''_1 \in D$, $(c_0, d'_1 1_A + d_1 1_{A^c}) \sim_{h_t} (c_0, d''_1 1_A + d_1 1_{A^c})$, where 1_A and 1_{A^c} are indicator functions, and the addition and multiplication are as in the space of random variables. Null events are those that are considered to have zero likelihood of happening.

Axiom 5 (Consistency). *For all* $d_1 = (c_1, \mathscr{F}, d_2)$ *and* $d'_1 = (c'_1, \mathscr{F}, d'_2) \in D$ *and* $\mathscr{F} = \{A_1, \ldots, A_n\} \in \mathbf{F}$, *if* $(c_1(\omega), d_2(\omega)) \succsim_{h_t \times A_i} (c'_1(\omega), d'_2(\omega))$, *for all* $\omega \in A_i$ *and* $i = 1, \ldots, n$,

[5] See the appendix for definition of the topology.

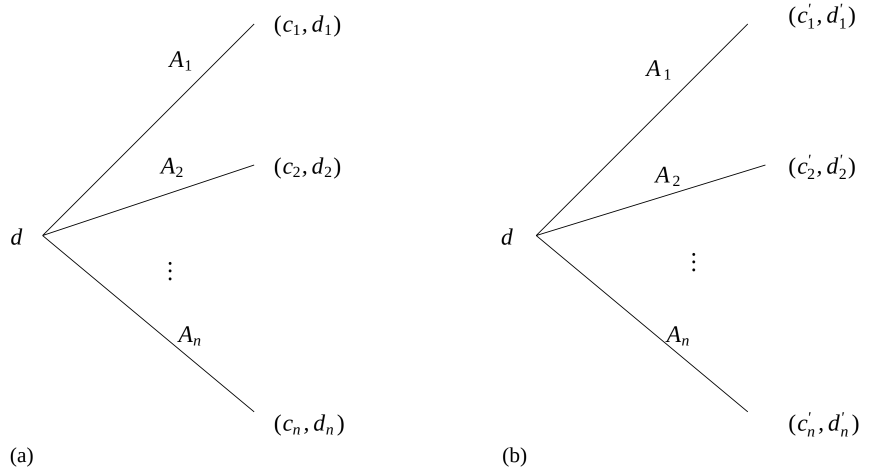

Fig. 2. Consistency.

then for any $c_0 \in R_+$,

$$(c_0, d_1) \succsim_{h_t} (c_0, d_1').$$

Moreover, the latter ordering is strict if for some A_i *that is not null,* $(c_1(\omega), d_2(\omega)) \succ_{h_t \times A_i} (c_1'(\omega), d_2'(\omega))$ *for all* $\omega \in A_i$.

The intuition behind this axiom is readily explained with Fig. 2. In Fig. 2(a), when event A_i happens, the realized consumption–information profile is (c_i, d_i). Fig. 2(b) has a similar interpretation. Suppose that for all $i = 1, \ldots, n$, when event A_i is realized, (c_i, d_i) is preferred to (c_i', d_i'). Consistency then requires that ex-ante $(c, [\sum_{i=1}^n c_i 1_{A_i}, \mathscr{F}, \sum_{i=1}^n d_i 1_{A_i}])$ is preferred to $(c, [\sum_{i=1}^n c_i' 1_{A_i}, \mathscr{F}, \sum_{i=1}^n d_i' 1_{A_i}])$ for any $c \in R_+$. In words, if ex-post d is preferred to d' and the information revealed over the period is identical, then ex-ante d must also be preferred to d'.

The intuitive appeal of the consistency axiom seems obvious. As will be seen below this axiom guarantees that the conditional preferences connect in a time-consistent fashion. It is well understood that in a dynamic optimization problem, if the intertemporal preference of the decision maker is not time-consistent, a plan chosen today may be regretted later on, and, if given the opportunity, the plan will be abandoned in favor of another one, causing inconsistency in choice over time.

For the following theorem, we need some preliminary definitions and notations. A function $V(h_t) : R_+ \times D \to R$ is said to represent the conditional preference $\succsim_{h_t}$ if, for all (c_0, d_1) and $(c_0', d_1') \in R_+ \times D$,

$$(c_0, d_1) \succsim_{h_t} (c_0', d_1')$$

if and only if

$$V(h_t, (c_0, d_1)) \geqslant V(h_t, (c_0', d_1')).$$

Given a function $V(h_t): R_+ \times D \to R$ that represents the conditional preference $\succsim_{h_t}$, define a companion function $\tilde{V}(h_t, d_1): \Omega \to R$ by, for each $d_1 \in D$ given by

$$d_1 : \omega \to (c_1(\omega), \mathscr{F}_1(\omega), d_2(\omega)), \quad \text{where } \mathscr{F}_1 = \{A_1, \dots, A_n\} \in \mathbf{F},$$

$$\tilde{V}(h_t, d_1)(\omega) = V(h_t \times A_i, (c_1(\omega), d_2(\omega))), \quad \text{if } \omega \in A_i.$$

$\tilde{V}(h_t, d_1)$ can be regarded as the ex-post evaluation of d_1 after the uncertainty in the current period is realized. A function $\mu : \mathscr{B}(R) \to R$ is called a *certainty equivalent* if (a) $\mu(x) = x$ for all $x \in R$, and (b) $\mu(\tilde{x}) \geqslant \mu(\tilde{y})$ if $\tilde{x} \geqslant \tilde{y}$.

Theorem 4.1. *A family of conditional preferences* $\{\succsim_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ *satisfies Axioms* 1–5 *if and only if it can be represented by a family of continuous functions* $\{V(h_t) : h_t \subset \Omega^t, t \geqslant 1\}$ *on* $R_+ \times D$ *such that*

$$V[h_t, (c_0, d_1)] = W(c_0, \mu(h_t, \tilde{V}(h_t, d_1))),$$

where $\mu(h_t, \cdot)$ *is a continuous certainty equivalent such that* $\mu(h_t) = \mu(h_t \times \Omega)$, *and* $W : R_+ \times R \to R$ *is continuous and strictly increasing.*[6]

The main outcome of this theorem is a description of how a family of conditional preferences that satisfies Axioms 1–5 are connected. The theorem states that they are connected through two aggregators.[7] The function W is the intertemporal aggregator, which describes, for deterministic consumption profiles, how conditional utility derived from future consumptions is aggregated with that derived from current consumption. For example, for a deterministic consumption–information profile $(c_0, d_1) = (c_0, c_1, \dots)$, $V(c_0, d_1) = W(c_0, W(c_1, (\cdots)))$. The certainty equivalent μ is the state aggregator. It aggregates utilities derived from state-contingent consumption–information profiles, taking into consideration the fact that preferences are constantly updated in light of new information. These two aggregators together allow for a weak separation of preference from belief, in the sense that for deterministic profiles, the utility is completely determined by the function W, and that μ is the aggregator that embodies the belief component of the preference and the rule of updating. Because of this structure of (W, μ), our theorems on updating rules in the next section are directed to μ only.

Consider next an axiom that is similar to Uncertainty separability, but is with respect to deterministic consumption–information profiles.

Axiom 6 (Future independence). *For all* $h_t \subset \Omega^t$, *all* c_0, c_1, c'_0 *and* $c'_1 \in R_+$ *and deterministic consumption–information profiles* $C = (c_2, c_3, \dots)$ *and* $C' = (c'_2, c'_3, \dots) \in D$, $(c_0, c_1, C) \succsim_{h_t} (c'_0, c'_1, C)$ *if and only if* $(c_0, c_1, C') \succsim_{h_t} (c'_0, c'_1, C')$.

[6] Consistency, uncertainty separability, deterministic information independence and stationarity imply the usual monotonicity.

[7] See a similar theorem in [6,41,42].

If we add this axiom to Axioms 1–5, the time aggregator W can be significantly simplified.

Theorem 4.2. *A family of conditional preferences* $\{\succcurlyeq_{h_t} : h_t \subset \Omega^t, t \geqslant 0\}$ *satisfies Axioms* 1–6 *if and only if it can be represented by a family of continuous functions of the form*

$$V[h_t, (c_0, d_1)] = u(c_0) + \beta\mu(h_t, \tilde{V}(h_t, d_1)),$$

where $\mu(h_t, \cdot)$ *is a continuous certainty equivalent such that* $\mu(h_t) = \mu(h_t \times \Omega)$, *and* $u : R_+ \to R$ *is strictly increasing and continuous. Furthermore*, u *is unique upto affine transforms.*

Without loss of generality, we will assume that $u(0) = 0$.

5. Updating rules

We have axiomatized the class of conditional preferences that will serve as the primitives for our investigation of updating rules. In this section, we will study three updating rules.

5.1. Bayes rule

Axiom 7 (Timing indifference).[8] *For all partitions of* Ω, $\{A_1, \ldots, A_n\}$ *and* $\{B_1, \ldots, B_m\}$, *all two-period consumption–information profiles of the form* $(0, d_1)$ *and* $(0, d_1')$ *with* $d_1 = (0, \mathscr{F}^0, d_{2i})$ *if* $\omega_1 \in A_i$ *and* $d_{2i}(\omega_2) = c_{ij}$ *if* $\omega_2 \in B_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$; $d_1' = (0, \mathscr{F}^0, d_{2j}')$ *if* $\omega_1 \in B_j$ *and* $d_{2j}'(\omega_2) = c_{ij}$ *if* $\omega_2 \in A_i$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, *and all* $h_t \subset \Omega^t$, *we have* $(0, d_1) \sim_{h_t} (0, d_1')$.

The intuition behind this axiom can be readily explained. Suppose that $(0, d_1)$ and $(0, d_1')$ are as described in the axiom. The consumption processes embedded in these two two-period profiles can be illustrated with Fig. 3. There are two events A and B. In Fig. 3(a), event A happens first and event B follows. In Fig. 3(b), event B happens first and then event A follows. It can be easily verified that the state-contingent consumptions at time 2 in Fig. 3(a) are identical to those in Fig. 3(b). The information filtrations embedded in both $(0, d_1)$ and $(0, d_1)$ are the same. They are the trivial information partition, $\mathscr{F}^0$, which means that, as far as updating is concerned, there is no new information revealed over the period, and hence nothing is to be learned.[9] There are no consumptions at times 0 and 1 in both of the trees. The only difference between Figs. 3(a) and (b) is that the timing of resolution of

[8] See [5,6] for a similar axiom in the lottery framework.

[9] To be precise, the filtration should be described by a process $(\mathscr{F}_1, \mathscr{F}_2)$ as in the example in Section 3. However, since both $(0, d_1)$ and $(0, d_1)$ are two-period profiles, the only relevant information is given by the partition at time 1. In the example in Section 3, this is the partition given by $\mathscr{F}_1$. The partition at time 2 is irrelevant for $(0, d_1)$ and $(0, d_1)$ because there are no more consumptions after time 2.

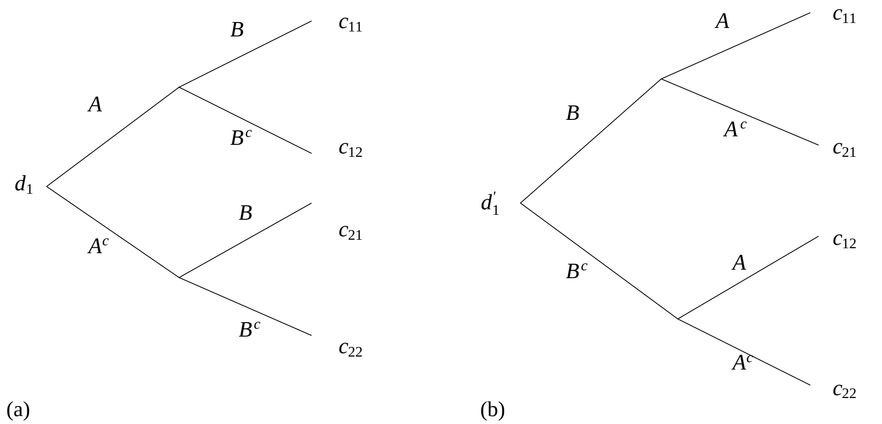

Fig. 3. Timing indifference.

uncertainty in the two trees are reversed. Axiom 7 says that in this situation, the two consumption–information profiles should be ranked indifferent.

The following theorem is a preparation for the main result of this section, the Bayes rule. It is, however, of independent interest.

Theorem 5.1.[10] *Suppose that the family* $\{\succsim_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ *of conditional preferences satisfies Axioms* 1–6. *Then it satisfies timing indifference if and only if there exists a unique family of probability measures* $\{P(h_t) : h_t \subset \Omega^t, t \geqslant 1\}$ *on* Ω *and a family of continuous, strictly increasing functions* $\{\psi_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ *such that the family of certainty equivalents* $\{\mu(h_t) : h_t \subset \Omega^t, t \geqslant 1\}$ *in Theorem* 4.2 *are given by, for any* $\tilde{x} \in \mathscr{B}(R)$,

$$\mu(h_t, \tilde{x}) = \psi_{h_t}^{-1}(E_{P(h_t)}[\psi_{h_t}(\tilde{x})]),$$

where $P(h_t, \cdot)$ *and* ψ_{h_t} *satisfy*,

$$\psi_{h_t}(0) = 0, \quad P(h_t, \cdot) = P(h_t \times \Omega, \cdot), \quad \psi_{h_t} = \psi_{h_t \times \Omega}$$

and for all $\tilde{x} \in \mathscr{B}(R)$,

$$\beta\psi_{h_t}^{-1}(E_{P(h_t)}[\psi_{h_t}(\tilde{x})]) = \psi_{h_t}^{-1}(E_{P(h_t)}[\psi_{h_t}(\beta\tilde{x})]).$$

Therefore, *by Theorem* 4.2,

$$V(h_t, (c_0, d_1)) = u(c_0) + \beta\psi_{h_t}^{-1}(E_{P(h_t)}[\psi_{h_t}(\tilde{V}(h_t, d_1))]).$$

Now that we have a theorem characterizing the certainty equivalents $\mu(h_t)$, we turn to its implication for updating. Let (c_0, d_1) be a consumption–information profile and $\{c_t\}$ and $\{\mathscr{F}_t\}$ be the consumption and information components of it,

[10] See [5,6] for a related result in the lottery framework.

respectively. Recall that $\mathscr{G}_t = \sigma(c_s, s \leqslant t)$, the filtration generated by the consumption process, and $\mathscr{H}_t = \sigma(\mathscr{G}_t, \mathscr{F}_t)$.

Theorem 5.2 (Bayes rule). *Suppose that the family* $\{\succcurlyeq_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ *of conditional preferences satisfies Axioms 1–7. Let* $\mathscr{F} = \{\mathscr{F}_t\}$ *be an increasing sequence of σ-algebras and* $R_+ \times D_{\mathscr{F}}$ *be the subset of consumption–information profiles* (c_0, d_1) *such that the information filtration embedded in* d_1 *coincides with* $\mathscr{F}$ *and the consumption process embedded in* d_1 *is uniformly bounded. Then there exists a unique probability measure* P_0 *on* $(\Omega^\infty, \mathscr{F}_\infty)$ *where* $\mathscr{F}_\infty = \sigma(\Omega_t : t = 1, \ldots)$ *such that for any* $(c_0, d_1) \in D_{\mathscr{F}}$ *and any elementary history* h_t,[11]

$$V(h_t, (c_t, d_{t+1})) = u(c_t) + \beta \psi_{h_t}^{-1}(E_{P_0}[\psi_{h_t}(\tilde{V}(h_t, d_{t+1})) | \mathscr{F}_t]),$$

where (c_t, d_{t+1}) *is the continuation of* (c_0, d_1) *at time t. In other words, on* $D_{\mathscr{F}}$, *the family of conditional preferences has a belief component described by the initial probability measure* P_0 *and the belief updates according to the Bayes rule.*

To illustrate Theorem 5.2 with the standard intertemporally additive expected utility, let $\psi_{h_t}(x) = x$ for all h_t. Then Theorem 5.2 says

$$V(h_t, (c_t, d_{t+1})) = E_{P_0}\left[\sum_{s=0}^{\infty} \beta^s u(c_{t+s}) \middle| \mathscr{H}_t\right].$$

In other words, the individual who is described by this family of conditional preferences behaves as if he has an initial prior P_0 and updates according to the Bayes rule.

To close this subsection, we clarify the role of ψ_{h_t} in Theorems 5.1 and 5.2. Let $(0, d_1) = (0, (\tilde{c}_1, \mathscr{F}))$ be a one-period consumption–information profile. By Theorem 5.1,

$$V(h_t, (0, d_1)) = \beta \psi_{h_t}^{-1}(E_{P(h_t)}[\psi_{h_t}(u(\tilde{c}_1))]).$$

It should be clear from this expression that the more concave ψ_{h_t} is the more risk averse the conditional preference is. Thus ψ_{h_t} can be viewed as a (state-contingent) risk aversion parameter of the conditional preference.

5.2. Dempster–Shafer rule

The Dempster–Shafer rule for updating non-additive probability measures first appeared in Dempster [11,12] and Shafer [38,39] in the statistics literature. Our axiomatization of the rule is based on the following timing indifference axiom.

Axiom 8 (Comonotonic timing indifference). *For all* $h_t \subset \Omega^t$, *all partitions of* Ω, $\{A_1, \ldots, A_n\}$ *and* $\{B_1, \ldots, B_m\}$, *and all two-period consumption–information profiles*

[11] An history $h_t \in \mathscr{F}_t$ is called elementary if there does not exist another history $h'_t = F'_1 \times \cdots \times F'_t \in \mathscr{F}_t$ such that $h'_t \subset h_t$ and $h'_t \neq h_t$.

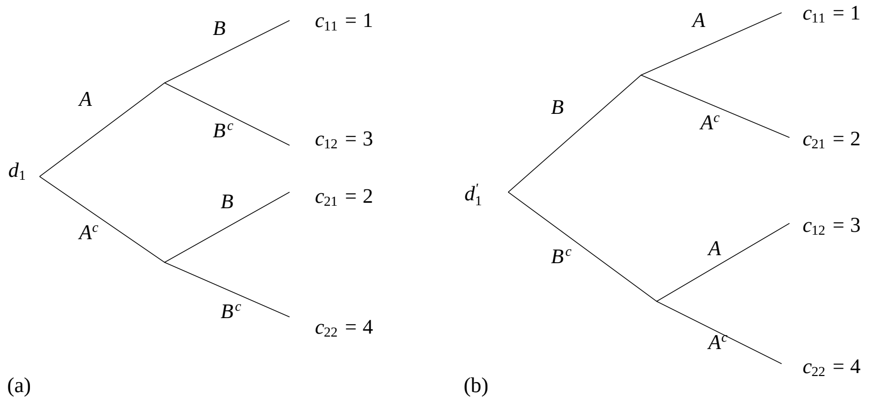

Fig. 4. Comonotonic timing indifference.

$(0, d_1)$ and $(0, d_1')$ such that $d_1(\omega_1) = (0, \mathscr{F}^0, d_{2i})$ if $\omega_1 \in A_i$ and $d_{2i}(\omega_2) = c_{ij}$ if $\omega_2 \in B_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$; $d_1'(\omega_1) = (0, \mathscr{F}^0, d_{2j}')$ if $\omega_1 \in B_j$ and $d_{2j}'(\omega_2) = c_{ij}$ if $\omega_2 \in A_i$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, we have $(0, d_1) \succsim_{h_t} (0, d_2')$, provided that $c_{i1} \leqslant \cdots \leqslant c_{im}$ and $c_{1j} \leqslant \cdots \leqslant c_{nj}$ for all i and j.

To understand the intuition formally expressed in Axiom 8, let us focus first on the difference between Axioms 7 and 8, which is the additional condition that $c_{i1} \leqslant \cdots \leqslant c_{im}$ and $c_{1j} \leqslant \cdots \leqslant c_{nj}$ for all i and j. This condition is related to the notion of comonotonicity introduced by Schmeidler [36]. Recall that any two random variables, $\tilde{x}$ and $\tilde{y}$, are said to be comonotonic if for all ω and $\omega' \in \Omega$ such that $\tilde{x}(\omega) \neq \tilde{x}(\omega')$ and $\tilde{y}(\omega) \neq \tilde{y}(\omega')$, $[\tilde{x}(\omega) - \tilde{x}(\omega')][\tilde{y}(\omega) - \tilde{y}(\omega')] > 0$. Now consider the two two-period consumption processes in Fig. 4. Here $c_{i1} < c_{i2}$ and $c_{1j} < c_{2j}$ for $i, j = 1$, 2, which is the condition of Axiom 8 with strict inequalities. If the two subtrees at time 1 in Fig. 4(a) are viewed as two random variables, then they are comonotonic. The same is also true for the subtrees in Fig. 4(b). Thus another way of stating the additional condition in Axiom 8 is that the two subtrees at time 1 are comonotonic.[12]

Comonotonicity is a convenient way of describing the relationship among better-than sets. Let $\tilde{c} = (c_1, A_1, c_2, A_2, \ldots, c_n, A_n)$ be a state-contingent consumption with $c_1 < c_2 < \cdots < c_n$. The better-than sets are, by definition,

$$A_n = \{\tilde{c} \geqslant c_n\}, A_n \cup A_{n-1} = \{\tilde{c} \geqslant c_{n-1}\}, \ldots, A_n \cup \cdots \cup A_1 = \Omega = \{\tilde{c} \geqslant c_1\}.$$

Observe that two random variables $\tilde{x}$ and $\tilde{y}$ are comonotonic if and only if the following two conditions hold: (a) $\tilde{x}$ and $\tilde{y}$ assume the same number of distinct values, say $x_1 < x_2 < \cdots < x_n$, $y_1 < y_2 < \cdots < y_n$, and (b) they have the same better-than

[12] We have just made the connection between the additional condition in Axiom 8 with comonotonicity when the inequalities in the condition are strict. The general case with weak inequalities can be viewed as the limit of the case with strict inequalities.

sets, $\{\omega\in\Omega:\tilde{x}(\omega)\geqslant x_i\}=\{\omega\in\Omega:\tilde{y}(\omega)\geqslant y_i\}$ for all i. Thus if $\{c_{ij}\}$ satisfy the comonotonicity condition, then the better-than sets in the two subtrees in Fig. 4(a) are identical (B^c and Ω). Since in Fig. 4(a) the payoff in each branch of the lower subtree is greater than that in the corresponding branch of the upper subtree, the lower sub-tree in $(0,d_1)$ has higher utility o the individual than the upper sub-tree. Suppose that the utility of the upper and lower sub-trees of $(0,d_1)$ are V_1 and V_2, respectively, with $V_1<V_2$. Then, at time 0 looking one-period ahead, the better-than sets for $(0,d_1)$ are $A^c=\{\tilde{V}\geqslant V_2\}$ and $\Omega=\{\tilde{V}\geqslant V_1\}$, where $\tilde{V}$ is the state-contingent utility at time 1. All together, the better-than sets in $(0,d_1)$ are $\{A^c,\Omega\}$ and $\{B^c,\Omega\}$. It can be readily verified that $(0,d_1')$ has the same better-than sets. The only difference is that in Fig. 4(b), the better-than sets of the subtrees are A^c and Ω, and those at time 0 are B^c and Ω. Thus the comonotonicity condition in Axiom 8 can be restated as: $(0,d_1)$ and $(0,d_1')$ have the same collection of better-than sets, but in reverse timing as described above.

Turn now to the relevance of better-than sets for an individual's choice behavior. For comparison, consider first the case where an individual has precise knowledge of the objective probability law. In this case, he is able to assign probabilities to all sets of the form $\{\tilde{c}=c_i\}$, $i=1,\dots,n$. Moreover, the assignment is independent of the payoff to the individual on the events. Now return to the case where the individual understands that there is an objective probability law that governs the realization of uncertainty, but does not have the precise knowledge about this law. In this case, precise likelihood can no longer be assigned to events like $\{\tilde{c}=c_i\}$. The revealed likelihood assignment is likely to be dependent on the payoff of $\tilde{c}$ on the event to the individual. In this situation, since better-than sets are characterized by worst-case payoffs, they seem to be a natural input that the individual would need, in combination with his vague assessment of the likelihood of events, to evaluate a uncertain consumption prospect.

Let $\{A_1,\dots,A_n\}\in\mathbf{F}$ and $\tilde{x}$ be a random variable on Ω that takes values $x_1<\cdots<x_n$ on the partition. Let $B_i=\bigcup_{j=i}^n A_j$, $i=1,\dots,n$, be the better-than sets. If v is a monotonic set function such that $v(\emptyset)=0$ and $v(\Omega)=1$, then the Choquet integral [8] of $\tilde{x}$ with respect to v is defined as

$$E_v[\tilde{x}]\equiv\int\tilde{x}\,dv=\sum_{i=1}^n[v(B_i)-v(B_{i+1})]x_i,$$

where $B_{n+1}=\emptyset$. It reduces to the standard integral when v is a probability measure.

Theorem 5.3. *Suppose that the family $\{\succcurlyeq_{h_t}:h_t\subset\Omega^t,t\geqslant 1\}$ of conditional preferences satisfies Axioms 1–6. Then $\{\succcurlyeq_{h_t}\}$ satisfies comonotonic timing indifference if and only if there exist a unique family of monotonic set functions $\{v(h_t,\cdot)\}$ and a family of continuous, strictly increasing functions $\{\psi_{h_t}\}$ such that, for any $\tilde{x}\in\mathscr{B}(R)$,*

$$\mu(h_t,\tilde{x})=\psi_{h_t}^{-1}(E_{v(h_t)}[\psi_{h_t}(\tilde{x})]),$$

where $v(h_t, \cdot)$ *and* ψ_{h_t} *satisfy*,

$$\psi_{h_t}(0) = 0, \quad v(h_t, \cdot) = v(h_t \times \Omega, \cdot), \quad \psi_{h_t} = \psi_{h_t \times \Omega}$$

and, *for all* $\tilde{x} \in \mathscr{B}(R)$,

$$\beta \psi_{h_t}^{-1}(E_{v(h_t)}[\psi_{h_t}(\tilde{x})]) = \psi_{h_t}^{-1}(E_{v(h_t)}[\psi_{h_t}(\beta \tilde{x})]).$$

Therefore, *by Theorem* 4.2,

$$V(h_t, (c_0, d_1)) = u(c_0) + \beta \psi_{h_t}^{-1}(E_{v(h_t)}[\psi_{h_t}(\tilde{V}(h_t, d_1))]).$$

Parallel to the previous subsection, we consider the implication of this theorem for updating.

Theorem 5.4 (Dempster–Shafer rule). *Suppose that the family* $\{\succcurlyeq_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ *of conditional preferences satisfies Axioms* 1–6 *and* 8. *Let* $\mathscr{F} = \{\mathscr{F}_t\}$ *be an increasing sequence of* σ*-algebras and* $R_+ \times D_{\mathscr{F}}$ *be the subset of consumption–information profiles* (c_0, d_1) *such that the information filtration embedded in* d_1 *coincides with* $\mathscr{F}$. *Then there exists a monotonic set function* v_0 *on* $(\Omega^\infty, \mathscr{F}_\infty)$ *where* $\mathscr{F}_\infty = \sigma(\Omega_t : t = 1, \ldots)$ *such that for any* $(c_0, d_1) \in D_{\mathscr{F}}$ *and any elementary event* $h_t \in \mathscr{F}_t$,

$$V(h_t, (c_t, d_{t+1})) = u(c_t) + \beta \psi_{h_t}^{-1}(E_{v_0}[\psi_{h_t}(\tilde{V}(h_t, d_{t+1})) | \mathscr{F}_t]),$$

where (c_t, d_{t+1}) *is the continuation of* (c_0, d_1) *at time* t *and*, *for any event* $B \subset \Omega$ *and* $h_t \in \mathscr{F}_t$, *and* $v_0(B|h_t)$ *is given by the Dempster–Shafer rule*,

$$v_0(B|h_t) = \frac{v_0([h_t \times B] \cup [(h_t)^c \times \Omega]) - v_0((h_t)^c \times \Omega)}{1 - v_0((h_t)^c \times \Omega)}.$$

Dempster–Shafer rule has been well studied in the statistics literature (see, for example, [38]). Its incorporation into a preference/behavior framework, however, has not been easy, mainly because of the time-consistency problem raised in [16]. Theorem 5.4 provides an alternative perspective on the issue. Consider first the result in [16]. Let $c_2 : \Omega^2 \to R_+$ be an act and let the initial preference be represented by a Choquet expected utility,

$$U(c_2) = \int u(c_2)\, d\bar{v}, \tag{1}$$

where $\bar{v}$ is a capacity over Ω^2. Suppose that conditional any event, the preference can still be represented by a Choquet expected utility (see [21, Proposition 3.2] for conditions under which this is true). Now we require that the preference be dynamically consistent so that,

$$U(c_2) = \int \left[\int u(c_2(\omega_1, \omega_2))\ \bar{v}(\omega_1, d\omega_2) \right] \bar{v}(d\omega_1), \tag{2}$$

where $\bar{v}(\omega_1, \cdot)$ is the conditional capacity derived from $\bar{v}$ and the inner integral is the conditional Choquet expected utility. Then by a mathematical result of Graf [23], (1) and (2) imply that v must be a probability measure. Thus we arrive at a version of the result of Epstein and LeBreton [16] that dynamically consistent preferences must be Bayesian.

Consider next Theorem 5.4 applied, with $t = 0$, $\psi_{h_t}(x) = x$ and $\beta = 1$, to a two-period consumption–information profile of the form $(0, d_1)$ where

$$d_1 : \omega_1 \to (0, \mathscr{F}_1(\omega_1), d_2(\omega_1)), \quad \mathscr{F}_1 = \{A_1, \ldots, A_n\}, \quad d_2(\omega_1, \omega_2) = c_2(\omega_1, \omega_2).$$

This profile is the counterpart of the act c_2 in our consumption–information profile framework. For such a profile,

$$V(0, d_1) = \int \int u(c_2(\omega_1, \omega_2)) \; v(\omega_1, d\omega_2) \; v(d\omega_1), \tag{3}$$

where $v(\omega_1, d\omega_2) = v(A_i, d\omega_2)$ if $\omega_1 \in A_i$. Eq. (3) is formally identical to Eq. (2). Because of the recursive structure of Eq. (3), dynamic consistency is retained. However, there is no requirement of Eq. (1) in our framework. In other words, we do not require that conditional preference be the restriction of the prior preference. As a result, v_0 need not be a probability measure.

In a sense, the discussion above is a restatement of the result in [16]. It says if one starts with conditional preferences and construct a v_0 on subsets of Ω^2 so that v_0 and $v(\cdot|h_t)$ are consistent with the Dempster–Shafer rule, and if the conditional preferences are time-consistent, then v_0 will not be consistent with the initial preference at time zero.

5.3. Generalized Bayes rule

The generalized Bayes rule has been studied in the statistics literature [44]. The axiomatization we provide below is based on a notion of pessimism by Wakker [43].

Axiom 9 (Pessimism). *Let $(0, d_x)$ with $d_x = (\tilde{x}, \mathscr{F}^0)$ and $(0, d_y)$ with $d_y = (\tilde{y}, \mathscr{F}^0)$ be two one-period consumption–information profiles, where $\tilde{x}$ and $\tilde{y} \in \mathscr{B}(R_+)$ assume values*

$$x_1 \leqslant \cdots \leqslant x_{i_1} \leqslant \cdots \leqslant x_{i_2} \leqslant \cdots \leqslant x_N$$

and

$$y_1 \leqslant \cdots \leqslant y_{i_1} \leqslant \cdots \leqslant y_{i_2} \leqslant \cdots \leqslant y_N$$

on non-null events $A_1, A_2, \ldots, A_N$, respectively, such that $x_i = y_i$ for $i_1 \leqslant i < i_2$ and $x_{i_2} > y_{i_2}$. Let $(0, d'_x)$ with $d'_x = (\tilde{x}', \mathscr{F}^0)$ and $(0, d'_y)$ with $d'_y = (\tilde{y}', \mathscr{F}^0)$ be another two one-period consumption–information profiles, where $\tilde{x}'$ and $\tilde{y}' \in \mathscr{B}(R_+)$ assume values

$$x'_1 \leqslant \cdots \leqslant x'_{i_1} \leqslant \cdots \leqslant x'_{i_2} \leqslant \cdots \leqslant x'_N$$

and

$$y'_1 \leqslant \cdots \leqslant y'_{i_1} \leqslant \cdots \leqslant y'_{i_2} \leqslant \cdots \leqslant y'_N$$

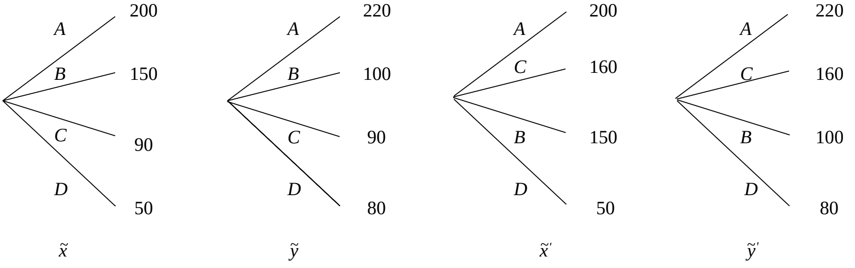

Fig. 5. Pessimism.

on $A'_1, A'_2, \ldots, A'_N$, *respectively, such that* (a) $A'_{i_1} = A_{i_2}$, $A'_{i_2} = A_{i_1}$, *and* $A'_i = A_i$ *for* $i \neq i_1$ *and* i_2; *and* (b) $x'_{i_1} = x_{i_2}$, $y'_{i_1} = y_{i_2}$, $x'_i = y'_i$ *for* $i_1 < i \leqslant i_2$, *and* $x'_i = x_i$, $y'_i = y_i$ *for* $i > i_2$ *and* $i < i_1$. *For all* $h_t \subset \Omega^t$ *and* $t \geqslant 1$, *if* $(0, d_x) \sim_{h_t} (0, d_y)$, *then* $(0, d'_x) \succcurlyeq_{h_t} (0, d'_y)$.

While the formalism in the axiom is a bit involved, the sense of pessimism it describes is very simple. It says that an individual is pessimistic if he assigns more likelihood to the lower outcomes. This is best seen for conditional preferences that satisfy Axioms 1–6 and 8. Let $\tilde{x}$, $\tilde{y}$, $\tilde{x}'$ and $\tilde{y}'$ be as described in the axiom. Fig. 5 describes a case with four outcomes. Suppose that, as in Theorem 5.3 with $\psi_{h_t}(x) = x$,

$$V(h_t, (0, d_x)) = \beta \sum_{i=1}^{n} [v(h_t, B_i) - v(h_t, B_{i+1})]u(x_i),$$

where $B_i = \bigcup_{j=i}^{n} A_j$, $i = 1, \ldots, N$. Borrowing an intuition from expected utility, $v(h_t, B_i) - v(h_t, B_{i+1})$ can be called the implied/revealed likelihood assigned to the outcome x_i. Then,

$$V(h_t, (0, d_x)) = V(h_t, (0, d_y))$$

is equivalent to

$$\sum_{i=1}^{n} [v(h_t, B_i) - v(h_t, B_{i+1})][u(x_i) - u(y_i)] = 0.$$

Similarly,

$$V(h_t, (0, d'_x)) \geqslant V(h_t, (0, d'_y))$$

is equivalent to

$$\sum_{i=1}^{n} [v(h_t, B'_i) - v(h_t, B'_{i+1})][u(x'_i) - u(y'_i)] \geqslant 0,$$

where $B'_i = \bigcup_{j=i}^{n} A'_j$, $i = 1, \ldots, N$. A subtraction yields

$$(v(h_t, B_{i_2}) - v(h_t, B_{i_2+1}))[u(x_{i_2}) - u(y_{i_2})] \leqslant (v(h_t, B_{i_1}) - v(h_t, B_{i_1+1}))[u(x_{i_2}) - u(y_{i_2})],$$

which is true if and only if,

$$v(h_t, B_{i_2}) - v(h_t, B_{i_2+1}) \leqslant v(h_t, B_{i_1}) - v(h_t, B_{i_1+1}).$$

That is, when the number of outcomes higher than x_{i_2} is larger in $\tilde{x}'$ than in $\tilde{x}$, x_{i_2} is assigned higher likelihood in $\tilde{x}'$ than in $\tilde{x}$, which is precisely what the intuitive definition of pessimism suggests.

An axiom of optimism can be symmetrically defined. The following theorem is established using Axiom 9. Pessimism is captured by the minimization over a set of probability measures. There is also a version of the theorem for optimism by symmetry.

Theorem 5.5. *Suppose that the family $\{\succsim_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ of conditional preferences satisfies Axioms* 1–6 *and* 8 *so that Theorem* 5.3 *holds. Then $\{\succsim_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ satisfies Axiom* 9 *if and only if there exist a unique family of closed and convex subsets $\{\mathbf{P}(h_t) : h_t \subset \Omega^t, t \geqslant 1\}$ of probability measures on Ω and a family of strictly increasing and continuous functions $\{\psi_{h_t} : h_t \subset \Omega^t, t \geqslant 1\}$ such that, for all $\tilde{x} \in \mathscr{B}(R)$,*

$$\mu(h_t, \tilde{x}) = \psi_{h_t}^{-1}\left(\min_{p \in \mathbf{P}(h_t)} E_p[\psi_{h_t}(\tilde{x})]\right),$$

where ψ_{h_t} has the property described in Theorem 5.1 *and $\mathbf{P}(h_t) = \mathbf{P}(h_t \times \Omega)$. Therefore, by Theorem* 4.2,

$$V(h_t, (c_0, d_1)) = u(c_0) + \beta \psi_{h_t}^{-1}\left(\min_{p \in \mathbf{P}(h_t)} E_p[\psi_{h_t}(\tilde{V}(h_t, d_1))]\right).$$

We now examine updating. First we introduce the generalized Bayes rule. Let $\{\mathscr{F}_t\}_{t=1}^{\infty}$ be a filtration and $\mathscr{F}_\infty$ be the σ-algebra generated by $\{\Omega_t\}_{t=1}^{\infty}$ on Ω^∞. Let

$$\{\mathscr{P}_t(h_t) : h_t \in \mathscr{F}_t, h_t \text{ is elementary}, \ t = 0, 1, \ldots\}$$

be a family of sets of probability measures on $(\Omega^\infty, \mathscr{F}_\infty)$. Each probability measure $P \in \mathscr{P}_t(h_t)$ can be viewed as a probability measure conditional on the history h_t. In the case where each of the $\mathscr{P}_t(h_t) = \{P_t(h_t)\}$ is a singleton set and

$$P_t(h_t) = P_0(\cdot | h_t),$$

the whole family of probability measures are related through the Bayes rule, and we can say that the family updates according to the Bayes rule. In the general case, the family is said to update according to the *generalized Bayes rule* if

(a) $P \in \mathscr{P}_t(h_t)$ implies that

$$P(\cdot | F_{t+1}) \in \mathscr{P}_{t+1}(h_t \times F_{t+1}), \tag{4}$$

for any F_{t+1} such that $h_t \times F_{t+1} \in \mathscr{F}_{t+1}$ is elementary; and

(b) $P_{t+1}(\omega^{t+1})$ is a $\mathscr{F}_{t+1}$-measurable selection from $\mathscr{P}_{t+1}(\omega^{t+1})$, where by definition $\mathscr{P}_{t+1}(\omega^{t+1}) = \mathscr{P}_{t+1}(h_t \times F_{t+1})$ for any $\omega^{t+1} \in h_t \times F_{t+1}$, and $P_t \in \mathscr{P}_t(h_t)$ imply

$$P \in \mathscr{P}_t(h_t), \tag{5}$$

where P is defined by

$$P(A) \equiv \int \int 1_A P_{t+1}(\omega^t, \omega_{t+1}, d\omega^\infty) P_t(d\omega_{t+1}), \tag{6}$$

for any $\omega^t \in h_t$.

Expression (4) is called the forward inclusion. It states that any probability measure in $\mathscr{P}(h_t)$, when updated according to the Bayes rule in light of new information F_{t+1}, falls into the set $\mathscr{P}(h_t \times F_{t+1})$. Similarly, expression (5) is called the backward inclusion, which states that when conditional probability measures $P_{t+1}(\omega^{t+1})$, each drawn from $\mathscr{P}_{t+1}(\omega^{t+1})$, and an unconditional $P_t(\omega^t)$ (relative to $P_{t+1}(\omega^{t+1})$) are given and when they are used to define a probability measure through the Kolmogorov backward equation, the probability measure falls into the set $\mathscr{P}(\omega^t)$. These forward and backward inclusions are straightforward extensions of the standard Kolmogorov forward and backward equations.

We emphasize that a family

$$\{\mathscr{P}_t(h_t) : h_t \in \mathscr{F}_t, h_t \text{ is elementary}, \ t = 0, 1, \ldots\}$$

satisfies the generalized Bayes rule if it satisfies *both* the forward and backward inclusions, (4) and (5). In the case where each $\mathscr{P}_t(h_t)$ is a singleton for all t and h_t, by standard probability theory, the family satisfies the forward inclusion if and only if it satisfies the backward inclusion. The following two-period example shows, however, that for a general family these two inclusions may not always be satisfied at the same time.[13] Suppose a fair coin is flipped twice and the flips may not be independent. Let H_i, $i = 1$ and 2 denote the event that the ith flip comes up head. Let

$$\mathscr{P}_0 = \{P : P(H_1) = P(H_2) = 1/2, P(H_1 \cap H_2) = p, p \in [0, 1/2]\}.$$

Applying Bayes rule to each P in the set yields

$$\mathscr{P}_1(H_1) = \{P(\cdot|H_1) : P(H_2|H_1) \in [0, 1]\},$$

$$\mathscr{P}_1(T_1) = \{P(\cdot|T_1) : P(H_2|T_1) \in [0, 1]\}.$$

The family of sets of probabilities, $\mathscr{P}_0$, $\mathscr{P}_1(H_1)$ and $\mathscr{P}_1(T_1)$, satisfy the forward inclusion condition. However, they do not satisfies the backward inclusion condition, because integrating yields

$$\mathscr{P}_0' = \{P : P(H_1) = 1/2, P(H_2) = q, q \in [0, 1], P(H_1 \cap H_2) = p, p \in [0, 1/2]\},$$

which is not equal to $\mathscr{P}_0$.

Return to the implication of Theorem 5.5 for updating. Let $\mathscr{P}$ be a set of probability measures on $(\Omega^\infty, \mathscr{F}_\infty)$. Define, for any bounded $\mathscr{F}_\infty$-measurable

[13] The example is adapted from Seidenfeld and Wasserman [40].

function f

$$E_{\mathscr{P}}[f|\mathscr{F}_t](\omega^t) = \inf\{E_P[f|\mathscr{F}_t](\omega^t) : P\in\mathscr{P}\} = \inf\{E_P[f(\omega^t)] : P\in\mathscr{P}_t(\omega^t)\}. \quad (7)$$

Theorem 5.6 (Generalized Bayes rule). *Suppose that the conditions of Theorem* 5.5 *hold. Let* $\mathscr{F} = \{\mathscr{F}_t\}_{t=1}^{\infty}$ *be an increasing sequence of* σ*-algebras and* $R_+ \times D_{\mathscr{F}}$ *be the subset of consumption–information profiles* (c_0, d_1) *such that the filtration embedded in* d_1 *coincides with* $\mathscr{F}$ *and the consumption process embedded in* d_1 *is uniformly bounded. Then* (a) *there exists a unique set of probability measures* $\mathscr{P}_0$ *on* $(\Omega^{\infty}, \mathscr{F}_{\infty})$ *such that the family*

$$\{\mathscr{P}_t(h_t) : h_t\in\mathscr{F}_t \text{ is elementary}, \ t = 0, 1, \ldots\}, \quad (8)$$

where

$$\mathscr{P}_t(h_t) = \{Q : Q = P(\cdot|\mathscr{F}_t), h_t\in\mathscr{F}_t, P\in\mathscr{P}_0\} \quad (9)$$

satisfies the generalized Bayes rule; (b) *for any* $(c_0, d_1)\in R\times D_{\mathscr{F}}$ *and any elementary history* h_t,

$$V(h_t, (c_t, d_{t+1})) = u(c_t) + \beta\psi_{h_t}^{-1}(E_{\mathscr{P}_0}[\psi_{h_t}(\tilde{V}(h_t, d_{t+1}))|\mathscr{F}_t]),$$

where (c_t, d_{t+1}) *is the continuation of* (c_0, d_1) *at time t*; *and* (c) *if further* $\psi_{h_t}(x) = x$, *then*

$$V(h_t, (c_t, d_{t+1})) = E_{\mathscr{P}_0}\left[\sum_{s=0}^{\infty}\beta^s u(c_{t+s})\middle|\mathscr{H}_t\right]. \quad (10)$$

In words, the theorem says that an individual with conditional preferences satisfying Axioms 1–6, 8 and 9 behaves as if, on $D_{\mathscr{F}}$, he has a belief component described by a set of priors, $\mathscr{P}_0$, on $(\Omega^{\infty}, \mathscr{F}_{\infty})$ and uses the generalized Bayes rule to update his belief over time. Furthermore, the generalized Bayes rule is operationally equivalent to updating each prior in $\mathscr{P}_0$ by the traditional Bayes rule (see (9)).

6. Applications

Epstein and Wang [17,18] develop an intertemporal asset pricing model under Knightian uncertainty. One of the key ingredients of the model is the representative agent's utility function:

$$V_t(c) = u(c_t) + \beta\inf\left\{\int V_{t+1}(c)\,dP : P\in\mathbf{P}(\omega)\right\}, \quad (11)$$

where $\mathbf{P}(\omega)$ is a closed convex set of probability measures on Ω for each state ω. Epstein and Wang introduce this utility function as an intertemporal extension of the multi-prior expected utility developed by Golboa and Schmeidler [20]. They then go on to study the effect of Knightian uncertainty on asset pricing. While there are axiomatizations of multi-prior expected utility in atemporal settings [4,20], axiomatization of such utility in an intertemporal setting has not been available. Theorem 5.5 of this paper can be viewed as providing such an axiomatization. In the same spirit, Theorem 5.3 can be viewed as an axiomatization of intertemporal Choquet expected utility, extending Schmeidler [37], Gilboa [19], Nakamura [32], Sarin and Wakker [33], and Chew and Karni [7]. We view these as another contribution of this paper.[14]

One interesting issue associated with extending multi-prior expected utility to an intertemporal setting is its appropriate formulation. Epstein and Wang [17] propose Eq. (11) as one possible formulation. They also discuss an alternative formulation

$$V_t(c) = \inf\left\{ E_P\left[\sum_{s=t}^{\infty} \beta^{(s-t)} u(c_s) \right] : P \in \mathscr{P}(\omega) \right\}, \tag{12}$$

where $\mathscr{P}$ is a correspondence from Ω to subsets of probability measures on Ω^{∞}. As they point out, unlike the time-additive expected utility, the two formulations (11) and (12) are not necessarily equivalent in general. The natural issue is then: which formulation is more appropriate?

It would be difficult to address this issue according to the respective intuitive appeal of the two formulations. The recursive formulation (11) ensures that the preference is dynamically consistent. It makes the utility function amenable to the application of dynamic programming technique. Formulation (12), on the other hand, is more convenient in certain applications. For example, to study the effect of learning on asset price/return dynamics in an environment with Knightian uncertainty, one may wish to specify a multi-prior expected utility preference as in (12) and examine how the set of priors, $\mathscr{P}$, evolves over time and its effect on equilibrium asset prices.

Fortunately, under some conditions, the two formulations are equivalent. Theorem 5.6 provides one such equivalence result. It starts with the recursive formulation and states that under certain conditions it has an equivalent formulation as in (12). A complementary result would be to start with (12) and establish its equivalence to the recursive formulation. The following theorem is such an complementary equivalence result.

Theorem 6.1. *Let $\mathscr{F} = \{\mathscr{F}_t\}_{t=1}^{\infty}$ be an increasing sequence of σ-algebras on $(\Omega^{\infty}, \mathscr{F}_{\infty})$ where $\mathscr{F}_{\infty} = \sigma(\Omega_t, t = 1, 2\ldots)$. Let $R_+ \times D_{\mathscr{F}}$ be the subset of consumption–information profiles (c_0, d_1) such that the information filtration embedded in d_1 coincides with $\mathscr{F}$ and the consumption process embedded in d_1 is uniformly bounded. Suppose that $\mathscr{P}$ is a closed set of probability measures on $(\Omega^{\infty}, \mathscr{F}_{\infty})$. Let*

$$\mathscr{P}_t(\omega^t) = \{Q : Q(\cdot) = P(\cdot|\mathscr{F}_t)(\omega^t), P \in \mathscr{P}\}$$

[14] See [26] for an alternative axiomatization.

be sets of probability measures on $(\Omega^\infty, \mathscr{F}_\infty)$. *If the family* $\{\mathscr{P}_t\}$ *satisfies the generalized Bayes rule, then*

$$E_{\mathscr{P}}\left[\sum_{s=0}^{\infty}\beta^s u(c_{t+s})\middle|\mathscr{H}_t\right] = E_{\mathscr{P}}\left[u(c_t) + \beta E_{\mathscr{P}}\left(\sum_{s=0}^{\infty}\beta^s u(c_{t+s+1})\middle|\mathscr{H}_{t+1}\right)\middle|\mathscr{H}_t\right].$$

An immediate corollary of the discussion above is that the agent in [17] can be viewed as using the generalized Bayes rule of updating. Thus another application of this paper is that it also provides the axiomatic foundation in the learning aspect for the preference used in [17].

Acknowledgments

The author is grateful to Lorenzo Garlappi, Jacob Sagi, Jiang Wang, Jiankang Zhang, and especially Larry Epstein for their comments and discussions, and to the Social Sciences and Humanities Research Council of Canada for financial support.

Appendix A. Proofs and supporting technical details

Definition of the Domain D. We begin with t-period consumption–information profiles. The space of t-period consumption–information profiles is constructed recursively. Let $D_1 = \mathscr{B}(R_+)$. For each $t > 1$, define

$$D_t = \mathscr{B}(R_+ \times \mathbf{F} \times D_{t-1}),$$

where, for notational convenience, $\mathscr{B}(R_+ \times \mathbf{F} \times D_{t-1})$ denotes the subset of $\mathscr{B}(R_+ \times 2^\Omega \times D_{t-1})$ such that if $(c_1, \mathscr{F}_1, d_2)$ is in the subset, then $\mathscr{F}_1 \in \mathbf{F}$. Elements of $R_+ \times D_t$ are called t-period consumption–information profiles.

Elements of D are defined using these t-period consumption–information profiles. Let $f_1 : D_2 \to D_1$ be defined by, for any $d \in D_2$ with $d : \omega \to (c_1(\omega), \mathscr{F}_1(\omega), d_2(\omega))$,

$$f_1(d)(\omega) = c_1(\omega), \quad \text{for each } \omega \in \Omega.$$

Inductively, for $t > 1$, define $f_t : D_{t+1} \to D_t$ by, for $d \in D_{t+1}$ with $d : \omega \to (c_1(\omega), \mathscr{F}_1(\omega), d_2(\omega))$,

$$f_t(d)(\omega) = (c_1(\omega), \mathscr{F}_1(\omega), f_{t-1}(d_2(\omega))), \quad \text{for all } \omega \in \Omega.$$

Intuitively, what mapping f_t does is to transform a t-period consumption–information profile into a $(t-1)$-period one by cutting off the consumption and the information of the last period in the t-period profile. We define D as the limit of a sequence of finite-horizon profiles with the property that each $(t+1)$-period profile in the sequence is consistent with the preceding t-period profile. That is, we define

$$D = \{(d^1, d^2, \dots) : d^t \in D_t \text{ and } d^t = f_t(d^{t+1}),\ t \geqslant 1\}.$$

All t-period consumption–information profiles can be naturally embedded in $R_+ \times D$. Specifically, let d^t be an element of D_t. It can be extended to a $(t+k)$-period profile by attaching at the end of each branch of it a k-period consumption–information profile whose consumption process is zero and whose information filtration consists of only trivial partitions. With this extension, d^t becomes an element of D_{t+k}. Since k is arbitrary, d_t corresponds naturally to an infinite sequence of finite-horizon profiles such that any one of them grows out its predecessor. Thus d^t becomes an element of D.

Now we define the topology for D. We start with the topology on D_t. For a generic topological space Y, the topology on $\mathscr{B}(Y)$ is the standard pointwise convergence topology. On $\mathbf{F}$, define the topology by the metric

$$\rho(\mathscr{F},\mathscr{G}) = \sum_{\omega\in\Omega}\sum_{\omega'\in\Omega}\left|\sum_{i=1}^{\#(\mathscr{F})}\frac{1}{\#(F_i)}\sum_{\omega''\in F_i}1_{\{\omega\}}(\omega'')1_{F_i}(\omega') - \sum_{j=1}^{\#(\mathscr{G})}\frac{1}{\#(G_j)}\sum_{\omega''\in G_i}1_{\{\omega\}}(\omega'')1_{G_j}(\omega')\right|,$$

where $\#(\mathscr{F})$ and $\#(F_i)$ denote the number of elements in $\mathscr{F}$ and F_i respectively, and 1_F is the indicator function. This metric induces the pointwise convergence topology introduced by Cotter [9] on set of σ-algebras. Intuitively, two partitions $\mathscr{F}$ and $\mathscr{G}$ are different if there are at least two subsets $F_i\in\mathscr{F}$ and $G_j\in\mathscr{G}$ such that $F_i\cap G_j\neq\emptyset$ and $F_i\neq G_j$. In that case, there exist ω and ω' such that $\omega\in F_i\cap G_j$ and $\omega'\in F_i$, but $\omega'\notin G_j$, or $\omega'\notin F_i$, but $\omega'\in G_j$. For this pair of ω and ω', $\rho(\mathscr{F},\mathscr{G})>0$. Conversely, if $\rho(\mathscr{F},\mathscr{G})>0$, then reversing the argument above implies that $\mathscr{F}$ and $\mathscr{G}$ are not identical. Thus, conforming to the intuition, if $\mathscr{F}_n$ is a sequence of partitions and $\rho(\mathscr{F}_n,\mathscr{F})\to 0$, then $\mathscr{F}_n$ "converges" to $\mathscr{F}$ because for large n, $\rho(\mathscr{F}_n,\mathscr{F})=0$. Now for each $t>1$, give $R_+\times\mathbf{F}\times D_{t-1}$ the product topology. Then the topology on D_t can be defined recursively. Finally, we give D the product topology.

Proof of Theorem 3.1. First, we define a mapping Θ from D to $\mathscr{B}(R_+\times\mathbf{F}\times D)$. Let $d\in D$. By definition

$$d=(d^1,d^2,\dots)\in\Pi_{t=1}^{\infty}D_t,\quad \text{with } f_t(d^{t+1})=d^t.$$

Recall that the latter consistency condition implies that $d^t(\omega)=(c_1(\omega),\mathscr{F}_1(\omega),d_2^t(\omega))$ for some common $c_1\in\mathscr{B}(R_+)$ and $\mathscr{F}_1\in\mathbf{F}$ for all t. Define $\Theta(d)$ by for each $\omega\in\Omega$,

$$\Theta(d)(\omega)=(c_1(\omega),\mathscr{F}_1(\omega),d_2(\omega)),\quad d_2(\omega)=(d_2^2(\omega),d_2^3(\omega),\dots,d_2^t(\omega),\dots).$$

To ensure that Θ is well defined, we need to show that $d_2(\omega)\in D$ for each $\omega\in\Omega$. That is, for each $\omega\in\Omega$, $f_{t-1}(d_2^{t+1}(\omega))=d_2^t(\omega)$. Fix $\omega\in\Omega$. By assumption, for all $t>1$, $f_t(d^{t+1})=d^t$. This implies that

$$f_t(d^{t+1})(\omega)=(c_1(\omega),\mathscr{F}_1(\omega),f_{t-1}(d_2^{t+1}(\omega)))=d^t(\omega)=(c_1(\omega),\mathscr{F}_1(\omega),d_2^t(\omega)),$$

and hence $f_{t-1}(d_2^{t+1}(\omega))=d_2^t(\omega)$ as desired. Arguing in reverse order shows that Θ is one-to-one and onto. For continuity, suppose that $d^n\to d$. Then $d^{t+1,n}\to d^{t+1}$ for each

t. This is equivalent to $(c_1^n(\omega), \mathscr{F}_1^n(\omega), d_2^{t,n}(\omega)) \to (c_1(\omega), \mathscr{F}_1(\omega), d_2^t(\omega))$ for all $\omega \in \Omega$. Thus $(d_2^{2,n}(\omega), d_2^{3,n}(\omega), \ldots) \to (d_2^2(\omega), d_2^3(\omega), \ldots)$ and hence

$$(c_1^n(\omega), \mathscr{F}_1^n(\omega), (d_2^{2,n}(\omega), d_2^{3,n}(\omega), \ldots)) \to (c_1(\omega), \mathscr{F}_1(\omega), (d_2^2(\omega), d_2^3(\omega), \ldots)).$$

Therefore Θ is continuous. Arguing in reverse order establishes the continuity of Θ^{-1}. □

The information filtration embedded in d_1. Let $d_1 \in D$. Fix a t and $(\omega_1, \ldots, \omega_t)$. Suppose

$$d_1(\omega_1) = (c_1(\omega_1), \mathscr{F}_1(\omega_1), d_2(\omega_1)).$$

Since $\mathscr{F}_1 = \{F_1, \ldots, F_{n_1}\}$ is a partition, there is a unique $F_{\omega_1} \in \mathscr{F}_1$ such that $\omega_1 \in F_{\omega_1}$. Next suppose

$$d_2(\omega_1)(\omega_2) = (c_2(\omega_1, \omega_2), \mathscr{F}_2(\omega_1, \omega_2), d_3(\omega_1, \omega_2)).$$

Since $\mathscr{F}_2(\omega_1) = \{F_{\omega_1,1}, \ldots, F_{\omega_1,n(\omega_1)}\}$ is a partition, there is a unique $F_{(\omega_1,\omega_2)} \in \mathscr{F}_2(\omega_1)$ such that $\omega_2 \in F_{(\omega_1,\omega_2)}$. Inductively, there exists a unique sequence $F_{\omega_1}, \ldots, F_{(\omega_1,\ldots,\omega_t)}$ such that $(\omega_1, \ldots, \omega_t) \in F_{\omega_1} \times \cdots \times F_{(\omega_1,\ldots,\omega_t)}$. The collection of such sets $F_{\omega_1} \times \cdots \times F_{(\omega_1,\ldots,\omega_t)}$ as $(\omega_1, \ldots, \omega_t)$ runs through Ω^t is a partition of Ω^t, which naturally extends to a corresponding partition of Ω^∞. Denote this partition by $\mathscr{F}_t$. Let t run through $1, 2, \ldots$. Then $\{\sigma(\mathscr{F}_t) : t \geqslant 1\}$ is the information filtration embedded in d_1.

Proof of Theorem 4.1. By Debreu [10], each conditional preference $\succcurlyeq_{h_t}$ on $R_+ \times D$ can be represented by an utility function $V(h_t, (c_0, d_1))$ on $R_+ \times D$. Due to Axiom 4, the ranking of deterministic consumption profiles are independent of past information histories. Thus we can normalize the utility functions by monotonic transforms such that for any deterministic consumption–information profile (c_0, d_1),

$$V(h_t, (c_0, d_1)) = V(h'_t, (c_0, d_1)),$$

for any $h_t = A_1 \times \cdots \times A_t$ and $h'_t = B_1 \times \cdots \times B_t$. By stationarity, we further normalize the utility functions such that $V(h_t, (c_0, d_1)) = V(h_t \times \Omega^s, (c_0, d_1))$.

Define a function $\hat{V}(h_t, d_1) : D \to R$ by

$$\hat{V}(h_t, d_1) = f(V(h_t, (0, d_1))),$$

where f is the unique strictly increasing function such that for all $c \in R_+$,

$$V(c, 0) = f(V(0, c)),$$

where $(c, 0)$ and $(0, c)$ are the deterministic consumption profile which has a one-time consumption of c today and tomorrow, respectively, and $V(c, 0)$ is the conditional utility at time 0 of a one-time current consumption c at time 0. By convention no historical information is recorded at time 0. Intuitively, $\hat{V}(h_t, d_1)$ is the utility of d_1 at time $t+1$ evaluated just before the uncertainty in the period between time

t and $t+1$ is realized. To illustrate, let $(c_0, d_1) = (c_0, (\tilde{c}_1, \mathscr{F}))$ be a one-period consumption–information profile and

$$V(h_t, (c_0, d_1)) = u(c_0) + \beta E[u(\tilde{c}_1)|h_t].$$

Then $V(c_0) = u(c_0), f(x) = x/\beta$ and

$$\hat{V}(h_t, d_1) = E[u(\tilde{c}_1)|h_t].$$

Now define for any $c_0 \in R_+$ and any real number v,

$$W(h_t, (c_0, v)) = V(h_t, (c_0, d_1)),$$

for any d such that $v = \hat{V}(h_t, d_1)$. If d_1 and d_1' are such that $v = \hat{V}(h_t, d_1) = \hat{V}(h_t, d_1')$, then it follows from uncertainty separability that

$$V(h_t, (c_0, d_1)) = V(h_t, (c_0, d_1')).$$

Thus the function W is well defined. Continuity and monotonicity of W are straightforward.

We now show that $W(h_t, (c_0, v))$ is independent of h_t. Let (c_0, d_1) be a deterministic consumption–information profile. By the normalization,

$$V(A_1 \times \cdots \times A_t, (c_0, d_1)) = V(B_1 \times \cdots \times B_t, (c_0, d_1)).$$

Thus

$$W(A_1 \times \cdots \times A_t, (c_0, v)) = W(B_1 \times \cdots \times B_t, (c_0, v)).$$

That is, $W(h_t, (c_0, d_1))$ is independent of h_t.

Next, let $t \geqslant 0$ and $h_t \subset \Omega^t$. Define $\mu(h_t) : \mathscr{B}(\Omega) \to R$ by

$$\mu(h_t, \tilde{x}) = \hat{V}(h_t, d_1),$$

for any $d_1 \in D$ such $\tilde{x} = \tilde{V}(h_t, d_1)$. We claim that $\mu(h_t)$ is well defined. Suppose that d_1 and d_1' are such that

$$\tilde{x} = \tilde{V}(h_t, d_1) = \tilde{V}(h_t, d_1'),$$

and that $d_1(\omega) = (c_1(\omega), \mathscr{F}(\omega), d_2(\omega))$ and $d_1'(\omega) = (c_1'(\omega), \mathscr{G}(\omega), d_2'(\omega))$ where $\mathscr{F} = \{A_1, \dots, A_n\}$ and $\mathscr{G} = \{B_1, \dots, B_m\}$. Then we have, for all i and j and $\omega \in A_i \cap B_j$,

$$V[h_t \times A_i, (c_1(\omega), d_2(\omega))] = V[h_t \times B_j, (c_1'(\omega), d_2'(\omega))]. \tag{13}$$

By Axiom 1, for each ω, there exist constant numbers $y(\omega)$ and $z(\omega)$ (which are viewed as deterministic consumption–information profiles) such that

$$(c_1(\omega), d_2(\omega)) \sim_{h_t \times A_i} y(\omega) \quad \text{and} \quad (c_1'(\omega), d_2'(\omega)) \sim_{h_t \times B_j} z(\omega).$$

Thus, by Axiom 4 and the normalization,

$$\begin{aligned} V[h_t \times A_i, (c_1(\omega), d_2(\omega))] &= V[h_t \times A_i, y(\omega)] = V[h_t \times B_j, y(\omega)] \\ &= V[h_t \times B_j, (c_1'(\omega), d_2'(\omega))] \\ &= V[h_t \times B_j, z(\omega)] = V[h_t \times A_i, z(\omega)]. \end{aligned}$$

Then for any $c_0 \in R_+$

$$\begin{aligned} V[h_t,(c_0,d_1)] &= V[h_t,(c_0,(c_1,\mathscr{F},d_2))] = V[h_t,(c_0,(y,\mathscr{F},0))] \\ &= V[h_t,(c_0,(y,\mathscr{G},0))] = V[h_t,(c_0,(z,\mathscr{G},0))] \\ &= V[h_t,(c_0,(c_1',\mathscr{G},d_2'))] = V[h_t,(c_0,d_1')], \end{aligned}$$

where the second, fourth and fifth equalities are by the consistency axiom, and the third equality is by Axiom 4. Since c_0 is arbitrary, it follows that

$$\hat{V}[h_t,d_1] = \hat{V}[h_t,d_1'].$$

Thus $\mu(h_t)$ is well defined. Stationarity clearly implies that $\mu(h_t) = \mu(h_t \times \Omega)$.

To show $\mu(h_t,\tilde{x})$ is a certainty equivalent, let $(c_0,d_1) = (c_0,(c_1,\mathscr{F}^0))$ be a one-period consumption–information profile. Observe that

$$\begin{aligned} \hat{V}(h_t,d_1) &= \mu(h_t,\tilde{V}(h_t,d_1)) \\ &= \mu(h_t,V(h_t \times \Omega,(c_1,0)). \end{aligned}$$

By stationarity and the normalization,

$$V(h_t \times \Omega,(c_1,0)) = V(h_t,(c_1,0)) = V(c_1).$$

By definition of $\hat{V}(h_t,d_1)$,

$$\hat{V}(h_t,d_1) = f[V(h_t,(0,(c_1,\mathscr{F}^0))] = V(c_1).$$

Thus $\mu(h_t,V(c_1)) = V(c_1)$, which is the property (a) of a certainty equivalent.

For property (b), let d_1 be given by $d_1(\omega) = (c_1(\omega),\mathscr{F}^0,0)$ and $\tilde{x} = \tilde{V}(h_t,d_1)$ so that $\tilde{x}(\omega) = V(h_t \times \Omega,(c_1(\omega),0))$. Similarly, let d_1' be given by $d'(\omega) = (c_1'(\omega),\mathscr{F}^0,0)$ and $\tilde{y}(\omega) = V(h_t \times \Omega,(c_1'(\omega),0))$. If $\tilde{x} \geqslant \tilde{y}$, then, by consistency,

$$\begin{aligned} \mu(h_t,\tilde{x}) &= \mu(h_t,\tilde{V}(h_t,d_1)) = \hat{V}[h_t,d_1] \\ &\geqslant \hat{V}[h_t,d_1'] = \mu(h_t,\tilde{V}(h_t,d_1')) = \mu(h_t,\tilde{y}). \qquad \square \end{aligned}$$

Proof of Theorem 4.2. This theorem follows from Koopmans [28] or Gorman [22]. $\square$

Proof of Theorem 5.1. The proof is exactly identical to that of Theorem 5.3 with only one change: under Timing Indifference, property (A6) is replaced by

$$d_A(m_{\mathscr{F}}(x_1,\dots,x_n),m_{\mathscr{F}}(y_1,\dots,y_n)) \sim_{F_1\times\cdots\times F_t} d_{\mathscr{F}}(m_A(x_1,y_1),\dots,m_A(x_n,y_n)). \quad (\text{A6}')$$

Under (A6$'$), by Nakamura [32, Theorem 1], $v(h_t,\cdot)$ in the proof of Theorem 5.3 is in fact a probability measure. $\square$

Proof of Theorem 5.2. Denote $(\omega_1,\dots,\omega_t)$ by ω^t. Let $P(h_t,\cdot)$, where $h_t \in \mathscr{F}_t$ is elementary, be the probability measures whose existence is guaranteed by Theorem 5.1. For any $A \subset \Omega$ and $\omega^t \in h_t \in \mathscr{F}_t$, let $P(\omega^t,A) = P(h_t,A)$. Define, for each t and ω^t, a probability measure $P_t(\omega^t)$ on $(\Omega^\infty, \mathscr{F}_\infty)$ by, for any $A = F_{t+1} \times \cdots \times F_T$,

$$P_t(\omega^t, A) = \int \left(\int \left(\cdots \left(\int 1_A P(\omega^{T-1}, d\omega_T) \right) \cdots \right) P(\omega^{t+1}, d\omega_{t+2}) \right) \times P(\omega^t, d\omega_{t+1}). \tag{14}$$

Kolmogorov Theorem ensures that $P_t(\omega^t)$, $t = 0, 1, \ldots,$ are well defined. The probability measure P_0 is the desired initial probability. By construction,

$$P_0[\cdot|\mathscr{F}_t](\omega^t) = P_t(\omega^t, \cdot), \quad P_t(\omega^t, \cdot|\mathscr{F}_{t+1})(\omega^{t+1}) = P_{t+1}(\omega^{t+1}, \cdot)$$

and

$$P_t(\omega^t, A \times \Omega^\infty) = P(\omega^t, A) = P(h_t, A),$$

for any $A \subset \Omega$ and $\omega^t \in h_t$. Thus $\{P(h_t)\}$ are conditionals of P_0. $\square$

Proof of Theorem 5.3. Fix h_t. First we introduce some simplifying notations. Let $\mathscr{F} = \{A_1, \ldots, A_n\}$ be a partition of Ω. Denote by $d_{\mathscr{F}}(x_1, \ldots, x_n)$ the random variable whose value in state $\omega \in A_i$ is x_i. We will also use it to denote the one-period consumption–information profile whose current consumption is zero and whose consumption at time 1 in state $\omega \in A_i$ is x_i. The meaning will always be clear from the context. Note that for one-period consumptions, updating in the forthcoming period is irrelevant by Axiom 4. Let $A \subset \Omega$. For the partition $\mathscr{F} = \{A, A^c\}$ we will also write $d_{\mathscr{F}}(x, y)$ simply as $d_A(x, y)$. For partitions $\mathscr{F} = \{A_1, \ldots, A_n\}$ and $\mathscr{F}_i = \{B_{i1}, \ldots, B_{im}\}$, $i = 1, \ldots, n$, denote by $d_{\mathscr{F}} = (d_{\mathscr{F}_1}(x_{11}, \ldots, x_{1m}), \ldots, d_{\mathscr{F}_n}(x_{n1}, \ldots, x_{nm}))$ the two-period consumption profiles

$$(0, (0, \mathscr{F}^0, d_2)), \quad d_2(\omega) = d_{\mathscr{F}_i}(x_{i1}, \ldots, x_{im}) \quad \text{for } \omega \in A_i, \; i = 1, \ldots, n.$$

Consider the restriction of $\succcurlyeq_{h_t}$ to one-period consumption–information profiles. To simplify notations, write $V[F_1 \times \cdots \times F_t, (0, d_{\mathscr{F}}(x_1, \ldots, x_n))]$ as $V[F_1 \times \cdots \times F_t, d_{\mathscr{F}}(x_1, \ldots, x_n)]$ when no confusion arises.

We shall first verify that if Axioms 1–6 and 8 hold, then the ordering $\succcurlyeq_{h_t}$ has the following properties:

(A1) For each $d_{\mathscr{F}}(x_1, \ldots, x_n)$, there exist x and $y \in R_+$ such that

$$d_\Omega(x) \succ_{h_t} d_{\mathscr{F}}(x_1, \ldots, x_n) \succ_{h_t} d_\Omega(y).$$

(A2) If $d_A(y, z) \succcurlyeq_{h_t} d_{\mathscr{F}}(x_1, \ldots, x_n) \succcurlyeq_{h_t} d_A(x, z)$, then $d_{\mathscr{F}}(x_1, \ldots, x_n) \sim_{h_t} d_A(a, z)$ for some $a \in R_+$.

(A3) If A is not null[15] and $\{x, y\} \leqslant z$, then $x \leqslant y$ if and only if $d_A(y, z) \succcurlyeq_{h_t} d_A(x, z)$; if A is not universal[16] and $\{x, y\} \geqslant z$, then $x \leqslant y$ if and only if $d_A(z, y) \succcurlyeq_{h_t} d_A(z, x)$.

(A4) If $x \leqslant y$ and $A \subset B$, then $d_A(x, y) \succcurlyeq d_B(x, y)$.

(A5) Every strictly bounded standard sequence is finite.[17]

---

[15] Recall that an event $A \subset \Omega$ is null if for all $x, y, z \in R_+$, $d_A(x, z) \sim_{F_1 \times \cdots \times F_t} d_A(y, z)$.

[16] An event $A \subset \Omega$ is universal if for all $x, y, z \in R$, $d_A(x, y) \sim d_A(x, z)$.

[17] Let N be any set of consecutive integers. Given an event A which is neither null nor universal, a standard sequence is defined as a set $\{a_i \in R_+ : i \in N\}$ for which there exist a and $b \in R_+$ such that $a \neq b$ and either $\{a, b\} \leqslant a_i$ and $d_A(a, a_i) \sim_{h_t} d_A(b, a_{i+1})$ for all $i \in N$, or $a_i \leqslant \{a, b\}$ and $d_A(a_i, a) \sim_{h_t} d_A(a_{i+1}, b)$ for all $i \in N$.

(A6) If $x_1 \leqslant \cdots \leqslant x_n$ and $y_1 \leqslant \cdots \leqslant y_n$ with $x_i \leqslant y_i$ for all i, then

$$d_A(m_{\mathscr{F}}(x_1, \dots, x_n), m_{\mathscr{F}}(y_1, \dots, y_n)) \sim_{h_t} d_{\mathscr{F}}(m_A(x_1, y_1), \dots, m_A(x_n, y_n)),$$

where $m_{\mathscr{F}}(x_1, \dots, x_n)$ is the constant such that

$$V[h_t, (0, m_{\mathscr{F}}(x_1, \dots, x_n))] = V[h_t, (0, d_{\mathscr{F}}(x_1, \dots, x_n))]$$

or

$$m_{\mathscr{F}}(x_1, \dots, x_n) = u^{-1}[V[h_t, (0, d_{\mathscr{F}}(x_1, \dots, x_n))]/\beta]. \tag{15}$$

Note that $m_{\mathscr{F}}(x_1, \dots, x_n)$ is the constant consumption that is realized *one period from now*.

Properties (A1)–(A4) follow from Axioms 1,4 and 5. Recall that consistency implies the usual monotonicity.

For (A5), let $\{x_n, n \in N\}$ be a standard sequence. Then, without loss of generality, there exist two real numbers p and $q \in R_+$ such that

$$V[h_t, d_A(x_n, p)] = V[h_t, d_A(x_{n+1}, q)]. \tag{16}$$

Assume first that $p > q$. Then by monotonicity, $x_n < x_{n+1}$ for all n. We wish to verify that if this standard sequence is strictly bounded in the sense that $a < x_n < b$ for some $a < b$, then the sequence must be finite. Suppose the contrary. Then x_n converges to a real number $x_0 \leqslant b$. Taking limit in (16) and applying the continuity of V, we have $V[h_t, d_A(x_0, p)] = V[h_t, d_A(x_0, q)]$, which contradicts the fact that A is not universal and hence A^c is not null. The case that $p < q$ can be verified similarly.

For (A6), we show first that the certainty equivalent operator, $\mu[h_t]$, satisfies

$$\mu[h_t, \beta\tilde{x}] = \beta\mu[h_t, \tilde{x}] \tag{17}$$

for all $\tilde{x} \in \mathscr{B}(R_+)$. In the following derivation, we make heavy use of the expressions

$$V[h_t, (c, d)] = u(c) + \beta\mu(h_t, \tilde{V}[h_t, d]) \tag{18}$$

and

$$V[h_t, d_{\mathscr{F}}(x_1, \dots, x_n)] = \beta\mu[h_t, u(\tilde{x})]. \tag{19}$$

Now let $\tilde{x} \in \mathscr{B}(R_+)$ be a random variable that assumes values $x_1 < \cdots < x_n$ on $A_1, \dots, A_n$ respectively. Let $\mathscr{F} = \{A_1, \dots, A_n\}$. Let

$$d_1(\omega) = (0, \mathscr{F}^0, d_2(\omega)), \quad d_2(\omega) = d_{\mathscr{F}}(u^{-1}(x_i), \dots, u^{-1}(x_i)), \quad \text{if } \omega \in A_i$$

and

$$d_1'(\omega) = (0, \mathscr{F}^0, d_2'(\omega)), \quad d_2'(\omega) = d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_n)), \quad \text{if} \omega \in A_i.$$

Observe that

$$\begin{aligned} &V(h_t, d_{\mathscr{F}}(d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_1)), \dots, d_{\mathscr{F}}(u^{-1}(x_n), \dots, u^{-1}(x_n)))) \\ &\quad = V(h_t, (0, d_1)) = \beta\mu(h_t, \tilde{V}(h_t, d_1)). \end{aligned} \tag{20}$$

where in the second equality we have used (18), noting that the argument of $V(h_t, \cdot)$ is a two-period consumption–information profile. For $\omega \in A_i$,

$$\tilde{V}(h_t, d_1)(\omega) = V(h_t \times \Omega, d_{\mathscr{F}}(u^{-1}(x_i), \dots, u^{-1}(x_i)))$$
$$= \beta\mu(h_t \times \Omega, d_{\mathscr{F}}(u^{-1}(x_i), \dots, u^{-1}(x_i))) = \beta\mu(h_t \times \Omega, x_i) = \beta x_i, \quad (21)$$

where the third equality is by (19). By a similar argument,

$$V(h_t, d_{\mathscr{F}}(d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_n)), \dots, d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_n))))$$
$$= V(h_t, (0, d_1')) = \beta\mu(h_t, \tilde{V}(h_t, d_1')) \quad (22)$$

and for $\omega \in A_i$,

$$\tilde{V}(h_t, d_2)(\omega) = V(h_t \times \Omega, d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_n)))$$
$$= V(h_t, d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_n))) = \beta\mu(h_t, \tilde{x}), \quad (23)$$

where the second equality is by Stationarity. Now by (20)–(23)

$$\beta\mu(h_t, \beta\tilde{x}) = V(h_t, d_{\mathscr{F}}(d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_1)), \dots, d_{\mathscr{F}}(u^{-1}(x_n), \dots, u^{-1}(x_n))))$$
$$= V(h_t, d_{\mathscr{F}}(d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_n)), \dots, d_{\mathscr{F}}(u^{-1}(x_1), \dots, u^{-1}(x_n))))$$
$$= \beta\mu(h_t, \beta\mu(h_t, \tilde{x})) = \beta^2\mu(h_t, \tilde{x}),$$

where the second equality is by comonotonic timing indifference, and last equality from $\mu(h_t)$ being a certainty equivalent. Thus (17) is shown.

Now, let

$$\tilde{f}(\omega) = \begin{cases} V(h_t \times \Omega, d_{\mathscr{F}}(x_1, \dots, x_n)) & \text{if } \omega \in A, \\ V(h_t \times \Omega, d_{\mathscr{F}}(y_1, \dots, y_n)) & \text{otherwise.} \end{cases}$$

Then

$$V(h_t, d_A(m_{\mathscr{F}}(x_1, \dots, x_n), m_{\mathscr{F}}(y_1, \dots, y_n)))$$
$$= V(h_t, d_A(u^{-1}[V(h_t \times \Omega, d_{\mathscr{F}}(x_1, \dots, x_n))/\beta],$$

$$u^{-1}[V(h_t \times \Omega, d_{\mathscr{F}}(y_1, \dots, y_n))/\beta])$$
$$= \mu[h_t, \tilde{f}] = \frac{1}{\beta}[\beta\mu(h_t, \tilde{f})]$$
$$= \frac{1}{\beta} V(h_t, d_A(d_{\mathscr{F}}(x_1, \dots, x_n), d_{\mathscr{F}}(y_1, \dots, y_n)))$$
$$= \frac{1}{\beta} V(h_t, d_{\mathscr{F}}(d_A(x_1, y_1), \dots, d_A(x_n, y_n)))$$
$$= V(h_t, d_{\mathscr{F}}(u^{-1}[V(h_t \times \Omega, d_A(x_1, y_1))/\beta], \dots, u^{-1}[V(h_t \times \Omega, d_A(x_n, y_n))/\beta]))$$
$$= V(h_t, d_{\mathscr{F}}(m_A(x_1, y_1)), \dots, m_A(x_n, y_n))))$$

where first equality is by (15), the second equality is by (19) and (17), the fourth equality is by (18), the fifth equality is by comonotonic timing indifference, the sixth equality is by (18), and the last equality is by (15). Thus (A6) holds.

Now by Nakamura [32, Theorem 1], there exist a strictly monotonic function g_{h_t}, unique under affine transform, and a unique normalized monotonic set function $v(h_t)$ such that $d_{\mathscr{F}}(x_1, \ldots, x_n) \succcurlyeq_{h_t} d_{\mathscr{G}}(y_1, \ldots, y_m)$ if and only if

$$\int g_{h_t}[u(\tilde{x})]\, dv(h_t) \geqslant \int g_{h_t}[u(\tilde{y})]\, dv(h_t).$$

Thus, for any $\tilde{x}$ and $\tilde{y} \in \mathscr{B}(R_+)$,

$$V(h_t, d_{\mathscr{F}}(x_1, \ldots, x_n)) \geqslant V(h_t, d_{\mathscr{G}}(y_1, \ldots, y_m))$$

if and only if

$$\int g_{h_t}[u(\tilde{x})]\, dv(h_t) \geqslant \int g_{h_t}[(\tilde{y})]\, dv(h_t),$$

which implies that there exists a strictly increasing function ψ_{h_t} such that

$$\psi_{h_t}(V(h_t, d_{\mathscr{F}}(x_1, \ldots, x_n))/\beta) = \int g_{h_t}(u(\tilde{x}))\, dv(h_t).$$

However, $V(h_t, d_{\mathscr{F}}(x_1, \ldots, x_n))/\beta = \mu(h_t, u(\tilde{x})]$ by (19). Thus

$$\psi_{h_t}(\mu(h_t, u(\tilde{x}))) = \int g_{h_t}(u(\tilde{x}))\, dv(h_t).$$

Since $\mu(h_t, \cdot)$ is a certainty equivalent, the above equation implies that

$$\psi_{h_t}[u(x)] = g_{h_t}(u(x)).$$

Since g_{h_t} is unique under affine transform, we can normalize ψ_{h_t} so that $\psi_{h_t}(0) = 0$. Returning to $\mu(h_t)$, we have

$$\mu(h_t, \tilde{y}) = \psi_{h_t}^{-1} \int \psi_{h_t}(\tilde{y})\, dv(h_t).$$

We have shown the if the conditional preferences satisfies Axioms 1–6 and 8, then the certainty equivalents are of the form claimed in the theorem. The converse is straightforward to verify. □

Proof of Theorem 5.4. To derive the Dempster–Shafer rule from Theorem 5.3, we first define an unconditional non-additive prior. We then tie it to the family of set functions, $\{v(h_t)\}$, in Theorem 5.3, which are viewed as non-additive conditionals. For simplicity, we examine the two-period case. The more general case is the same. Define the unconditional non-additive prior v on Ω^2 by, for any $A \times B \subset \Omega^2$,

$$v(A \times B) = V(0, d_1),$$

where

$$d_1 : \omega_1 \to (0, \mathscr{F}(\omega_1), d_2(\omega_1)), \quad \mathscr{F} = \{A, A^c\},$$

$$d_2(\omega_1, \omega_2) = \begin{cases} 1 & \text{if } \omega_1 \in A \text{ and } \omega_2 \in B, \\ 0 & \text{otherwise.} \end{cases}$$

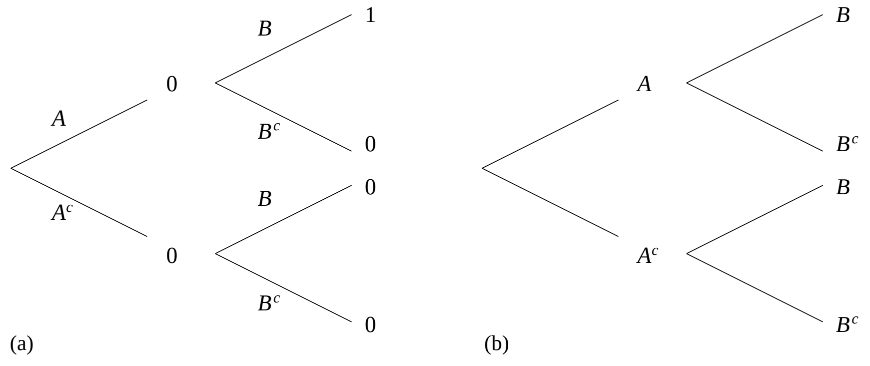

Fig. 6. Simple consumption–information profile.

This definition looks more complicated than it really is. The $(0, d_1)$ corresponds to the two-period trees in Fig. 6. Now let $\mathscr{F} = \{A_1, \ldots, A_N\}$. For subsets of Ω^2 of the form $\bigcup_{i=1}^n A_i \times B_i$ where $A_i \times B_i$ are disjoint and $N \geqslant n$, define

$$v\left(\bigcup_{i=1}^{n} A_i \times B_i\right) = V(0, d_1),$$

where

$$d_1 : \omega_1 \to (0, \mathscr{F}(\omega_1), d_2(\omega_1)), \quad d_2(\omega_1, \omega_2) = \begin{cases} 1 & \text{if } \omega_1 \in A_i \text{ and } \omega_2 \in B_i, \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, n$.[18]

For F_1 and $B \subset \Omega$, it is readily verified that

$$v(F_1 \times B) = v(F_1)v(F_1, B).$$

Thus

$$v(F_1, B) = v(F_1 \times B)/v(F_1).$$

It is formally the same as the Bayes rule. It is also readily verified that

$$v([F_1 \times B] \cup [F_1^c \times \Omega]) = (1 - v(F_1^c \times \Omega))v(F_1, B) + v(F_1^c \times \Omega).$$

Thus

$$v(F_1, B) = \frac{v([F_1 \times B] \cup [F_1^c \times \Omega]) - v(F_1^c \times \Omega)}{(1 - v(F_1^c \times \Omega))}.$$

This last expression is the Dempster–Shafer rule. □

Proof of Theorem 5.5. This theorem follows from Wakker [43, Theorem 7.3] and the standard representation theorem for Choquet integration with respect to a convex capacity. See for example [2]. □

[18] This v on subsets of Ω^2 corresponds to the likelihood relation in the literature. See [16,30].

Proof of Theorem 5.6. We first construct $\mathscr{P}_0$. Fix a $t \geqslant 0$. Let $P(\omega^s, \cdot)$, $s = t, t+1, \ldots,$ be a sequence of $\mathscr{F}_s$-measurable selection from $\mathbf{P}(h_s)$, i.e.,

$$P(\omega^s, \cdot) \in \mathbf{P}(h_s) \quad \text{and for each } A \subset \Omega \quad P(\omega^s, A) \quad \text{is } \mathscr{F}_s\text{-measurable.}$$

By Kolmogorov Theorem, there exists a probability measure $P_t(\omega^t, \cdot)$ on $(\Omega^\infty, \mathscr{F}_\infty)$ such that

$$P_t(\omega^t, A_{t+1} \times \cdots \times A_s) = \int \cdots \int 1_{A_{t+1} \times \cdots \times A_s}(\omega_{t+1}, \ldots, \omega_s) P(\omega^{s-1}, d\omega_s) \cdots \times P(\omega^t, d\omega_{t+1}),$$

for any $s > t$ and $A_{t+1} \times \cdots \times A_s \subset \Omega^{s-t}$. Let $\mathscr{P}_t(\omega^t)$, $t \geqslant 0$, denote the set of all such measures. By construction, if $P \in \mathscr{P}_t(\omega^t)$, then

$$P(\cdot | \mathscr{F}_{t+1})(\omega^{t+1}) \in \mathscr{P}_{t+1}(\omega^{t+1}). \tag{24}$$

That is, if P is any probability measure in $\mathscr{P}_t(\omega^t)$, then its conditionals fall into $\mathscr{P}_{t+1}(\omega^{t+1})$, which is the forward inclusion. Conversely, if $P_{t+1}(\omega^{t+1})$ is a $\mathscr{F}_{t+1}$-measurable selection from $\mathscr{P}_{t+1}(\omega^{t+1})$ and $P_t(\omega^t) \in \mathbf{P}(h_t)$, then

$$P \in \mathscr{P}_t(\omega^t), \quad P(A) \equiv \int \int 1_A P_{t+1}(\omega^t, \omega_{t+1}, d\omega^\infty) P_t(\omega^t, d\omega_{t+1}), \tag{25}$$

which is the backward inclusion. Thus the family just constructed, $\{\mathscr{P}_t(\omega^t)\}$, satisfies the generalized Bayes rule. $\mathscr{P}_0$ is the desired set of priors on $(\Omega^\infty, \mathscr{F}_\infty)$.

The second claim is straightforward. For the third claim, we will prove the case of "finite-horizon" consumption–information profiles. The infinite-horizon case can be shown by a limiting argument, assuming that the limits on both sides of the equation exist. We prove the case of two-period consumption–information profiles. The more general case is the same, but involves more notation. We also assume $t = 0$. The case for general t is the same. Let $(c_0, d_1) = (c_0, (c_1, \mathscr{F}_1, d_2))$ with $\mathscr{F}_1 = \{F_1, \ldots, \mathscr{F}_n\}$ and $d_2 = (c_2, \mathscr{F}_2, 0)$. Note that $\mathscr{F}_2$ is irrelevant for the evaluation of the utility. So there is no need to specify it. By Theorem 5.5,

$$V(c_0, d_1) = u(c_0) + \beta \min \left\{ \int \tilde{V}(d_1)(\omega_1) P(d\omega_1) : P \in \mathbf{P}_0 \right\},$$

$$\begin{aligned} \tilde{V}(d_1)(\omega_1) &= V(F_i, (c_1(\omega_1), d_2(\omega_1))) \\ &= u(c_1(\omega_1)) + \beta \min \left\{ \int u(c_2(\omega_1, \omega_2)) P(d\omega_2) : P \in \mathbf{P}(F_i) \right\}. \end{aligned}$$

Since both $\mathbf{P}_0$ and $\mathbf{P}(F_i)$ are closed, there exist $P_0^* \in \mathbf{P}_0$ and $P_1^*(\omega_1) \in \mathbf{P}(F_i)$ for $\omega_1 \in F_i$ such that

$$V(c_0, d_1) = u(c_0) + \beta \int \tilde{V}(d_1)(\omega_1) P_0^*(d\omega_1),$$

$$\begin{aligned} \tilde{V}(d_1)(\omega_1) &= V(F_i, (c_1(\omega_1), d_2(\omega_1))) \\ &= u(c_1(\omega_1)) + \beta \int u(c_2(\omega_1, \omega_2)) P_1^*(\omega_1, d\omega_2), \quad \text{if } \omega_1 \in F_i. \end{aligned}$$

Let $\{P_t^*(\omega^t)\}$ be a sequence of $\mathscr{F}_t$-measurable selection from $\mathbf{P}(h_t)$, for $t \geqslant 2$. Let P be the probability measure on $(\Omega^\infty, \mathscr{F}_\infty)$ generated by $\{P_t^*\}_{t=0}^\infty$ via the Kolmogorov Theorem. Then $P \in \mathscr{P}_0$ and

$$\begin{aligned} V(F_i, (c_1(\omega_1), d_2(\omega_1))) &= u(c_1(\omega_1)) + \beta \int u(c_2(\omega_1, \omega_2)) P_1^*(\omega_1, d\omega_2) \\ &= E_P[u(c_1) + \beta u(c_2) | \mathscr{H}_1], \end{aligned}$$

$$\begin{aligned} V(c_0, d_1) &= u(c_0) + \beta E_P[\tilde{V}(d_1)] \\ &= u(c_0) + \beta E_P[u(c_1) + \beta u(c_2) | \mathscr{H}_0]. \end{aligned}$$

Thus, the LHS of Eq. (10) is greater than its RHS. On the other hand, by Theorem 5.5 and the construction of $\mathscr{P}_0$, its LHS is always less than its RHS. Therefore, the equality holds. □

Proof of Theorem 6.1. Note first that

$$E_{\mathscr{P}}\left[\sum_{s=0}^\infty \beta^s u(c_{t+s}) \middle| \mathscr{H}_t\right] \geqslant E_{\mathscr{P}}\left[u(c_t) + \beta E_{\mathscr{P}}\left(\sum_{s=0}^\infty \beta^s u(c_{t+s+1}) \middle| \mathscr{H}_{t+1}\right) \middle| \mathscr{H}_t\right].$$

For the reverse inequality, let $P(\omega^{t+1}) \in \hat{\mathscr{P}}_{t+1}(\omega^{t+1})$ be such that

$$E_{P(\omega^{t+1})}\left(\sum_{s=0}^\infty \beta^s u(c_{t+s+1}) \middle| \mathscr{H}_{t+1}\right) = E_{\mathscr{P}}\left(\sum_{s=0}^\infty \beta^s u(c_{t+s+1}) \middle| \mathscr{H}_{t+1}\right),$$

and let $P(\omega^t) \in \mathscr{P}_t(\omega^t)$ be such that

$$\begin{aligned} &E_{\mathscr{P}}\left[u(c_t) + \beta E_{\mathscr{P}}\left(\sum_{s=0}^\infty \beta^s u(c_{t+s+1}) \middle| \mathscr{H}_{t+1}\right) \middle| \mathscr{H}_t\right] \\ &\quad = E_{P(\omega^t)}\left[u(c_t) + \beta E_{P(\omega^{t+1})}\left(\sum_{s=0}^\infty \beta^s u(c_{t+s+1}) \middle| \mathscr{H}_{t+1}\right) \middle| \mathscr{H}_t\right]. \end{aligned}$$

Since $\{\mathscr{P}_t\}$ satisfies the generalized Bayes rule, the probability measure defined by

$$P(A) = \int \int 1_A P(\omega^t, \omega_{t+1}, d\omega^\infty) P(\omega^t, d\omega_{t+1})$$

is in $\mathscr{P}_t$. Thus the reverse inequality holds. □

References

[1] E. Anderson, L. Hansen, T. Sargent, Robustness, detection and the price of risk, Working paper, University of Chicago, 1999.

[2] B. Anger, Representation of capacities, Math. Ann. 229 (1977) 245–258.

[3] C.F. Camerer, M. Weber, Recent developments in modelling preferences: uncertainty and ambiguity, J. Risk Uncertainty 8 (1992) 167–196.

[4] R. Casadesus-Masanell, P. Klibanoff, E. Ozdenoren, Maxmin expected utility over savage acts with a set of priors, J. Econom. Theory 92 (2000) 35–65.
[5] S.H. Chew, L. Epstein, The structure of preferences and attitudes towards the timing of the resolution of uncertainty, Internat. Econom. Rev. 30 (1989) 103–117.
[6] S.H. Chew, L. Epstein, Recursive utility under uncertainty, in: A. Kahn, N. Yannelis (Eds.), Equilibrium Theory with an Infinite Number of Commodities, Springer, Berlin, Heidelberg, New York, 1991, pp. 352–369.
[7] S.H. Chew, E. Karni, Choquet expected utility with a finite state space: commutativity and act-independence, J. Econom. Theory 62 (1994) 469–479.
[8] G. Choquet, Theory of capacities, Ann. Inst. Fourier 5 (1953–4) 131–295.
[9] K. Cotter, Similarity of information and behavior with a pointwise convergence topology, J. Math. Econom. 15 (1986) 25–38.
[10] G. Debreu, Representation of a preference ordering by a numerical function, in: R. Thrall, C. Coombs, R. David (Eds.), Decision Processes, Wiley, New York, 1954.
[11] A.P. Dempster, Upper and lower probabilities induced from a multivalued mapping, Ann. Math. Statist. 38 (1967) 525–539.
[12] A.P. Dempster, A generalization of Bayesian inference, J. Roy. Statist. Soc. Ser. B 30 (1968) 205–247.
[13] J. Dow, S. Werlang, Laws of large numbers for non-additive probabilities, Working paper, London Business School, 1994.
[14] D. Ellsberg, Risk, ambiguity, and the savage axioms, Quart. J. Econom. 75 (1961) 643–669.
[15] L. Epstein, S. Zin, Substitution, risk aversion and the temporal behavior of consumption and asset returns: a theoretical framework, Econometrica 57 (1989) 937–969.
[16] L. Epstein, M. Le Breton, Dynamically consistent beliefs must be Bayesian, J. Econom. Theory 61 (1993) 1–22.
[17] L. Epstein, T. Wang, Intertemporal asset pricing under knightian uncertainty, Econometrica 62 (1994) 283–322.
[18] L. Epstein, T. Wang, Uncertainty, risk-neutral measures and security price booms and crashes, J. Econom. Theory 67 (1995) 40–80.
[19] I. Gilboa, Expected utility theory with purely subjective non-additive probabilities, J. Math. Econom. 16 (1987) 65–88.
[20] I. Gilboa, D. Schmeidler, Maxmin expected utility theory with non-unique prior, J. Math. Econom. 18 (1989) 141–153.
[21] I. Gilboa, D. Schmeidler, Updating ambiguous beliefs, J. Econom. Theory 59 (1993) 33–49.
[22] W.M. Gorman, The structure of utility functions, Rev. Econom. Stud. 35 (1968) 367–390.
[23] S. Graf, A Radon–Nikodym theorem for capacities, J. Math. 320 (1980) 192–214.
[24] L. Hansen, T. Sargent, T. Tallarini, Robust permanent income and pricing, Rev. Econom. Stud. 66 (1999) 873–907.
[25] T. Herron, T. Seidenfeld, L. Wasserman, Divisive conditioning: further results on dilation, Philos. Sci. 64 (1997) 411–444.
[26] P. Klibanoff, Dynamic choice with uncertainty aversion, Working paper, Northwestern University, 1995.
[27] F. Knight, Risk, Uncertainty and Profit, Houghton, Mifflin, Boston, 1921.
[28] T. Koopmans, Stationary ordinal utility and impatience, Econometrica 28 (1960) 287–309.
[29] D. Kreps, E. Porteus, Temporal resolution of uncertainty and dynamic choice theory, Econometrica 46 (1978) 185–200.
[30] M. Machina, D. Schmeidler, A more robust definition of subjective probability, Econometrica 60 (1992) 745–780.
[31] M. Marinacci, Limit laws for non-additive probabilities and their frequentist interpretation, J. Econom. Theory 84 (1999) 145–195.
[32] Y. Nakamura, Subjective expected utility with non-additive probabilities on finite state spaces, J. Econom. Theory 51 (1990) 346–366.
[33] R. Sarin, P. Wakker, A simple axiomatization of nonadditive expected utility, Econometrica 60 (1998) 1255–1272.

[34] R. Sarin, P. Wakker, Revealed likelihood and knightian uncertainty, J. Risk Uncertainty 16 (1998) 223–250.
[35] L. Savage, Foundations of Statistics, Wiley, New York, 1954.
[36] D. Schmeidler, Integral representation without additivity, Proc. Amer. Math. Soc. 97 (1986) 255–261.
[37] D. Schmeidler, Subjective probability and expected utility without additivity, Econometrica 57 (1989) 571–587.
[38] G. Shafer, A Mathematical Theory of Evidence, University of Princeton Press, New Jersey, 1976.
[39] G. Shafer, Allocations of probability, Ann. Probab. 7 (1979) 827–839.
[40] T. Seidenfeld, L. Wasserman, Dilation for convex sets of probabilities, Ann. Statist. 21 (1993) 1139–1154.
[41] C. Skiadas, Subjective probability under additive aggregation of conditional preferences, J. Econom. Theory 76 (1997) 242–271.
[42] C. Skiadas, Recursive utility and preferences for information, Econom. Theory 12 (1998) 293–312.
[43] P. Wakker, Testing and characterizing properties of nonadditive measures through violations of the sure-thing principle, Econometrica 69 (2001) 1039–1059.
[44] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall, London, 1991.
[45] T. Wang, A class of dynamic risk measures, Working paper, University of British Columbia, 1999.
[46] L. Wasserman, J. Kadane, 'Bayes' Theorem for Choquet capacities, Ann. Statist. 18 (1990) 1328–1339.