# Updating Rules for Non-Bayesian Preferences.

Tan Wang*

December, 1999

**Abstract**

We axiomatize updating rules for preferences that are not necessarily in the expected utility class. Two sets of results are presented. The first is the axiomatization and representation of conditional preferences. The second consists of the axiomatization of three updating rules: the traditional Bayesian rule, the Dempster-Shafer rule, and the generalized Bayesian rule. The last rule can be regarded as the updating rule for the multi-prior expected utility (Gilboa and Schmeidler (1989)). The operational merit of it is that it is equivalent to updating each prior by the traditional Bayesian rule.

*Tan Wang is with the Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z2. tan.wang@commerce.ubc.ca. The author is grateful to the Social Sciences and Humanities Research Council of Canada for financial support.

# 1 Introduction

The traditional approach to updating is the Bayesian rule. This approach is justified by the axiomatic treatment of Savage (1954), where it is shown that, in situations of uncertainty, if a decision maker's preference satisfies a certain set of axioms, his preference can be represented by an expected utility with respect to a subjective probability measure and that probability measure represents the decision maker's belief about the likelihood of events. Moreover, in light of new information, the decision maker updates his belief according the Bayesian rule. This Savage paradigm has been the foundation of much of the economic theories under uncertainty. At the same time, however, the Savage paradigm has been challenged by behavior exhibited in Ellsberg paradox (Ellsberg (1961)), which seems to question the very notion of representing a decision maker's belief by a probability measure and hence by implication the validity of the Bayesian rule.

Such discrepancy between the theories and the empirical facts has been the driving force behind various attempts to extend the Savage paradigm. The earliest attempt dates back at least to Dempster (1967, 1968). Since then, the extensions have been developed along two fronts. One is the axiomatization of preferences that can accommodate behavior such as that seen in Ellsberg paradox. Schmeidler (1989) axiomatizes a class of preferences as the integral of a utility function with respect to a non-additive probability. Schmeidler's axiomatization is in the Anscome-Aumann framework. Gilboa (1987), Nakamura (1990), and Sarin and Wakker (1992) develop the same class of preferences in the Savage setting. Gilboa and Schmeidler (1989) provides a theory of expected utility with multi-priors.[1] The second front is on developing updating rules for beliefs/preferences that cannot be represented by probability measures/expected utility. The existing literature includes Dempster (1967, 1968), Shafer (1976, 1979), and more recently, Gilboa and Schmeidler (1993).

In this paper we address the issue of how people update their beliefs when their preferences do not necessarily fall into the class of expected utility. The motivation for this paper comes from two sources. First, while there has been progress in the decision theory literature, the results on updating rules for non-expected utility type of preferences are not completely satisfactory. For instance, Machina and Schmeidler (1992) and Epstein and Le Breton (1993) extend the notion of subjective beliefs and updating to non-expected utility. The beliefs, however, are still represented by probability measures and the updating rule is still Bayesian as in the Savage paradigm. Dempster (1967, 1968) and Shafer (1976, 1979) generalize the Bayesian rule. However, the Dempster-Shafer

[1] For more references to this literature, see Camerer and Weber (1992) and Sarin and Wakker (1998) and the references therein.

rule they propose lacks rigorous axiomatic foundation. Gilboa and Schmeidler (1993) are the first to study axiomatically updating with general preferences. However, most preferences in the class they study do not support a separation of preference and belief, making it difficult to interpret intuitively the notion of updating. The second source of motivation comes from the recent advance in asset pricing literature. Epstein and Wang (1994, 1995) develop an intertemporal asset pricing model under Knightian uncertainty. In the model, the agent's preference is represented by a multi-period version of the multi-prior expected utility developed by Gilboa and Schmeidler (1989). The evolution of the agent's belief is modeled 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. The learning/updating issue is not formally addressed. Hansen, Sargent and Tallarini (1999) and Anderson, Hansen and Sargent (1999) introduce 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 describing the state of the economy and his preference for his decision rule to be robust to the misspecification. One potential justification for these models to abstract away from the issue of learning is that the Knightian uncertainty or the potential error in model specification is taken by the agent as the state of affairs, or the models are the reduced form of a model with learning.[2] While this is sometimes justifiable, for a more complete rational expectations model, it seems desirable to allow the agent to learn and update, especially if one is to study the impact of learning on asset price/return dynamics.

We present two sets of results. Section 4 contains the first set: the axiomatization and representation of conditional preferences. The second set of results consists of three updating rules: the traditional Bayesian rule (Section 5.1), the Dempster-Shafer rule (Section 5.2), and the generalized Bayesian rule (Section 5.3). The last rule can be regarded as the updating rule for the multi-prior expected utility. The operational merit of it is that it is equivalent to updating each prior by the traditional Bayesian rule.

The rest of the paper is organized as follows: Section 2 contains a brief overview of the methodology. In Section 3, we introduce the set of multi-period consumption-information profiles. These consumption-information profiles correspond to the acts in Savage setting. Each profile has two components, one is the consumption profile, which is standard; the other is the information profile, which describes the information flow according to which the preference is updated. Section 6 discusses some of the potential applications. Proofs and supporting technical details are collected in

[2] Dow and Werlang (1994) show that if uncertainty is persistent, learning and updating will not completely eliminate uncertainty.

the Appendix.

## 2 Overview of the Methodology and Related Literature

Following Savage (1954) and Gilboa and Schmeidler (1993), our approach is axiomatic. We differ, however, in the assumed primitives. In the literature, there exist two strands of research on updating rules for preferences more general than expected utility. One strand maintains the probability framework. Machina and Schmeidler (1992) show that notion of subjective probability can be extended to non-expected utility preferences. Epstein and Le Breton (1993) further show that the updating rule for such non-expected utility preferences must be Bayesian if dynamic consistency is to be ensured. The second strand of research goes beyond the probability framework. Gilboa and Schmeidler (1993) study the updating rules for general preferences. While different in the class of preferences dealt with, these two strands of literature share a common feature in the primitives assumed. They both start with an initial preference. Axioms are imposed on the initial preference. The implied updating rules are then derived. In Machina and Schmeidler (1992) and Epstein and Le Breton (1993), a subjective probability or belief component of the preference is first separated from the initial preference. It is then shown to update according to the Bayesian rule. In Gilboa and Schmeidler (1993), updating rules are defined using the initial preference. No belief component is separated from the initial preference. In the case of Choquet expected utility, the Dempster-Shafer rule is derived. This paper starts in a different direction. It takes as primitives the family of conditional preferences. The premise is that the updating rule is encoded in this family of conditional preferences, in the connection between the current and future conditional preferences, in particular. We use a set of axioms to analyze that connection and extract the updating rule encoded.

In addition to the presumed preferences, the second difference in the assumed primitives pertains to the objects of choice. Traditionally, the objects of choice are one period acts. This perhaps is the natural consequence of starting with initial preferences. At a more fundamental level, it reflects the distinction between the payoff approach, where even in a multi-period setting an object of choice is described only by its payoff vector, and the temporal lottery approach, where both the payoff vector and the timing of resolution of uncertainty are important (Kreps and Porteus (1978)). In a multi-period uncertain environment, it seems intuitive or rational that if a decision maker anticipates new information at a future time, he would evaluate the entire (multi-period) act by a backward induction based on the incoming information. This implies that the evaluation should depend on not only the payoffs but also the information flow of the multi-period act. Thus, following the approach first started by Skiadas (1997, 1998), we mold consumption and information flows together as the objects of choice.

Our methodology is perhaps best motivated from application's point of view. First, for intertemporally additive expected utility, due to the law of iterated expectation,

$$V_0 = E_0\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right] = E_0\left[u(c_0) + \beta E_1[V_1]\right]. \tag{1}$$

It says that the unconditional formulation is equivalent to the recursive conditional formulation. For more general preferences, however, law of iterated expectation need not hold, and the recursive formulation is not necessarily equal to the unconditional formulation. It is well-understood that when a preference cannot be represented by an intertemporally additive expected utility, which happens, for instance, if either the preference is not intertemporally additive as in the case of general recursive utility[3] or the preference is not additive across states as in the case expected utility with non-additive probability,[4] in order to ensure dynamic consistency and independence of unrealized events, the preference has to be formulated recursively. In this setting, current utility is obtained by aggregating utility from current consumption and future conditional utility derived from future consumptions, leading naturally to taking conditional preferences are primitives. Secondly, for an intertemporal expected utility, uncertainty can be described either by an unconditional probability measure or by a consistent family of conditional probability measures. Either way is equivalent with the other. For starting with an unconditional probability measure, updating by Bayesian rule leads to a consistent family of conditional probability measures. Conversely, starting with a consistent family of conditional probability measures, one can construct a unique unconditional probability by Kolmogorov theorem. This equivalence combined with the law of iterated expectation implies that the intertemporally additive expected utility enjoys a property called timing indifference, which means that an individual is indifferent to earlier or later resolution of uncertainty.[5] Fundamentally, it is this property that ensures the equivalence in (1) and partially justifies the payoff approach. When the equivalence fails for the more general preferences, the definition of objects of choice necessitates appropriate specification of the information flows.

[3]See Epstein and Zin (1989).

[4]See Schmeidler (1989), Gilboa (1987), Gilboa and Schmeidler (1989) in static setting and Epstein and Wang (1994,1995) in dynamic setting.

[5]See Kreps and Porteus (1978), Chew and Epstein (1991), and Skiadas (1998).

## 3 Consumption-Information Profiles

Let $\Omega$ be a finite set, which is taken as the state space for a generic period. The full state space is $\Omega^\infty$. Let $\mathcal{B}(R_+)$ be the space of bounded functions from $\Omega$ to $R_+$. An element of $\mathcal{B}(R_+)$ is viewed as the state-contingent consumptions in a period. Let $\mathbf{F}$ denote the collection of all possible partitions of $\Omega$. Each member of $\mathbf{F}$ can be viewed as the information revealed in a period.

As motivated in Section 2, the consumption-information profiles that we will construct below have two components: a consumption component and an information component. Intuitively, the consumption profile is a $t$-period tree as illustrated in Figure 1 for the case of $t = 2$. Here $(c_0, d_1)$

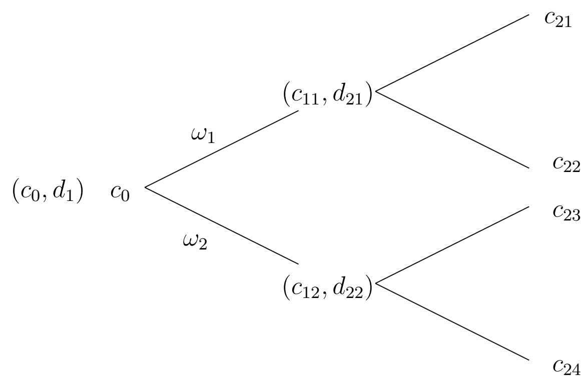


Figure 1: Two-Period Consumption Profile.

denotes the consumption profile, $c_0$, $c_{11}$, $c_{12}$, $c_{21}$, $c_{22}$, $c_{23}$, and $c_{24}$, the time-state-contingent consumptions, and $d_{21}$ and $d_{21}$, the continuation of $d_1$ in state $\omega_1$ and $\omega_2$ respectively. The information component describes the evolution of information or the resolution of uncertainty over time. It is typically described by a filtration, i.e., a sequence of increasing $\sigma$-algebras on $\Omega^\infty$.

We begin with $t$-period consumption-information profiles. For any space $Y$, let $\mathcal{B}(Y)$ denote the space of bounded functions $\tilde{x} : \Omega \to Y$. The space of $t$-period consumption-information profiles is constructed recursively. Let $D_1 = \mathcal{B}(R_+) \times \mathbf{F}$. For each $t > 1$, define

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

A typical element of $D_t$ is denoted by

$$d = (\tilde{d}, \mathcal{F}),$$

where $\tilde{d} : \omega \to (c_1(\omega), d_2(\omega))$ maps $\omega$ to $(c_1(\omega), d_2(\omega))$. Elements of $R_+ \times D_t$ are called $t$-period consumption-information profiles. For example, let $(c_0, d_1) \in R_+ \times D_2$ be a two-period consumption-information profile. Suppose

$$d_1 = (\tilde{d}_1, \mathcal{F}), \qquad \text{with } \ \mathcal{F} = \{A, A^c\}, \qquad \tilde{d}_1 : \omega_1 \to (c_1(\omega_1), d_2(\omega_1)),$$

$$c_1(\omega_1) = \begin{cases} c_{11} & \text{if } \omega_1 \in A \\ c_{12} & \text{if } \omega_1 \in A^c \end{cases}, \qquad d_2(\omega_1, \omega_2) = \begin{cases} c_{21} & \text{if } \omega_1 \in A \text{ and } \omega_2 \in B \\ c_{22} & \text{if } \omega_1 \in A \text{ and } \omega_2 \in B^c \\ c_{23} & \text{if } \omega_1 \in A^c \text{ and } \omega_2 \in E \\ c_{24} & \text{if } \omega_1 \in A^c \text{ and } \omega_2 \in E^c \end{cases}.$$

Then $(c_0, d_1)$ corresponds to the two-period tree in Figure 1.

As will become clear in Section 4, the component $\mathcal{F}$ in $(\tilde{d}, \mathcal{F})$ is the information partition with which the individual updates his preference/belief at the beginning of the next period after the uncertainty in the current period has realized. As time elapses, the information filtration embedded in $d \in D_t$ gradually realizes step by step.[6] This is the information based on which the individual updates his preference. It needs not coincide with the objective information filtration of the economy where the individual resides. For instance, in a world with switching regimes, the current dividend can be high due to a high economic regime. However, the individual may not be able to confirm the switching of regime and hence may not update his belief.[7] Thus we do not require

$$D_t = \Big\{(\tilde{d}, \mathcal{F}) \in \mathcal{B}(R_+ \times D_{t-1}) \times \mathbf{F}, \quad \tilde{d} \text{ is } \mathcal{F}\text{-measurable}\Big\}$$

in order to reflect that the individual may decide not to update his belief.

In a $t$-period consumption-information profiles, consumption ends after $t$ period. We would of course like the space of consumption-information profiles to contain these $t$-period consumption-information profiles. We would also like the space to contain those profiles that extend indefinitely into the future. For that purpose, we introduce the mappings $f_t$. Let $f_1 : D_2 \to D_1$ be defined by, for any $d = (\tilde{d}, \mathcal{F}) \in D_2$ with $\tilde{d} : \omega \to (c_1(\omega), d_2(\omega))$,

$$f_1(\tilde{d}, \mathcal{F}) = \Big(f_1(\tilde{d}), \mathcal{F}\Big) \quad \text{and} \quad f_1(\tilde{d})(\omega) = c_1(\omega),$$

where $c_1(\omega)$ denotes the value of $c_1$ in state $\omega$. Inductively, for $t > 1$, $f_t : D_{t+1} \to D_t$ is defined by,

[6] The Appendix contains the information on how to extract the information filtration for updating that is embedded in $d \in D_t$.

[7] The discrepancy between the objective and subjective information filtration may reflect some sort of friction in information acquisition on the part of the individual.

for any $d = (\tilde{d}, \mathcal{F}) \in D_{t+1}$ with $\tilde{d} : \omega \to (c_1(\omega), d_2(\omega))$,

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

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 in the last period and eliminate the information partition in the second last period of the $t$-period profile.[8] We define an infinite consumption-information profile as the limit of a sequence of finite profiles with the property that each $(t+1)$-period profile in the sequence is consistent with the preceeding $t$-period profile. Thus, we define the space $D$ of consumption-information profiles as

$$D = \{(d_1, d_2, \ldots) : d_t \in D_t \text{ and } d_t = f_t(d_{t+1}),\ t \geq 1\}.$$

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 zero profile. 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 profiles such that any one of them grows out its predecessor. Thus $d_t$ becomes an element of $D$.

The space $R_+ \times D$ will be the domain on which conditional preferences are defined. We endow $D$ with the pointwise convergence topology. More specifically, for any space $Y$, the topology on $\mathcal{B}(Y)$ is the standard pointwise convergence topology. On $\mathbf{F}$, define the topology by the metric

$$\rho(\mathcal{F}, \mathcal{G}) = \sum_{\omega \in \Omega} \sum_{\omega' \in \Omega} \left| \sum_{i=1}^{\#(\mathcal{F})} \frac{1}{\#(F_i)} \sum_{\omega'' \in F_i} 1_{\{\omega\}}(\omega'') 1_{F_i}(\omega') - \sum_{j=1}^{\#(\mathcal{G})} \frac{1}{\#(G_j)} \sum_{\omega'' \in G_i} 1_{\{\omega\}}(\omega'') 1_{G_j}(\omega') \right|, \quad (2)$$

where $\#(\mathcal{F})$ and $\#(F_i)$ denote the number of elements in $\mathcal{F}$ and $F_i$ respectively, and $1_F$ is the indicator function. This metric induces the pointwise convergence topology introduced by Cotter (1986) on set of $\sigma$-algebras. Intuitively, two partitions $\mathcal{F}$ and $\mathcal{G}$ are different if there are at least two subsets $F_i \in \mathcal{F}$ and $G_j \in \mathcal{G}$ such that $F_i \cap G_j \neq \emptyset$ and $F_i \neq G_j$. In that case, there exist $\omega$ and $\omega'$ 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 $\omega$ and $\omega'$, the term inside the absolute value sign in equation (2) is strictly positive, implying $\rho(\mathcal{F}, \mathcal{G}) > 0$. Conversely, if $\rho(\mathcal{F}, \mathcal{G}) > 0$, then reversing the argument above implies that $\mathcal{F}$ and $\mathcal{G}$ are not identical. Thus, conforming to the intuition, if $\mathcal{F}_n$ is a sequence of partitions and

[8] As will be seen later, this information partition becomes irrelevant once the consumption in the last period is cut off.

$\rho(\mathcal{F}_n, \mathcal{F}) \to 0$, then $\mathcal{F}_n$ "converges" to $\mathcal{F}$ because for large $n$, $\rho(\mathcal{F}_n, \mathcal{F}) = 0$. Now for each $t \geq 1$, give $\mathcal{B}(R_+ \times D_{t-1})$ the standard pointwise convergence topology, and $D_t = \mathcal{B}(R_+ \times D_{t-1}) \times \mathbf{F}$ the product topology. Finally, we give $D$ the product topology.

As it is, the definition of space $D$, although intuitive, is not convenient to use. The following theorem provides some structure to it.

**Theorem 3.1** *$D$ is homeomorphic to $\mathcal{B}(R_+ \times D) \times \mathbf{F}$.*

The main merit of this theorem is that it allows us to write $d = (d_1, d_2, \ldots)$ as

$$d = (\tilde{d}, \mathcal{F}), \quad \text{where } \tilde{d} : \omega \to (c_1(\omega), d_2(\omega)) \in R_+ \times D.$$

That is, we can view $d$ as a random variable whose value in each state $\omega$ is a (infinite) consumption-information profile. This structure of elements of $D$ will be useful in the subsequent analysis.

In addition to what is explained earlier, our space $D$ differs from what exist in the literature in some other respects. In Kreps and Porteus (1978), Epstein and Zin (1989) and Chew and Epstein (1991), the space $D$ consists of multi-period lotteries, i.e., trees with a probability attached to each of its branches. Modeling consumption profiles as multi-period lotteries implicitly assumes that the probabilities associated with various events have already been evaluated. In an uncertainty world, by definition, probabilities are not given. To allow for the derivation of subjective probability as in Savage (1954), or non-probabilistically sophisticated preferences, it is imperative that we model the space $D$ at a more primitive level by removing the assumption of exogenously given probabilities. Wang (1999) also models consumption profiles as multi-period trees without probabilities. However, the information profiles are not modeled. Our consumption-information profiles are closest to the acts in Skiadas (1997, 1998). However, Skiadas (1997, 1998) requires that the consumption profile be adapted to the information profile and does not exploit the recursive structure as in our construction of $D_t$ and in Theorem 3.1.

# 4 Conditional Preferences

The objective of this section is to axiomatize and provide the numerical representation for a class of conditional preferences. This class of conditional preferences will provide the basis for our later study of updating rules.

Let $t \geq 1$ and $F_1 \times \cdots \times F_t \subset \Omega^t$ be a sequence of past events. A conditional preference $\succeq_{F_1 \times \cdots \times F_t}$ given $F_1 \times \cdots \times F_t$ 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., $\{\succeq_{F_1 \times \cdots \times F_t} : F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 1\}$.

**Axiom 1** (**Continuity**) *For all $F_1 \times \cdots \times F_t \subset \Omega^t$ and all sequences $\{(c_n, d_n)\}$ and $\{(c'_n, d'_n)\} \in R_+ \times D$ with $(c_n, d_n) \to (c, d)$ and $(c'_n, d'_n) \to (c', d')$, if $(c_n, d_n) \succeq_{F_1 \times \cdots \times F_t} (c'_n, d'_n)$ for all $n$, then $(c, d) \succeq_{F_1 \times \cdots \times F_t} (c', d')$.*

**Axiom 2** (**Risk Separability**): *For all $F_1 \times \cdots \times F_t \subset \Omega^t$, $(c, d)$ and $(c', d') \in R_+ \times D$, $(c, d) \succeq_{F_1 \times \cdots \times F_t} (c, d')$ if and only if $(c', d) \succeq_{F_1 \times \cdots \times F_t} (c', d')$.*

**Axiom 3** (**Deterministic Information Independence**) *For all $A_1 \times \cdots \times A_t \subset \Omega^t$, $B_1 \times \cdots \times B_t \subset \Omega^t$, and deterministic consumption-information profiles, $(c, d)$ and $(c', d')$, $(c, d) \succeq_{A_1 \times \cdots \times A_t} (c', d')$ if and only if $(c, d) \succeq_{B_1 \times \cdots \times B_t} (c', d')$.*

These three Axioms are straightforward to interpret. Continuity is a technical property. Risk Separability says that if two consumption-information profiles have identical current consumption, then the ranking of these two profiles should be independent of the common current consumption. Deterministic Information Independence Axiom says that if there is no uncertainty associated with the consumption profiles, then the ranking of the consumption profiles should be independent of how the preferences are updated. In other words, all conditional preferences rank deterministic consumption profiles the same.[9]

To state the next property we need the following definitions. An event $A \subset \Omega$ is said to be null if for all $(c, d)$, $(c', d')$ and $(c'', d'') \in R_+ \times D$, $(0, (d' 1_A + d 1_{A^c}, \mathcal{F})) \sim (0, (d'' 1_A + d 1_{A^c}, \mathcal{F}))$, where $\mathcal{F} = \{A, A^c\}$ and the addition and multiplication are as in the space of random variables. An event $A \subset \Omega$ is said to be universal if for all $(c, d)$, $(c', d')$ and $(c'', d'') \in R_+ \times D$, $(0, (d 1_A + d' 1_{A^c}, \mathcal{F})) \sim (0, (d 1_A + d'' 1_{A^c}, \mathcal{F}))$. Clearly, $A$ is null if and only if $A^c$ is universal. Null events are those that are considered to have zero likelihood of happening.

[9]Due to this axiom, although in an one-period consumption-information profile, there is an information component, it is irrelevant. This is because when $(c_0, d_1) \in R_+ \times D_t$ evolves to the last period,

$$(c_{t-1}(\omega_1, \ldots, \omega_{t-1}), d_t(\omega_1, \ldots, \omega_{t-1})) \in R_+ \times D_1,$$

the consumption will end when the uncertainty in this last period is realized. There is no further consumption. Therefore, how preference is updated further after that is irrelevant.

**Axiom 4** *(**Consistency**) For all $(c_i, d_i)$ and $(c_i', d_i') \in R_+ \times D$, $i = 1, \ldots, n$, and all partitions $\mathcal{F} = \{A_1, \ldots, A_n\}$ of $\Omega$, if $(c_i, d_i) \succeq_{F_1 \times \cdots \times F_t \times A_i} (c_i', d_i')$, for $i = 1, \ldots, n$, then for any $c \in R_+$,*

$$\left(c, \left[\sum_{i=1}^{n} (c_i, d_i) 1_{A_i}, \mathcal{F}\right]\right) \succeq_{F_1 \times \cdots \times F_t} \left(c, \left[\sum_{i=1}^{n} (c_i', d_i') 1_{A_i}, \mathcal{F}\right]\right).$$

*Moreover, the latter ordering is strict if $(c_i, d_i) \succ_{F_1 \times \cdots \times F_t \times A_i} (c_i', d_i')$ for some $A_i$ that is not null.*

The intuition behind this axiom is readily explained with Figure 2. For Figure 2(a), when

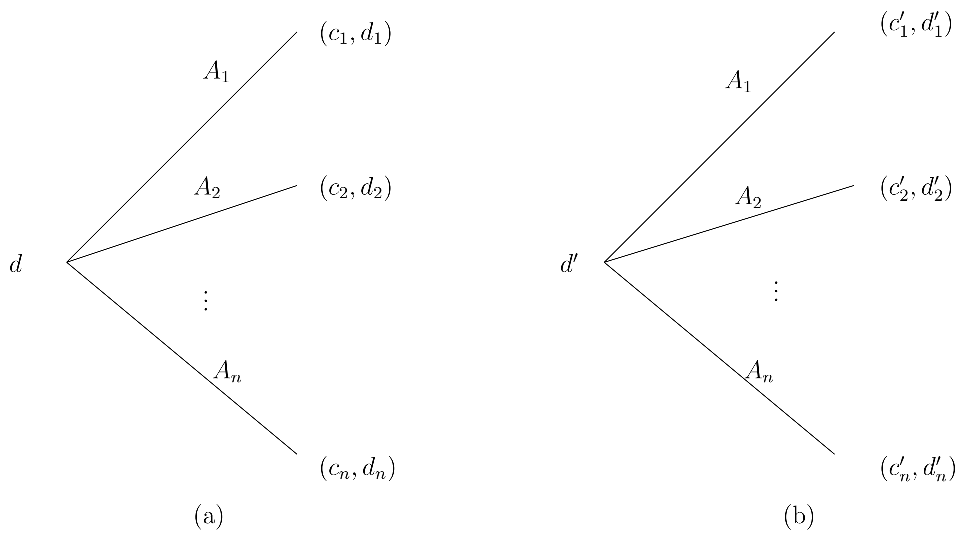


Figure 2: Consistency

event $A_i$ happens, the realized consumption profile is $(c_i, d_i)$, which is itself a $(T-1)$-period profile yet to be fully realized over the next $T-1$ periods. Figure 2(b) has a similar interpretation. Suppose that for all $i = 1, \ldots, n$, $(c_i, d_i) \succeq_{F_1 \times \cdots \times F_t \times A_i} (c_i', d_i')$. That is, when event $A_i$ is realized at the end of period one, $(c_i, d_i)$ is preferred to $(c_i', d_i')$. Consistency then requires that $(c, [\sum_{i=1}^{n} (c_i, d_i) 1_{A_i}, \mathcal{F}]) \succeq_{F_1 \times \cdots \times F_t} (c, [\sum_{i=1}^{n} (c_i', d_i') 1_{A_i}, \mathcal{F}])$ today for any $c \in R_+$. In other words, if ex-post $d$ is preferred to $d'$, then ex-ante $d$ must also be preferred.

The intuitive appeal of the Consistency Axiom seems obvious. As will be seen in Section 4 this axiom guarantees that the conditional preferences aggregate in a time-consistent fashion. It is well understood that in a dynamic optimization problem, if the objective function is not time-consistent, a strategy chosen today may be regretted later on. That is, if given the opportunity, the strategy chosen earlier will be abandoned in favor of another one, causing inconsistency in choices over time.

Consistency can also be viewed as a generalization of monotonicity. Indeed, when combined with Stationarity, Risk Separability and Deterministic Information Independence Axioms, it implies the usual monotonicity property.

**Axiom 5** (**Stationarity**) *For all* $(c, d)$, $(c', d') \in R_+ \times D$ *and* $F_1 \times \cdots \times F_t \subset \Omega^t$, $(c, d) \succeq_{F_1 \times \cdots \times F_t}$ $(c', d')$ *if and only if* $(c, d) \succeq_{F_1 \times \cdots \times F_t \times \Omega} (c', d')$.

This axiom says that if there is no information revealed in the next period, the conditional preference remains unchanged.

Now we are ready to present the first representation theorem. First some preliminary definitions and notations. A numerical function $V_t[F_1 \times \cdots \times F_t] : D \to R$ is said to represent the conditional preference $\succeq_{F_1 \times \cdots \times F_t}$ if, for all $(c, d)$ and $(c', d') \in R_+ \times D$,

$$(c, d) \succeq_{F_1 \times \cdots \times F_t} (c', d')$$

if and only if

$$V[F_1 \times \cdots \times F_t, (c, d)] \geq V[F_1 \times \cdots \times F_t, (c', d')].$$

Given a numerical function $V_t[F_1 \times \cdots \times F_t]$ that represents the conditional preference $\succeq_{F_1 \times \cdots \times F_t}$, define a companion numerical function $V(F_1 \times \cdots \times F_t, d)$ on $D$ by

$$V(F_1 \times \cdots \times F_t, d) = f[V(F_1 \times \cdots \times F_t, (0, d))],$$

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

$$V(c) = f[V(F_1 \times \cdots \times F_t, (0, c))].$$

Here $V(c)$ is the conditional utility of one time consumption at time 0 when by convention no historical information is recorded, and $(0, c)$ is the deterministic consumption profile whose consumption at time 1 is $c$ and whose consumption at any other time is zero. Intuitively, $V(F_1 \times \cdots \times F_t, d)$ is the utility of $d$ at time $t+1$ evaluated just before the uncertainty in the period between time $t$ and $t+1$ is realized. To illustrate, let $(c, d) = (c, (\tilde{c}_1, \mathcal{F}))$ be an one-period consumption profile and

$$V(F_1 \times \cdots \times F_t, (c, d)) = u(c) + \beta E\left[u(\tilde{c}_1) | F_1 \times \cdots \times F_t\right].$$

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

$$V(F_1 \times \cdots \times F_t, d) = E\left[u(\tilde{c}_1) | F_1 \times \cdots \times F_t\right].$$

Next, for each

$$d = (\tilde{d}, \mathcal{F}), \quad \tilde{d} : \omega \to (c_1(\omega), d_2(\omega)), \quad \mathcal{F} = \{A_1, \ldots, A_n\} \in \mathbf{F},$$

in $D$, define $\tilde{V}[F_1 \times \cdots \times F_t, d] : \Omega \to R$ by

$$\tilde{V}[F_1 \times \cdots \times F_t, d](\omega) = V[F_1 \times \cdots \times F_t \times A_i, (c_1(\omega), d_2(\omega))], \quad \text{if } \omega \in A_i. \tag{3}$$

$\tilde{V}[F_1 \times \cdots \times F_t, d]$ can be regarded as the ex-post evaluation of $d$ after the uncertainty in the current period is realized. A function $\mu : \mathcal{B}(R) \to R$ is called a certainty equivalent if (a) $\mu(x) = x$ for all $x \in R$, and (b) $\mu(\tilde{x}) \geq \mu(\tilde{y})$ if $\tilde{x} \geq \tilde{y}$.

**Theorem 4.1** *A family of conditional preferences* $\{\succeq_{F_1 \times \cdots \times F_t} : F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 1\}$ *satisfies Axioms 1–5 if and only if it can be represented by a family of continuous functions*

$$\Big\{V(F_1 \times \cdots \times F_t) : F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 1\Big\}$$

*on* $R_+ \times D$ *such that*

$$V[F_1 \times \cdots \times F_t, d] = \mu(F_1 \times \cdots \times F_t, \tilde{V}(F_1 \times \cdots \times F_t, d)) \tag{4}$$

*and*

$$V[F_1 \times \cdots \times F_t, (c, d)] = W(c, V[F_1 \times \cdots \times F_t, d]), \tag{5}$$

*where* $\mu(F_1 \times \cdots \times F_t, \cdot)$ *is a continuous certainty equivalent and* $W : R_+ \times R \to R$ *is continuous and strictly increasing.*

Theorem 4.1 is our basic aggregation theorem. Of particular interest is the structure of aggregation it provides. The function $W$ is the time aggregator. It describes for deterministic consumption profiles, how utility of future consumptions is aggregated with that derived from current consumption. The certainty equivalent $\mu$ 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.

Consider next an axiom that is similar to Risk Separability, with respect to deterministic losses.

**Axiom 6** (**Future Independence**)*: For all* $F_1 \times \cdots \times F_t \subset \Omega^t$*, all* $x_1$*,* $x_2$*,* $x_1'$ *and* $x_2' \in R$ *and deterministic losses* $Y = (y_1, y_2, \ldots)$ *and* $Y' = (y_1', y_2', \ldots) \in D$*,* $(x_1, x_2, Y) \succeq_{F_1 \times \cdots \times F_t} (x_1', x_2', Y)$ *if and only if* $(x_1, x_2, Y') \succeq_{F_1 \times \cdots \times F_t} (x_1', x_2', Y')$*.*

If we add this Axiom to Axioms 1-5, the time aggregator can be significantly simplified.

**Theorem 4.2** *A family of conditional preferences* $\{\succeq_{F_1\times\cdots\times F_t}: F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 0\}$ *satisfies Axioms 1–6 if and only if it can be represented by a family of continuous functions*

$$V[F_1 \times \cdots \times F_t, (c, d)] = u(c) + \beta\mu(F_1 \times \cdots \times F_t, \tilde{V}(F_1 \times \cdots \times F_t, d)). \tag{6}$$

*where* $\mu(F_1 \times \cdots \times F_t, \cdot)$ *is a continuous certainty equivalent 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

The purpose of this section is to axiomatize three updating rules. As briefly mentioned in the introduction, the starting point of our approach to updating is to take conditional preferences as the primitives. The premise is that updating rules are encoded in the (evolution of) conditional preferences. Together with other factors, the updating rules determine how future conditional preferences are aggregated to current conditional preferences. In this respect, Theorems 4.1 and 4.2 provide the basic structure of the aggregation in time and state dimensions. It should be clear that it is the aggregation along the state dimension that carries the information on the updating rules. Our study of updating rules will thus focus on that aggregation. The axioms introduced in this section will be directed at the aggregators $\mu(F_1 \times \cdots \times F_t, \cdot)$.

It should be noted at the outset that the aggregators $\mu(F_1 \times \cdots \times F_t, \cdot)$ carry more information than just about updating rules. For instance, they contain information about the individual's attitude toward risk and uncertainty.[10] When combined with the time aggregator $W$, it can also help determine the individual's attitude toward intertemporal substitution. We will focus only on the updating aspect.

[10]See Epstein (1999) on attitudes toward risk and uncertainty in a Savage setting.

### 5.1 Bayesian Updating Rule

First we axiomatize the ubiquitous Bayesian rule. Bayesian rule is important not only because of its wide applications, but also because, in the context of this paper, it serves as a benchmark for the other two updating rules that we will axiomatize later.

Let $\mathcal{F}^0$ denote the trivial partition $\{\Omega\}$. If $\mathcal{F}^0$ is the information partition, there is no new information revealed over the period and hence nothing to be learned.

**Axiom 7** *(***Strong Timing Indifference***): Let $F_1 \times \cdots \times F_t \subset \Omega^t$, and $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_m\}$ be two partitions of $\Omega$. For all (two-period) consumption-information profiles of the form $(0, d_1)$ and $(0, d_1')$ with $d_1 = (\tilde{d}_1, \mathcal{F}^0)$, $\tilde{d}_1(\omega_1) = (0, d_{2i})$ if $\omega_1 \in A_i$ and $\tilde{d}_{2i}(\omega_2) = c_{ij}$ if $\omega_2 \in B_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$; $d_1' = (\tilde{d}_1', \mathcal{F}^0)$, $\tilde{d}_1'(\omega_1) = (0, d_{2j}')$ if $\omega_1 \in B_j$ and $\tilde{d}_{2j}'(\omega_2) = c_{ij}$ if $\omega_2 \in A_i$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, we have $(0, d_1) \sim_{F_1 \times \cdots \times F_t} (0, d_1')$.*

The basic intuition behind this axiom can be readily explained with Figure 3. There are two

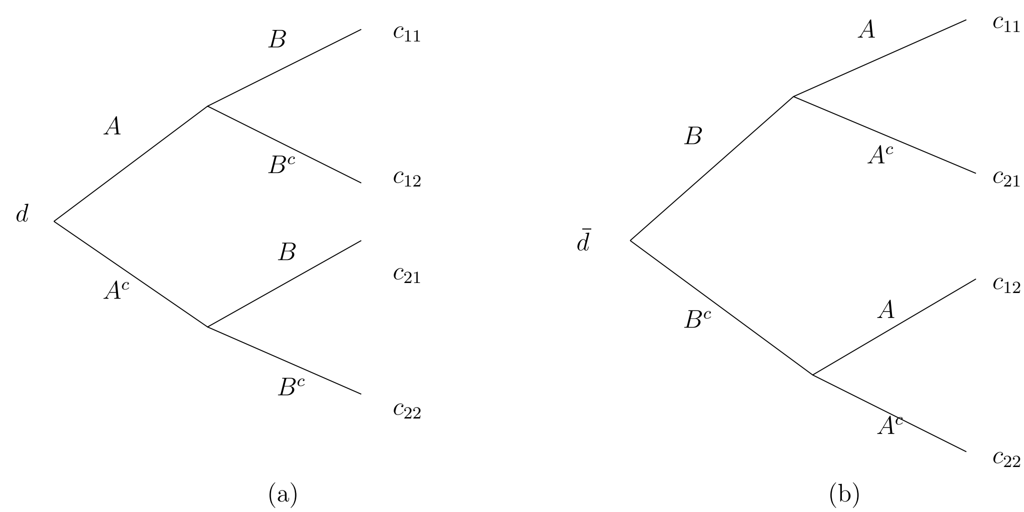


Figure 3: Strong Timing Indifference

events $A$ and $B$. In Figure 3(a), event $A$ transpires first and event $B$ follows. In Figure 3(b), event $B$ happens first and then event $A$ follows. Thus the timing of resolution of uncertainty in Figure 3(a) and (b) are reversed. There are no consumptions at time 0 and 1. It can be easily verified that the state-contingent consumptions are identical in these two consumption-information profiles.

Axiom 7 says that in this situation, the two consumption-information profiles should be ranked as indifferent.

**Theorem 5.1** *Suppose that the family* $\{\succeq_{F_1\times\cdots\times F_t}: F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 1\}$ *of conditional preferences satisfies Axioms 1-6. Then* $\succeq_{F_1\times\cdots\times F_t}$ *satisfies Strong Timing Indifference if and only if there exist a probability measure* $P(F_1 \times \cdots \times F_t, \cdot)$ *and a strictly increasing function* $\psi_{F_1\times\cdots\times F_t}$ *such that the certainty equivalent in Theorem 4.2 is given by*

$$\begin{aligned} & \mu(F_1 \times \cdots \times F_t, \tilde{V}(F_1 \times \cdots \times F_t, d)) \\ = \;& \psi^{-1}_{F_1\times\cdots\times F_t}\left(\int \psi_{F_1\times\cdots\times F_t}(\tilde{V}(F_1 \times \cdots \times F_t, d))\; dP(F_1 \times \cdots \times F_t)\right), \end{aligned} \tag{7}$$

*where* $\psi_{F_1\times\cdots\times F_t}$ *satisfies* $\psi_{F_1\times\cdots\times F_t}(0) = 0$ *and, for all* $\tilde{x} \in \mathcal{B}(R)$,

$$\beta\psi^{-1}_{F_1\times\cdots\times F_t}\left[\int \psi_{F_1\times\cdots\times F_t}[\tilde{x}]\; d\nu(F_1 \times \cdots \times F_t)\right] = \psi^{-1}_{F_1\times\cdots\times F_t}\left[\int \psi_{F_1\times\cdots\times F_t}[\beta\tilde{x}]\; d\nu(F_1 \times \cdots \times F_t)\right].$$

To explain the implication of Theorem 5.1 for updating, it is helpful to clarify first the role $\psi_{F_1\times\cdots\times F_t}$ in equation (7). Let $d = (\tilde{c}, \mathcal{F}) \in D_1$ be an one-period consumption-information profile. Then

$$\tilde{V}(F_1 \times \cdots \times F_t, d)(\omega) = u(c(\omega)).$$

Applying (7) yields

$$\mu(F_1 \times \cdots \times F_t, u(\tilde{c})) = \psi^{-1}_{F_1\times\cdots\times F_t}\left(\int \psi_{F_1\times\cdots\times F_t} \circ u(\tilde{c})\; d\nu(F_1 \times \cdots \times F_t)\right). \tag{8}$$

It should be clear from this expression that the more concave $\psi_{F_1\times\cdots\times F_t}$ is the more risk averse the certainty equivalent $\mu(F_1 \times \cdots \times F_t, \cdot)$ is.[11] Thus $\psi_{F_1\times\cdots\times F_t}$ can be viewed as the (state-contingent) risk aversion parameter of the conditional preference. That the aggregator $\mu(F_1 \times \cdots \times F_t)$ has such a (state-contingent) risk aversion parameter should not come as a surprise. After all, as explained in the beginning of this section, $\mu(F_1 \times \cdots \times F_t)$ carries the information not only about the updating rule but also about other behavioral characteristics of the conditional preferences. In order to focus on the updating rule encoded in the aggregator, however, we impose

[11] See Epstein and Zin (1989), for example.

**Assumption 5.2** *(**Time-State Invariant Risk Aversion**) $\psi_{F_1\times\cdots\times F_t}(x) = x$, for all $F_1 \times \cdots \times F_t \subset \Omega^t$ and $t \geq 1$ .*

Combining Theorems 4.2 and 5.1 and this assumption together we have

**Theorem 5.3** *The family $\{\succeq_{F_1\times\cdots\times F_t} : F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 1\}$ of conditional preferences satisfies Axioms 1-7 and Assumption 5.2 if and only if there exist probability measures $P(F_1 \times \cdots \times F_t, \cdot)$ and a strictly increasing function $u$ with $u(0) = 0$ such that*

$$V(F_1 \times \cdots \times F_t, (c, d)) = u(c) + \beta \int \tilde{V}(F_1 \times \cdots \times F_t, d)\ dP(F_1 \times \cdots \times F_t). \tag{9}$$

Turn now to updating. We show that the probability measures $P(F_1 \times \cdots \times F_t)$ in Theorems 5.1 and 5.3 are the (subjective) conditionals of some initial probability measure on the state space $\Omega^\infty$ and hence (9) is the familiar intertemporally additive expected utility. In other words, any family of conditional preferences that satisfies Axiom 1-7 and Assumption 5.2 has a belief component described by that initial probability measure and the belief updates according to the Bayesian updating rule. To that end, fix a filtration $\{\mathcal{F}_t\}_{t=1}^T$. In this paper, a filtration is defined as a sequence of partitions of $\Omega^T$ of the form: $\mathcal{F}_t = \{F_1 \times \cdots \times F_t \times \Omega^{T-t}\}$, $t = 1, \ldots, T$. We now construct a probability measure $P_T$ on $\Omega^T$ such that $P(F_1 \times \cdots \times F_t)$ are its conditionals. Let

$$P(\omega_1, \ldots, \omega_t, A) = P(F_1 \times \cdots \times F_t, A) \quad \text{for} \quad (\omega_1, \ldots, \omega_t) \in F_1 \times \cdots \times F_t \in \mathcal{F}_t, \tag{10}$$

for any $A \subset \Omega$ and define the probability measure $P_T$ on $\Omega^T$ by, for any $A \in (\Omega^T, \mathcal{F}_T)$,

$$P_T(A) = \int \left( \int \left( \cdots \left( \int 1_A P(\omega_1, \ldots, \omega_{T-1}, d\omega_T) \right) \cdots \right) P(\omega_1, d\omega_2) \right) P(d\omega_1). \tag{11}$$

Then for any $A \subset \Omega$ and $(\omega_1, \ldots, \omega_t) \in F_1 \times \cdots \times F_t \subset \Omega^t$,

$$P_T(A|\mathcal{F}_t)(\omega_1, \ldots, \omega_t) = P(\omega_1, \ldots, \omega_t, A) = P(F_1 \times \cdots \times F_t, A).$$

Thus $P(\omega_1, \ldots, \omega_t, A)$ are the conditionals of $P_T$ given the filtration $\{\mathcal{F}_t\}$. Let $(c_0, d_1) \in R_+ \times D_T$ by any $T$-period consumption-information profile such that the filtration embedded in $d_1$ is the same as $\{\mathcal{F}_t\}_{t=1}^T$.[12] Replacing $P(F_1 \times \cdots \times F_t, \cdot)$ in (9) with $P_T(\cdot|\mathcal{F}_t)$, we have

$$V(F_1 \times \cdots \times F_t, (c_t(\omega^t), d_{t+1}(\omega^t))) = u(c_t(\omega^t)) + \beta E^{P_T}[\tilde{V}(F_1 \times \cdots \times F_t, d_{t+1}(\omega^t))|\mathcal{F}_t],$$

[12] In the Appendix we show how the information filtration embedded in $d$ can be extracted.

for $\omega^t = (\omega_1, \ldots, \omega_t) \in F_1 \times \cdots \times F_t \subset \Omega^t$. Applying the law of iterated expectation,

$$V(c_0, d_1) = E^{P_T}\left(\sum_{t=0}^{T} \beta^t u(c_t)\right),$$

as is to be shown.

### 5.2 Dempster-Shafer Rule

The Dempster-Shafer rule for updating non-additive probability measures first appeared in Dempster (1967, 1968) and Shafer (1976, 1979) in statistics literature. Our axiomatization of the rule is based on the following timing indifference axiom.

**Axiom 8** *(***Timing Indifference***): Let $F_1 \times \cdots \times F_t \subset \Omega^t$, and $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_m\}$ be two partitions of $\Omega$. For all (two-period) consumption-information profiles of the form $(0, d_1)$ and $(0, d_1')$ with $d_1 = (\tilde{d}_1, \mathcal{F}^0)$, $\tilde{d}_1(\omega_1) = (0, d_{2i})$ if $\omega_1 \in A_i$ and $\tilde{d}_{2i}(\omega_2) = c_{ij}$ if $\omega_2 \in B_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$; $d_1' = (\tilde{d}_1', \mathcal{F}^0)$, $\tilde{d}_1'(\omega_1) = (0, d_{2j}')$ if $\omega_1 \in B_j$ and $\tilde{d}_{2j}'(\omega_2) = c_{ij}$ if $\omega_2 \in A_i$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, we have $(0, d_1) \succeq_{F_1 \times \cdots \times F_t} (0, d_2')$, provided $c_{i1} \leq \cdots \leq c_{im}$ and $c_{1j} \leq \cdots \leq c_{nj}$ for all $i$ and $j$.*

This is a weaker timing indifference axiom than Axiom 7. The first part of the definition describes two two-period consumption-information profiles $(0, d_1)$ and $(0, d_1')$ just as in Axiom 7 that have identical period-two state-contingent consumptions and zero consumptions at time 0 and 1, and whose timing of resolution of uncertainty is reversed of each other. The difference lies in the additional requirement on the ordering of period-two state-contingent consumptions. To understand what this additional condition asks for, consider two two-period risks as in Figure 4. Focus on the better-than sets. Recall that, given a random variable $\tilde{c}$ representing state-contingent consumptions and a number $z$ representing a level of consumption, the better-than set is given by $\{\omega : \tilde{c}(\omega) \geq z\}$, i.e., the set of states in which the realized consumption is less than $z$. Without further assumptions on $c_{ij}$, the better-than sets from the two two-period consumption-information profiles can different in general. In situations of uncertainty, better-than sets are the easiest to deal with, because for $z_1 < z_2$, $\{\omega : \tilde{c}(\omega) \geq z_1\} \supset \{\omega : \tilde{c}(\omega) \geq z_2\}$. That is, all better-than sets are nested according to set inclusion. As a minimal requirement, it seems sensible to require the

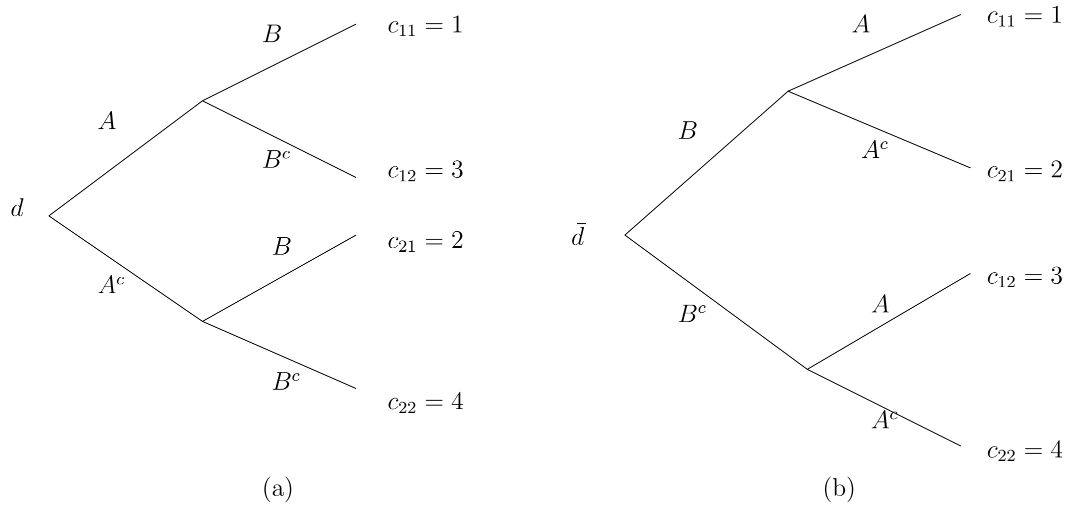


Figure 4: Timing Indifference

individual be able to rank the likelihood of nested events. This monotonicity requirement is far less imposing than requiring that the individual be able to rank the likelihood of two events which are not nested, which becomes even more imposing when the likelihood has to be additive as in a probability measure. Thus in comparing two consumption-information profiles when the timing of events is switched, it is desirable to control for the difference in the better-than sets. It turns out that the two consumption-information profiles have identical better-than sets if and only if they satisfy the additional condition in Timing Indifference Axiom. To see this, refer again back to Figure 4. Observe first that if $c_{ij}$ satisfies the condition in the axiom, then the better-than sets in any time-one subtree are $B^c$ and $\Omega$ in $d$, and $A^c$ and $\Omega$ in $d'$. Observe next that since the payoff in each branch of the lower subtree is greater than that in the corresponding branch of the upper subtree, the lower sub-trees in both $d$ and $d'$ have higher utility to the individual than the upper sub-trees. Suppose that the utility of the upper and lower sub-trees of $d$ are $V_1$ and $V_2$, respectively, with $V_1 < V_2$. Then, at time 0 and looking one-period ahead, the better-than sets for $d$ are $A^c = \{V \geq V_2\}$ and $\Omega = \{V \geq V_1\}$. Since, at any sub-tree of $d$, the one-period ahead better-than sets are $B^c$ and $\Omega$, the possible better-than sets in $d$ are $A^c$, $B^c$ and $\Omega$. It can be readily verified that $d'$ has the same better-than sets. Therefore, if $c_{i1} < c_{i2}$ and $c_{1j} < c_{2j}$ for $i, j = 1, 2$, then $d$ and $d'$ have not only the same state-contingent losses, but also the same better-than sets. The converse is also true.

**Remark**: It turns out that $d$ and $d'$ have the same collection of better-than sets can also be described by the notion of comonotonicity introduced by Schmeidler (1986). Let $\tilde{x}$ and $\tilde{y}$ be two one-period risks. $\tilde{x}$ and $\tilde{y}$ are said to be comonotonic if for all $\omega$ and $\omega' \in \Omega$ such that $\tilde{x}(\omega) \neq \tilde{x}(\omega')$ or $\tilde{y}(\omega) \neq \tilde{y}(\omega')$, we have $[\tilde{x}(\omega) - \tilde{x}(\omega')][\tilde{y}(\omega) - \tilde{y}(\omega')] > 0$. It is easy to see that $\tilde{x}$ 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 (2) $\{\omega \in \Omega : \tilde{x}(\omega) \geq x_i\} = \{\omega \in \Omega : \tilde{y}(\omega) \geq y_i\}$ for all $i$ (Schmeidler (1986)). Now in this terminology, $c_{i1} < c_{i2}$ and $c_{1j} < c_{2j}$ for $i, j = 1$, 2, which is the condition of Axiom 8 with strict inequalities, is equivalent to that the two one-period state-contingent consumptions of the first tree are comonotonic plus that the same is true for the second tree. The general case with weak inequalities can be viewed as the limit of the case with strict inequalities.

To relax individual's likelihood ranking to only nested events, we need a more general type of integrals—Choquet integrals (Choquet (1953/4)). Let $A_1, \ldots, A_n$ be a partition of the state space $\Omega$, $\tilde{x}$ be a random variable on $\Omega$ that takes values $x_1 < \cdots < x_n$ on the partition and $B_i = \cup_{j=i}^{n} A_j$, $i = 1, \ldots, n$, be the better-than sets. Let $\nu$ be a monotonic set function such that $\nu(\emptyset) = 0$ and $\nu(\Omega) = 1$. The Choquet integral of $\tilde{x}$ with respect to $\nu$ is defined as (Schmeidler (1986))

$$\int \tilde{x} \, d\nu = \sum_{i=1}^{n} [\nu(B_i) - \nu(B_{i+1})] x_i, \tag{12}$$

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

**Theorem 5.4** *Suppose that the family* $\{\succeq_{F_1 \times \cdots \times F_t} : F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 1\}$ *of conditional preferences satisfies Axioms 1-6. Then* $\succeq_{F_1 \times \cdots \times F_t}$ *satisfies Timing Indifference if and only if there exist a monotonic set function* $\nu(F_1 \times \cdots \times F_t, \cdot)$ *and a strictly increasing function* $\psi_{F_1 \times \cdots \times F_t}$*,*

$$\begin{aligned} & \mu(F_1 \times \cdots \times F_t, \tilde{V}(F_1 \times \cdots \times F_t, d)) \\ = \; & \psi^{-1}_{F_1 \times \cdots \times F_t} \left( \int \psi_{F_1 \times \cdots \times F_t}(\tilde{V}(F_1 \times \cdots \times F_t, d)) \, d\nu(F_1 \times \cdots \times F_t) \right), \end{aligned} \tag{13}$$

*where* $\psi_{F_1 \times \cdots \times F_t}$ *satisfies* $\psi_{F_1 \times \cdots \times F_t}(0) = 0$ *and, for all* $\tilde{x} \in \mathcal{B}(R)$*,*

$$\beta \psi^{-1}_{F_1 \times \cdots \times F_t} \left[ \int \psi_{F_1 \times \cdots \times F_t}[\tilde{x}] \, d\nu(F_1 \times \cdots \times F_t) \right] = \psi^{-1}_{F_1 \times \cdots \times F_t} \left[ \int \psi_{F_1 \times \cdots \times F_t}[\beta \tilde{x}] \, d\nu(F_1 \times \cdots \times F_t) \right].$$

The function $\psi_{F_1\times\cdots\times F_t}$ again has the interpretation as the risk aversion parameter of the certainty equivalent $\mu(F_1 \times \cdots \times F_t, \cdot)$.[13] We assume that $\psi_{F_1\times\cdots\times F_t}(x) = x$. Combining this assumption with Theorems 4.2 and 5.4 we have

**Theorem 5.5** *The family $\{\succeq_{F_1\times\cdots\times F_t}: F_1\times\cdots\times F_t \subset \Omega^t, t \geq 1\}$ of conditional preferences satisfies Axioms 1-6, 8 and Assumption 5.2 if and only if there exist monotonic set functions $\nu(F_1\times\cdots\times F_t, \cdot)$ and strictly increasing function $u$ such that $u(0) = 0$ and,*

$$V(F_1 \times \cdots \times F_t, (c, d)) = u(c) + \beta \int \tilde{V}(F_1 \times \cdots \times F_t, d)\ d\nu(F_1 \times \cdots \times F_t). \tag{14}$$

If we are to explore the updating rule encoded in the conditional preferences, it seems necessary that the conditional preferences embody a component of belief about the likelihood of events.[14] The formal definition will not be given of what is meant by a preference having a belief or liklihood evaluation component. The reader is referred to Machina and Schmeidler (1992), Epstein and Le Breton (1993) and Wang (1999) for the formal definitions. Restricting to the case of Choquet integrals as in (12), however, it is intuitively clear that if the certainty equivalent, $\mu(F_1\times\cdots\times F_t, \cdot)$, is represented by a Choquet integral as in (14), then $\nu(F_1 \times \cdots \times F_t, \cdot)$ represents the belief or likelihood evaluation of the conditional preferences. Therefore, our study of the Dempster-Shafer updating rule will be focused on $\nu(F_1 \times \cdots \times F_t, \cdot)$.

To derive the Dempster-Shafer updating rule from equation (14) of Theorem 5.5, we need to define an unconditional non-additive prior and examine how the conditional non-additive probabilities $\nu(F_1 \times \cdots \times F_t)$ are related to the unconditional prior. For simplicity, we examine the two-period case. Define the unconditional non-additive prior $\nu$ on $\Omega^2$ by, for any $A \times B \subset \Omega^2$,

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

where $d_1 = (\tilde{d}_1, \mathcal{F})$ with $\mathcal{F} = \{A, A^c\}$,

$$\tilde{d}_1 : \omega_1 \to (0, d_2(\omega_1)), \quad d_2(\omega_1, \omega_2) = \begin{cases} 1 & \text{if } \omega_1 \in A \text{ and } \omega_2 \in B \\ 0 & \text{otherwise} \end{cases}.$$

[13] See Epstein (1999) for separating risk aversion from uncertainty aversion for preferences that cannot be represented by an expected utility, and more generally, for non-probabilistically sophisticated preferences. The basic intuition is that if there is a subclass of events whose probabilities are objectively known to the decision maker and if, for acts involving only this subclass of events, the decision maker's preference can be represented by, say an expected utility, then any behavior characteristics implied must pertain to risk. Then it should be clear that when restricted to that subclass of events, the function $\psi_{F_1\times\cdots\times F_t}$ determines the risk aversion of the certainty equivalent just as in the case of Subsection 5.1. It is in this sense that we call it the risk aversion parameter.

[14] See Gilboa and Schmeidler (1993) for updating rules for general preferences which do not necessarily have a belief component.

This definitions looks more complicated than what it really is. The $(0, d_1)$ corresponds to the two-period tree in Figure 1 with $c_{21}$ equal to one and the rest all equal to zero. For sets in $\Omega^2$ of the form $\cup_{i=1}^n A_i \times B_i$ where $A_i$ are disjoint,

$$\nu(\cup_{i=1}^n A_i \times B_i) = V(0, d_1),$$

where $d_1 = (\tilde{d}_1, \mathcal{F})$ with $\mathcal{F} = \{A_1, \ldots, A_n\}$,

$$\tilde{d}_1 : \omega_1 \to (0, d_2(\omega_1)), \quad d_2(\omega_1, \omega_2) = \begin{cases} 1 & \text{if } \omega_1 \in A_i \text{ and } \omega_2 \in B_i \\ 0 & \text{otherwise} \end{cases}.$$

Now to see how the conditional non-additive probabilities $\nu(F, \cdot)$, $F \subset \Omega$, at time 1 are related to the unconditional non-additive probability, $\nu(\cdot)$, at time 0, let $A$ and $B \subset \Omega$. First it is readily verified that

$$\nu(A \times B) = \nu(A)\nu(A, B).$$

Thus

$$\nu(A, B) = \nu(A \times B)/\nu(A).$$

It is formally the same as the Bayesian rule. Next it is again readily verified that

$$\nu([A \times B] \cup A^c) = (1 - \nu(A^c))\nu(A, B) + \nu(A^c).$$

Thus

$$\nu(A, B) = \frac{\nu([A \times B] \cup A^c) - \nu(A^c)}{(1 - \nu(A^c))}.$$

This last expression is called the Dempster-Shafer updating rule.

The main drawback of the Dempster-Shafer updating rule, in the context of this paper, is perhaps its lack of consistency with the conditionals. While the conditionals appear in the preference representation as the Choquet integrators, the unconditional does not play a similar role, unless of course it is a probability measure. One may then question whether it really represents belief in the sense of Savage (1954) and Machina and Schmeidler (1992) if it does not play a role in the preference. This question leads naturally to the updating rule that we will study in the next subsection.

### 5.3 Generalized Bayesian Updating Rule

In this subsection we axiomatize a subclass of the conditional preferences characterized by Axioms 1-6 and 8 and examine an updating rule called the generalized Bayesian updating rule.[15]

The additional axiom that we now introduce is based on a notion of pessimism by Wakker (1999).

**Axiom 9** (**Pessimism**) *Let $\tilde{x}$ and $\tilde{y} \in \mathcal{B}(\Omega)$ be two one-period consumption profiles that assume*

$$x_1 \leq \cdots \leq x_{i_1} \leq \cdots \leq x_{i_2} \leq \cdots \leq x_N, \quad \text{and}$$
$$y_1 \leq \cdots \leq y_{i_1} \leq \cdots \leq y_{i_2} \leq \cdots \leq y_N$$

*on non-null events $A_1$, $A_2$, $\ldots$, $A_N$, respectively, such that $x_{i_1} = y_{i_1}$ and $x_{i_2} > y_{i_2}$. Let $\tilde{x}'$ and $\tilde{y}' \in \mathcal{B}(\Omega)$ be another two one-period consumption profiles that assume*

$$x_1 \leq \cdots \leq x_{i_2} \leq \cdots \leq x'_{i_2} \leq \cdots \leq x_N, \quad \text{and}$$
$$y_1 \leq \cdots \leq y_{i_2} \leq \cdots \leq y'_{i_2} \leq \cdots \leq y_N$$

*on $A_1$, $A_2$, $\ldots$, $A_N$, respectively, such that $x'_{i_2} = y'_{i_2}$. For all $F_1 \times \cdots \times F_t \subset \Omega^t$ and $t \geq 1$, if*

$$\mu(F_1 \times \cdots \times F_t, \tilde{x}) = \mu(F_1 \times \cdots \times F_t, \tilde{y}),$$

*then*

$$\mu(F_1 \times \cdots \times F_t, \tilde{x}') \geq \mu(F_1 \times \cdots \times F_t, \tilde{y}').$$

Intuitively, an individual is pessimistic if he assigns more likelihood to the lower outcomes. This rank-dependent assignment of likelihood is the intuition described in the axiom above. To see it, let $\tilde{x}$, $\tilde{y}$, $\tilde{x}'$ and $\tilde{y}'$ be as described in axiom. Figure 5 describes a case of four outcomes. Suppose that, as in Theorem 5.4,

$$\mu(F_1 \times \cdots \times F_t, \tilde{x}) = \sum_{i=1}^{n} [\nu(F_1 \times \cdots \times F_t, B_i) - \nu(F_1 \times \cdots \times F_t, B_{i+1})]\, u(x_i), \tag{15}$$

[15] Walley (1991) studies the rule for general (static) multi-prior expected utility functions, a class broader than we study here. In dynamic setting, the generalized Bayesian rule may cause inconsistency for the broader class of utility functions. See the discussion at the end of this subsection.

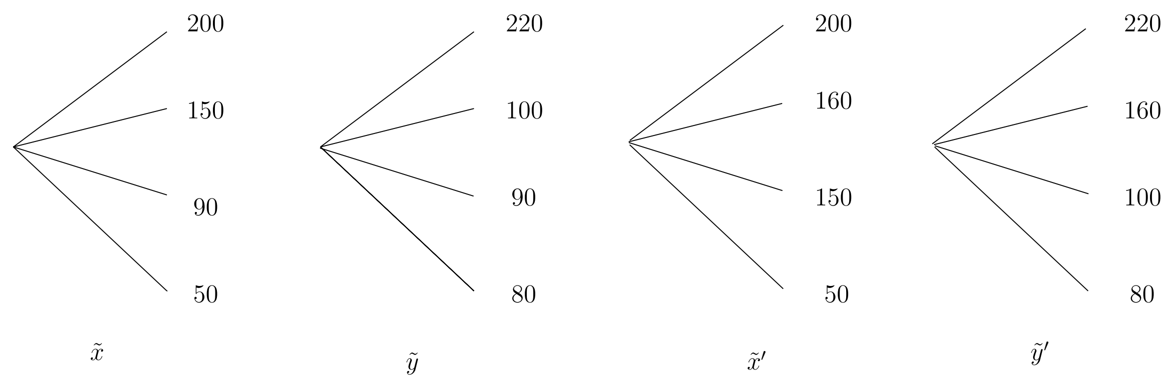


Figure 5: Pessimism

where $B_i = \bigcup_{j=i}^{n} A_j$, $i = 1, \ldots, N$, are the better-than sets. In this expression, borrowing from an intuition from expected utility, $\nu(F_1 \times \cdots \times F_t, B_i) - \nu(F_1 \times \cdots \times F_t, B_{i+1})$ can be viewed as the implied likelihood assigned to the outcome $\tilde{x} = x_i$. Using (15),

$$\mu(F_1 \times \cdots \times F_t, \tilde{x}) = \mu(F_1 \times \cdots \times F_t, \tilde{y})$$

is equivalent to

$$\sum_{i=1}^{n} [\nu(F_1 \times \cdots \times F_t, B_i) - \nu(F_1 \times \cdots \times F_t, B_{i+1})] [u(x_i) - u(y_i)] = 0.$$

Similarly,

$$\mu(F_1 \times \cdots \times F_t, \tilde{x}') \geq \mu(F_1 \times \cdots \times F_t, \tilde{y}')$$

is equivalent to

$$\sum_{i=1}^{n} [\nu(F_1 \times \cdots \times F_t, B_i) - \nu(F_1 \times \cdots \times F_t, B_{i+1})] [u(x_i') - u(y_i')] \geq 0.$$

A subtraction yields

$$\begin{aligned} & (\nu(F_1 \times \cdots \times F_t, B_{i_2}) - \nu(F_1 \times \cdots \times F_t, B_{i_2+1})) [u(x_{i_2}) - u(y_{i_2})] \\ \leq \; & (\nu(F_1 \times \cdots \times F_t, B_{i_1}) - \nu(F_1 \times \cdots \times F_t, B_{i_1+1})) [u(x_{i_2}) - u(y_{i_2})], \end{aligned}$$

which is true if and only if,

$$\nu(F_1 \times \cdots \times F_t, B_{i_2}) - \nu(F_1 \times \cdots \times F_t, B_{i_2+1}) \leq \nu(F_1 \times \cdots \times F_t, B_{i_1}) - \nu(F_1 \times \cdots \times F_t, B_{i_1+1}).$$

That is, when there are more higher, or fewer lower, outcomes than $x_{i_2}$ in $\tilde{x}'$ than in $\tilde{x}$, it is assigned higher likelihood in $\tilde{x}'$ than in $\tilde{x}$, which is exactly the intuitive definition of pessimism.

An axiom of optimism can be symmetrically defined. Since pessimism or optimism speaks directly to the likelihood assigned to the lower or higher outcomes, the ranking of the outcomes is important in the assignment of likelihood. This is partially reflected in the definition of Choquet integral where the better-than sets $B_i$, rather than the "level sets" $A_i$, are used. The following theorem is for the case of pessimism. As is seen in the theorem, the pessimism is captured by the min over a set of probability measures (see (i) of Theorem 5.8 also). There is also a version of the theorem for optimism by symmetry.

**Theorem 5.6** *Suppose that the family* $\{\succeq_{F_1\times\cdots\times F_t}: F_1 \times \cdots \times F_t \subset \Omega^t, t \geq 1\}$ *of conditional preferences satisfies Axioms 1-6 and 8 so that as in Theorem 5.4,* $\mu(F_1\times\cdots\times F_t, \tilde{V}(F_1\times\cdots\times F_t, d))$ *is given by (13). Then* $\succeq_{F_1\times\cdots\times F_t}$ *satisfies Axiom 9 if and only if there exists a closed and convex subset* $\mathbf{P}(F_1 \times \cdots \times F_t)$ *of probability measures on* $\Omega$ *such that, for all* $\tilde{x} \in \mathcal{B}(R_+)$,

$$\mu(F_1 \times \cdots \times F_t, \tilde{x}) = \psi^{-1}_{F_1\times\cdots\times F_t}\left(\min_{p\in\mathbf{P}(F_1\times\cdots\times F_t)} \int \psi_{F_1\times\cdots\times F_t}(\tilde{x})\ dp\right). \tag{16}$$

Combining Theorems 4.2 and 5.6 and Assumption 5.2 together we have

**Theorem 5.7** *The family* $\{\succeq_{F_1\times\cdots\times F_t}: F_1\times\cdots\times F_t \subset \Omega^t, t \geq 1\}$ *of conditional preferences satisfies Axioms 1-6, 8, 9 and Assumption 5.2 if and only if there exist closed and convex subsets* $\mathbf{P}(F_1 \times \cdots \times F_t)$ *of probability measures on* $\Omega$ *and strictly increasing function* $u$ *with* $u(0) = 0$ *such that*

$$V(F_1 \times \cdots \times F_t, (c, d)) = u(c) + \beta \min_{p\in\mathbf{P}(F_1\times\cdots\times F_t)} \int \tilde{V}(F_1 \times \cdots \times F_t, d)\ dp. \tag{17}$$

We now examine the implication of this theorem for updating. Suppose that, as in the theorem, each $\nu(F_1 \times \cdots \times F_t)$ corresponds to a convex and closed set $\mathbf{P}(F_1 \times \cdots \times F_t)$ of probability measures on $\Omega$. Let $\{\mathcal{F}_t\}$ be a filtration. Consider the state space $\Omega^\infty$ endowed with the $\sigma$-algebra $\mathcal{F}_\infty = \sigma(\mathcal{F}_t, t \geq 1)$. Fix a $s \geq 0$. Let $P(\omega_1, \ldots, \omega_t, \cdot)$, $t = s, \ldots$, be a sequence of $\mathcal{F}_t$-measurable selection from $\mathbf{P}(F_1 \times \cdots \times F_t)$, i.e.,

$$P(\omega_1, \ldots, \omega_t, \cdot) \in \mathbf{P}(F_1 \times \cdots \times F_t), \quad \text{if} \quad (\omega_1, \ldots, \omega_t) \in F_1 \times \cdots \times F_t \in \mathcal{F}_t. \tag{18}$$

For each $t \geq s$, define a conditional finite dimensional measure by

$$\begin{aligned} & P_t(\omega_1, \ldots, \omega_s, A_{s+1} \times \cdots \times A_t) \\ = & \int \cdots \int 1_{A_{s+1}\times\cdots\times A_t}(\omega_{s+1}, \ldots, \omega_t) P(\omega_1, \ldots, \omega_{t-1}, d\omega_t) \cdots P(\omega_1, \ldots, \omega_s, d\omega_{s+1}). \end{aligned} \quad (19)$$

This sequence of conditional finite dimensional measure is consistent. By Kolmogorov Theorem, there exists a probability measure $P_s(\omega_1, \ldots, \omega_s, \cdot)$ on $(\Omega^\infty, \mathcal{F}_\infty)$ such that its restriction to $\Omega^{t-s}$ is equal to $P_t(\omega_1, \ldots, \omega_s, \cdot)$. Let $\mathcal{P}_s(\omega_1, \ldots, \omega_s)$, $s \geq 0$, denote the set of all such measures. Let $\omega^s = (\omega_1, \ldots, \omega_s)$. By construction, if $P_s(\omega^s, \cdot) \in \mathcal{P}_s(\omega^s)$, then

$$P_s(\omega^s, \cdot | \mathcal{F}_{t+1})(\omega^{s+1}) \in \mathcal{P}_{s+1}(\omega^{s+1}).$$

That is, if $P$ is any probability measure in $\mathcal{P}_s(\omega^s)$, then its conditionals fall into $\mathcal{P}_{s+1}(\omega^{s+1})$. Conversely, if $P_{s+1}(\omega^{s+1})$ is a $\mathcal{F}_{s+1}$-measurable selection from $\mathcal{P}_{s+1}(\omega^{s+1})$ and $P_s(\omega^s) \in \mathbf{P}_s(F_1 \times \cdots \times F_s)$, then $P_{s+1}(\omega^{s+1})$ and $P_s(\omega^s)$ together via equation (19) with $t = s+1$ define a probability measure in $\mathcal{P}_s(\omega^s)$, which in particular implies that $\mathcal{P}_{s+1}(\omega^{s+1})$ consists only of conditionals from $\mathcal{P}_s(\omega^s)$. When the family $\mathcal{P}_t(\omega^t)$ satisfies such relationship, it is said to update according to the generalized Bayesian rule.

To simplify notation, we shall write $\mathcal{P}_s(\omega_1, \ldots, \omega_s)$ as $\mathcal{P}_s(F_1 \times \cdots \times F_s)$, for $(\omega_1, \ldots, \omega_s) \in F_1 \times \cdots \times F_s$, when it is more convenient. This is justified because if $(\omega_1, \ldots, \omega_s)$ and $(\omega'_1, \ldots, \omega'_s)$ are both in $F_1 \times \cdots \times F_s$, then $\mathcal{P}_s(\omega_1, \ldots, \omega_s) = \mathcal{P}_s(\omega'_1, \ldots, \omega'_s)$.

**Theorem 5.8** *Suppose that the conditions of Theorem 5.7 hold. Let $\{\mathcal{F}_t\}$ be a filtration on $\Omega^\infty$ and $\{\nu(F_1 \times \cdots \times F_t) : F_1 \times \cdots \times F_t \in \mathcal{F}_t, t \geq 1\}$ be the (sub-) family of conditional non-additive probability measures associated with the conditional preferences, $\{\mathbf{P}(F_1 \times \cdots \times F_t) : F_1 \times \cdots \times F_t \in \mathcal{F}_t, t \geq 1\}$ be the set of probability measures on $\Omega$ in Theorem 5.6, and $\{\mathcal{P}_t(F_1 \times \cdots \times F_t) : F_1 \times \cdots \times F_t \in \mathcal{F}_t, t \geq 1\}$ be the set of probability measures on $\Omega^\infty$ defined above through (18) and (19).*

*(i) For any $t$, $A \subset \Omega$, and $\omega^t \equiv (\omega_1, \ldots, \omega_t) \in \Omega^t$, if $\omega^t \in F_1 \times \cdots \times F_t$,*

$$\begin{aligned} \nu(F_1 \times \cdots \times F_t, A) &= \min \{P(A) : P \in \mathbf{P}(F_1 \times \cdots \times F_t)\} \quad (20) \\ &= \min \left\{ E^P \left[1_{F_1\times\cdots\times F_t\times A} | \mathcal{F}_t\right](\omega^t) : P \in \mathcal{P}_0 \right\}. \quad (21) \end{aligned}$$

*(ii) For any $d_1 \in D_T$ such that the filtration embedded in $d_1$ coincides with $\{\mathcal{F}_t\}$ and any $F_1 \times \cdots \times F_t \in \mathcal{F}_t$,*

$$\begin{aligned} V(F_1 \times \cdots \times F_t, d_{t+1}(\omega^t)) &= \min\left\{ E^P\left[\sum_{s=t+1}^{T} \beta^s u(c_s)\right] : P \in \mathcal{P}_t(\omega^t)\right\} \qquad (22) \\ &= \min_{p \in \mathcal{P}_t(\omega^t)} E^p \left( u(c_{t+1}(\omega^{t+1}) + \min_{P \in \mathcal{P}_{t+1}(\omega^{t+1})} E^P\left[\sum_{s=t+2}^{T} \beta^s u(c_s)\right]\right). \end{aligned}$$

This theorem is the foundation of the generalized Bayesian updating rule. In plain English, it says that if a family of conditional preferences satisfies the conditions of the theorem, then the conditional preferences evolve as if there is an initial set of probability measures on $(\Omega^\infty, \mathcal{F}_\infty)$ that represents the belief component of the preferences, and over time that belief is updated according to the generalized Bayesian rule.

From application's perspective, it is sometimes more convenient to begin with the specification of the initial set of priors. For example, to study the effect of learning on asset price/return dynamics in an environment with Knightian uncertainty, one may wish specify a multi-prior expected utility type of preference,

$$\min_{P \in \mathcal{P}} E^P\left[\sum_{t=0}^{\infty} \beta^s u(c_t)\right],$$

and examine how the set of priors, $\mathcal{P}$, evolves over time and its effect on the pricing kernel. The following theorem is complementary to Theorem 5.8 in that respect.

Let $\mathcal{P}$ be a closed and convex set of probability measures on a measurable space $(X, \mathcal{G})$. $\mathcal{P}$ is said to be strongly super-additive if the set function,

$$\gamma(A) = \min\{P(A) : P \in \mathcal{P}\}, \quad A \in \mathcal{G}$$

is convex, i.e., $\gamma(A) + \gamma(B) \le \gamma(A \cup B) + \gamma(A \cap B)$, for any $A$ and $B \in \mathcal{G}$.

**Theorem 5.9** *Let $\hat{\mathcal{P}}$ be a closed, convex and strongly supper-additive set of probability measures on $(\Omega^\infty, \mathcal{F}_\infty)$. For each $F_1 \times \cdots \times F_t \in \mathcal{F}_t$, denote by $\hat{\mathcal{P}}_t(F_1 \times \cdots \times F_t)$ the set of probability measures on $(\Omega^\infty, \mathcal{F}_\infty)$ obtained by conditioning the probability measures in $\hat{\mathcal{P}}$ on the event $F_1 \times \cdots \times F_t$. Let*

$\mathbf{P}(F_1 \times \cdots \times F_t)$ *be the set of probability measures on* $\Omega$ *given by*

$$\mathbf{P}(F_1 \times \cdots \times F_t) = \left\{ P : P(A) = \hat{P}(A \times \Omega^\infty), \hat{P} \in \hat{\mathcal{P}}_t(F_1 \times \cdots \times F_t) \right\}.$$

*Then both* $\hat{\mathcal{P}}_t(F_1 \times \cdots \times F_t)$ *and* $\mathbf{P}(F_1 \times \cdots \times F_t)$ *are closed, convex and strongly supper-additive. Construct* $\mathcal{P}_t(F_1 \times \cdots \times F_t)$ *via equations (18) and (19). Then*

$$\min_{P \in \mathcal{P}_t(\omega^t)} E^P \left[ \sum_{s=t+1}^{T} \beta^s u(c_s) \right] = \min_{p \in \mathcal{P}_t(\omega^t)} E^p \left[ u(c_{t+1}(\omega^{t+1})) + \beta \min_{P \in \mathcal{P}_{t+1}(\omega^{t+1})} E^P \left[ \sum_{s=t+2}^{T} \beta^s u(c_s) \right] \right].$$

Two implications of Theorems 5.8 and 5.9 are worth highlighting. (a) These two theorems can be viewed as a generalization of the law of iterated expectation in probability theory. They establish a equivalence between the static and recursive formulation of multi-prior expected utility when the set of priors is chosen appropriately, resembling that for intertemporally additive expected utility mentioned in Section 2. (b) They also point to a potential non-equivalence for general multi-prior expected utility functions: if we start with an initial set of priors, update with the generalized Bayesian rule to construct $\hat{\mathcal{P}}_t(F_1 \times \cdots \times F_t)$ and $\mathbf{P}(F_1 \times \cdots \times F_t)$ as in Theorem 5.9, and then use either $\hat{\mathcal{P}}_t(F_1 \times \cdots \times F_t)$ or $\mathbf{P}(F_1 \times \cdots \times F_t)$ formulate a recursive multi-prior expected utility, it may not be the same as the static formulation of multi-prior expected utility with the initial set of priors. If the equivalence of static and recursive is considered desirable, then such potential non-equivalence should raise the caution about the applicability of the generalized Bayesian rule to the broader class of multi-prior expected utility.

# 6 Potential Applications

While the main objective of our paper is to explore updating rules for non-Bayesian preferences, the results of this paper have wider applications beyond updating. In this section we will list a few such potential applications.

The immediate ones are the axiomatization of preferences used asset pricing models.

(**A**): Epstein and Wang (1994, 1995) develop an intertemporal asset pricing model where the agent's utility function is given by

$$V_t(c) = u(c_t) + \beta \min \left\{ \int V_{t+1}(c)\, dP : \; P \in \mathcal{P}(\omega) \right\},$$

where $\mathcal{P}(\omega)$ is a closed convex set of probability measures on $\Omega$, given the current state $\omega$. No axiomatic basis for the utility function was given. Theorem 5.7 provides an axiomatization for such multi-period multi-prior expected utility functions. Theorems 5.8, 5.9 and the discussion at the end of the last subsection address the consistency issue raised in Epstein and Wang (1994, p.293).

(**B**): Chen and Epstein (1999) provide a continuous-time asset pricing model that incorporate Knightian uncertainty. A special case of the utility function of the representative agent is given by

$$V_t(c) = \mathrm{essinf}_{P\in\mathcal{P}} E^P \left[ \int_t^T e^{-\beta(s-t)} u(c_s) ds \middle| \mathcal{F}_t \right],$$

where $\mathcal{P}$ is a set of Brownian measures that are absolutely continuous with each other. Theorems 5.8 and 5.9 provide an axiomatic foundation for such utility functions, except perhaps for the fact that Chen and Epstein model is in continuous time and allows for infinite states. Using Theorems 5.8 and 5.9, one can readily extend a version of Chen and Epstein (1999) to a model with learning in which uncertainty prevail even in the presence of learning and Chen and Epstein (1999) appears as the reduced form.

(**C**): More generally, Theorems 5.8 and 5.9 axiomatize the dynamic multi-prior expected utility preferences.

(**D**): Theorem 5.1 can be viewed as an axiomatic treatment of a class of preference with time-varying risk aversion, as in Barberis, Huang and Santos (1999) and Campbell and Cochrane (1999). Barberis, Huang and Santos (1999) develop their preference alternatively from the prospect and prior outcomes influence theories in psychology. Preference in Campbell and Cochrane (1999) is based on habit formation.

A potentially interesting application of the theory of this paper is in the area of learning and its impact on asset prices. The recent literature on the effect of learning on asset prices, such as Brennan (1997), Brennan and Xia (1998), assumes a probability framework. It would be interesting to study the same issue in the non-Bayesian framework as in this paper.

## A Proofs and Supporting Technical Details

**Extraction of information filtration**: Let $(c_0, d_1) \in R_+ \times D_T$ be a $T$-period profile. To extract the information filtration embedded in $d_1$, fix a $t \le T$ and $(\omega_1, \ldots, \omega_t)$. Let $d_1 = (\tilde{d}_1, \mathcal{F}^1)$. Since $\mathcal{F}^1 = \{F_1, \ldots, F_{n_1}\}$ is a partition, there is a unique $j_1$ such that $\omega_1 \in F_{j_1}$. Then there exist a $(c_{1,j_1}, d_{2,j_1}) \in R_+ \times D_{T-1}$ such that

$$\tilde{d}_1(\omega_1) = (c_{1,j_1}, d_{2,j_1}) = (c_{1,j_1}, (\tilde{d}_{2,j_1}, \mathcal{F}^2_{j_1})), \quad \mathcal{F}_{2,j_1} = \{F_{j_1,1}, \ldots, F_{j_1,n(j_1)}\}.$$

In turn, there is a unique $j_2 \le n(j_1)$ such that $\omega_2 \in F_{j_1,j_2}$. Continue inductively, there exists a unique sequence $F_{j_1}$, …, $F_{j_1,\ldots,j_t}$ such that $(\omega_1, \ldots, \omega_t) \in F_{j_1} \times \cdots \times F_{j_1,\ldots,j_t}$. The collection of such sets $F_{j_1} \times \cdots \times F_{j_1,\ldots,j_t}$ as $(\omega_1, \ldots, \omega_t)$ runs through $\Omega^t$ is a partition of $\Omega^t$, which naturally extends to a unique partition of $\Omega^T$. Denote this partition by $\mathcal{F}_t$. Then $\mathcal{F}_1$, …, $\mathcal{F}_T$ is the information filtration embedded in $d$.

**Proof of Theorem 3.1**: First we define mapping $\Theta$ from $D$ to $\mathcal{B}(R_+ \times D) \times \mathbf{F}$. Let $d \in D$. By definition

$$d = (d_1, d_2, \ldots) \in \Pi_{t=1}^{\infty} D_t, \qquad \text{with } f_t(d_{t+1}) = d_t.$$

Recall that for each $t \ge 1$, $d_t = (\tilde{d}_t, \mathcal{F})$ for some $\mathcal{F} \in \mathbf{F}$ common to all $t$ and $\tilde{d}_t \in \mathcal{B}(R_+ \times D_{t-1})$. ($D_0 = \emptyset$ by convention). Let

$$\tilde{d}_{t+1}(\omega) = (c(\omega), \bar{d}_t(\omega))$$

for $\omega \in \Omega$. Define

$$\Theta(d) = d' = (\tilde{d}', \mathcal{F}),$$

where $\tilde{d}' \in \mathcal{B}(R_+ \times D)$ is defined by

$$\tilde{d}'(\omega) = (c(\omega), \bar{d}(\omega)) = \left(c(\omega), (\bar{d}_1(\omega), \bar{d}_2(\omega))\right).$$

To ensure that $\Theta$ is well-defined, we need to show that $\bar{d}(\omega) \in D$ for each $\omega \in \Omega$. That is, for each $\omega \in \Omega$, $f_t(\bar{d}_{t+1}(\omega)) = \bar{d}_t(\omega)$. Fix $\omega \in \Omega$. By assumption,

$$f_t(d_{t+1}) = d_t. \tag{23}$$

The left hand side of this equation is

$$f_t(\tilde{d}_{t+1}, \mathcal{F}) = \left(f_t(\tilde{d}_{t+1}), \mathcal{F}\right), \quad \text{and} \quad f_t(\tilde{d}_{t+1})(\omega) = \left(c(\omega), f_{t-1}(\bar{d}_t(\omega))\right).$$

The right hand side of the equation is $d_t = (\tilde{d}_t, \mathcal{F})$ and $\tilde{d}_t(\omega) = (c(\omega), \bar{d}_{t-1}(\omega))$. Thus we have $f_{t-1}(\bar{d}_t(\omega)) = \bar{d}_{t-1}(\omega)$ as desired. Thus $\Theta$ is well-defined. Arguing in reverse order shows that $\Theta$ is one-to-one and onto. For continuity, suppose that $d^n = (\tilde{d}^n, \mathcal{F}^n) \to d = (\tilde{d}, \mathcal{F})$. Then $d^n_{t+1} \to d_{t+1}$ for each $t$ and $\mathcal{F}^n \to \mathcal{F}$. This is equivalent to $(c^n(\omega), \bar{d}^n_t(\omega)) \to (c(\omega), \bar{d}_t(\omega))$ for all $\omega \in \Omega$ and $\mathcal{F}^n \to \mathcal{F}$. Thus $(\bar{d}^n_1(\omega), \bar{d}^n_2(\omega), \ldots) \to (\bar{d}_1(\omega), \bar{d}_2(\omega), \ldots)$ and hence

$$(c^n(\omega), (\bar{d}^n_1(\omega), \bar{d}_2(\omega), \ldots)) \to (c(\omega), (\bar{d}_1(\omega), \bar{d}_2(\omega), \ldots)).$$

Therefore $\Theta$ is continuous. Arguing in reverse order establishes the continuity of $\Theta^{-1}$. ∎

**Proof of Theorem 4.1**: By Debreu (1954), each conditional preference $\succeq_{F_1 \times \cdots \times F_t}$ on $R_+ \times D$ can be represented by a numerical function $V(F_1 \times \cdots \times F_t, (c, d))$ on $R_+ \times D$. Due to Axiom 3, the ranking of deterministic consumption profiles are independent of past information histories. Thus we can normalize the numerical representation by monotonic transforms such that for any deterministic consumption profile $(c, d)$,

$$V(A_1 \times \cdots \times A_t, (c, d)) = V(B_1 \times \cdots \times B_t, (c, d)),$$

for any $A_1 \times \cdots \times A_t$ and $B_1 \times \cdots \times B_t$. Without loss of generality, we also normalized $V_t(F_1 \times \cdots \times F_t)$ such that $V(F_1 \times \cdots \times F_t, (c, 0)) = u(c)$ for a strictly increasing function $u : R_+ \to R$. By stationarity, we further normalize the conditional utility functions such that $V(F_1 \times \cdots \times F_t, (c, d)) = V(F_1 \times \cdots \times F_t \times \Omega^s, (c, d))$.

Let $t \geq 0$ and $F_1 \times \cdots \times F_t \subset \Omega^t$. Define $\mu(F_1 \times \cdots \times F_t) : \mathcal{B}(\Omega) \to R$ by

$$\mu(F_1 \times \cdots \times F_t, \tilde{x}) = V(F_1 \times \cdots \times F_t, d),$$

for any $d \in D$ such $\tilde{x} = \tilde{V}(F_1 \times \cdots \times F_t, d)$. We first show that $\mu(F_1 \times \cdots \times F_t)$ is well-defined. Suppose that $d = (\tilde{d}, \mathcal{F})$ and $d' = (\tilde{d}', \mathcal{G})$, where $\mathcal{F} = \{A_1, \ldots, A_n\}$ and $\mathcal{G} = \{B_1, \ldots, B_m\}$, are such that

$$\tilde{x} = \tilde{V}(F_1 \times \cdots \times F_t, d) = \tilde{V}(F_1 \times \cdots \times F_t, d').$$

Then we have, for all $i$ and $j$ and $\omega \in A_i \cap B_j$,

$$V[F_1 \times \cdots \times F_t \times A_i, (c(\omega), d_1(\omega))] = V[F_1 \times \cdots \times F_t \times B_j, (c'(\omega), d'_1(\omega))]. \tag{24}$$

For each $\omega$ there exist deterministic consumption profiles $y(\omega)$ and $z(\omega)$ (we do not need to specify the information profiles due to Axiom 3) such that

$$(c(\omega), d_1(\omega)) \sim_{F_1 \times \cdots \times F_t \times A_i} y(\omega) \quad \text{and} \quad (c'(\omega), d'_1(\omega)) \sim_{F_1 \times \cdots \times F_t \times B_j} z(\omega).$$

Thus, by Axiom 3 and the normalization,

$$
\begin{aligned}
& V[F_1 \times \cdots \times F_t \times A_i, (c(\omega), d_1(\omega))] && (25)\\
= \;& V[F_1 \times \cdots \times F_t \times A_i, y(\omega)] = V[F_1 \times \cdots \times F_t \times B_j, y(\omega)] && (26)\\
= \;& V[F_1 \times \cdots \times F_t \times B_j, (c'(\omega), d_1'(\omega))] && (27)\\
= \;& V[F_1 \times \cdots \times F_t \times B_j, z(\omega)] = V[F_1 \times \cdots \times F_t \times A_i, z(\omega)]. && (28)
\end{aligned}
$$

Using these equations, we claim that Consistency Axiom implies $V(F_1 \times \cdots \times F_t, d) = V(F_1 \times \cdots \times F_t, d')$. Suppose the contrary: $V(F_1 \times \cdots \times F_t, d) > V(F_1 \times \cdots \times F_t, d')$. By Consistency Axiom, (25)-(28) imply that

$$V(F_1 \times \cdots \times F_t, d) = V[F_1 \times \cdots \times F_t, \tilde{y}] > V[F_1 \times \cdots \times F_t, \tilde{z}] = V(F_1 \times \cdots \times F_t, d'),$$

which by Consistency again implies that for some $i$ and $\omega \in A_i$,

$$V[F_1 \times \cdots \times F_t \times A_i, y(\omega)] > V[F_1 \times \cdots \times F_t \times A_i, z(\omega)].$$

This contradicts (25)-(28).

An immediate implication of the above argument is:

**Axiom 10** *(Strong Consistency) For all $(c_i, d_i)$ and $(c_i', d_i') \in R_+ \times D$, $i = 1, \ldots, n$, and all partitions $\mathcal{F} = \{A_1, \ldots, A_n\}$ and $\mathcal{G} = \{B_1, \ldots, B_m\}$ of $\Omega$, if*

$$V[F_1 \times \cdots \times F_t \times A_i, (c_i, d_i)] \geq V[F_1 \times \cdots \times F_t \times B_j, (c_j', d_j')],$$

*for all $\omega \in A_i \cap B_j$, $i = 1, \ldots, n$ and $j = 1, \ldots, m$, then*

$$V\left[F_1 \times \cdots \times F_t, \left(\left[\sum_{i=1}^{n} (c_i, d_i) 1_{A_i}\right], \mathcal{F}\right)\right] \geq V\left[F_1 \times \cdots \times F_t, \left(\left[\sum_{j=1}^{m} (c_j', d_j') 1_{B_j}\right], \mathcal{G}\right)\right].$$

What the above argument demonstrates is that Strong Consistency is implied by Consistency and Deterministic Information Independence.

To show $\mu(F_1 \times \cdots \times F_t, \tilde{x})$ is a certainty equivalent, let $d = ((c, 0), \mathcal{F}^0)$, where $\mathcal{F}^0 = \{\Omega\}$ is the trivial partition. Observe that

$$
\begin{aligned}
V(F_1 \times \cdots \times F_t, d) &= \mu(F_1 \times \cdots \times F_t, \tilde{V}(F_1 \times \cdots \times F_t, d))\\
&= \mu(F_1 \times \cdots \times F_t, V(F_1 \times \cdots \times F_t \times \Omega, (c, 0)).
\end{aligned}
$$

On the other hand, by Stationarity and the normalization,

$$\begin{aligned} u(c) &= f[V(F_1 \times \cdots \times F_t \times \Omega, (0, [(c,0), \mathcal{F}^0])] = V(F_1 \times \cdots \times F_t, d) \\ u(c) &= V(F_1 \times \cdots \times F_t \times \Omega, (c,0)) = V(F_1 \times \cdots \times F_t, (c,0)). \end{aligned}$$

Since $c$ is arbitrary, we have $\mu(F_1 \times \cdots \times F_t, c) = c$, which is the property (a) of certainty equivalent. For property (b), let

$$\tilde{x} = \tilde{V}(F_1 \times \cdots \times F_t, d)$$

so that for $\omega \in A_i$,

$$\tilde{x}(\omega) = V(F_1 \times \cdots \times F_t \times A_i, (c(\omega), d_1(\omega))).$$

Similarly, let $\tilde{y}$ be such that

$$\tilde{y}(\omega) = V(F_1 \times \cdots \times F_t \times B_i, (c'(\omega), d_1'(\omega))).$$

If $\tilde{x} \geq \tilde{y}$, then, by Strong Consistency,

$$\begin{aligned} & \mu(F_1 \times \cdots \times F_t, \tilde{x}) = \mu(F_1 \times \cdots \times F_t, \tilde{V}(F_1 \times \cdots \times F_t, (\tilde{d}, \mathcal{F}))) \\ = \; & V(F_1 \times \cdots \times F_t, (\tilde{d}, \mathcal{F})) \geq V(F_1 \times \cdots \times F_t, (\tilde{d}', \mathcal{G})) \\ = \; & \mu(F_1 \times \cdots \times F_t, \tilde{V}(F_1 \times \cdots \times F_t, (\tilde{d}', \mathcal{G}))) = \mu(F_1 \times \cdots \times F_t, \tilde{y}). \end{aligned}$$

Define for any $c \in R_+$ and any real number $v$,

$$W_t(F_1 \times \cdots \times F_t, (c, v)) = V(F_1 \times \cdots \times F_t, (c, d)),$$

for any $d$ such that $v = V(F_1 \times \cdots \times F_t, d)$. We show that the function $W_t$ is well-defined. If $d$ and $d'$ are such that $v = V(F_1 \times \cdots \times F_t, d) = V(F_1 \times \cdots \times F_t, d')$, then it follows from Risk Separability that

$$V(F_1 \times \cdots \times F_t, (c, d)) = V(F_1 \times \cdots \times F_t, (c, d')).$$

Continuity and monotonicity of $W_t$ are straightforward.

We now show that $W_t(F_1 \times \cdots \times F_t, c, v)$ is independent of $F_1 \times \cdots \times F_t$ and $t$. Let $(c, d)$ and $(c', d')$ be two deterministic consumption profiles. By the normalization,

$$V(A_1 \times \cdots \times A_t, (c, d)) = V(B_1 \times \cdots \times B_t, (c, d)).$$

Thus

$$W_t(A_1 \times \cdots \times A_t, (c, v)) = W_t(B_1 \times \cdots \times B_t, (c, v)).$$

That is, $W_t(F_1 \times \cdots \times F_t, (c, d))$ is independent of $F_1 \times \cdots \times F_t$. That $W_t$ is independent of $t$ follows from Stationarity. ∎

**Proof of Theorem 4.2**: This theorem follows from Koopmans (1960) or Gorman (1968). ∎

**Proof of Theorem 5.1**: The proof is exactly identical to that of Theorem 5.4 with only one change: under Strong Timing Indifference, property (A6) is replaced by

(A6') $d_A(m_{\mathcal{F}}(x_1, \ldots, x_n), m_{\mathcal{F}}(y_1, \ldots, y_n)) \sim_{F_1 \times \ldots \times F_t} d_{\mathcal{F}}(m_A(x_1, y_1), \ldots, m_A(x_n, y_n))$.

Under (A6'), by Theorem 1 of Nakamura (1990), $\nu(F_1 \times \cdots \times F_t, \cdot)$ in the proof of Theorem 5.4 is in fact a probability measure. ∎

**Proof of Theorem 5.4**: Fix $F_1 \times \cdots \times F_t$. First we introduce some simplifying notations. Let $\mathcal{F} = \{A_1, \ldots, A_n\}$ be a partition of $\Omega$. Denote by

$$d_{\mathcal{F}}(x_1, \ldots, x_n)$$

the (one-period) consumption profile whose current consumption is zero and whose consumption at time 1 in state $A_i$ is $x_i$. Note that for one-period consumptions, updating in the forthcoming period is irrelevant. So if $d = (\tilde{d}, \mathcal{G}) \in D$ and $\tilde{d} = d_{\mathcal{F}}(x_1, \ldots, x_n)$, we will simply write $d$ as $d_{\mathcal{F}}(x_1, \ldots, x_n)$. Let $A \subset \Omega$. For the partition $\mathcal{F} = \{A, A^c\}$ we will also write $d_{\mathcal{F}}(x, y)$ simply as $d_A(x, y)$. For partitions $\mathcal{F} = \{A_1, \ldots, A_n\}$ and $\mathcal{F}_i = \{B_{i1}, \ldots, B_{im}\}$, denote by

$$d_{\mathcal{F}} = (d_{\mathcal{F}_1}(x_{11}, \ldots, x_{1m}), \ldots, d_{\mathcal{F}_n}(x_{n1}, \ldots, x_{nm})) = [(d_{\mathcal{F}_1}(x_{11}, \ldots, x_{1m}), \ldots, d_{\mathcal{F}_n}(x_{n1}, \ldots, x_{nm})), \mathcal{F}^0]$$

the two-period consumption profiles whose current and time 1 consumptions are zero and there is no updating at time 1 (because $\mathcal{F}^0$ is trivial). Consider the restriction of $\succeq_{F_1 \times \cdots \times F_t}$ on the space $\mathcal{B}(R)$ of one-period consumption-information profiles. To simplify notations, write $V[F_1 \times \ldots \times F_t, (0, d_\pi(x_1, \ldots, x_n))]$ as $V[F_1 \times \ldots \times F_t, d_{\mathcal{F}}(x_1, \ldots, x_n)]$ when no confusion arises.

We shall first verify that if Axioms 1-6 hold, then the ordering has the following properties:

(A1) For each $\tilde{x} \in \mathcal{B}(R)$, there are $x$ and $y \in R$ such that $x \succ_{F_1 \times \ldots \times F_t} \tilde{x} \succ_{F_1 \times \ldots \times F_t} y$.

(A2) If $d_A(y,z) \succeq_{F_1\times\ldots\times F_t} \tilde{x} \succeq_{F_1\times\ldots\times F_t} d_A(x,z)$, then $\tilde{x} \sim_{F_1\times\ldots\times F_t} d_A(a,z)$ for some $a \in R$.

(A3) If $A$ is not null[16] and $\{x,y\} \leq z$, then $x \leq y$ if and only if $d_A(y,z) \succeq_{F_1\times\ldots\times F_t} d_A(x,z)$; if $A$ is not universal[17] and $\{x,y\} \geq z$, then $x \leq y$ if and only if $d_A(z,y) \succeq_{F_1\times\ldots\times F_t} d_A(z,x)$.

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

(A5) Every strictly bounded standard sequence is finite.[18]

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

$$d_A(m_{\mathcal{F}}(x_1,\ldots,x_n), m_{\mathcal{F}}(y_1,\ldots,y_n)) \sim_{F_1\times\ldots\times F_t} d_{\mathcal{F}}(m_A(x_1,y_1),\ldots,m_A(x_n,y_n)),$$

where $m_{\mathcal{F}}(x_1,\ldots,x_n)$ is the constant risk that is ranked the same as $d_{\mathcal{F}}(x_1,\ldots,x_n)$. That is,

$$V[F_1\times\ldots\times F_t,(0,m_{\mathcal{F}}(x_1,\ldots,x_n))] = V[F_1\times\ldots\times F_t,(0,d_{\mathcal{F}}(x_1,\ldots,x_n))],$$

or

$$m_{\mathcal{F}}(x_1,\ldots,x_n) = u^{-1}\left[V[F_1\times\ldots\times F_t,(0,d_{\mathcal{F}}(x_1,\ldots,x_n))]/\beta\right]. \tag{29}$$

Note that $m_{\mathcal{F}}(x_1,\ldots,x_n)$ is the constant "loss" that is realized *one period from now.*

Properties (A1), (A2), (A3) and (A4) follow from consistency and continuity. 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

$$V[F_1\times\ldots\times F_t, d_A(x_n,p)] = V[F_1\times\ldots\times F_t, d_A(x_{n+1},q)]. \tag{30}$$

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 $x_0 \leq b$. Taking limit in (30) and applying the continuity of $V$, we have $V[F_1\times\ldots\times F_t, d_A(x_0,p)] = V[F_1\times\ldots\times F_t, d_A(x_0,q)]$, 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.

[16] An event $A \subset \Omega$ is null if for all $x$, $y$, $z \in R$, $d_A(x,z) \sim_{F_1\times\ldots\times F_t} d_A(y,z)$.

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

[18] 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 $\{a,b\} \leq a_i$ and $d_A(a,a_i) \sim_{F_1\times\ldots\times F_t} d_A(b,a_{i+1})$ for all $i \in N$, or $a_i \leq \{a,b\}$ and $d_A(a_i,a) \sim_{F_1\times\ldots\times F_t} d_A(a_{i+1},b)$ for all $i \in N$.

For (A6), we show first that the certainty equivalent operator, $\mu[F_1 \times \ldots \times F_t]$, satisfies

$$\mu[F_1 \times \ldots \times F_t, \beta\tilde{x}] = \beta\mu[F_1 \times \ldots \times F_t, \tilde{x}] \tag{31}$$

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

$$V[F_1 \times \ldots \times F_t, (c, d)] = u(c) + \beta\mu\left(F_1 \times \ldots \times F_t, \tilde{V}[F_1 \times \ldots \times F_t, d]\right), \tag{32}$$

and

$$V[F_1 \times \ldots \times F_t, d_{\mathcal{F}}(x_1, \ldots, x_n)] = \beta\mu[F_1 \times \ldots \times F_t, u(\tilde{x})]. \tag{33}$$

Now let $\tilde{x} \in \mathcal{B}(R)$ be an random variable that assumes values $x_1 < \cdots < x_n$ on $A_1, \ldots, A_n$ respectively. Let $\mathcal{F} = \{A_1, \ldots, A_n\}$. Let

$$d_1 = (\tilde{d}, \mathcal{F}^0), \quad \tilde{d}(\omega) = d_{\mathcal{F}}(u^{-1}(x_i), \ldots, u^{-1}(x_i)), \quad \text{if } \omega \in A_i,$$

and

$$d_2 = (\tilde{d}, \mathcal{F}^0), \quad \tilde{d}(\omega) = d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_n)), \quad \text{if } \omega \in A_i,$$

Observe that

$$\begin{aligned} & V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_1)), \ldots, d_{\mathcal{F}}(u^{-1}(x_n), \ldots, u^{-1}(x_n)))) \\ = \; & V(F_1 \times \ldots \times F_t, (0, d_1)) = \beta\mu(F_1 \times \ldots \times F_t, \tilde{V}(F_1 \times \ldots \times F_t, d_1)) \end{aligned}$$

where in the second equality we have used (32), noting that the argument of $V(F_1 \times \ldots \times F_t, \cdot)$ is a two-period consumption-information profile. For $\omega \in A_i$,

$$\begin{aligned} & \tilde{V}(F_1 \times \ldots \times F_t, d_1)(\omega) = V(F_1 \times \ldots \times F_t \times \Omega, d_{\mathcal{F}}(u^{-1}(x_i), \ldots, u^{-1}(x_i))) \\ = \; & \beta\mu(F_1 \times \ldots \times F_t \times \Omega, d_{\mathcal{F}}(u^{-1}(x_i), \ldots, u^{-1}(x_i))) = \beta\mu(F_1 \times \ldots \times F_t \times \Omega, x_i) = \beta x_i, \end{aligned}$$

where the third equality is by (33). By a similar argument,

$$\begin{aligned} & V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_n)), \ldots, d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_n)))) \\ = \; & V(F_1 \times \ldots \times F_t, (0, d_2)) = \beta\mu(F_1 \times \ldots \times F_t, \tilde{V}(F_1 \times \ldots \times F_t, d_2)) \end{aligned}$$

and for $\omega \in A_i$,

$$\begin{aligned} & \tilde{V}(F_1 \times \ldots \times F_t, d_2)(\omega) = V(F_1 \times \ldots \times F_t \times \Omega, d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_n))) \\ = \; & V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_n))) = \beta\mu(F_1 \times \ldots \times F_t, \tilde{x}), \end{aligned}$$

where the second equality is by Stationarity. Now

$$
\begin{aligned}
& \beta\mu(F_1 \times \ldots \times F_t, \beta\tilde{x}) \\
= \; & V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_1)), \ldots, d_{\mathcal{F}}(u^{-1}(x_n), \ldots, u^{-1}(x_n)))) \\
= \; & V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_n)), \ldots, d_{\mathcal{F}}(u^{-1}(x_1), \ldots, u^{-1}(x_n)))) \\
= \; & \beta\mu(F_1 \times \ldots \times F_t, \beta\mu(F_1 \times \ldots \times F_t, \tilde{x})) = \beta^2\mu(F_1 \times \ldots \times F_t, \tilde{x}).
\end{aligned}
$$

where the second equality is by Timing Indifference, and last equality from $\mu(F_1 \times \ldots \times F_t)$ being a certainty equivalent. Thus (31) is shown.

Now, let

$$
\tilde{f}(\omega) = \begin{cases} V(F_1 \times \ldots \times F_t \times \Omega, d_{\mathcal{F}}(x_1, \ldots, x_n)) & \text{if } \omega \in A, \\ V(F_1 \times \ldots \times F_t \times \Omega, d_{\mathcal{F}}(y_1, \ldots, y_n)) & \text{otherwise.} \end{cases}
$$

Then

$$
\begin{aligned}
& V(F_1 \times \ldots \times F_t, d_A(m_{\mathcal{F}}(x_1, \ldots, x_n), m_{\mathcal{F}}(y_1, \ldots, y_n))) \\
= \; & V\Big(F_1 \times \ldots \times F_t, d_A(u^{-1}\left[V(F_1 \times \ldots \times F_t \times \Omega, d_{\mathcal{F}}(x_1, \ldots, x_n))/\beta\right], \\
& \qquad u^{-1}\left[V(F_1 \times \ldots \times F_t \times \Omega, d_{\mathcal{F}}(y_1, \ldots, y_n))/\beta\right]\Big) \\
= \; & \mu[F_1 \times \ldots \times F_t, \tilde{f}] = \frac{1}{\beta}[\beta\mu(F_1 \times \ldots \times F_t, \tilde{f})] \\
= \; & \frac{1}{\beta}V(F_1 \times \ldots \times F_t, d_A(d_{\mathcal{F}}(x_1, \ldots, x_n), d_{\mathcal{F}}(y_1, \ldots, y_n))) \\
= \; & \frac{1}{\beta}V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(d_A(x_1, y_1), \ldots, d_A(x_n, y_n))) \\
= \; & V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(u^{-1}[V(F_1 \times \ldots \times F_t \times \Omega, d_A(x_1, y_1))/\beta], \\
& \qquad \ldots, u^{-1}[V(F_1 \times \ldots \times F_t \times \Omega, d_A(x_n, y_n))/\beta])) \\
= \; & V(F_1 \times \ldots \times F_t, d_{\mathcal{F}}(m_A(x_1, y_1)), \ldots, m_A(x_n, y_n))))
\end{aligned}
$$

where first equality is by (29), the second equality is by (33), the fourth equality is by (32), the fifth equality is by Timing Indifference, the sixth equality is by (32), and the last equality is by (29). Thus (A6) holds.

Now by Theorem 1 of Nakamura (1990), there exist a strictly monotonic function $g_{F_1 \times \cdots \times F_t}$ and a monotonic set function $\nu(F_1 \times \cdots \times F_t)$ such that $d_{\mathcal{F}}(x_1, \ldots, x_n) \succeq d_{\mathcal{G}}(y_1, \ldots, y_m)$ if and only if

$$\int g_{F_1\times\cdots\times F_t}[u(\tilde{x})]\ d\nu(F_1\times\cdots\times F_t) \geq \int g_{F_1\times\cdots\times F_t}[u(\tilde{y})]\ d\nu(F_1\times\cdots\times F_t).$$

Thus, for any $\tilde{x}$ and $\tilde{y}\in\mathcal{B}(R)$,

$$V(F_1\times\cdots\times F_t, d_{\mathcal{F}}(x_1,\ldots,x_n)) \geq V(F_1\times\cdots\times F_t, d_{\mathcal{G}}(y_1,\ldots,y_m))$$

if and only if

$$\int g_{F_1\times\cdots\times F_t}[u(\tilde{x})]\ d\nu(F_1\times\cdots\times F_t) \geq \int g_{F_1\times\cdots\times F_t}[(\tilde{y})]\ d\nu(F_1\times\cdots\times F_t),$$

which implies that there exists a strictly increasing function $\psi_{F_1\times\cdots\times F_t}$ such that

$$\psi_{F_1\times\cdots\times F_t}(V(F_1\times\cdots\times F_t, d_{\mathcal{F}}(x_1,\ldots,x_n))/\beta) = \int g_{F_1\times\cdots\times F_t}(u(\tilde{x}))\ d\nu(F_1\times\cdots\times F_t).$$

However, $V(F_1\times\cdots\times F_t, d_{\mathcal{F}}(x_1,\ldots,x_n))/\beta = \mu(F_1\times\cdots\times F_t, u(\tilde{x})]$ by (33). Thus

$$\psi_{F_1\times\cdots\times F_t}(\mu(F_1\times\cdots\times F_t, u(\tilde{x}))) = \int g_{F_1\times\cdots\times F_t}(u(\tilde{x}))\ d\nu(F_1\times\cdots\times F_t).$$

Since $\mu(F_1\times\cdots\times F_t,\cdot)$ is a certainty equivalent, the above equation implies that

$$\psi_{F_1\times\cdots\times F_t}[u(x)] = g_{F_1\times\cdots\times F_t}(u(x)).$$

Returning to $\mu(F_1\times\cdots\times F_t)$, we have

$$\mu(F_1\times\cdots\times F_t,\tilde{y}) = \psi^{-1}_{F_1\times\cdots\times F_t}\int \psi_{F_1\times\cdots\times F_t}(\tilde{y})\ d\nu(F_1\times\cdots\times F_t). \quad \blacksquare$$

**Proof of Theorem 5.6**: This theorem follows from Theorem 7.3 of Wakker (1999) and the standard representation theorem for Choquet integration with respect to a convex capacity. See for example Anger (1977). ∎

**Proof of Theorem 5.8**: (i) The first expression follows from Theorems 5.4 and 5.6. The second follows from the construction of $\mathbf{P}$.

(ii) We prove the first equation for the case of $T=2$. The more general case is the same, but involves more notation. Let $d_1=(\tilde{d}_1,\mathcal{F}_1)$ with $\mathcal{F}_1=\{F_1,\ldots,\mathcal{F}_n\}$. By Theorem 5.5,

$$\begin{aligned} V(d_1) &= \int \tilde{V}(d_1)(\omega_1)\nu(d\omega_1) \\ \tilde{V}(d_1)(\omega_1) &= V(F_i,(c_1(\omega_1),d_2(\omega_1))) = u(c_1(\omega_1)) + \beta\int c_2(\omega_1,\omega_2)\nu(F_1,d\omega_2), \quad \text{if } \omega_1\in F_i. \end{aligned}$$

By Theorem 5.7, these two expression can be written as

$$\begin{aligned} V(d_1) &= \min\left\{\int \tilde{V}(d_1)(\omega_1)P(d\omega_1) : P \in \mathbf{P}_0\right\} \\ \tilde{V}(d_1)(\omega_1) &= V(F_i,(c_1(\omega_1),d_2(\omega_1))) = u(c_1(\omega_1)) + \beta \min\left\{\int c_2(\omega_1,\omega_2)P(d\omega_2) : P \in \mathbf{P}(F_i)\right\}, \end{aligned}$$

where $\mathbf{P}_0$ is the closed convex subset of probability measures that is associated with the conditional non-additive probability $\nu$ at time zero. Since both $\mathbf{P}_0$ and $\mathbf{P}(F_i)$ are closed, there exist $P^* \in \mathbf{P}_0$ and $P^*(\omega_1) \in \mathbf{P}(F_i)$ for $\omega_1 \in F_i$ such that

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

Let $P$ be the probability measure on $\Omega^2$ associated with $P^*$ and $P^*(\omega_1)$ through equation (19). Then $P \in \mathbf{P}$. Thus, the LHS of (22) is greater than the RHS. But the LHS of (22) is always less than the RHS. Therefore, the equality holds.

The second equation follows from the first and Theorem 5.7. ∎

**Proof of Theorem 5.9**: By the remark in Wasserman and Kadane (1990) or note 11 of Walley (1991, p.551), if $\mathcal{P}$ is closed, convex and strongly supper-additive, so is $\mathcal{P}_t(F_1 \times \cdots \times F_t)$ for any $F_1 \times \cdots \times F_t$. The strong supper-additivity of $\mathbf{P}(F_1 \times \cdots \times F_t)$ follows from that of $\mathcal{P}_t(F_1 \times \cdots \times F_t)$. The closedness and convexity of $\mathbf{P}(F_1 \times \cdots \times F_t)$ is straightforward. The rest follows from Theorem 5.8. ∎

## References


Anderson, E., L. Hansen and T. Sargent (1999): "Robustness, Detection and the Price of Risk," working paper, University of Chicago.

Anger, B. (1977): "Representation of Capacities," *Math. Ann.* 229, pp.245-258.

Barberis, N., M. Huang and T. Santos (1999): "Prospect Theory and Asset Prices," working paper, University of Chicago.

Brennan, M. (1997): "The Role of Learning in Dynamic Portfolio Decisions," working paper, UCLA.

Brennan, M. and Y. Xia (1998): "Stock Price Volatility, Learning, and the Equity Premium," working paper, UCLA.

Camerer, C.F. and M. Weber (1992): "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty*, 8, pp.167-196.

Campbell, J. and J. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy,* 107, pp.205-251.

Chen, Z. and L. Epstein (1999): "Ambiguity, Risk and Asset Returns in Continuous Time," working paper, University of Rochester.

Chew, S.H and L. Epstein (1989): "The Structure of Preferences and Attitudes towards the Timing of the Resolution of Uncertainty," *International Economic Review*, 30, 1, pp.103-117.

Chew, S.H and L. Epstein (1991): "Recursive Utility under Uncertainty," in Kahn, A. and Yannelis, N. (eds) *Equilibrium Theory with an Infinite Number of Commodities.* Berlin Heidelberg New York: Springer, pp.352-369.

Cotter, K. (1986):"Similarity of Information and Behavior with a Pointwise Convergence Topology," *Journal of Mathematical Economics* 15, pp.25-38.

Debreu, G. (1954): "Representation of a Preference Ordering by a Numerical Function," in *Decision Processes,* R. Thrall, C. Coombs and R. David, eds., New York, John Wiley.

Dempster, A.P. (1967): "Upper and Lower Probabilities Induced from a Multivalued Mapping," *Annals of Mathematical Statistics* 38, pp.525-539.

Dempster, A.P. (1968): "A Generalization of Bayesian Inference," *Journal of Royal Statistics Society, Series B.* 30, pp.205-247.

Dow, J. and S. Werlang (1994): "Laws of Large Numbers for Non-additive Probabilities", working paper, London Business School.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics,* 75, pp.643-669.

Epstein, L. (1999): "A Definition of Uncertainty Aversion," *Review of Economic Studies,* 66, pp.579-608.

Epstein, L. and S. Zin (1989): "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: a Theoretical Framework," *Econometrica*, 57, pp.937-969.

Epstein, L. (1993) and M. Le Breton: "Dynamically Consistent Beliefs Must Be Bayesian," *Journal of Economic Theory,* 61, 1, pp.1-22.

Epstein L. and T. Wang (1994): "Intertemporal Asset Pricing under Knightian Uncertainty," *Econometrica*, 62, pp.283-322.

Epstein L. and T. Wang (1995): "Uncertainty, Risk-Neutral Measures and Security Price Booms and Crashes," *Journal of Economic Theory*, 67, pp.40-80.

Gilboa, I. (1987): "Expected Utility Theory with Purely Subjective Non-Additive Probabilities," *Journal of Mathematical Economics,* 16, pp.65-88.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility Theory with Non-Unique Prior," *Journal of Mathematical Economics,* 18, pp.141-153.

Gilboa, I. and D. Schmeidler (1993): "Updating Ambiguous Beliefs," *Journal of Economic Theory,* 59, 1, pp.33-49.

Gorman, W.M. (1968): "The Structure of Utility Functions," *Review of Economic Studies,* Vol.35, pp.367-390.

Hansen, L., T. Sargent and T. Tallarini (1999): "Robust Permanent Income and Pricing," forthcoming *Review of Economic Studies.*

Huber, Peter (1981): *Robust Statistics*. New York: John Wiley.

Koopmans, T. (1960): "Stationary Ordinal Utility and Impatience," *Econometrica*, 28, pp.287-309.

Kreps, D. and E. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, pp.185-200.

Machina, M. and D. Schmeidler (1992): "A More Robust Definition of Subjective Probability," *Econometrica*, 60, pp.745-780.

Nakamura, Y. (1990): "Subjective Expected Utility with Non-additive Probabilities on Finite State Spaces," *Journal of Economic Theory*, 51, pp.346-366.

Sarin, R. and P. Wakker (1998): "A Simple Axiomatization of Nonadditive Expected Utility," *Econometrica,* 60, pp.1255-1272.

Sarin, R. and P. Wakker (1998): "Revealed Likelihood and Knightian Uncertainty," *Journal of Risk and Uncertainty,* 16, pp.223-250.

Savage, L. (1954): *Foundations of Statistics,* Wiley: New York.

Schmeidler, D. (1986): "Integral Representation without Additivity," *Proceedings of the American Mathematical Society,* 97, pp.255-261.

Schmeidler, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica,* 57, pp.571-587.

Shafer, G. (1976): *A Mathematical Theory of Evidence*, University of Princeton Press: New Jersey.

Shafer, G. (1979): "Allocations of Probability," *Annals of Probability*, 7, pp.827-839.

Skiadas, C. (1997): "Subjective Probability under Additive Aggregation of Conditional Preferences," *Journal of Economic Theory*, 76, pp.242-271.

Skiadas, C. (1998): "Recursive Utility and Preferences for Information," *Economic Theory*, 12, pp.293-312.

Wakker, P. (1999): "Testing and Characterizing Properties of Non-Additive Measures through Violations of the Sure-Thing Principle," working paper, University of Tilburg.

Walley, P. (1991): *Statistical Reasoning with Imprecise Probabilities,* Chapman and Hall.

Wang, T. (1999): "A Class of Dynamic Risk Measures," working paper, University of British Columbia.

Wasserman, L. and J. Kadane (1990): "Bayes' Theorem for Choquet Capacities," *Annals of Statistics*, 18, pp.1328-1339.