Learning May Increase Perceived Uncertainty:

A Model of Confidence Erosion under Knightian Uncertainty*

by

Kiyohiko G. Nishimura

Policy Board
Bank of Japan

and

Hiroyuki Ozaki

Faculty of Economics
Keio University

May 19, 2008

Abstract

This paper shows that new information may increase perceived uncertainty if uncertainty is characterized as Knightian uncertainty. We consider ε-contamination of confidence: an economic agent is (1-ε) $\times$ 100% certain that uncertainty she faces is characterized by a particular stochastic model, but that she has a fear that, with ε $\times$100% chance, her conviction is wrong and she is left ignorant about the "true" model. In this situation, if the economic agent follows Bayesian procedure or its variant, which is considered as rational in the theory of economics, her confidence erodes after having new information, if initial confidence is not strong enough.

*A former version of this paper was circulated under the title of "Economics of Self-feeding Fear." The research reported here was conducted when the first author was at the Faculty of Economics of the University of Tokyo before joining the Policy Board of the Bank of Japan. This study is partially supported by a grant from Economic and Social Research Institute, Cabinet Office, Government of Japan. However, the views expressed here are strictly personal and in no way represent those of the Bank of Japan, Economic and Social Research Institute, the University of Tokyo, or Keio University. The authors are grateful to Atsushi Kajii and audiences at the University of Tokyo for their helpful comments.

1. Introduction and Summary

Economic agents including policy makers face various uncertainties when they make decision. Here we must distinguish two different kinds of uncertainty. The first one, which is often called *risk*, is formulated as a known probability distribution with possibly unknown parameters learnable from past experience of, say, stock prices and the GDP growth rate in the near future. The second one is another kind of uncertainty, which deserves the name of true fundamental uncertainty. This is uncertainty that cannot be reduced to a known distribution, often called *Knightian uncertainty* in recognition of the writing of Frank Knight. Not only are they uncertain about the future value of stock prices and GDP growth in a known probability distribution, but also they do not have clear knowledge of their probability distribution itself. To put it differently, they do not have clear confidence in the "stochastic model" that they often use to describe economic activities in the real world.

One way to cope with the uncertain world is to gather information about unknown economic conditions, and to learn about underlying parameters from it. Thus, learning, which is often formulated as Bayesian learning, is considered to reduce the magnitude of uncertainty. In fact, if the uncertainty agents face is risk, Bayesian learning is shown to reduce the magnitude of uncertainty appropriately defined.

The purpose of this paper is to show this is not always the case if the uncertainty is Knightian.[1] In contrast to risk, Knightian uncertainty is characterized as a set of distributions, instead of a single distribution. Hence, learning is characterized by an update process of the set of distributions after each of random sampling. In this paper, ε-cotamination of confidence is taken as an example of Knightian uncertainty. Suppose that an economic agent is $(1-\varepsilon)\times 100\%$ certain that uncertainty she faces is characterized by a particular dynamic stochastic model, but that she has fear that, with $\varepsilon \times 100\%$ chance, her conviction is wrong and she is left ignorant about the "true" model. We call this situation ε-contamination of confidence and $1-\varepsilon$ can be

[1]In the statistics literature, Seidenfeld and Wasserman (1993) presented necessary and sufficient conditions that dilation of uncertainty (which corresponds to erosion of confidence discussed later) take place in the case of the "no-narrowing" Bayes rule if uncertainty is formulated as a set of distributions (that is, Knightian uncertainty). However, these conditions are hard to explain and thus they are difficult to apply in economic problems of our interest. The contribution of this paper is, firstly, to show that confidence erosion can happen under relatively simple, not-so-implausible conditions in the case of ε-contamination, and secondly, to present sufficient conditions under which dilation still occurs in the "range-narrowing" maximum-likelihood rule.

taken as a measure of confidence. It is a convenient characterization of fundamental uncertainty economic agents face[2] and it has axiomatic foundation[3].

Suppose further that the economic agent follows Bayesian procedure or its variant, which is considered rational in the theory of economics.[4] Then, we show that her confidence erodes (i.e., the degree of confidence $1 - \varepsilon$ decreases) after having a new observation, if the initial degree of confidence in the stochastic model is not strong compared with the new observation's "informational value"[5]. The reason of confidence erosion is that new information brings in a new dynamic possibility which is not previously considered seriously.

This paper is organized as follows. In Section 2, we present a simple example of confidence erosion in the learning model developed by Rothschild (1974). A general model of confidence erosion is presented in Sections 3 through 5. In Section 3, we formulate stochastic environment and the decision maker's objective function, and define "dilation of uncertainty", that is, a phenomenon that "new observation reduces confidence." Section 4 defines and examines two "sensible" updating rules: the maximum-likelihood and multi-prior Bayesian[6]. Section 5 contains the main results: In the case of ε-contamination, if the initial degree of confidence is not strong compared with a new observation's "informational value", dilation of uncertainty occurs regardless of whether the maximum-likelihood rule or the Bayesian one is utilized.

2. An Example: Rothschild's Learning Model

Let us consider a case considered by Rothschild (1974), which has been one of the most well-known examples in the economics of learning. An unemployed worker is searching for a

[2]The concept of ε-contamination defined in this paper is used in Nishimura and Ozaki (2004) who examine search behavior under the Knightian uncertainty.

[3]Nishimura and Ozaki (2006) show that if economic agents' behavior is in concord with several axioms, then their perceived uncertainty can be characterized as ε-contamination of confidence. Their axioms are not at all singular. Thus, their results suggest that ε-contamination of confidence may commonly be observed.

[4]We consider the maximum-likelihood rule and a multi-prior Bayesian one since they seem intuitive and sensible. After having new observation, the maximum-likelihood rule chooses, among all distributions in the set characterizing the Knightian uncertainty, those that put the highest probability on the occurrence of an actual observation, and updates the chosen distributions by using the Bayes rule. The multi-prior Bayesian rule updates all distributions in the set by using the Bayes rule. Both rules are based on Bayesian ideas.

[5]See Theorem 2 of Section 5. The exact meaning of "informational value" will be clarified later in this paper. The result is surprising particularly in the case of the maximum-likelihood "update" rule, in which substantial "narrowing" of the range of probability measures seems to occur after obtaining a new observation through the maximum-likelihood principle.

[6]In fact, to our knowledge, there is no other update rule that has been discussed as widely and intensively as these rules in the literature.

job. Different firms offer different wages. She takes a job interview sequentially and gets one wage quotation each time. To make analysis simple and apparent, we consider a two-period model.[7]

In Rothschild's model, the unemployed worker is risk-neutral, and contemplates her optimal policy in terms of expected income. She does not know the wage distribution, and learns about the distribution from the wage observation. In particular, the unemployed worker assumes that the wage-offer distribution is a multinomial distribution with a support of $W = \{w_1, \dots, w_k\} \subseteq \mathbb{R}$. However, she does not know probability p_i of a particular w_i.

It is then assumed that the unemployed worker thinks that the probability of p_i's is distributed according to a Dirichlet distribution over a set $\mathcal{P}$,

$$\mathcal{P} = \left\{ \boldsymbol{p} = (p_1, \dots, p_k) \in \mathbb{R}^k \,\middle|\, (\forall i)\ p_i > 0 \text{ and } \sum_{i=1}^{k} p_i = 1 \right\},$$

whose density function is

$$f(\boldsymbol{p}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} p_1^{\alpha_1 - 1} \cdots p_{k-1}^{\alpha_{k-1}-1} (1 - \textstyle\sum_{i=1}^{k-1} p_i)^{\alpha_k - 1},$$

where $\boldsymbol{\alpha} \in \mathbb{R}^k_{++}$ is a parameter vector and $\Gamma(\cdot)$ is the gamma function. The mean of each marginal, p_i $(i = 1, \dots, k)$, is given by

$$E[p_i] = \frac{\alpha_i}{\sum_{\ell=1}^{k} \alpha_\ell}. \tag{1}$$

Suppose that the decision-maker observed a wage offer w_i in the first period. Then, by DeGroot (1970, p.174, Theorem 1), the posterior distribution of w_j's, updated by Bayes' rule upon observing w_i, turns out to be the Dirichlet distribution with the parameter vector

$$\boldsymbol{\alpha_i'} = (\alpha_1, \dots, \alpha_{i-1}, \alpha_i + 1, \alpha_{i+1}, \dots, \alpha_k). \tag{2}$$

The learning process of the unemployed worker has the following interpretation. Suppose that the agent has a "prior" wage distribution which is multinomial with parameters $\boldsymbol{p}^0 = (p_1^0, \dots, p_k^0)$ over the wage offer in the second period, where for each j, p_j^0 is a probability of w_j's occurrence and it is defined by $p_j^0 \equiv E[p_j]$. Then, from (1), her "prior" second-period

[7] Rothschild (1974) considers an infinite horizon. We deviate from his work in this respect, in order to make our argument simple and transparent.

expected wage income will be

$$\sum_{j=1}^{k} w_j p_j^0 = \sum_{j=1}^{k} w_j E[p_j] = \frac{\sum_{j=1}^{k} w_j \alpha_j}{\sum_{\ell=1}^{k} \alpha_\ell} \,. \tag{3}$$

Then, the worker gets the wage offer w_i for some i in the first period. Upon observing w_i, she revises her prior distribution, $\boldsymbol{p}^0$, to the posterior one, $\boldsymbol{p_i'} = (p_1'(w_i), \ldots, p_k'(w_i))$, where $p_j'(w_i) = E[p_j|w_i]$. Then, with some calculation[8], her "prior" second-period expected income (3) is revised to the "posterior" second-period expected income *given* the observation of the first period:

$$\sum_{j=1}^{k} w_j p_j'(w_i) = \sum_{j=1}^{k} w_j E[p_j|w_i] = \frac{\sum_{j\neq i} w_j \alpha_j}{\sum_{\ell=1}^{k} \alpha_\ell + 1} + \frac{w_i(\alpha_i + 1)}{\sum_{\ell=1}^{k} \alpha_\ell + 1} = \frac{\sum_{j=1}^{k} w_j \alpha_j + w_i}{\sum_{\ell=1}^{k} \alpha_\ell + 1} \,.$$

The unemployed worker then uses this posterior second-period expected wage income in contemplating her optimal strategy: whether to stop searching now or to go on to the next period.

The above example of Rothschild assumes that the unemployed worker is *perfectly* certain that the wage distribution is a multinomial one and the distribution of the wage-occurrence probability is a Dirichlet distribution. However, there is no *a priori* rationale that the worker assumes this particular combination.

Let us now deviate from Rothschild's specification, and consider a case in which the unemployed worker is *almost* certain that the true distribution is the multinomial distribution with the known $\boldsymbol{p}^0 = (p_1^0, \ldots, p_k^0)$, but that she is not completely certain about that. Thus, she fears that, with $\varepsilon \times 100\%$ probability, the true distribution is different from this multinomial distribution, and moreover, she may not have any information about the true parameter values if $\boldsymbol{p}^0$ is not the true one. In other words, the unemployed worker is almost $((1-\varepsilon) \times 100\%)$ certain about the wage distribution but has a $\varepsilon \times 100\%$ fear that she is wrong and left ignorant about the true distribution.[9] In this setting, it is natural to call ε as a measure to gauge

[8]Letting $E[\cdot|w_i]$ be the posterior mean, (1) and the paragraph containing (2) imply that

$$(\forall j \neq i)\;\; E[p_j|w_i] = \frac{\alpha_j}{\sum_{\ell=1}^{k} \alpha_\ell + 1} \quad \text{and} \quad E\left[p_i|w_i\right] = \frac{\alpha_i + 1}{\sum_{\ell=1}^{k} \alpha_\ell + 1} \,.$$

[9]Nishimura and Ozaki (2006) show that, if the decision maker's behavior is consistent with certain plausible axioms, her decision making is characterized as maximizing the minimum of her expected utility over multiple priors that are characterized by ε-contamination of confidence explained in the text. (Their representation result is in the Choquet framework, but the basic argument can easily be interpreted in this way.) The set of axioms they presented are an extension of Schmeidler (1989)'s axioms.

ignorance, or equivalently, $(1-\varepsilon)$ as the degree of confidence.[10]

Since the unemployed worker is risk-neutral and thus maximizes expected income, her situation is the same as that of a decision-maker facing (a restricted form of) ε*-contamination* of the distribution.[11] Formally, let $\varepsilon \in (0,1)$ and let $\mathcal{P} \times \mathcal{P}$ be a set of pairs of $\boldsymbol{p}$ in the first period and $\boldsymbol{p'}$ in the second period[12]:

$$\mathcal{P} \times \mathcal{P} = \left\{ \, (\boldsymbol{p}, \boldsymbol{p'}) \,\middle|\, \boldsymbol{p}, \boldsymbol{p'} \in \mathcal{P} \right\} .$$

Then, the ε-contamination of $(\boldsymbol{p}, \boldsymbol{p'}) = (\boldsymbol{p}^0, \boldsymbol{p}^0)$ considered in this section, $\left\{(\boldsymbol{p}^0, \boldsymbol{p}^0)\right\}^{\varepsilon}$, is

$$\left\{(\boldsymbol{p}^0, \boldsymbol{p}^0)\right\}^{\varepsilon} = \left\{ \, (1-\varepsilon)\,(\boldsymbol{p}^0, \boldsymbol{p}^0) + \varepsilon(\boldsymbol{q}, \boldsymbol{q'}) \,\middle|\, (\boldsymbol{q}, \boldsymbol{q'}) \in \mathcal{P} \times \mathcal{P} \right\} .$$

We now examine what happens to the degree of confidence when new observation arrives. However, in order to proceed with our analysis, we should specify the decision maker's objective function and update procedure of priors in the case of the Knightian uncertainty or multiple probability distributions.

Firstly, it is known (see Schmeidler (1989) and Gilboa and Schmeidler (1989)) that in multiple-probability cases of this kind, if the decision-maker's behavior is in accordance with certain sensible axioms, then her behavior is characterized as being *uncertainty-averse*: when the decision-maker evaluates her position, she uses probability corresponding to the "worst" scenario. Following this line of argument, we assume that the unemployed worker is uncertainty-averse. Secondly, we assume that the decision maker uses Bayesian procedure to multiple priors by applying it to all probabilities in $\left\{(\boldsymbol{p}^0, \boldsymbol{p}^0)\right\}^{\varepsilon}$.[13]

Let us now consider this Bayesian process. Let (w_i, w'_j) denote an event that the first-period wage observation is w_i and the second-period one is w_j. Then, the probability of this

[10] ε-contamination has been widely used in statistics literature to specify a set of measures (see, for example, Berger, 1985). There, the sensitivity of an estimator to the assumed prior distribution ($(\boldsymbol{p}^0, \boldsymbol{p}^0)$ in the text) is the main concern in the context of Bayesian estimation problems. While we also specify a set of measures or Knightian uncertainty by ε-contamination, our main concern is not robustness of a specfic estimator but the set itself, which reflects the decision-maker's lack of confidence.

[11] In this section, we restrict contamination, $(\boldsymbol{q}, \boldsymbol{q'})$, to be a product probability measure to make a proof simple and intuitive. However, in general, contamination is not restricted to a product probability measure but it is allowed to be any probability measure defined over the product space. We consider these general cases in the formal analysis of Sections 3 to 5. See in particular Eq (13) in Section 5.

[12] In other words, $\mathcal{P} \times \mathcal{P}$ is the set of all product measures of the form: $\boldsymbol{p} \otimes \boldsymbol{p'}$ when we regard $\boldsymbol{p}$ and $\boldsymbol{p'}$ as probability measures on W. In the text, we denote $\boldsymbol{p} \otimes \boldsymbol{p'}$ by $(\boldsymbol{p}, \boldsymbol{p'})$.

[13] The case of maximum-likelihood rule will be discussed in Sections 3 and 4. Here we analyze the Bayesian rule since it is more tractable than the maximum-likelihood rule.

event measured by one element, $(1-\varepsilon)(\boldsymbol{p}^0, \boldsymbol{p}^0) + \varepsilon(\boldsymbol{q}, \boldsymbol{q'})$, of $\left\{(\boldsymbol{p}^0, \boldsymbol{p}^0)\right\}^\varepsilon$ is

$$\Pr\left(w_i, w_j'\right) = (1-\varepsilon)\, p_i^0 p_j^0 + \varepsilon q_i q_j'$$

and a corresponding second-period marginal probability is

$$\Pr\left(w_j'\right) = (1-\varepsilon)\, p_j^0 + \varepsilon q_j' \,.$$

And hence, the set of the prior second-period probabilities is given by

$$\left\{ (1-\varepsilon)\, \boldsymbol{p}^0 + \varepsilon \boldsymbol{q'} \;\middle|\; \boldsymbol{q'} \in \mathcal{P} \right\} . \tag{4}$$

Suppose as before that w_i is observed. The unemployed worker updates each element in the set of the prior second-period probabilities to their posterior, so that we have

$$\Pr\left(w_j' \middle| w_i\right) = \frac{\Pr(w_i, w_j')}{\Pr(w_i)} = \left(1-\varepsilon'\right) p_j^0 + \varepsilon' q_j' \tag{5}$$

where

$$\varepsilon' = \frac{\varepsilon q_i}{(1-\varepsilon)\, p_i^0 + \varepsilon q_i} \,. \tag{6}$$

The set of corresponding posteriors is the set of all these probabilities obtained by varying $\boldsymbol{q}$ and $\boldsymbol{q'}$.

Let

$$\bar{\varepsilon}' = \frac{\varepsilon}{(1-\varepsilon)\, p_i^0 + \varepsilon} \,.$$

Then, we have

$$\left(1-\varepsilon'\right) p_j^0 + \varepsilon' q_j' = \left(1-\bar{\varepsilon}'\right) p_j^0 + \bar{\varepsilon}' \left(\left(1 - \frac{\varepsilon'}{\bar{\varepsilon}'}\right) p_j^0 + \frac{\varepsilon'}{\bar{\varepsilon}'} q_j' \right) .$$

Since $\bar{\varepsilon}' \geq \varepsilon'$ and that $\mathcal{P}$ is the set of all conceivable $\boldsymbol{q'}$, we know

$$\left(1 - \frac{\varepsilon'}{\bar{\varepsilon}'}\right) \boldsymbol{p}^0 + \frac{\varepsilon'}{\bar{\varepsilon}'} \boldsymbol{q'} \in \mathcal{P} \,.$$

Consequently, the set of corresponding posteriors is a subset of

$$\left\{ \left(1-\bar{\varepsilon}'\right) \boldsymbol{p}^0 + \bar{\varepsilon}' \boldsymbol{q'} \;\middle|\; \boldsymbol{q'} \in \mathcal{P} \right\} . \tag{7}$$

Conversely, take one element of (7), $(1-\bar{\varepsilon}')\,\boldsymbol{p}^0+\bar{\varepsilon}'\bar{\boldsymbol{q}}'$. Then, it is always possible to find $\varepsilon' \in [0,\bar{\varepsilon}']$ (and ultimately $\boldsymbol{q} \in \mathcal{P}$) and $\boldsymbol{q}' \in \mathcal{P}$ satisfying that $(1-\bar{\varepsilon}')\,\boldsymbol{p}^0+\bar{\varepsilon}'\bar{\boldsymbol{q}}' = (1-\varepsilon')\,\boldsymbol{p}^0+\varepsilon'\boldsymbol{q}'$ and then $q_i \in [0,1]$ satisfying (6). Since the set of posterior distributions corresponding to (4) is characterized by (5) and (6) with $\boldsymbol{q}$ and $\boldsymbol{q}'$ varying (see the paragraph containing (5) and (6)), (7) is a subset of that set. Thus, all things considered, we conclude that the set of posteriors after w_i is observed is equal to (7).

Let us now compare the set of priors (4) and that of posteriors (7). The latter shows that the unemployed worker is now $(1-\bar{\varepsilon}')\times 100\%$ certain about $\boldsymbol{p}^0$: her fear that her conviction is wrong now increased from ε to $\bar{\varepsilon}'$ ($\bar{\varepsilon}' > \varepsilon$ as far as $p_i^0 < 1$). *The decision-maker's degree of confidence is decreased after the observation of w_i.* Note that there is no "surprise" justifying a decrease in confidence.

It is clear that dynamic feature of the Knightian uncertainty plays a crucial role to obtain this confidence erosion. Here, the Knightian uncertainty is dynamic in the sense that the decision-maker thinks that the true distribution may change over time.[14] Loosely speaking, the argument in the second to the last paragraph reveals that a new observation makes the decision-maker "find" a combination of probabilities over two periods leading to a posterior probability that is not considered by her before (probability outside her prior beliefs).

In this section, we have presented an example that new information reduces confidence of the decision-maker about the uncertain world. However, the argument we have employed is based on a specific example of a mutli-nomial distribution a la Rothschild. Thus, one may question the generality of the results. In the rest of this paper, we extend our model to a general setting and to show that the same result holds in general cases.

In the next section, we reformulate the basic problem of this section in a general framework of behavior under a dynamic Knightian uncertainty. We consider two updating rules commonly utilized in the literature for this kind of problems: maximum-likelihood and (generalized) Bayesian. The formal exposition of these updating rules is given in Section 4. In Section

[14] If the wage distribution of the first period is perfectly correlated with the one of the second period, then we cannot have self-feeding fear. Perfect correlation means that if the decision maker gets wage w_i in the first period then she gets w_i in the second period. In this case, uncertainty is completely resolved in the first period. However, so long as correlation is not perfect, we have a possibility of self-feeding fear.

5, we show that, under gerenral conditions, the same results as in this section holds true for general probability measures and for both updating rules in general ε-contamination cases: new information reduces the decision-maker's confidence.

3. The Two-Period Dynamic Model of Knightian Uncertainty

In order to make a formal analysis, we have to set up a dynamic model in which the decision-maker have multiple probability measures about her economic environment. In the following, we first specify stochastic environment and consider an update rule. We then incorporate the update rule into the decision-maker's objective function to represent evolution of her view of the world in the form of multiple probability measures over stochastic environment. We exclusively consider a two-period model. An extension to multi-period cases is straightforward but notationally cumbersome.

In the following, notations are somewhat involved, because of the complexity introduced by a dynamic Knghtian uncertainty: the decision-maker does not have perfect confidence not only about a "true" probability measure each period but also how it changes over periods. Consequently, the model, including the objective function and updating rules, is specified in an entire dynamic structure of the deicision-maker's stochastic environment.

Information Structure. Let W be a state space for each single period and let $\Omega = W \times W$ be the whole state space. A generic element of Ω is denoted by (w_1, w_2).

The information structure, which represents the basis of the decision-maker's view of the world, is exogenously given by a filtration $\mathfrak{F} = \langle \mathcal{F}_t \rangle_{t=0,1,2}$. We assume that $\mathcal{F}_0 = \{\phi, \Omega\}$, that $\mathcal{F}_1$ is represented by a finite partition of Ω of the form: $\langle E_i \times W \rangle_i$ for some finite partition $\langle E_i \rangle_{i=1}^m$ of W, and that $\mathcal{F}_2$ is represented by a finite partition of Ω of the form: $\langle E_i \times F_j \rangle_{i,j}$ for some finite partition $\langle F_j \rangle_{j=1}^n$ of W. Clearly, it holds that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2$. We further assume that $m \geq 2$.

We abuse a notation to denote by $(W, \langle E_i \rangle_i)$ the measurable space on which the algebra is generated by the partition $\langle E_i \rangle_i$ and we denote the set of all probability measures on it by $\mathcal{M}(W, \langle E_i \rangle_i)$. Similar notations apply to other cases in obvious manners.

Given $p \in \mathcal{M}(\Omega, \mathcal{F}_2)$, we denote by $p|_1$ its restriction on $(\Omega, \mathcal{F}_1)$. Although $p|_1$ is formally a measure on Ω, it can be naturally regarded as the one on $(W, \langle E_i \rangle_i)$ and in that case, $p|_1(\cdot) = p(\cdot \times W)$. Thus viewed, $p|_1$ can be considered as the *first-period marginal probability measure* of p. Similarly, we define the *second-period marginal probability measure*, $p|_2$, of p. That is, let $p|_2 \in \mathcal{M}(W, \langle F_j \rangle_j)$ be defined by $p|_2(\cdot) = p(W \times \cdot)$.

The decision-maker's view of the world is represented by not a single probability measure but a set of probability measures (Knightian uncertainty). Formally, we assume that the decision-maker's *Knightian uncertainty* is represented by $\mathcal{P} \subseteq \mathcal{M}(\Omega, \mathcal{F}_2)$.

Finally, let us now define "priors." Given $\mathcal{P} \subseteq \mathcal{M}(\Omega, \mathcal{F}_2)$, we define the *(prior) second-period marginal Knightian uncertainty*, $\mathcal{P}|_2$, as a set of second-period marginal probability measures such that

$$\mathcal{P}|_2 = \{ \, p|_2 \mid p \in \mathcal{P} \, \} \ .$$

Here, the adjective *prior* emphasizes the fact that this is a set of the second-period marginal probability measures before the decision-maker obtains an observation in the first period.

Income Process. An income in each period, denoted y_1 and y_2, is a function from $\Omega = W \times W$ into $\mathbb{R}$. We call (y_1, y_2) an *income process* if it is $\mathcal{F}$-adapted, that is, $(\forall t)$ y_t is $\mathcal{F}_t$-measurable. Given an income process (y_1, y_2), we write the value of y_2 as $y_2|_{w_1 \in E,\, w_2 \in F}$ if $(w_1, w_2) \in E \times F$ for some $E \times F \in \mathcal{F}_2$. The $\mathcal{F}$-adaptedness allows us to write the value of y_1 as $y_1|_{w_1 \in E}$ if $w_1 \in E$ for some E such that $E \times W \in \mathcal{F}_1$. We denote the set of $\mathcal{F}$-adapted income processes by $Y(\mathcal{F})$.

Updating Rules. Let p be a probability measure on $(\Omega, \mathcal{F}_2)$, that is, let $p \in \mathcal{M}(\Omega, \mathcal{F}_2)$. After observing E_i in the first period, the decision maker updates her probability measures.

Let us now first consider the ordinary Bayesian updating procedure. Given p and E_i such that $p(E_i \times W) > 0$, we denote by $p|_2(\cdot|E_i)$ the (posterior) probability measure on $(\Omega, \mathcal{F}_2)$ conditional on the occurrence of $E_i \times W$. Here, the adjective *posterior* signifies the fact that this is a probability measure after the decision-maker obtains an observation E_i. That is, $(\forall i, j)$ $p|_2(E_i \times F_j | E_i) = p(E_i \times F_j)/p(E_i \times W)$. By writing $p|_2(\cdot|E_i) = p|_2(E_i \times \cdot|E_i)$, $p|_2(\cdot|E_i)$ may be regarded as a probability measure on $(W, \langle F_j \rangle_j)$. (It should be noted here that $p|_2(\cdot) = p|_2(\cdot|W)$.) The *Bayesian procedure* is defined as a function: $(p, E_i) \mapsto p|_2(\cdot|E_i)$, which maps a pair of measure p on $(\Omega, \mathcal{F}_2)$ and an event E_i in the first period, to the measure on $(W, \langle F_j \rangle_j)$ according to the manner defined in this paragraph.

An updating rule we consider in this paper generalizes the function $p|_2$ in the ordinary Bayesian procedure to the case of multiple p's, that is, where there exists Knightian uncertainty. Formally, an *updating rule* is a function that maps a pair $(\mathcal{P}, E)$, where $\mathcal{P}$ is the decision-maker's Knightian uncertainty (a nonempty compact subset of $\mathcal{M}(\Omega, \mathcal{F}_2)$) and E is an $\langle E_i \rangle_i$-measurable event such that $(\forall p \in \mathcal{P})\ p(E \times W) > 0$, to a set of (posterior) probability measures, which is a nonempty compact subset of $\mathcal{M}(W, \langle F_j \rangle_j)$. We denote an updating rule by ϕ and its specific value by $\phi(\mathcal{P}, E)$. (This seemingly cumbersome notation is necessary for taking account of dynamic Knightian uncertainty, as we will see later in this and following sections.)

There is one natural restriction on sensible updating rules. When $\mathcal{P}$ happens to be a singleton, they should coincide with Bayes' rule:

$$\phi(\{p\}, E) = \{\, p|_2(\cdot|E) \,\}\,. \tag{8}$$

Objective Function. Let us now turn to the issue of formulating the objective function of the decision-maker. As in the previous section, we assume that the minimum of the "expected" life-time income, V, is her objective function to be maximized, which is given by:

$$V(y_1, y_2) = \min_{p \in \mathcal{P}} \sum_{i=1}^{m} \left[(y_1|_{w_1 \in E_i}) + \beta \min_{q \in \phi(\mathcal{P}, E_i)} \sum_{j=1}^{n} \big(y_2|_{w_1 \in E_i,\, w_2 \in F_j}\big)\ q(F_j) \right] p(E_i \times W)\,, \tag{9}$$

where $(y_1, y_2) \in Y(\mathcal{F})$, ϕ is a updating rule, $\beta\,(> 0)$ is a discount factor and $\mathcal{P}$ is the decision-maker's Knightian uncertainty, which is a subset of $\mathcal{M}(\Omega, \mathcal{F}_2)$. In order that this definition is meaningful, $\mathcal{P}$ must be a nonempty compact subset of $\mathcal{M}(\Omega, \mathcal{F}_2)$ satisfying $(\forall p \in \mathcal{P})(\forall i)\ p(E_i \times W) > 0$.

Preferences represented by special cases of Eq (9), where the updating rules are further specified, are axiomatized by Epstein and Schneider (2003) and Wang (2003) (see next section).

Dilation of the Knightian Uncertainty. We now define "dilation" of the Knightian uncertainty. Let $\mathcal{P} \in \mathcal{M}(\Omega, \mathcal{F}_2)$ be the Knightian uncertainty that the decision-maker faces and let ϕ

be her update rule. The *dilation of the Knightian uncertainty* takes place upon the occurrence of $E \in \langle E_i \rangle_i$ if the set of posterior probability measures generated by the update rule is strictly "greater" than the set of prior probability measures, or equivalently if it holds that

$$\phi(\mathcal{P}, E) \supset \mathcal{P}|_2$$

where the set-inclusion is strict. In this case, the set of prior probability measures does not shrink but dilates: the decision-maker faces larger uncertainty than before obtaining the observation.[15]

In contrast, if the opposite strict set-inclusion holds for some $E \in \langle E_i \rangle_i$, we describe it as the *contraction of the Knightian uncertainty* upon the occurrence of E. In this case, the decision maker faces smaller uncertainty than before obtaining the observation.

4. The (Generalized) Bayesian and Maximum-Likelihood Rules

We consider two updating rules which have been extensively studied in the literature.[16]

[15] In the statistics literature, the dilation is defined with respect to lower- and upper-probabilities. To be more precise, let $\mathcal{P} \subseteq \mathcal{M}(\Omega, \mathcal{F}_2)$ and let $B \in \mathcal{F}_2$ be such that $(\forall p \in \mathcal{P})\ p(B) > 0$. Then, define the *lower-probability*, denoted $\underline{\mathcal{P}}$, by

$$(\forall A \in \mathcal{F}_2) \quad \underline{\mathcal{P}}(A) = \inf_{p \in \mathcal{P}} p(A)$$

and define the *conditional lower-probability*, denoted $\underline{\mathcal{P}}(\cdot|B)$, by

$$(\forall A \in \mathcal{F}_2) \quad \underline{\mathcal{P}}(A|B) = \inf_{p \in \mathcal{P}} p(A \cap B)/p(B)\,.$$

The upper-probability $\overline{\mathcal{P}}$ and the conditional upper-probability $\overline{\mathcal{P}}(\cdot|B)$ are defined symmetrically. Each of these "probabilities" turns out to be non-additive probability measure, or capacity. It is said that B *dilates* A if the following holds:

$$\underline{\mathcal{P}}(A|B) < \underline{\mathcal{P}}(A) \le \overline{\mathcal{P}}(A) < \overline{\mathcal{P}}(A|B)\,. \tag{10}$$

For this concept of dilation and study of its properties, see Seidenfeld and Wasserman (1993). Herron, Seidenfeld and Wasserman (1997) contains some additional analysis. Walley (1991) extensively studies the lower- and upper-probabilities.

Seidenfeld and Wasserman (1993) derives a necessary and sufficient condition for the dilation to take place in the sense of (10), for cases including the ε-contamination. Their condition, however, is based on a particular event A, not on set of measures, so that its application to economic models is rather difficult if not impossible.

In Section 5, we derive a sufficient condition for the dilation to take place for the ε-contamination in the sense defined in the text. Our definition is more general than (10) since it is applied directly to a set of measures, not to a particular event A. We consider the maximum-likelihood update rule as well as the generalized Bayesian update rule (see the next section) while (10) is related only to the generalized Bayesian rule. Further, we consider dynamic nature of Knightian uncertainty explicitly to derive economic intuition behind the dilation.

[16] See Dempster (1967, 1968); Shafer (1976); Fagin and Halpern (1990); Gilboa and Schmeidler (1993); and Denneberg (1994).

The generalized *Bayesian rule* (henceforth, the *GB rule*)[17] is denoted by ϕ_{GB} and is defined by

$$(\forall \mathcal{P} \subseteq \mathcal{M}(\Omega, \mathcal{F}_2))(\forall E \in \langle E_i \rangle_i) \quad \phi_{GB}(\mathcal{P}, E) = \{\, p|_2(\cdot|E) \,|\, p \in \mathcal{P} \,\}\,. \tag{11}$$

This means that the decision-maker updates all probability measures according to the ordinary Bayesian procedure. In particular, she does not discard any of these measures after the observation. It is evident that the procedure we employed in Section 2 corresponds to this rule. When ϕ is specified by ϕ_{GB}, the decision maker's objective function becomes

$$V(y_1, y_2) = \min_{p \in \mathcal{P}} \sum_{i=1}^{m} \left[(y_1|_{w_1 \in E_i}) + \beta \min_{p \in \mathcal{P}} \sum_{j=1}^{n} \left(y_2|_{w_1 \in E_i,\, w_2 \in F_j} \right) p|_2(F_j|E_i) \right] p(E_i \times W)\,.$$

A preference-theoretic foundation of this updating rule is given by Epstein and Schneider (2003). They axiomatize the preference relation represented by (9) with $\mathcal{P}$ being "rectangular" and ϕ being the GB rule (see Epstein and Schneider (2003) for details including the concept of rectangularity).

To define the *Maximum-Likelihood rule* (henceforth, the *ML rule*)[18], let $\mathcal{P}^*$ be defined by

$$(\forall E \in \langle E_i \rangle_i) \quad \mathcal{P}^*(E) = \arg\max \{\, p|_1(E) \mid p \in \mathcal{P} \,\}\,.$$

Then, the ML rule is defined by

$$(\forall \mathcal{P} \subseteq \mathcal{M}(\Omega, \mathcal{F}_2))(\forall E \in \langle E_i \rangle_i) \quad \phi_{ML}(\mathcal{P}, E) = \{\, p|_2(\cdot|E) \,|\, p \in \mathcal{P}^*(E) \,\}\,. \tag{12}$$

A preference-theoretic foundation of this updating rule is given by Wang (2003). He axiomatizes the preference relation represented by (9) with $\mathcal{P}$ being the core of some convex probability capacity and ϕ being the GB rule and the ML rule (see Wang (2003) for details including the concept of probability capacity).[19]

Both the GB and ML rules satisfy the requirement we impose on updating rules, (8).

---

[17]The generalized Bayesian rule is originally proposed as an update rule for a non-additive measure. More precisely, the rule was developed for $\mathcal{P}$ which is characterized as the *core* of a non-additive measure (Fagin and Halpern, 1990; Denneberg, 1994). The text use of the rule is its natural extension to the case of a more general $\mathcal{P}$.

[18]The maximum-likelihood rule is originally proposed as an updating rule for a non-additive measure (Dempster, 1967, 1968; Shafer, 1976). Later, Gilboa and Schmeidler (1993) showed that this rule is identical to the maximum-likelihood updating rule, which we extend to the case of a more general $\mathcal{P}$ in the text.

[19]For a related work which provides some axiomatic foundation to the ML rule, see Gilboa and Schmeidler (1993).

Lemma 1. *Assume that* $\mathcal{P} = \{p\}$ *for some* $p \in \mathcal{M}(\Omega, \mathcal{F}_2)$ *such that* $(\forall i)\ p(E_i \times W) \neq 0$*. Then,*

$$(\forall i) \quad \phi_{GB}(\mathcal{P}, E_i) = \phi_{ML}(\mathcal{P}, E_i) = \{\, p|_2(\cdot|E_i)\,\}\,.$$

Proof. For the GB rule, the claim is immediate from (11). For the ML rule, the claim is also immediate from (12) and the fact that $(\forall i)\ \mathcal{P}^*(E_i) = \{p\}$. ■

This lemma shows that the both rules extend Bayes' rule to the case where the prior is not unique. Finally, it immediately follows from the definition that

$$(\forall \mathcal{P})(\forall i) \quad \phi_{ML}(\mathcal{P}, E_i) \subseteq \phi_{GB}(\mathcal{P}, E_i)\,.$$

That is, the "degree of (Knightian) uncertainty" in the posteriors implied by the ML rule is no more than that implied by the GB rule.

5. ε-contamination and Dilation of the Knightian Uncertainty

In this section, we consider a case where the decision-maker's Knightian uncertainty, $\mathcal{P}$, is specified by a general ε-contamination. Here ε-contamination is "general", since we do not restrict it to be of a product probability measure. We give a simple and easily verifiable condition under which dilation takes place. Using this condition, we then show that if ε-contamination under consideration is a restricted one, that is, one of a product of probability measures (as in the case of Section 2), the decision-maker *always* experiences dilation of uncertainty, *regardless of whether the updating rule is GB or ML.*

Formally, let p^0 be a probability measure on $(\Omega, \mathcal{F}_2)$ such that $(\forall i)\ p^0(E_i \times W) > 0$, and let $\varepsilon \in (0, 1)$. We assume that the decision-maker's $\mathcal{P}\,(\subseteq \mathcal{M}(\Omega, \mathcal{F}_2))$ is characterized by the *ε-contamination* of p^0, such that

$$\mathcal{P} = \left\{p^0\right\}^{\varepsilon} \equiv \left\{\ (1-\varepsilon)p^0 + \varepsilon q \;\middle|\; q \in \mathcal{M}(\Omega, \mathcal{F}_2)\ \right\}\ . \tag{13}$$

In the following analysis, the one-period counterpart of the two-period ε-contamination (13) turns out to be important. Applying the same idea to the one-period case, we define for each $\varepsilon \in (0, 1)$ and each $E \in \langle E_i \rangle_i$, the ε-contamination of $p^0|_2(\cdot|E)\,(\in \mathcal{M}(W, \langle F_j \rangle_j))$ by

$$\left\{p^0|_2(\cdot|E)\right\}^{\varepsilon} \equiv \left\{\ (1-\varepsilon)p^0|_2(\cdot|E) + \varepsilon q_2 \;\middle|\; q_2 \in \mathcal{M}(W, \langle F_j \rangle_j)\ \right\}\ .$$

The following lemma shows that the second-period "restriction" of the ε-contamination of p^0 is the same as the ε-contamination of the second-period "restriction" of p^0. In a sense, the "operator" of taking ε-contamination and the "operator" of taking second-period "restriction" or marginal are interchangeable with respect to p^0, which is a probability measure on $(\Omega, \mathcal{F}_2)$ such that $(\forall i)\ p^0(E_i \times W) > 0$.

Formally, $\left\{p^0\right\}^\varepsilon\big|_2$, the (prior) second-period marginal Knightian uncertainty of the ε-contamination of p^0, is equal to $\left\{p^0|_2\right\}^\varepsilon$, the ε-contamination of the (prior) second-period marginal probability measure $p^0|_2 = p^0|_2(\cdot|W)$:

Lemma 2. *Taking ristriction (or marginal), $\cdot|_2$, and taking ε-contamination, $\{\cdot\}^\varepsilon$, are interchangeable with respect to p^0: that is, $(\forall \varepsilon \in (0,1))\quad \left\{p^0\right\}^\varepsilon\big|_2 = \left\{p^0|_2\right\}^\varepsilon$.*

Proof. To show $\left\{p^0\right\}^\varepsilon\big|_2 \subseteq \left\{p^0|_2\right\}^\varepsilon$, let $p_2 \in \left\{p^0\right\}^\varepsilon\big|_2$. Then, there exists $p \in \left\{p^0\right\}^\varepsilon$ such that $p_2 = p(W \times \cdot)$. That $p \in \left\{p^0\right\}^\varepsilon$ in turn implies that there exists $q \in \mathcal{M}(\Omega, \mathcal{F}_2)$ such that $p = (1-\varepsilon)p^0 + \varepsilon q$. Hence, $p_2 = p(W \times \cdot) = (1-\varepsilon)p^0|_2(\cdot) + \varepsilon q|_2(\cdot)$. This shows that $p_2 \in \left\{p^0|_2\right\}^\varepsilon$ because $q|_2(\cdot) \in \mathcal{M}(W, \langle F_j \rangle_j)$.

To show $\left\{p^0\right\}^\varepsilon\big|_2 \supseteq \left\{p^0|_2\right\}^\varepsilon$, let $p_2 \in \left\{p^0|_2\right\}^\varepsilon$. Then, there exists $q_2 \in \mathcal{M}(W, \langle F_j \rangle_j)$ such that $p_2 = (1-\varepsilon)p^0|_2 + \varepsilon q_2$. Let $q_1 \in \mathcal{M}(W, \langle E_i \rangle_i)$ and let $p = (1-\varepsilon)p^0 + \varepsilon(q_1 \times q_2)$. Then, $p \in \left\{p^0\right\}^\varepsilon$ and $p|_2 = (1-\varepsilon)p^0|_2 + \varepsilon q_2 = p_2$, and hence, $p_2 \in \left\{p^0\right\}^\varepsilon\big|_2$. ■

We now present a result characterizing posterior second-period (marginal) Knightian uncertainty derived by the two update rules in the case of ε-contamination.

Theorem 1. *Let $\varepsilon \in (0,1)$ and let $E \in \langle E_i \rangle_i$. Then,*

$$\phi_{GB}\left(\left\{p^0\right\}^\varepsilon, E\right) = \phi_{ML}\left(\left\{p^0\right\}^\varepsilon, E\right) = \left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'}$$

where

$$\varepsilon' = \varepsilon'(\varepsilon, E) \equiv \frac{\varepsilon}{(1-\varepsilon)p^0|_1(E) + \varepsilon} > \varepsilon.$$

Proof. (a) ***The GB rule.*** Define $\mathcal{R} \subseteq \mathcal{M}(W, \langle F_j \rangle_j)$ by

$$\mathcal{R} = \left\{ \frac{(1-\varepsilon)p^0|_1(E)}{(1-\varepsilon)p^0|_1(E) + \varepsilon q_1(E)}\, p^0|_2(\cdot|E) + \frac{\varepsilon q_1(E)}{(1-\varepsilon)p^0|_1(E) + \varepsilon q_1(E)}\, q_2 \;\middle|\; \begin{array}{l} q_1 \in \mathcal{M}(W, \langle E_i \rangle_i)\,, \\ q_2 \in \mathcal{M}(W, \langle F_j \rangle_j) \end{array} \right\}.$$

We first show that

$$\phi_{GB}\left(\left\{p^0\right\}^{\varepsilon}, E\right) = \mathcal{R}\,. \tag{14}$$

By definition of ϕ_{GB}, it holds that

$$\phi_{GB}\left(\left\{p^0\right\}^{\varepsilon}, E\right) = \left\{\, p|_2(\cdot|E) \mid p \in \left\{p^0\right\}^{\varepsilon} \,\right\} = \left\{ \frac{p(E \times \cdot)}{p(E \times W)} \;\middle|\; p \in \left\{p^0\right\}^{\varepsilon} \right\} \tag{15}$$

$$= \left\{ \frac{(1-\varepsilon)p^0|_1(E)}{(1-\varepsilon)p^0|_1(E) + \varepsilon q(E \times W)}\, p^0|_2(\cdot|E) + \frac{\varepsilon}{(1-\varepsilon)p^0|_1(E) + \varepsilon q(E \times W)}\, q(E \times \cdot) \;\middle|\; q \in \mathcal{M}(\Omega, \mathcal{F}_2) \right\},$$

where we invoked the fact that $p^0(E \times \cdot) = p^0|_1(E) \cdot p^0|_2(\cdot|E)$. Eq (15) shows that $\mathcal{R} \subseteq \phi_{GB}\left(\left\{p^0\right\}^{\varepsilon}, E\right)$ since $q_1 \times q_2 \in \mathcal{M}(\Omega, \mathcal{F}_2)$.

To show that the opposite inclusion also holds, let $p \in \phi_{GB}\left(\left\{p^0\right\}^{\varepsilon}, E\right)$. Then, there exists $q \in \mathcal{M}(\Omega, \mathcal{F}_2)$ such that

$$p = \frac{(1-\varepsilon)p^0|_1(E)}{(1-\varepsilon)p^0|_1(E) + \varepsilon q(E \times W)}\, p^0|_2(\cdot|E) + \frac{\varepsilon}{(1-\varepsilon)p^0|_1(E) + \varepsilon q(E \times W)}\, q(E \times \cdot)$$

by (15). When $q(E \times W) = 0$, it follows that $p = p^0|_2(\cdot|E)$, and hence, $p \in \mathcal{R}$ (let q_1 be such that $q_1(E) = 0$ in the definition of $\mathcal{R}$). When $q(E \times W) \neq 0$, let $q_1 = q|_1$ and $q_2 = q|_2(\cdot|E)$, which is now well-defined, in the definition of $\mathcal{R}$. Then, $q_1 \in \mathcal{M}(W, \langle E_i \rangle_i)$ and $q_2 \in \mathcal{M}(W, \langle F_j \rangle_j)$, and hence, $p \in \mathcal{R}$. Thus we have proved that (14) holds true.

We next show that

$$\left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'} = \mathcal{R}$$

which completes the proof in the case of the GB rule.

It immediately follows that $\left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'} \subseteq \mathcal{R}$ (let q_1 be such that $q_1(E) = 1$). To show that the opposite inclusion also holds, let $p \in \mathcal{R}$. Then, there exist $q_1 \in \mathcal{M}(W, \langle E_i \rangle_i)$ and $q_2 \in \mathcal{M}(W, \langle F_j \rangle_j)$ such that

$$p \;=\; \frac{(1-\varepsilon)p^0|_1(E)}{(1-\varepsilon)p^0|_1(E) + \varepsilon q_1(E)}\, p^0|_2(\cdot|E) + \frac{\varepsilon q_1(E)}{(1-\varepsilon)p^0|_1(E) + \varepsilon q_1(E)}\, q_2$$

$$= \quad (1-\varepsilon')p^0|_2(\cdot|E) + \varepsilon' \left\{(1-\tilde{\varepsilon})p^0|_2(\cdot|E) + \tilde{\varepsilon} q_2\right\}$$

where

$$\tilde{\varepsilon} = \frac{(1-\varepsilon)p^0|_1(E)q_1(E) + \varepsilon q_1(E)}{(1-\varepsilon)p^0|_1(E) + \varepsilon q_1(E)} .$$

Since $(1-\tilde{\varepsilon})p^0|_2(\cdot|E) + \tilde{\varepsilon} q_2 \in \mathcal{M}(W, \langle F_j \rangle_j)$ by the fact that $\tilde{\varepsilon} \in [0,1]$, it follows that $p \in \left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'}$ as desired.

(b) ***The ML Rule.*** We only need to show that $\left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'} \subseteq \phi_{ML}\left(\left\{p^0\right\}^{\varepsilon}, E\right)$ since the opposite inclusion holds by (a) and the fact that $\phi_{ML} \subseteq \phi_{GB}$ always holds.

To prove this, first note (see (12)) that $(\forall E \in \langle E_i \rangle_i)$ we have

$$\left(\left\{p^0\right\}^{\varepsilon}\right)^* (E) = \left\{ (1-\varepsilon)p^0 + \varepsilon q \;\middle|\; q \in \mathcal{M}(\Omega, \mathcal{F}_2) \text{ and } q(E \times W) = 1 \right\} ,$$

which in turn implies that

$$\phi_{ML}\left(\left\{p^0\right\}^{\varepsilon}, E\right) = \left\{ p|_2(\cdot|E) \;\middle|\; p \in \left(\left\{p^0\right\}^{\varepsilon}\right)^* (E) \right\}$$

$$= \left\{ \frac{(1-\varepsilon)p^0|_1(E)}{(1-\varepsilon)p^0|_1(E)+\varepsilon} p^0|_2(\cdot|E) + \frac{\varepsilon}{(1-\varepsilon)p^0|_1(E)+\varepsilon} q(E \times \cdot) \;\middle|\; q \in \mathcal{M}(\Omega, \mathcal{F}_2) \text{ and } q(E \times W) = 1 \right\} .$$

Let $p_2 \in \left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'}$. Then, there exists $q_2 \in \mathcal{M}(W, \langle F_j \rangle_j)$ such that $p_2 = (1-\varepsilon')p^0|_2(\cdot|E) + \varepsilon' q_2$. Let q_1 be the element of $\mathcal{M}(W, \langle E_i \rangle_i)$ such that $q_1(E) = 1$. Then, $q_1 \times q_2 \in \mathcal{M}(\Omega, \mathcal{F}_2)$, $(q_1 \times q_2)(E \times W) = 1$ and $p_2 = (1-\varepsilon')p^0|_2(\cdot|E) + \varepsilon' q_1(E) q_2(\cdot)$. Therefore, $p_2 \in \phi_{ML}\left(\left\{p^0\right\}^{\varepsilon}, E\right)$ as desired.

(c) ***To show*** $\varepsilon' > \varepsilon$***.*** Since we have assumed that $(\forall i : i = 1, \ldots, m)$ $p^0(E_i \times W) > 0$ and $m \geq 2$ in Section 3, it follows that $(\forall i : i = 1, \ldots, m)$ $p^0(E_i \times W) = p^0|_1(E_i) < 1$. Therefore, it holds that $\varepsilon' > \varepsilon$. ■

Let us now define a measure of the "informational value" of the observation E with respect to p^0, the "pre-contamination" probability measure. Let $E \in \langle E_i \rangle_i$ and let $\delta(E) \in [0,1]$ be defined by

$$\delta(E) = \max_{j=1,\ldots,n} \left| \, p^0|_2(F_j|E) - p^0|_2(F_j) \, \right| .$$

The real number $\delta(E)$ is the maximum of the "probability change" due to the observation E with respect to the pre-contamination probability measure p^0, which can be considered as a measure of the informational value of the observation E for p^0.

The next theorem shows that, if ε, the degree of contamination of p^0, is sufficiently large with respect to $\delta(E)$, the observation E's information value with respect to p^0, then the dilation takes place.

Theorem 2. *Let $\mathcal{P}$ be given by $\left\{p^0\right\}^{\varepsilon}$ and let $E \in \langle E_i\rangle_i$. Suppose that the degree of contamination of p^0 is sufficiently large compared with the informational value of the observation E with respect to p^0, that is, suppose that the following inequality holds:*

$$\varepsilon > \frac{p^0|_1(E)}{(1-p^0|_1(E))\min_j p^0|_2(F_j)}\,\delta(E)\,. \tag{16}$$

Then, the dilation occurs in the sense that it holds that

$$\left[\ \phi_{GB}\left(\left\{p^0\right\}^{\varepsilon}, E\right) = \phi_{ML}\left(\left\{p^0\right\}^{\varepsilon}, E\right) = \ \right] \left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'} \supset \left\{p^0|_2\right\}^{\varepsilon} \left[\ = \left\{p^0\right\}^{\varepsilon}\big|_2\ \right],$$

where the inclusion is strict and ε' is as defined in Theorem 1.

Proof. Note that the two equalities in the left-hand side were established by Theorem 1 and the equality in the right-hand side was established by Lemma 2, and hence, the theorem claims that the strict inclusion holds.

We first prove $\left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'} \supseteq \left\{p^0|_2\right\}^{\varepsilon}$, and then shows that inclusion is strict.

(a) Proof of $\left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'} \supseteq \left\{p^0|_2\right\}^{\varepsilon}$. Let $p_2 \in \left\{p^0|_2\right\}^{\varepsilon}$. Then, there exists $q_2 \in \mathcal{M}(W, \langle F_j\rangle_j)$ such that $p_2 = (1-\varepsilon)p^0|_2 + \varepsilon q_2$. Therefore, we have

$$\begin{aligned} p_2 &= (1-\varepsilon')p^0|_2(\cdot|E) + \varepsilon'\left(\frac{1-\varepsilon}{\varepsilon'}\,p^0|_2 - \frac{1-\varepsilon'}{\varepsilon'}\,p^0|_2(\cdot|E) + \frac{\varepsilon}{\varepsilon'}\,q_2\right) \\ &= (1-\varepsilon')p^0|_2(\cdot|E) + \varepsilon'\mu\,, \end{aligned} \tag{17}$$

where

$$\mu \equiv \frac{1-\varepsilon}{\varepsilon'}\,p^0|_2 - \frac{1-\varepsilon'}{\varepsilon'}\,p^0|_2(\cdot|E) + \frac{\varepsilon}{\varepsilon'}\,q_2\,.$$

It immediately follows that μ is an (additive) signed measure such that $\mu(\phi) = 0$ and $\mu(W) = 1$. If $\mu \geq 0$, then $\mu \in \mathcal{M}(W, \langle F_j \rangle_j)$ and hence $p_2 \in \left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'}$ implying $\left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'} \supseteq \left\{p^0|_2\right\}^{\varepsilon}$.

In the remaining of this subsection, we prove that $\mu \geq 0$. Note that if

$$(\forall F \in \langle F_j \rangle_j) \quad \frac{1-\varepsilon}{\varepsilon'} p^0|_2(F) - \frac{1-\varepsilon'}{\varepsilon'} p^0|_2(F|E) \geq 0\,,$$

then we have $\mu \geq 0$ since $q_2 \geq 0$. Therefore, it is sufficient to show the above relation.

If $\delta(E) = 0$, it is straightforward to show

$$\begin{aligned} & \frac{1-\varepsilon}{\varepsilon'} p^0|_2(F) - \frac{1-\varepsilon'}{\varepsilon'} p^0|_2(F|E) \\ = \; & \frac{1-\varepsilon}{\varepsilon'} \left(p^0|_2(F) - p^0|_2(F|E)\right) - \frac{\varepsilon-\varepsilon'}{\varepsilon'} p^0|_2(F|E) \\ = \; & \frac{\varepsilon'-\varepsilon}{\varepsilon'} p^0|_2(F|E) \;\geq\; 0\,, \end{aligned}$$

since $\delta(E) = \max_j \left|\, p^0|_2(F_j|E) - p^0|_2(F_j) \,\right| = 0$ and $\varepsilon' \geq \varepsilon$.

If $\delta(E) > 0$, we have

$$\begin{aligned} & \frac{1-\varepsilon}{\varepsilon'} p^0|_2(F) - \frac{1-\varepsilon'}{\varepsilon'} p^0|_2(F|E) \qquad (18) \\ = \; & (1-\varepsilon)\left[\left(\frac{1-\varepsilon}{\varepsilon} p^0|_1(E) + 1\right) p^0|_2(F) - \frac{1}{\varepsilon} p^0|_1(E) p^0|_2(F|E)\right] \\ \geq \; & (1-\varepsilon)\left[\left(\frac{1-\varepsilon}{\varepsilon} p^0|_1(E) + 1\right) p^0|_2(F) - \frac{1}{\varepsilon} p^0|_1(E)\left(p^0|_2(F) + \delta(E)\right)\right] \\ = \; & (1-\varepsilon)\left[\left(1 - p^0|_1(E)\right) p^0|_2(F) - \frac{\delta(E)}{\varepsilon} p^0|_1(E)\right] \\ \geq \; & (1-\varepsilon)\left[\left(1 - p^0|_1(E)\right) \min_j p^0|_2(F_j) - \frac{\delta(E)}{\varepsilon} p^0|_1(E)\right] \\ > \; & (1-\varepsilon)\left[\left(1 - p^0|_1(E)\right) \min_j p^0|_2(F_j) - \delta(E)\, p^0|_1(E) \left(\frac{p^0|_1(E)}{(1-p^0|_1(E)) \min_j p^0|_2(F_j)}\, \delta(E)\right)^{-1}\right] \\ = \; & 0\,, \end{aligned}$$

where the first equality holds by the definition of ε'; the first inequality holds by the definition of δ; the second inequality holds by the min operator; and the strict inequality holds by (16) and the assumptions that $\delta(E) > 0$ and $p^0|_1(E) > 0$. This completes the first half of the proof.

(b) Proof of strict inclusion. Let $F \in \langle F_j \rangle_j$ be such that $p^0|_2(F) > 0$ and let $\hat{p}_2 \in \left\{p^0|_2(\cdot|E)\right\}^{\varepsilon'}$ be such that $\hat{p}_2(F) = (1-\varepsilon')p^0|_2(F|E)$. We show $\hat{p}_2 \notin \left\{p^0|_2\right\}^{\varepsilon}$.

If $\delta(E) = 0$, we have for any $p_2 \in \{p^0|_2\}^\varepsilon$

$$\begin{aligned} p_2(F) &\geq (1-\varepsilon)p^0|_2(F) \\ &= (1-\varepsilon')p^0|_2(F) + (\varepsilon' - \varepsilon)\, p^0|_2(F) \\ &> (1-\varepsilon')p^0|_2(F) \\ &= (1-\varepsilon')p^0|_2(F|E) \;=\; \hat{p}_2(F)\,, \end{aligned}$$

where the strict inequality holds since $\varepsilon' > \varepsilon$ (Theorem 1) and $p^0|_2(F) > 0$ by the assumption of F, and its next equality holds since $p^0|_2(F) = p^0|_2(F|E)$ by the assumption that $\delta(E) = 0$. Therefore, we have $\hat{p}_2 \notin \{p^0|_2\}^\varepsilon$.

If $\delta(E) > 0$, we have for any $p_2 \in \{p^0|_2\}^\varepsilon$

$$\begin{aligned} p_2(F) &\geq (1-\varepsilon')p^0|_2(F|E) + \varepsilon' \left(\frac{1-\varepsilon}{\varepsilon'}\, p^0|_2(F) - \frac{1-\varepsilon'}{\varepsilon'}\, p^0|_2(F|E) \right) \\ &> (1-\varepsilon')p^0|_2(F|E) = \hat{p}_2(F) \end{aligned}$$

where the first inequality follows (17) and the second is implied by (18). Consequently, we have $\hat{p}_2 \notin \{p^0|_2\}^\varepsilon$. ■

This theorem shows that the dilation occurs when the degree of confidence in p^0 is small (*i.e.*, ε is large) compared with the informational value of the observation with respect to p^0 (*i.e.*, $\delta(E)$).

An important special case is the one in which we have $p^0 = p_1^0 \otimes p_2^0$ for some $p_1^0 \in \mathcal{M}(W, \langle E_i \rangle_i)$ and $p_2^0 \in \mathcal{M}(W, \langle F_j \rangle_j)$, that is, p^0 is a product of two probability measures. An example of this case is analyzed in Section 2. In this example, there is no informational value in observation E with respect to p^0. To see this, note that we have $p^0|_2(F_j|E) = p^0|_2(F_j) = p_2^0(F_j)$ for all F_j. It is clear that we have $\delta(E) = 0$ for all events E . Theorem 2 implies the following corollary in this case.

Corollary 1. *Suppose that* $p^0 = p_1^0 \otimes p_2^0$ *for some* $p_1^0 \in \mathcal{M}(W, \langle E_i \rangle_i)$ *and* $p_2^0 \in \mathcal{M}(W, \langle F_j \rangle_j)$. *Also, suppose that* $\mathcal{P}$ *is given by* $\{p^0\}^\varepsilon$. *Then, for any* $E \in \langle E_i \rangle_i$, *it holds that* $\phi_{GB}(\{p^0\}^\varepsilon, E_i) = \phi_{ML}(\{p^0\}^\varepsilon, E_i) \supset \mathcal{P}|_2$, *where the inclusion is strict.*

Proof. This follows immediately from Thoerem 2 since $\delta(E) = 0$ when $p^0 = p_1^0 \otimes p_2^0$ for some p_1^0 and p_2^0. ■

This corollary shows a striking result. In the case of ε-contamination of a product of probability measures, the GB rule and even ML rule, which are considered to have some behavioral foundation and thus to be sensible in the multiple prior framework, actually *increase*, rather than decrease, the degree of the Knightian uncertainty. In a sense, new information worsens the decision-maker's confused state of confidence, rather than improves it.

REFERENCES

Berger, J. (1985): *Statistical Decision Theory*, 2nd edition, Springer, New York.

Dempster, A. P. (1967): "Upper and Lower Probabilities Induced by A Multivalued Map," *Annals of Mathematical Statistics* 38, 325-339.

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

Denneberg, D. (1994): "Conditioning (Updating) Non-Additive Measures," *Annals of Operations Research* 52, 21-42.

Epstein, L. G. and M. Schneider (2003): "Recursive Multiple-Priors," *Journal of Economic Theory* 113, 1-31.

Fagin, R. and J. Y. Halpern (1990): "A New Approach to Updating Beliefs," *Proceedings of the 6th Conference on Uncertainty in AI.*

Ghirardato, P. (1997): "On Independence for Non-Additive Measures, with a Fubini Theorem," *Journal of Economic Theory* 73, 261-291.

Ghirardato, P. and M. Marinacci (2002): "Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory* 102, 251-289.

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

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

Herron, T., T. Seidenfeld and L. Wasserman (1997): "Divisive Conditioning: Further Results on Dilation," *Philosophy of Science* 64, 411-444.

Nishimura, K. G. and H. Ozaki (2004): "Search and Knightian Uncertainty," *Journal of Economic Theory* 119, 299-333.

Nishimura, K. G. and H. Ozaki (2006): "An Axiomatic Approach to ε-contamination," *Economic Theory* 27, 333-340.

Rothschild, M. (1974): "Searching for the Lowest Price When the Distribution of Prices is Unknown," *Journal of Political Economy* 82, 689-711.

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

Seidenfeld, T. and L. Wasserman (1993): "Dilation for Convex Sets of Probabilities," *Annals of Statistics* 21, 1139-1159.

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

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

Wang, T. (2003): "Conditional Preferences and Updating," *Journal of Economic Theory* 108, 286-321.