The emergence of “fifty-fifty” probability judgements in a conditional Savage world

Alexander Zimper

ERSA working paper 291

http://econrsa.org/home/index.php?option=com_docman&task=doc_download&gid=435&Itemid=67

June 2012

The emergence of "fifty-fifty" probability judgements in a conditional Savage world

Alexander Zimper*

May 25, 2012

Abstract

This paper models the empirical phenomenon of persistent "fifty-fifty" probability judgements within a dynamic non-additive Savage framework. To this purpose I construct a model of Bayesian learning such that an agent's probability judgement is characterized as the solution to a Choquet expected utility maximization problem with respect to a conditional neo-additive capacity. Only for the non-generic case in which this capacity degenerates to an additive probability measure, the agent's probability judgement coincides with the familiar estimate of a Bayesian statistician who minimizes a quadratic (squared error) loss function with respect to an additive posterior distribution. In contrast, for the generic case in which the capacity is non-additive, the agent's probability judgements converge through Bayesian learning to the unique fuzzy probability measure that assigns a 0.5 probability to any uncertain event.

Keywords: Non-additive measures, Learning, Decision analysis, Economics

*Department of Economics, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa. Tel: 0027 12 420-3522. Fax: 0027 86 691-2749. E-mail: alexander.zimper@up.ac.za

1 Introduction

Let us consider two opposite benchmark cases of how agents form and revise probability judgements. On the one hand, there is the statistically sophisticated agent whose learning process is described by additive Bayesian estimates that are updated in accordance with Bayes' rule. On the other hand, there is the statistically ignorant agent who attaches a fifty percent chance to any uncertain event whereby he sticks to this probability judgement regardless of any new information. Such persistent "fifty-fifty" answers are well-documented within the literature on focal point answers in economic surveys (Hurd 2009; Kleinjans and Van Soest 2010; Manski and Molinari 2010; van Santen et al. 2012) as well as within the psychological literature (Bruine de Bruin et al. 2000; Wakker 2004; Camerer 2007). Wakker (2010) interprets "fifty-fifty" judgements as an extreme case of cognitive likelihood insensitivity; more specifically, he writes:

> "Likelihood insensitivity can be related to regression to the mean. It is not a statistical artifact resulting from data analysis with noise, though, but it is a psychological phenomenon, describing how people perceive and weight probabilities in decisions. In weighting probabilities, a regression to the mean takes place, with people failing to discriminate sufficiently between intermediate probabilities and taking them all too much as the same ("50-50", "don't know"). (p. 228).

Given that likelihood insensitivity corresponds, in theory, to a very large domain of possible probability judgements, the question arises why "fifty-fifty" judgements are the predominant empirical expression of likelihood insensitivity. This paper proposes a decision theoretic explanation of this "fifty-fifty" phenomenon. To this purpose I construct a model of non-additive Bayesian learning of probability judgements which formally encompasses the sophisticated as well as the ignorant agent as complementary special cases. Because my formal approach can be completely interpreted in terms of behavioral axioms, it aims at "opening the black box of decision makers instead of modifying functional forms" (Rubinstein 2003, p. 1207).

My point of departure is the formal description of probability judgements as preference-maximizing acts in a Savage (1954) world. Let the state space be given as

$$\Omega = \Theta \times \mathbb{I} \tag{1}$$

with generic element $\omega \equiv (\theta, i)$. $\Theta = (0, 1)$ denotes the parameter space of all possible "true" probabilities. $\mathbb{I}$ denotes the information space. Endow Θ with the Euclidean metric and denote by $\mathcal{B}$ the corresponding Borel σ-algebra on Θ. Furthermore, fix some topology on $\mathbb{I}$ and denote by $\mathcal{I}$ the corresponding Borel σ-algebra. The relevant event space $\mathcal{F}$ is then defined as the standard product algebra $\mathcal{B} \otimes \mathcal{I}$. Define by $\tilde{\theta}(\theta, i) = \theta$ the $\mathcal{F}$-measurable random variable that gives for every state of the world the true probability and by $I = \Theta \times A$ with $A \in \mathcal{I}$ some information in $\mathcal{F}$. For a given constant $x \in (0,1)$ define now the $\mathcal{F}$-measurable Savage act $f_x : \Omega \to \mathbb{R}$ such that $f_x(\omega) = x - \theta$. We interpret f_x as the Savage act of "making the probability judgement x" whereby the outcome of this act in state $\omega = (\theta, i)$ is defined as the difference between x and the true probability θ. If the agent satisfies Savage's (1954) axioms, then there exists a von Neumann Morgenstern (vNM) utility function $u : \mathbb{R} \to \mathbb{R}$, unique up to a positive affine transformation, and a unique additive probability measure μ on $(\Omega, \mathcal{F})$ such for all Savage acts g, f and for all $I \in \mathcal{F}$:

$$f \succeq_I g \Leftrightarrow E\left[u(f(\omega)), \mu\left(\tilde{\theta} \mid I\right)\right] \geq E\left[u(g(\omega)), \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{2}$$

That is, under the Savage axioms the agent's preferences over acts conditional on information I have an expected utility (=EU) representation with respect to the subjective additive conditional probability measure $\mu\left(\tilde{\theta} \mid I\right)$. In a next step, let us assume that the agent's vNM utility function is given as a negative quadratic function implying for probability judgements

$$u(f_x(\omega)) = -(x - \theta)^2. \tag{3}$$

The solution to the maximization problem

$$x_I^* = \arg \sup_{x \in (0,1)} E\left[-\left(x - \tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right], \tag{4}$$

i.e., the agent's revealed probability judgement, then coincides with the classical Bayesian point estimate that results from the minimization of the expected value of a quadratic (=squared error) loss function with respect to the posterior distribution $\mu\left(\tilde{\theta} \mid I\right)$, (cf., Girshick and Savage 1951; James and Stein 1961).

The familiar Bayesian estimate (4) will be recovered in my model as the special case that describes the revealed probability judgement of the sophisticated agent. In order to describe both, the sophisticated and the ignorant, agents within a unified model, I generalize the above framework to a Savage world in which the agents are Choquet expected utility (=CEU) rather than EU maximizers. Behavioral axioms that give rise to a CEU representation were first presented in Schmeidler (1989) within the Anscombe and Aumann (1963) framework, which assumes preferences over objective probability distributions. Subsequently, Gilboa (1987) as well as Sarin and Wakker (1992) have presented behavioral CEU axiomatizations within the Savage (1954) framework, assuming a purely subjective notion of likelihood. From a mathematical perspective, CEU theory is an application of fuzzy measure theory such that the integration with respect to a fuzzy (=non-additive) probability measure is characterized by a comonotonic, positive homogeneous and monotonic functional (cf., Schmeidler 1986; Grabisch et al. 1995; Murofushi and Sugeno 1989, 1991; Narukawa and Murofushi 2003, 2004; Sugeno et al. 1998; Narukawa et al. 2000, 2001; Narukawa 2007). From the perspective of behavioral decision theory, CEU theory is formally equivalent to cumulative prospect theory (=CPT) (Tversky and Kahneman 1992; Wakker and Tversky 1993; Basili and Chateauneuf 2011) whenever CPT is restricted to gains. CPT, in turn, extends the celebrated concept of original prospect theory by Kahneman and Tversky (1979) to the case of several possible gain values in a way that satisfies first-order stochastic dominance.

If an agent has CEU preferences in a conditional Savage world, we obtain as a generalization of (2) that for all Savage acts g, f and for all $I \in \mathcal{F}$

$$f \succeq_I g \Leftrightarrow E^C \left[u\left(f\left(\omega\right)\right), \kappa\left(\tilde{\theta} \mid I\right) \right] \geq E^C \left[u\left(g\left(\omega\right)\right), \kappa\left(\tilde{\theta} \mid I\right) \right] \tag{5}$$

where E^C denotes the Choquet expectation operator with respect to a subjective non-additive conditional probability measure $\kappa\left(\cdot \mid I\right)$. Under the assumption of a quadratic vNM function (3), the CEU agent's revealed probability judgement is then given as the solution to the maximization problem

$$x_I^C = \arg \sup_{x \in (0,1)} E^C \left[-\left(x - \tilde{\theta}\right)^2, \kappa\left(\tilde{\theta} \mid I\right) \right]. \tag{6}$$

Unlike (4) the CEU maximization problem (6) does, in general, not allow for an analytically convenient solution because E^C is non-linear and, while being continuous, it is no longer differentiable everywhere. As a consequence, the global maximum of (6) is no longer uniquely characterized by a first-order condition (FOC).

To derive an analytical solution to problem (6), I am going to restrict attention to a convenient subclass of conditional non-additive probability measures. First, I restrict attention to non-additive probability measures given as neo-additive capacities in the sense of Chateauneuf et al. (2007). Neo-additive capacities reduce the potential complexity of non-additive probability measures in a very parsimonious way (i.e., two additional parameters only) such that important empirical features (e.g., inversely S-shaped probability transformation functions) are portrayed (cf. Chapter 11 in Wakker 2010). Second, I assume that these neo-additive capacities are updated in accordance with the Generalized Bayesian update rule (Pires 2002; Eichberger et al. 2006; Siniscalchi 2011). Among the many perceivable Bayesian update rules for non-additive probability measures, the Generalized Bayesian update rule has convenient technical and empirical features.

Having derived a comprehensive analytical solution (Theorem 1) to the CEU maximization problem (6), I study a model of Bayesian learning where the agent can observe an arbitrarily large amount of statistically relevant (i.i.d. sample) information. As this paper's main conceptual findings (Theorem 2), I demonstrate:

1. Only for the non-generic case in which the neo-additive capacity reduces to an additive probability measure, the agent's probability judgement converges almost surely to the event's true probability.

2. For the generic case in which the neo-additive capacity is non-additive, the agent's probability judgement about any uncertain event converges everywhere to 0.5.

The remainder of the analysis proceeds as follows. Section 2 recalls concepts from Choquet decision theory. Section 3 presents the analytical solution to the CEU maximization problem. The Bayesian learning model is constructed in Section 4. Section 5 concludes.

2 Preliminaries

2.1 Neo-additive capacities and Choquet integration

Fix the measurable space $(\Omega, \mathcal{F})$ and a set of null events $\mathcal{N} \subset \mathcal{F}$. A fuzzy probability measure $\kappa : \mathcal{F} \rightarrow [0,1]$ satisfies

(i) $\kappa(A) = 0$ for $A \in \mathcal{N}$,

(ii) $\kappa(A) = 1$ for A such that $\Omega \backslash A \in \mathcal{N}$,

(iii) $\kappa(A) \leq \kappa(B)$ for A, B such that $A \subset B$.

For reasons of analytical tractability we focus on non-additive probability measures defined as neo-additive capacities (Chateauneuf et al. 2007).

Definition. *Fix some parameters $\delta, \lambda \in [0,1]$. A neo-additive capacity $\nu : \mathcal{F} \rightarrow [0,1]$ is defined as*

$$\nu(A) = \delta \cdot \nu_\lambda(A) + (1-\delta) \cdot \mu(A) \tag{7}$$

for all $A \in \mathcal{F}$ such that μ is some additive probability measure satisfying

$$\mu(A) = \begin{cases} 0 & \text{if } A \in \mathcal{N} \\ 1 & \text{if } \Omega \backslash A \in \mathcal{N} \end{cases} \tag{8}$$

and the non-additive probability measure ν_λ is defined as follows

$$\nu_\lambda(A) = \begin{cases} 0 & \text{iff } A \in \mathcal{N} \\ \lambda & \text{else} \\ 1 & \text{iff } \Omega \backslash A \in \mathcal{N}. \end{cases} \tag{9}$$

Throughout this paper I restrict attention to sets of null-events $\mathcal{N}$ such that $A \in \mathcal{N}$ iff $\mu(A) = 0$. If $0 < \mu(A) < 1$, we call A essential. As a consequence, the neo-additive capacity (7) simplifies to

$$\nu(A) = \delta \cdot \lambda + (1 - \delta) \cdot \mu(A) \tag{10}$$

for essential A. The parameter δ is interpreted as an ambiguity or insensitivity parameter whereas the value of λ determines whether $\nu(A)$ over- or underestimates the additive probability $\mu(A)$.

The Choquet integral of a bounded $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{R}$ with respect to capacity κ is defined as the following Riemann integral extended to domain Ω (Schmeidler 1986):

$$E^C[f, \kappa] = \int_{-\infty}^{0} (\kappa(\{\omega \in \Omega \mid f(\omega) \geq z\}) - 1)\, dz + \int_{0}^{+\infty} \kappa(\{\omega \in \Omega \mid f(\omega) \geq z\})\, dz. \tag{11}$$

Proposition 1. *Let $f : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable function with bounded range. The Choquet expected value (11) of f with respect to a neo-additive capacity (7) is given as*

$$E^C[f, \nu] = \delta(\lambda \sup f + (1 - \lambda) \inf f) + (1 - \delta) E[f, \mu]. \tag{12}$$

Proof: By an argument in Schmeidler (1986), it suffices to restrict attention to a non-negative function f so that

$$E^C[f, \nu] = \int_{0}^{+\infty} \nu(\{\omega \in \Omega \mid f(\omega) \geq z\})\, dz, \tag{13}$$

which is equivalent to

$$E^C[f, \nu] = \int_{\inf f}^{\sup f} \nu(\{\omega \in \Omega \mid f(\omega) \geq z\})\, dz \tag{14}$$

since f is bounded. We consider a partition P_n, $n = 1, 2, ...$, of Ω with members

$$A_n^k = \{\omega \in \Omega \mid a_{k,n} < f(\omega) \leq b_{k,n}\} \text{ for } k = 1, ..., 2^n \tag{15}$$

such that

$$a_{k,n} = [\sup f - \inf f] \cdot \frac{(k-1)}{2^n} + \inf f \tag{16}$$

$$b_{k,n} = [\sup f - \inf f] \cdot \frac{k}{2^n} + \inf f. \tag{17}$$

Define the step functions $a_n : \Omega \to \mathbb{R}$ and $b_n : \Omega \to \mathbb{R}$ such that, for $\omega \in A_n^k$, $k = 1, ..., 2^n$,

$$a_n(\omega) = a_{k,n} \quad (18)$$
$$b_n(\omega) = b_{k,n}. \quad (19)$$

Obviously,

$$E^C[a_n, \nu] \leq E^C[f, \nu] \leq E^C[b_n, \nu] \quad (20)$$

for all n and

$$\lim_{n\to\infty} E^C[b_n, \nu] - E^C[a_n, \nu] = 0. \quad (21)$$

That is, $E^C[a_n, \nu]$ and $E^C[b_n, \nu]$ converge to $E^C[f, \nu]$ for $n \to \infty$. Furthermore, observe that

$$\inf a_n = \inf f \text{ for all } n, \text{ and} \quad (22)$$
$$\sup b_n = \sup f \text{ for all } n. \quad (23)$$

Since $\lim_{n\to\infty} \inf b_n = \lim_{n\to\infty} \inf a_n$ and $E^C[b_n, \mu]$ is continuous in n, we have

$$\lim_{n\to\infty} E^C[b_n, \nu] = \delta\left(\lambda \lim_{n\to\infty} \sup b_n + (1-\lambda) \lim_{n\to\infty} \inf b_n\right) + (1-\delta) \lim_{n\to\infty} E^C[b_n, \mu] \quad (24)$$
$$= \delta(\lambda \sup f + (1-\lambda) \inf f) + (1-\delta) E^C[f, \mu]. \quad (25)$$

In order to prove Proposition 1, it therefore remains to be shown that, for all n,

$$E^C[b_n, \nu] = \delta(\lambda \sup b_n + (1-\lambda) \inf b_n) + (1-\delta) E^C[b_n, \mu]. \quad (26)$$

Since b_n is a step function, (14) becomes

$$E^C[b_n, \nu] = \sum_{A_n^k \in P_n} \nu\left(A_n^{2^n} \cup ... \cup A_n^k\right) \cdot (b_{k,n} - b_{k-1,n}) \quad (27)$$
$$= \sum_{A_n^k \in P_n} b_{k,n} \cdot \left[\nu\left(A_n^{2^n} \cup ... \cup A_n^k\right) - \nu\left(A_n^{2^n} \cup ... \cup A_n^{k-1}\right)\right], \quad (28)$$

implying for a neo-additive capacity

$$E^C[b_n, \nu] = \sup b_n\left[\delta\lambda + (1-\delta)\mu\left(A_n^{2^n}\right)\right] + \sum_{k=2}^{2^n-1} b_{k,n}(1-\delta)\mu\left(A_n^k\right) \quad (29)$$
$$+ \inf b_n\left[1 - \delta\lambda - (1-\delta)\sum_{k=2}^{2^n} \mu\left(A_n^k\right)\right]$$
$$= \delta\lambda \sup b_n + (1-\delta)\sum_{k=1}^{2^n} b_{k,n}\mu\left(A_n^k\right) + \inf b_n[\delta - \delta\lambda] \quad (30)$$
$$= \delta(\lambda \sup b_n + (1-\lambda) \inf b_n) + (1-\delta) E^C[b_n, \mu]. \quad (31)$$

□

2.2 Bayesian Updating of non-additive probability measures

Recall that a Savage act f maps Ω into a set of consequences $\mathbb{R}$, i.e., $f : \Omega \to \mathbb{R}$. For complements $I, \neg I \in \mathcal{F}$ and acts f, h, define the following Savage act

$$f_I h(\omega) = \begin{cases} f(\omega) & \text{for } \omega \in I \\ h(\omega) & \text{for } \omega \in \neg I. \end{cases} \tag{32}$$

Key to the expected utility representation (2) for preferences $\succeq$ over Savage acts is the sure thing principle stating that, for all Savage acts f, g, h, h' and all events $I \in \mathcal{F}$,

$$f_I h \succeq g_I h \;\Rightarrow\; f_I h' \succeq g_I h'. \tag{33}$$

CEU theory has been developed in order to accommodate paradoxes of the Ellsberg (1961) type which show that real-life decision-makers violate Savage's sure thing principle. Abandoning of the sure thing principle gives rise to several perceivable Bayesian update rules for non-additive probability measures (Gilboa and Schmeidler 1993; Sarin and Wakker 1998; Pires 2002; Eichberger et al. 2006; Siniscalchi 2011). To see this recall that a Bayesian update rule specifies how the ex ante preference ordering $\succeq$ determines, for all essential $I \in \mathcal{F}$, the ex post preference ordering $\succeq_I$. Consider, e.g., I, h-Bayesian update rules in the sense that there exists for every essential $I \in \mathcal{F}$ and every pair of Savage acts f, g some Savage act h such that

$$f_I h \succeq g_I h \Rightarrow f_I h' \succeq_I g_I h'' \text{ for all } h', h''. \tag{34}$$

In case the sure thing principle is satisfied, the specification of h in (34) does not matter for deriving ex post preferences. In the case of CEU preferences, however, different specifications of h in (34) result in different ways of updating ex ante CEU into ex post CEU preferences. That is, for the CEU framework there exist several perceivable ways of defining a conditional capacity $\kappa(\cdot \mid I)$ such that

$$f \succeq_I g \Leftrightarrow E^C\left[u(f), \kappa(\cdot \mid I)\right] \geq E^C\left[u(g), \kappa(\cdot \mid I)\right]. \tag{35}$$

For example, Gilboa and Schmeidler (1993) consider the benchmark cases of the optimistic and pessimistic update rules where h corresponds to the constant act giving the worst, respectively best, possible consequence. The corresponding definitions of conditional capacities are

$$\kappa(A \mid I) = \frac{\kappa(A \cap I)}{\kappa(I)}. \tag{36}$$

for the optimistic and

$$\kappa(A \mid I) = \frac{\kappa(A \cup \neg I) - \kappa(\neg I)}{1 - \kappa(\neg I)}. \tag{37}$$

for the pessimistic update rule.

In the present paper I restrict attention to the popular Generalized Bayesian update rule

$$\kappa(A \mid I) = \frac{\kappa(A \cap I)}{\kappa(A \cap I) + 1 - \kappa(A \cup \neg I)}, \tag{38}$$

which results when h in (34) is given as the conditional certainty equivalent of g on I, i.e., h is the constant act such that $g \sim_I h$ (Eichberger et al. 2006). Compared to (36) and (37), (38) is less extreme and arguably more realistic.[1]

Proposition 2. *An application of the Generalized Bayesian update rule (38) to a neo-additive capacity (38) results in the following conditional neo-additive capacity*

$$\nu(A \mid I) = \delta_I \cdot \lambda + (1 - \delta_I) \cdot \mu(A \mid I), \tag{39}$$

for essential $A, I \in \mathcal{F}$, whereby

$$\delta_I = \frac{\delta}{\delta + (1 - \delta) \cdot \mu(I)}. \tag{40}$$

Proof. Observe that

$$\begin{aligned} \nu(A \mid I) &= \frac{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap I)}{\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cap I) + 1 - (\delta \cdot \lambda + (1 - \delta) \cdot \mu(A \cup \neg I))} && (41) \\ &= \frac{\delta \cdot \lambda}{\delta + (1 - \delta) \cdot \mu(I)} + \frac{(1 - \delta) \cdot \mu(I)}{\delta + (1 - \delta) \cdot \mu(I)} \mu(A \mid I) && (42) \\ &= \delta_I \cdot \lambda + (1 - \delta_I) \cdot \mu(A \mid I) && (43) \end{aligned}$$

with δ_I given by (40).□

3 Solving the CEU Maximization Problem

Fix some information $I \in \mathcal{F}$ and consider the conditional neo-additive capacity space $(\nu(\cdot \mid I), \Omega, \mathcal{F})$ such that $\nu(\cdot \mid I)$ is given by (39). By Proposition 1, we have

$$\begin{aligned} & E^C\left[u(f_x), \nu\left(\tilde{\theta} \mid I\right)\right] && (44) \\ = \;& \delta_I \left(\lambda \sup_{x \in (0,1)} u(f_x) + (1 - \lambda) \inf_{x \in (0,1)} u(f_x)\right) + (1 - \delta_I) E\left[u(f_x), \mu\left(\tilde{\theta} \mid I\right)\right]. \end{aligned}$$

[1] Cohen at al. (2000) investigate in an experiment the question whether the pessimistic update rule or the Generalized Bayesian update rule is consistent with ambiguity averse subjects' choice behavior. Their experimental findings establish an approximate ratio of 2:1 in favor for the Generalized Bayesian update rule.

Under the assumption of a quadratic vNM utility function (3), we obtain

$$\sup u(f_x) = 0 \tag{45}$$

as well as

$$\inf u(f_x) = \begin{cases} -x^2 & \text{if } x \geq \frac{1}{2} \\ -(1-x)^2 & \text{if } x \leq \frac{1}{2} \end{cases} \tag{46}$$

Collecting terms gives us the following characterization of the agent's CEU function (44).

Proposition 3. *The agent's objective function*

$$E^C\left[-\left(x-\tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I\right)\right] \tag{47}$$

satisfies

$$\begin{aligned} & E^C\left[-\left(x-\tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I\right)\right] \\ = \;& \delta_I(1-\lambda)(1-x)^2 + (1-\delta_I) E\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \end{aligned} \tag{48}$$

if $x \in \left(0, \frac{1}{2}\right]$, *and*

$$\begin{aligned} & E^C\left[-\left(x-\tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I\right)\right] \\ = \;& \delta_I(1-\lambda)x^2 + (1-\delta_I) E\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \end{aligned} \tag{49}$$

if $x \in \left[\frac{1}{2}, 1\right)$; *with* δ_I *given by (40).*

To exclude the trivial case where (48) and (49) are constantly zero, I henceforth assume that

$$\text{either } \delta_I < 1 \text{ or } \delta_I = 1,\ \lambda < 1. \tag{50}$$

Theorem 1. *Let*

$$\begin{aligned} x_1 &\equiv \frac{\delta_I - \delta_I\lambda + (1-\delta_I) E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]}{\delta_I - \delta_I\lambda + 1 - \delta_I}, && (51) \\ x_2 &\equiv \frac{(1-\delta_I)}{(1-\delta_I\lambda)} E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]. && (52) \end{aligned}$$

The analytical solution

$$x_I^C = \arg \sup_{x\in(0,1)} E^C\left[-\left(x-\tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I\right)\right] \tag{53}$$

is comprehensively described as follows.

(a) *If*

$$x_1 < \frac{1}{2} \text{ and } x_2 > \frac{1}{2} \tag{54}$$

then

$$x_I^C = \arg \max_{\{x_1,x_2\}} E^C\left[-\left(x-\tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I\right)\right]. \tag{55}$$

(b) *If*

$$x_1 \leq \frac{1}{2} \text{ and } x_2 \leq \frac{1}{2} \tag{56}$$

then

$$x_I^C = x_1. \tag{57}$$

(c) *If*

$$x_1 \geq \frac{1}{2} \text{ and } x_2 \geq \frac{1}{2} \tag{58}$$

then

$$x_I^C = x_2. \tag{59}$$

(d) *If*

$$x_1 \geq \frac{1}{2} \text{ and } x_2 \leq \frac{1}{2} \tag{60}$$

then

$$x_I^C = \frac{1}{2}. \tag{61}$$

Proof. Step 1. Observe at first that

$$E\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \tag{62}$$

is locally uniformly integrably bounded because $\mu\left(\tilde{\theta} \mid I\right)$ is finite and $-\left(x-\tilde{\theta}\right)^2$ is continuous in x and measurable in $\tilde{\theta}$ as well as bounded. Similarly, the continuous and $\tilde{\theta}$-measurable partial derivative function

$$\frac{d\left(-\left(x-\tilde{\theta}\right)^2\right)}{dx} \tag{63}$$

is locally uniformly integrably bounded with respect to $\mu\left(\tilde{\theta} \mid I\right)$. As a consequence (cf. Theorem 16.8 in Billingsley 1995), (62) is continuously differentiable in x whereby

$$\frac{dE\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right]}{dx} = E\left[\frac{d\left(-\left(x-\tilde{\theta}\right)^2\right)}{dx}, \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{64}$$

Step 2. Focus on the function (48) and observe that it is, by assumption (50), strictly concave. Furthermore, (48) is, by (64), continuously differentiable with

$$\frac{d\left(\delta_I(1-\lambda)(1-x)^2 + (1-\delta_I) E\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right]\right)}{dx} \tag{65}$$

$$= \frac{d\left(\delta_I(1-\lambda)(1-x)^2\right)}{dx} + (1-\delta_I) E\left[\frac{d\left(-\left(x-\tilde{\theta}\right)^2\right)}{dx}, \mu\left(\tilde{\theta} \mid I\right)\right] \tag{66}$$

$$= \delta_I(1-\lambda)\cdot 2(1-x) + (1-\delta_I)\int_{\omega\in\Omega} -(2x-2\theta)\, d\mu\left(\tilde{\theta} \mid I\right). \tag{67}$$

At $x=0$ (67) is, by assumption (50), strictly greater zero implying

$$0 \neq \arg\sup_{x\in(0,1)} E^C\left[-\left(x-\tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I\right)\right]. \tag{68}$$

Consequently, there exists a maximum x_1 of function (48) on the interval $\left(0, \frac{1}{2}\right]$ which is either $x_1 \geq \frac{1}{2}$ characterized by the FOC or given as the boundary solution $x_1 = \frac{1}{2}$. From the FOC we obtain

$$\delta_I(1-\lambda)\cdot 2(1-x_1) + (1-\delta_I)\int_{\omega\in\Omega} -(2x_1-2\theta)\, d\mu\left(\tilde{\theta} \mid I\right) = 0 \tag{69}$$

$$\Leftrightarrow$$

$$\frac{\delta_I - \delta_I\lambda}{(\delta_I - \delta_I\lambda + 1 - \delta_I)} + \frac{(1-\delta_I)}{(\delta_I - \delta_I\lambda + 1 - \delta_I)} E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right] = x_1. \tag{70}$$

That is, whenever $x_1 \in \left(0, \frac{1}{2}\right]$ it is a local maximizer of the objective function (47).

Step 3. Turn now to the function (49), which is also strictly concave. (49) is continuously differentiable with

$$\frac{d\left(\delta_I\left(1-\lambda\right)x^2+\left(1-\delta_I\right)E\left[-\left(x-\tilde{\theta}\right)^2,\mu\left(\tilde{\theta}\mid I\right)\right]\right)}{dx} \tag{71}$$

$$= \frac{d\left(\delta_I\left(1-\lambda\right)x^2\right)}{dx}+\left(1-\delta_I\right)E\left[\frac{d\left(-\left(x-\tilde{\theta}\right)^2\right)}{dx},\mu\left(\tilde{\theta}\mid I\right)\right] \tag{72}$$

$$= -2\delta_I\left(1-\lambda\right)\cdot x+\left(1-\delta_I\right)\int_{\omega\in\Omega}-\left(2x-2\theta\right)d\mu\left(\tilde{\theta}\mid I\right). \tag{73}$$

Because (73) is strictly decreasing at $x=1$, we have

$$1\neq\arg\sup_{x\in(0,1)}E^C\left[-\left(x-\tilde{\theta}\right)^2,\nu\left(\tilde{\theta}\mid I\right)\right]. \tag{74}$$

Notice that there exists a maximum x_2 of function (49) on the interval $\left[\frac{1}{2},1\right)$ which is either $x_2\geq\frac{1}{2}$ characterized by the FOC or given as the boundary solution $x_2=\frac{1}{2}$. Solving the FOC gives

$$-2\delta_I\left(1-\lambda\right)\cdot x_2+\left(1-\delta_I\right)\int_{\omega\in\Omega}-\left(2x_2-2\theta\right)d\mu\left(\tilde{\theta}\mid I\right) = 0 \tag{75}$$

$$\Leftrightarrow$$

$$\frac{\left(1-\delta_I\right)}{\left(1-\delta_I\lambda\right)}E\left[\tilde{\theta},\mu\left(\tilde{\theta}\mid I\right)\right] = x_2, \tag{76}$$

implying that x_2 is a local maximizer of (47) iff $x_2\in\left[\frac{1}{2},1\right)$.

Step 4. If condition (54) holds, we have thus two local maximizers, x_1 and x_2, of (47), characterized by FOCs (70) and (76) respectively, and whichever is greater is the global maximizer for (47). This proves (a).

Step 5. If condition (56) holds, we have one local maximizer x_1 characterized by the FOC (70). Since $x_1\geq\frac{1}{2}$, with $\frac{1}{2}$ being the boundary maximum of (47) on the interval $\left[\frac{1}{2},1\right)$, x_1 is also the global maximizer for (47). This proves (b).

Step 6. If condition (58) holds, we have one local maximizer $x_2\geq\frac{1}{2}$ characterized by the FOC (76) , which is also the global maximizer for (47). This proves (c).

Step 7. If condition (60) holds, there is no local maximizer characterized by any FOC. Instead (47) takes on its maximum at the kink $x=\frac{1}{2}$. This proves (d).□

The following corollary establishes that the classical point estimate of Bayesian statistics—given as the expected parameter value with respect to a subjective posterior (additive) probability distribution—is nested in the solution of Theorem 1 whenever the neo-additive capacity reduces to an additive probability measure.

Corollary.

(a) *If* $\delta = 0$*, the analytical solution (53) becomes*

$$x_I^C = E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{77}$$

(b) *Consider the generic case* $E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right] \neq \frac{1}{2}$*. Only if* $\delta = 0$*, the analytical solution (53) becomes (77).*

Proof. Case (a) is trivial: If $\delta = 0$, then

$$E^C\left[-\left(x - \tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I\right)\right] = E\left[-\left(x - \tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \tag{78}$$

for which (77) is the global maximizer.

Ad case (b). Suppose, on the contrary, that

$$x_I^C = E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right] \neq \frac{1}{2}. \tag{79}$$

In that case, either $x_I^C = x_1$ or $x_I^C = x_2$ because x_I^C must coincide with some local maximizer characterized by the corresponding FOC. However, if $\delta > 0$, then

$$E\left[-\left(x - \tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \neq x_1 \text{ and } E\left[-\left(x - \tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \neq x_2, \tag{80}$$

by (50), implying

$$x_I^C \neq E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{81}$$

□

4 Bayesian Learning of Probability Judgements

Consider some measurable space $(\Omega', \mathcal{A})$ and fix an arbitrary essential $A \in \mathcal{A}$. The agent's probability judgement will refer to the occurrence of event A. To describe a standard framework of Bayesian learning of such probability judgements let us impose the following structure on the information space:

$$\mathbb{I} = \times_{k=1}^{\infty} S_k \tag{82}$$

such that, for all k,

$$S_k = \{A, \neg A\}. \tag{83}$$

S_k collects the possible outcomes of the k-th statistical trial according to which either A or the complement of A occurs. Denote by $\sigma(S_k)$ the σ-algebra generated by the powersets of S_1,...,S_k and define the following σ-algebra

$$\mathcal{F}_k = \{\Theta \times S \mid S \in \sigma(S_k)\}. \tag{84}$$

Observe that (84) constitutes a filtration, i.e., $\mathcal{F}_1 \subset \mathcal{F}_2 \subset .... \subset \mathcal{F}$ where $\mathcal{F}$ is generated by $\mathcal{F}_1, \mathcal{F}_2, ...$ Any event

$$I_n = \Theta \times \{s_1\} \times ... \times \{s_n\} \times S_{n+1} \times ... \in \mathcal{F} \tag{85}$$

with $s_j \in S_j$, $j = 1, ..., n$, is interpreted as possible sample information that the agent may have observed after the n-th statistical trial.

Consider now the additive probability space $(\mu, \Omega, \mathcal{F})$ with Ω defined in (1) and $\mathcal{F}$ defined above. Further suppose that the trials are, conditional on $\tilde{\theta}(\omega)$, i.i.d. such that A occurs in every trial with true probability $\tilde{\theta}(\omega)$. That is,

$$\mu\left(I_n \mid \tilde{\theta}\right) = \prod_{j=1}^{n} \pi_{\tilde{\theta}}(s_j) \tag{86}$$

such that

$$\pi_{\tilde{\theta}}(s_j) = \begin{cases} \tilde{\theta} & \text{if } s_j = A \\ 1 - \tilde{\theta} & \text{if } s_j = \neg A \end{cases} \tag{87}$$

By Bayes' rule, we obtain the posterior $\mu\left(\tilde{\theta} \mid I_n\right)$ such that, for any Borel set B in $\Theta = (0, 1)$,

$$\mu(B \mid I_n) = \frac{\int_{\theta \in B} \mu\left(I_n \mid \tilde{\theta}\right) d\mu\left(\tilde{\theta}\right)}{\mu(I_n)} \tag{88}$$

$$= \frac{\int_{\theta \in B} \prod_{j=1}^{n} \pi_{\tilde{\theta}}(s_j)\, d\mu\left(\tilde{\theta}\right)}{\int_{\theta \in (0,1)} \prod_{j=1}^{n} \pi_{\tilde{\theta}}(s_j)\, d\mu\left(\tilde{\theta}\right)}. \tag{89}$$

Recall that Doob's (1949) consistency theorem[2] implies that, for almost all true parameter values $\tilde{\theta}(\omega)$ belonging to the support of μ, the posterior distribution $\mu\left(\tilde{\theta} \mid I_n\right)$

[2]For a comprehensive discussion of Doob's theorem see Gosh and Ramamoorthi (2003) and Lijoi et al. (2004).

concentrates with μ-probability one around the true value $\tilde{\theta}(\omega)$ as n gets large, i.e.,

$$\mu\left(\tilde{\theta} \mid I_n\right) \to \mathbf{1}_B \tilde{\theta}(\omega), \ \mu\text{-a.s.} \tag{90}$$

for any Borel set B in $\Theta = (0,1)$ where $\mathbf{1}_B$ denotes the indicator function of B. Applied to the additive Bayesian estimate of the sophisticated agent (77), Doob's theorem gives us immediately the following convergence result.

Proposition 4. *Let $\mu\left(\tilde{\theta}\right)$ have full support on $(0,1)$. For the non-generic case $\delta = 0$, the agent's probability judgement about any essential event $A \in \mathcal{A}$ will almost surely converge to A's true probability if the sample size n gets large, i.e.,*

$$x^C_{I_n} \equiv E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I_n\right)\right] \to \tilde{\theta}(\omega), \ \mu\text{-a.s.} \tag{91}$$

Since Proposition 4 holds for any event in $\mathcal{A}$, the probability judgements about all events in $\mathcal{A}$ will converge through Bayesian learning towards an additive probability measure on the space $(\Omega', \mathcal{A})$ that resembles an "objective" probability measure. Whenever (91) holds, I speak of a (statistically) sophisticated agent.

The situation is different for the generic case of a non-additive capacity: Theorem 2 states this paper's main result according to which the generic agent converges to a (statistically) ignorant agent whose probability judgement is given as a fuzzy probability measure that attaches probability 0.5 to any uncertain event in $\mathcal{A}$.

Theorem 2. *For any $\delta \in (0,1]$ and $\lambda \in [0,1)$, the agent's probability judgement about any essential event $A \in \mathcal{A}$ will converge, for all $\omega \in \Omega$, to 0.5 if the sample size n gets large, i.e.,*

$$x^C_{I_n} \equiv \arg\sup_{(0,1)} E^C\left[-\left(x - \tilde{\theta}\right)^2, \nu\left(\tilde{\theta} \mid I_n\right)\right] \to \frac{1}{2}, \ \Omega\text{-everywhere.} \tag{92}$$

Proof. Step 1. Let

$$\theta_{\max} = \max\{\theta, 1-\theta\} \tag{93}$$

and observe that, for any I_1 with $s_1 \in \{A, \neg A\}$,

$$\mu(I_1) = \int_{(0,1)} \pi_{\tilde{\theta}}(s_1)\, d\mu\left(\tilde{\theta}\right) \tag{94}$$

$$\leq \int_{(0,1)} \theta_{\max} d\mu\left(\tilde{\theta}\right) \tag{95}$$

$$= \int_{\left(0,\frac{1}{2}\right)} (1-\theta)\, d\mu\left(\tilde{\theta}\right) + \int_{\left[\frac{1}{2},1\right)} \theta d\mu\left(\tilde{\theta}\right) \equiv c < 1. \tag{96}$$

Further observe that, for all I_n,

$$\mu(I_n) \leq c^n, \tag{97}$$

implying

$$\lim_{n\to\infty} \mu(I_n) = 0. \tag{98}$$

Step 2. Notice that (98) together with (40) implies

$$\lim_{n\to\infty} \delta_{I_n} \to 1 \tag{99}$$

for any I_n.

Step 3. Consider

$$x_{1,I_n} \equiv \frac{\delta_{I_n} - \delta_{I_n}\lambda + (1-\delta_{I_n}) E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I_n\right)\right]}{\delta_{I_n} - \delta_{I_n}\lambda + 1 - \delta_{I_n}}. \tag{100}$$

Because of

$$0 < E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I_n\right)\right] < 1 \tag{101}$$

for all $\omega \in \Omega$, we have

$$\underline{x}_{1,I_n} \leq x_{1,I_n} \leq \overline{x}_{1,I_n} \tag{102}$$

for all $\omega \in \Omega$ and all n, whereby

$$\underline{x}_{1,I_n} \equiv \frac{\delta_{I_n} - \delta_{I_n}\lambda}{\delta_{I_n} - \delta_{I_n}\lambda + 1 - \delta_{I_n}}, \tag{103}$$

$$\overline{x}_{1,I_n} \equiv \frac{\delta_{I_n} - \delta_{I_n}\lambda + (1-\delta_{I_n})}{\delta_{I_n} - \delta_{I_n}\lambda + 1 - \delta_{I_n}}. \tag{104}$$

By (99), we have

$$\lim_{n\to\infty} \underline{x}_{1,I_n} = \frac{1-\lambda}{1-\lambda+1-1}, \ \Omega\text{-everywhere} \tag{105}$$

$$= 1 \text{ for all } \lambda < 1, \tag{106}$$

as well as

$$\lim_{n\to\infty} \overline{x}_{1,I_n} = \frac{1-\lambda+1-1}{1-\lambda+1-1}, \ \Omega\text{-everywhere} \tag{107}$$

$$= 1 \text{ for all } \lambda < 1. \tag{108}$$

Consequently,

$$x_{1,I_\infty} \equiv \lim_{n\to\infty} x_{1,I_n} = 1, \ \Omega\text{-everywhere for all } \lambda < 1. \tag{109}$$

Step 4. Turn to

$$x_{2,I_n} \equiv \frac{(1-\delta_{I_n})}{(1-\delta_{I_n}\lambda)} E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I_n\right)\right] \tag{110}$$

and observe that, for all $\omega \in \Omega$ and all n,

$$\underline{x}_{2,I_n} \leq x_{2,I_n} \leq \overline{x}_{2,I_n} \tag{111}$$

with

$$\underline{x}_{2,I_n} \equiv 0, \tag{112}$$

$$\overline{x}_{2,I_n} \equiv \frac{(1-\delta_{I_n})}{(1-\delta_{I_n}\lambda)}. \tag{113}$$

Because of

$$\lim_{n\to\infty} \overline{x}_{2,I_n} = \frac{0}{1-\lambda}, \ \Omega\text{-everywhere} \tag{114}$$

$$= 0 \text{ for all } \lambda < 1, \tag{115}$$

we have

$$x_{2,I_\infty} \equiv \lim_{n\to\infty} x_{2,I_n} = 0, \ \Omega\text{-everywhere for all } \lambda < 1. \tag{116}$$

Step 5. Collecting results establishes

$$x_{1,I_\infty} > \frac{1}{2} \text{ and } x_{2,I_\infty} < \frac{1}{2}, \ \Omega\text{-everywhere} \tag{117}$$

so that part (d) of Theorem 1 implies the desired result (92).□

5 Concluding Remarks

Referring to the "new psychological concept" of cognitive likelihood insensitivity, Peter Wakker (2010) demands that "new mathematical tools have to be developed to analyze this phenomenon" (p. 227). The present paper has done exactly that: Based on technical tools from fuzzy measure theory, the agent's probability judgements have been formally described as the solution to a CEU maximization problem subject to Bayesian learning of neo-additive capacities. The conceptual main result establishes that all probability judgements converge to "fifty-fifty" answers whenever the agent's neo-additive capacity is not given as an additive probability measure. This generic convergence result might thus contribute towards an explanation of why "fifty-fifty" judgements are the predominant empirical phenomenon of likelihood insensitivity.

References

Anscombe, F.J., and R.J. Aumann (1963) "A definition of subjective probability," *Annals of American Statistics* **34**, 199-205.

Basili, M., and A. Chateauneuf (2011) "Extreme events and entropy: A multiple quantile utility model," *International Journal of Approximate Reasoning* **52**, 1095-1102.

Billingsley, P. (1995), *Probability and Measure*, John Wiley and & Sons, Inc.: New York.

Bruine de Bruin, W., Fischhoff, B, Millstein, S.G., and B.C. Halpern-Felsher (2000) "Verbal and numerical expressions of probability: Its a fifty-fifty chance," *Organizational Behavior and Human Decision Processes* **81**, 115-131.

Camerer, C.F. (2007) "Neuroeconomics: Using neuroscience to make economic predictions," *Economic Journal* **117**, C26–C42.

Chateauneuf, A., Eichberger, J., and S. Grant (2007) "Choice under uncertainty with the best and worst in mind: Neo-additive capacities," *Journal of Economic Theory* **127**, 538-567.

Cohen, M., Gilboa, I., Jaffray J.Y., and D. Schmeidler (2000) "An experimental study of updating ambiguous beliefs," *Risk, Decision, and Policy* **5**, 123–133.

Doob, J. L. (1949) "Application of the theory of martingales," In: *Le Calcul des Probabilites et ses Applications*, Colloques Internationaux du Centre National de la Recherche Scientifique 13, pp. 23–27. Paris: CNRS.

Eichberger, J., Grant, S., and D. Kelsey (2006) "Updating Choquet beliefs," *Journal of Mathematical Economics* **43**, 888-899.

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

Gilboa, I. (1987) "Expected utility with purely subjective non-additive probabilities," *Journal of Mathematical Economics* **16**, 65-88.

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

Girshick, M.A., and L.J. Savage (1951) "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, p.53-73.

Gosh, J.K., and R.V. Ramamoorthi (2003) *Bayesian Nonparametrics*, Springer Verlag Inc.: New York.

Grabisch, M., Nguyen, H.T., and E.A. Walker (1995) *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*, Kluwer Academic Publishers: Dordrecht.

Hurd, M.D. (2009) "Subjective probabilities in household surveys," *Annual Review of Economics* **1**, 543-562.

James, W., and C. Stein (1961) "Estimation with quadratic loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, p.361-379.

Kahneman, D., and A. Tversky (1979) "Prospect theory: An analysis of decision under risk," *Econometrica* **47**, 263-291.

Kleinjans, K.J., and A. Van Soest (2010) "Nonresponse and focal point answers to subjective probability questions," IZA Discussion Papers 5272, Institute for the Study of Labor (IZA).

Lijoi, A., Pruenster, I., and S. O. Walker (2004) "Extending Doob's consistency theorem to nonparametric densities," *Bernoulli* **10**, 651-663.

Manski, C.F, and F. Molinari (2010) "Rounding probabilistic expectations in surveys," *Journal of Business and Economic Statistics* **28**, 219-231.

Murofushi T., and M. Sugeno (1989) "An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems* **29**, 201-227.

Murofushi T., and M. Sugeno (1991) "A theory of fuzzy measures: Representations, the Choquet integral and null sets," *Journal of Mathematical Analysis and Applications* **159**, 532-549.

Narukawa, Y. (2007) "Inner and outer representation by Choquet integral," *Fuzzy Sets and Systems* **158**, 963-972.

Narukawa, Y. and T. Murofushi (2003) "Conditions for Choquet integral representation of the comonotonically additive and monotone functional," *Journal of Mathematical Analysis and Applications* **282**, 201–211.

Narukawa, Y. and T. Murofushi (2004) "Regular non-additive measure and Choquet integral," *Fuzzy Sets and Systems* **143**, 487–492.

Narukawa, Y., Murofushi, T., and M. Sugeno (2000) "Regular fuzzy measure and representation of comonotonically additive functionals," *Fuzzy Sets and Systems* **112**, 177–186.

Narukawa, Y., Murofushi, T., and M. Sugeno (2001) "Extension and representation of comonotonically additive functionals," *Fuzzy Sets and Systems* **121**, 217–226.

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

Rubinstein, A. (2003) "Economics and psychology? The case of hyperbolic discounting," *International Economic Review* **44**, 1207—1216.

Sarin, R., and P.P. Wakker (1992) "A simple axiomatization of nonadditive expected utility," *Econometrica* **60**, 1255-1272.

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

Savage, L.J. (1954) *The Foundations of Statistics*, John Wiley and & Sons, Inc.: New York, London, Sydney.

Schmeidler, D. (1986) "Integral representation without additivity," *Proceedings of the American Mathematical Society* **97**, 255-261.

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

Siniscalchi, M. (2011) "Dynamic choice under ambiguity," *Theoretical Economics* **6**, 379-421.

Sugeno, M., Narukawa, Y., and T. Murofushi (1998) "Choquet integral and fuzzy measures on locally compact space," *Fuzzy Sets and Systems* **99**, 205–211.

Tversky, A., and D. Kahneman (1992) "Advances in prospect theory: Cumulative representations of uncertainty," *Journal of Risk and Uncertainty* **5**, 297-323.

van Santen, P., Alessie, R., and A. Kalwij (2012) "Probabilistic survey questions and incorrect answers: Retirement income replacement rates," *Journal of Economic Behavior and Organization* **82**, 267-280.

Wakker, P.P (2004) “On the composition of risk preference and belief,” *Psychological Review* **111**, 236-241.

Wakker, P.P. (2010) *Prospect Theory: For Risk and Ambiguity* Cambridge University Press: Cambridge, UK.

Wakker, P.P, and A. Tversky (1993) “An axiomatization of cumulative prospect theory,” *Journal of Risk and Uncertainty* **7**, 147-176.