The emergence of "fifty-fifty" probability judgments through Bayesian updating under ambiguity*

Alexander Zimper†

December 9, 2012

Abstract

This paper explains the empirical phenomenon of persistent "fifty-fifty" probability judgments through a model of Bayesian updating under ambiguity. To this purpose I characterize an announced probability judgment as a Bayesian estimate given as the solution to a Choquet expected utility maximization problem with respect to a neo-additive capacity that has been updated in accordance with the Generalized Bayesian update rule. Only for the non-generic case, in which this capacity degenerates to an additive probability measure, the agent will learn the event's true probability if the number of i.i.d. data observations gets large. In contrast, for the generic case in which the capacity is not additive, the agent's announced probability judgment becomes a persistent "fifty-fifty" probability judgment after finitely many observations.

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

*I would like to thank two anonymous referees for their very helpful comments and suggestions. Financial support from ERSA (Economic Research Southern Africa) is gratefully acknowledged.

†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

1.1 Motivation

Let us consider two opposite benchmark cases of how agents form and revise probability judgments. On the one hand, there is the statistically sophisticated agent whose probability judgments are 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 judgment regardless of any new information. More precisely, I speak of a "fifty-fifty" probability judgment if for any given pair of uncertain complementary events $A, \neg A$ the agent announces that either event occurs with 0.5 probability.

The persistence of such "fifty-fifty" probability judgments 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" judgments 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 judgments, the question arises why "fifty-fifty" judgments 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 Bayesian estimation under ambiguity which I combine with a Bayesian learning model such that the agent can observe i.i.d. statistical trials whose outcomes are either A or $\neg A$. The resulting model of Bayesian learning under ambiguity turns out to encompass the announced probability judgments of sophisticated agents as the non-generic and the announced probability judgments of ignorant agents as the generic case.

Formally, my model is based on the assumption that decision makers are Choquet expected utility (=CEU) rather than expected utility (=EU) maximizers. Behavioral axioms that generalize EU to CEU theory 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 purely subjective Savage (1954) framework.

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. Because my formal approach thus admits for an interpretation in terms of behavioral axioms (Gilboa 1987; Chateauneuf et al. 2007; Eichberger et al. 2007), it contributes towards "opening the black box of decision makers instead of modifying functional forms" (Rubinstein 2003, p. 1207).

1.2 The set-up

My point of departure is the formal description of announced probability judgments about some family of complementary events $A, \neg A$ whereby the agent can observe the outcome of statistical trials resulting each time in the occurrence of either A or $\neg A$. To fix ideas consider the following illustrative example.

Example. Suppose that the agent knows the total (large) number of balls in a given urn as well as the fact that each ball has some color y, with $y \in \{blue, red, green\}$, but that he has no information whatsoever about the proportions of differently colored balls within this urn. A state $\omega' \in \Omega'$ stands for the fact that a specific ball is drawn. Fix some color y and define $A \equiv A_y$ as the event that a ball of color y will be drawn. In a next step, we allow the agent to observe statistical trials according to which balls are drawn from this urn and put back after their color has been revealed to the agent. Assume that these trials are i.i.d. such that, for example, the "true" probability $\theta \in (0,1)$ of A_y corresponds to the actual proportion of y-colored balls in the urn. If the agent is an ignorant agent in our sense, he will announce after finitely many observations that the chance of drawing versus not-drawing a ball of color y is "fifty-fifty". Now we can switch to a different color and we will observe the same "fifty-fifty" judgment after sufficiently many statistical trials. That is, the ignorant agent in our sense will eventually attach a fifty-fifty chance to all complementary event pairs $A_y, \neg A_y$ with $y \in \{blue, red, green\}$.

More generally, consider the measurable space $(\Omega', \mathcal{A}')$ and fix some event $A \in \mathcal{A}'$ such that (i) $\theta \in (0,1)$ denotes the true probability of A and (ii) the agent deems it impossible that A might have probability one or zero. Let $\mathcal{A} = \{\emptyset, A, \neg A, \Omega'\}$. The agent's probability judgments will then apply to some family of "binary" measurable spaces $\{(\Omega', \mathcal{A})\}_{A \in \mathcal{A}'}$.

Fix an arbitrary $(\Omega', \mathcal{A})$ and construct the measurable space $(\Omega, \mathcal{F})$ as follows. Define the state space

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

with generic element $\omega \equiv (\theta, i)$. The parameter space $\Theta = (0,1)$ collects all possible probabilities of event A.[1] Endow Θ with the Euclidean metric and denote by $\mathcal{B}$ the corresponding Borel σ-algebra on Θ. The information space $\mathbb{I}$ is given as the infinite sample space

$$\mathbb{I} = \times_{j=1}^{\infty} S_j \tag{2}$$

such that, for all j,

$$S_j = \{A, \neg A\}. \tag{3}$$

S_j collects the possible outcomes of the j-th statistical trial according to which either A or $\neg A$ occurs. Define $\mathcal{I}$ as the standard product algebra of $(\mathcal{A})_{j=1}^{\infty}$, i.e., the σ-algebra generated by the sets in

$$\left\{ \times_{j=1}^{\infty} B_j \mid B_j \in \mathcal{A} \text{ such that } B_j \neq \Omega' \text{ for finitely many } B_j \right\}. \tag{4}$$

Consequently, the agent will always only receive a finite amount of sample information. The relevant event space $\mathcal{F}$ is then defined as the standard product algebra of $\mathcal{B}$ and $\mathcal{I}$.

[1] In line with the assumption that the agent deems it impossible that A has probability one or zero, I consider $\Theta = (0,1)$ rather than $\Theta = [0,1]$. Mathematically, the restriction to $\Theta = (0,1)$ also has the advantage that I do not have to consider degenerate measures which put all probability mass on 0 or 1. (For example, for such measures Bayesian updating is not well-defined for all possible sample observations. Also see footnote 7.)

I assume that the agent is a Choquet Bayesian decision maker in the sense that his beliefs are characterized by (i) a unique—not necessarily additive—subjective probability measure on $(\Omega, \mathcal{F})$ and (ii) that he forms conditional—not necessarily additive—probability measures in accordance with some Bayesian update rule for CEU decision makers (see Section 2.2 for details).[2] More specifically, let us focus thoughts and assume that the agent's subjective probability measure is pinned down by his CEU preferences over Savage acts defined on the state space Ω. Recall that a Savage act f maps the state space into some set of consequences, i.e., $f : \Omega \rightarrow Z$, whereby I assume throughout the paper that $Z = [0, 1]$. For a given constant $x \in (0, 1)$ define the $\mathcal{F}$-measurable Savage act $f_x : \Omega \rightarrow [0, 1]$ such that $f_x(\theta, i) = |x - \theta|$. I interpret f_x as the Savage act that corresponds to the probability judgment x whereby the consequence of f_x in state (θ, i) is given as the Euclidean distance between the agent's probability judgment x and the true probability θ.

An *announced probability judgment* x^* is then formally characterized by the preference maximizing act f_{x^*} over all f_x, $x \in (0, 1)$. Through an announced probability judgment the agent thus expresses his probabilistic forecast about the occurrence of an uncertain event such as, e.g., tomorrow's weather conditions, next year's economic growth, or whether the agent will die at age $y \in \{65, ...100\}$.

In the remainder of this introduction I describe this preference maximization problem, at first, for the special case of a standard expected utility and, afterwards, for the general case of a Choquet expected utility maximizer.

1.3 The standard approach: Expected utility decision makers

Consider at first the special case of a CEU decision maker who reduces to a standard expected utility (=EU) decision maker. To focus thoughts let us assume that this EU decision maker is a subjective EU maximizer with preferences $\succeq$ over Savage acts on Ω that satisfy Savage's (1954) axioms. For this EU decision maker there exists a bounded utility function $u : Z \equiv [0, 1] \rightarrow \mathbb{R}$, unique up to a positive linear transformation, and a unique additive probability measure μ on $(\Omega, \mathcal{F})$ such that for all Savage acts f_x, f_y,

[2] Especially in the Statistics and Econometrics literature, the term "Bayesian agent" is usually associated with subjective additive probability measures. The situation is, in my opinion, somewhat different in Decision Theory where a "Bayesian agent" is distinguished from an, e.g., case-based or an heuristic agent, in that his likelihood considerations in the light of new information are formalized through some mathematical definition of subjective, not necessarily additive, conditional probability measures. To avoid terminological confusions, I speak of a Choquet Bayesian rather than a Bayesian agent whenever this agent is characterized by subjective, not necessarily additive, conditional probability measures.

$x, y \in (0,1)$:

$$f_x \succeq f_y \Leftrightarrow E\left[u\left(\left|x-\tilde{\theta}\right|\right), \mu\left(\tilde{\theta}\right)\right] \geq E\left[u\left(\left|y-\tilde{\theta}\right|\right), \mu\left(\tilde{\theta}\right)\right], \tag{5}$$

where $\tilde{\theta}(\theta, i) = \theta$ denotes the $\mathcal{F}$-measurable random variable that gives for every state (θ, i) of the world the true probability $\theta \in (0,1)$ and E denotes the expectation operator with respect to $\mu\left(\tilde{\theta}\right)$.

Let

$$I \in \{\Theta \times I' \mid I' \in \mathcal{I}\} \tag{6}$$

denote some sample information in $\mathcal{F}$. Denote by $\succeq_I$ an ex post preference ordering which we interpret as the agent's preferences over Savage acts after he has learnt information I. If the agent updates his preferences $\succeq$ to $\succeq_I$ in the standard Bayesian way, we obtain for an EU decision maker that, for all Savage acts f_x, f_y, $x, y \in (0,1)$:

$$f_x \succeq_I f_y \Leftrightarrow E\left[u\left(\left|x-\tilde{\theta}\right|\right), \mu\left(\tilde{\theta} \mid I\right)\right] \geq E\left[u\left(\left|y-\tilde{\theta}\right|\right), \mu\left(\tilde{\theta} \mid I\right)\right] \tag{7}$$

where

$$\mu(B \mid I) = \frac{\mu(B \cap I)}{\mu(I)} \tag{8}$$

for $B \in \{B' \times \mathbb{I} \mid B' \in \mathcal{B}\}$. The solution to the EU maximization problem

$$x_I^* = \arg \sup_{x \in (0,1)} E\left[u\left(\left|x-\tilde{\theta}\right|\right), \mu\left(\tilde{\theta} \mid I\right)\right] \tag{9}$$

stands for this EU decision maker's announced probability judgment in the light of sample information I.

Observe that the announced probability judgment (9) is the standard estimate of Bayesian statistics which minimizes the expectation of some loss function $l = -u$, defined over absolute errors, with respect to some posterior distribution (cf., e.g., Berger 1980). For example, if the loss function is quadratic, i.e.,

$$u(|x-\theta|) = -(x-\theta)^2, \tag{10}$$

it can be shown (cf. Section 3) that (9) becomes

$$\begin{aligned} x_I^* &= \arg \sup_{x \in (0,1)} E\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] && (11) \\ &= E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]. && (12) \end{aligned}$$

That is, under the assumption of a negative quadratic utility function the EU agent's announced probability judgment x_I^* coincides with the classical Bayesian estimate given as the expected parameter value with respect to the posterior distribution $\mu\left(\tilde{\theta} \mid I\right)$ (cf., Girshick and Savage 1951; James and Stein 1961; Section 35 in Billingsley 1995).

The example revisited for a negative quadratic utility function. Fix some color $y \in \{blue, red, green\}$ and suppose that $\mu\left(\tilde{\theta}\right)$ corresponds to a Beta distribution over $(0,1)$ with parameters $\alpha, \beta > 0$ so that, prior to any sample information, the agent's announced probability judgment (11) about drawing an y-colored ball, i.e., event A_y, is given as

$$E\left[\tilde{\theta}, \mu\left(\tilde{\theta}\right)\right] = \frac{\alpha}{\alpha + \beta}. \tag{13}$$

Denote by $I \equiv I_n^k$ the sample information according to which k balls of color y have been drawn in n (Bernoulli) trials. It can be easily shown (cf. Zimper 2011 and references therein) that the posterior distribution $\mu\left(\tilde{\theta} \mid I_n^k\right)$ is itself a Beta distribution with parameters $\alpha+k, \beta+n-k$ so that the agent's announced probability judgment (11) in the light of sample information I_n^k becomes

$$E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I_n^k\right)\right] = \frac{\alpha + k}{\alpha + \beta + n} \tag{14}$$

$$= \left(\frac{\alpha + \beta}{\alpha + \beta + n}\right) E\left[\tilde{\theta}, \mu\left(\tilde{\theta}\right)\right] + \left(\frac{n}{\alpha + \beta + n}\right)\frac{k}{n}. \tag{15}$$

Observe that the probability judgment (15) is a convex combination between the agent's prior probability judgment, $E\left[\tilde{\theta}, \mu\left(\tilde{\theta}\right)\right]$, and the observed proportion of drawn balls of color y, $\frac{k}{n}$, whereby the agent will attach an increasing weight to this observed proportion when the number of trials increases. By the law of large numbers, this EU maximizing agent's probability judgment will thus converge with certainty to the true probability of event A_y. Since this convergence result holds, by Doob's (1949) consistency theorem, for EU maximizing agents under fairly general conditions (see Section 4), the probability judgments of an ignorant agent in our sense cannot be modeled as the Bayesian estimates of an EU decision maker.

1.4 My approach: Choquet expected utility decision makers

As this paper's main conceptual contribution, I generalize the standard model (9) of Bayesian estimation to a model of Choquet Bayesian estimation. To focus thoughts suppose that the agent satisfies Gilboa's (1987) CEU axioms on preferences $\succeq$ over Savage acts on Ω so that there exists a bounded utility function $u : Z \equiv [0,1] \to \mathbb{R}$, unique up to a positive linear transformation, and a unique non-additive probability measure κ on $(\Omega, \mathcal{F})$ such that for all Savage acts f_x, f_y, $x, y \in (0,1)$:

$$f_x \succeq f_y \Leftrightarrow E^C\left[u\left(\left|x - \tilde{\theta}\right|\right), \kappa\left(\tilde{\theta}\right)\right] \geq E^C\left[u\left(\left|y - \tilde{\theta}\right|\right), \kappa\left(\tilde{\theta}\right)\right] \tag{16}$$

where E^C denotes the Choquet expectation operator with respect to κ (see Section 2 for formal definitions).

Furthermore, if the agent updates his preferences $\succeq$ to $\succeq_I$ in accordance with the Generalized Bayesian update rule (Pires 2002; Eichberger et al. 2007; Siniscalchi 2011), we obtain for an CEU decision maker that, for all Savage acts f_x, f_y, $x, y \in (0,1)$:

$$f_x \succeq_I f_y \Leftrightarrow E^C\left[u\left(\left|x-\tilde{\theta}\right|\right), \kappa\left(\tilde{\theta} \mid I\right)\right] \geq E^C\left[u\left(\left|y-\tilde{\theta}\right|\right), \kappa\left(\tilde{\theta} \mid I\right)\right] \tag{17}$$

such that

$$\kappa\left(B \mid I\right) = \frac{\kappa\left(B \cap I\right)}{\kappa\left(B \cap I\right) + 1 - \kappa\left(B \cup \neg I\right)} \tag{18}$$

for $B \in \{B' \times \mathbb{I} \mid B' \in \mathcal{B}\}$. I refer to the solution of the CEU maximization problem

$$x_I^C = \arg \sup_{x \in (0,1)} E^C\left[u\left(\left|x-\tilde{\theta}\right|\right), \kappa\left(\tilde{\theta} \mid I\right)\right] \tag{19}$$

as the CEU decision maker's announced probability judgment in the light of sample information I. Alternatively, I call (19) the Choquet Bayesian estimate with respect to $\kappa\left(\tilde{\theta} \mid I\right)$.

If κ reduces to an additive measure μ, the Choquet Bayesian estimate (19) becomes (9) so that standard Bayesian estimation is nested within Choquet Bayesian estimation. If κ is not additive, the CEU maximization problem (19) does, in general, not allow for an analytically convenient solution. In contrast to the additive expectation operator E in (9), the Choquet expectation operator E^C is non-linear and, while being continuous, it is no longer differentiable everywhere. Unlike the global maximum of (9), the global maximum of (19) is, for strictly concave and differentiable u, therefore no longer characterized by a first-order condition.

To simplify the maximization problem (19), I am going to restrict attention to non-additive probability measures described 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).

As a first formal result (Lemma), I comprehensively characterize the solution to the CEU maximization problem (19) for strictly concave and differentiable utility functions and neo-additive capacities. In a next step, I study a model of Bayesian learning in which the agent can observe i.i.d. sample information generated by the true probability of event A. For the non-generic case in which the neo-additive capacity reduces to an additive probability measure, the agent's announced probability judgment converges, by Doob's (1949) consistency theorem, to the event A's true probability. As this paper's main formal result (Theorem), I prove:

For the generic case in which the neo-additive capacity is not additive, the agent's announced probability judgment about any events $A, \neg A$ becomes a "fifty-fifty" judgment after finitely many data observations.

The remainder of the analysis proceeds as follows. Section 2 recalls concepts from Choquet decision theory. Section 3 presents the analytical solution to the Choquet Bayesian estimation problem. The Bayesian learning model is constructed in Section 4. Section 5 concludes. All formal proofs are relegated to the Appendix.

2 Preliminaries

2.1 Non-additive probability measures and Choquet integration

Fix the measurable space $(\Omega, \mathcal{F})$ and a set of null events $\mathcal{N} \subset \mathcal{F}$ such that (i) $\emptyset \in \mathcal{N}$, (ii) $B \in \mathcal{N}$ implies $B' \in \mathcal{N}$ for all $B' \in \mathcal{F}$ such that $B' \subset B$, and (iii) $B, B' \in \mathcal{N}$ implies $B \cup B' \in \mathcal{N}$. A non-additive (=fuzzy) probability measure $\kappa : \mathcal{F} \rightarrow [0,1]$ satisfies

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

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

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

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*

$$\nu(B) = \delta \cdot \nu_\lambda(B) + (1-\delta) \cdot \mu(B) \tag{20}$$

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

$$\mu(B) = \begin{cases} 0 & \text{if } B \in \mathcal{N} \\ 1 & \text{if } \Omega \backslash B \in \mathcal{N} \end{cases} \tag{21}$$

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

$$\nu_\lambda(B) = \begin{cases} 0 & \text{iff } B \in \mathcal{N} \\ \lambda & \text{else} \\ 1 & \text{iff } \Omega \backslash B \in \mathcal{N}. \end{cases} \tag{22}$$

I call an event $B \in \mathcal{F}$ *essential* if and only if $B \notin \mathcal{N}$ and $\Omega \backslash B \notin \mathcal{N}$. Throughout this paper I restrict attention to sets of null-events $\mathcal{N}$ such that $B \in \mathcal{N}$ iff $\mu(B) = 0$. Consequently, B is essential if and only if $0 < \mu(B) < 1$. Observe that the neo-additive capacity (20) simplifies to

$$\nu(B) = \delta \cdot \lambda + (1 - \delta) \cdot \mu(B) \tag{23}$$

for essential B. The parameter δ is interpreted as an ambiguity or insensitivity parameter whereas the value of λ determines whether $\nu(B)$ overestimates (i.e., $\lambda > \nu(B)$) or underestimates (i.e., $\lambda < \nu(B)$) the additive probability $\mu(B)$ whenever $\delta > 0$.

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

$$E^C[w, \kappa] = \int_{-\infty}^{0} (\kappa(\{\omega \in \Omega \mid w(\omega) \geq z\}) - 1)\, dz + \int_{0}^{+\infty} \kappa(\{\omega \in \Omega \mid w(\omega) \geq z\})\, dz. \tag{24}$$

For instance, if w takes on m different values such that $B_1, ..., B_m$ is the unique partition of Ω with $w(\omega_1) > ... > w(\omega_m)$ for $\omega_i \in A_i$, the Choquet expectation (24) becomes

$$E^C[w, \kappa] = \sum_{i=1}^{m} w(\omega_i) \cdot [\kappa(B_1 \cup ... \cup B_i) - \kappa(B_1 \cup ... \cup B_{i-1})]. \tag{25}$$

A formal proof of the following Observation can be found for finite valued w in Chateauneuf et al. (2007) and for bounded w in Zimper (2012).

Observation 1. *Let $w : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable function with bounded range. The Choquet expected value (24) of w with respect to a neo-additive capacity (20) is given as*

$$E^C[w, \nu] = \delta(\lambda \sup w + (1 - \lambda) \inf w) + (1 - \delta) E[w, \mu]. \tag{26}$$

2.2 Conditional non-additive probability measures

CEU theory generalizes EU theory in order to accommodate paradoxes of the Ellsberg (1961) type which show that real-life decision makers violate Savage's sure thing principle. Because CEU relaxes the sure thing principle, there exist several alternative Bayesian update rules for CEU decision makers which correspond to alternative definitions of conditional non-additive probability measures. The existing literature foremostly focuses on conditional non-additive probability measures derived from the *optimistic (=naive)*, the *pessimistic (=Dempster-Shafer)*, or the *Generalized (=full) Bayesian update rule* (cf. Gilboa and Schmeidler 1993; Sarin and Wakker 1998; Pires 2002; Eichberger, Grant, Kelsey 2007, 2010; Eichberger, Grant, Lefort 2009; Zimper and Ludwig 2009; Siniscalchi 2011; Zimper 2012). In what follows I briefly sketch why the relaxation of the sure thing principle gives rise to alternative definitions of conditional non-additive probability measures.[3]

Ex ante preferences over Savage acts, denoted $\succeq$, are interpreted as the decision maker's preferences before he receives any information. In contrast, ex post preferences over Savage acts, denoted $\succeq_I$, are interpreted as preferences conditional on information I, i.e., after the decision maker has observed the occurrence of some non-null event $I \in \mathcal{F}$. A Bayesian update rule specifies how ex post preferences $\succeq_I$ over Savage acts are derived from ex ante preferences $\succeq$ for all essential $I \in \mathcal{F}$. Define the following Savage act $f_I h : \Omega \rightarrow Z$

$$f_I h(\omega) = \begin{cases} f(\omega) & \text{for } \omega \in I \\ h(\omega) & \text{for } \omega \in \neg I \end{cases} \tag{27}$$

and recall the definition of Savage's (1954) sure thing principle:

Definition: Sure thing principle. *For all Savage acts f, g, h, h' and all $I \in \mathcal{F}$, the following condition holds for ex ante preferences:*

$$f_I h \succeq g_I h \;\Rightarrow\; f_I h' \succeq g_I h'. \tag{28}$$

If the sure thing principle (28) holds, then there exists one plausible Bayesian update rule; namely, for all f, g, h and all I,

$$f_I h \succeq g_I h \;\Rightarrow\; f \succeq_I g. \tag{29}$$

In words: If the decision maker prefers the consequences of f on I to the consequences of g on I given that f and g have common consequences on $\neg I$, then the decision maker should also prefer f to g after he has learnt that I has occurred. Under the assumption of Savage's (1954) axioms for $\succeq$ and $\succeq_I$, the update rule (29) gives rise to the familiar definition of a conditional additive probability measure (8).

CEU preferences that do not reduce to EU preferences, however, violate the sure thing principle to the effect that there exist f, g and I such that

$$f_I h \succ g_I h \text{ and } g_I h' \succ f_I h' \tag{30}$$

[3] Since any in-depth discussion of this topic is beyond the scope of the present paper, I would like to refer the interested reader to the mentioned literature.

for some h and h' with $h \neq h'$. Fix h and h' in (30) and observe that there exist two alternative ways of deriving ex post preferences $\succeq_I$ over f and g from (30); namely, either

$$f_I h \succeq g_I h \;\Rightarrow\; f \succeq_I g \tag{31}$$

or

$$g_I h' \succeq f_I h' \Rightarrow\; g \succeq_I f. \tag{32}$$

That is, if the sure principle does not hold, there exist several alternative Bayesian update rules in the sense that we can associate, for fixed f, g, with every $I \in \mathcal{F}$ an act $h^*(I)$ such that

$$f_I h^*(I) \succeq g_I h^*(I) \;\Rightarrow\; f \succeq_I g. \tag{33}$$

2.3 The Generalized Bayesian update rule

In the present paper I restrict attention to the Generalized Bayesian update rule, which appears to perform well in empirical investigations (Cohen at al. 2000) and which is popular in the literature because it avoids the extreme updating behavior of alternative update rules. More specifically, the Generalized Bayesian update rule defines $h^*(I)$ in (33) as the ($\succeq_I$)-*conditional certainty equivalent* of g; that is, $h^*(I)$ is the constant act such that $g \succeq_I h^*(I)$ as well as $h^*(I) \succeq_I g$.

Eichberger et al. (2007) prove that the Generalized Bayesian update rule gives rise to a conditional non-additive probability measure given as (18). A formal proof of the following result appears in Zimper and Ludwig (2009) and in Eichberger et al. (2010).

Observation 2. *An application of the Generalized Bayesian update rule (18) to a neo-additive capacity (20) results in the following conditional neo-additive capacity*

$$\nu(B \mid I) = \delta_I \cdot \lambda + (1 - \delta_I) \cdot \mu(B \mid I), \tag{34}$$

for essential $B \in \mathcal{F}$ and non-null $I \in \mathcal{F}$, whereby

$$\delta_I = \frac{\delta}{\delta + (1 - \delta) \cdot \mu(I)}. \tag{35}$$

By Observation 2, an application of the Generalized Bayesian update rule to a neo-additive capacity results in a conditional neo-additive capacity for which the ambiguity parameter has changed from δ to δ_I whereas the parameter λ remains unchanged. As a consequence, the Generalized Bayesian update rule is formally very convenient because it reduces the Bayesian updating dynamics of neo-additive capacities to additive Bayesian updating plus the updating of the ambiguity parameter. In particular, we have that $\delta_I > \delta$ whereby smaller additive probabilities of the observed information, $\mu(I)$, imply a greater updated ambiguity parameter δ_I. If we interpret the ambiguity parameter δ as the agent's lack of confidence in the additive probability measure μ, the formal relationship (35) implies that the agent's confidence will further decrease if he observes unlikely information.

Although this interpretation of (35) has intuitive appeal, it also implies that more and more sample information will increase the ambiguity parameter. This is because in a Bayesian framework any specific realization of large sample information is rather unlikely from an ex ante perspective and it will be the more unlikely the larger the sample is. The fact that ambiguity increases by (35) with more information will drive the Theorem in Section 4 where I consider a Bayesian learner who can observe large data samples.

3 Choquet Bayesian estimation

Consider a Savage act $f_x : \Omega \rightarrow [0,1]$ such that, for all $(\theta, i) \in \Omega$,

$$f_x(\theta, i) = |\theta - x| \tag{36}$$

whereby I interpret $x \in (0,1)$ as the decision maker's probability judgment associated with act f_x. The consequence of f_x in state (θ, i) is thus defined as the Euclidean distance between the true probability θ and the probability judgment x. To evaluate probability judgments I consider a bounded utility function $u : [0,1] \rightarrow \mathbb{R}_-$ which is strictly concave and strictly decreasing whereby I normalize $u(0) = 0$.[4] Furthermore, I assume that

$$u(f_x((\theta, i))) = u(|\theta - x|) \tag{37}$$

is continuously differentiable in x as well as measurable with respect to $\tilde{\theta}(\theta, i) = \theta$. For the first order derivative of u it holds that

$$u'(z) = \begin{cases} 0 & \text{if } z = 0 \\ < 0 & \text{if } z > 0 \end{cases} \tag{38}$$

That is, I assume that the maximum of $u(|\theta - x|)$ at $\theta = x$ is conveniently characterized by the first order condition (=FOC).

[4] A strictly concave u corresponds to the standard case of strictly risk averse decision makers. Interpreted in terms of the loss function $l = -u$ of Bayesian estimation, strict concavity of u (i.e., strict convexity of l) implies that large differences between the probability judgment and the true probability are over-proportionally penalized.

Fix some information $I = \Theta \times I' \in \mathcal{F}$ where $I' \in \mathcal{I}$ and consider the conditional neo-additive capacity space $(\Omega, \mathcal{F}, \nu(\cdot \mid I))$ such that $\nu(\cdot \mid I)$ is given by (34). By Observation 1, the CEU of the probability judgment x is given as

$$\begin{aligned} & E^C \left[u(f_x), \nu\left(\tilde{\theta} \mid I\right) \right] \quad (39) \\ = \;& \delta_I \left(\lambda \sup_{\theta \in (0,1)} u(f_x) + (1-\lambda) \inf_{\theta \in (0,1)} u(f_x) \right) + (1-\delta_I) E \left[u(f_x), \mu\left(\tilde{\theta} \mid I\right) \right]. \end{aligned}$$

The best probability judgment a decision maker could come up with is to correctly announce the true probability, i.e., $x = \theta$, implying

$$\sup_{\theta \in (0,1)} u(f_x) = 0. \quad (40)$$

The worst possible probability judgment corresponds to the maximal Euclidean distance between the true probability θ and the announced probability x, implying

$$\inf_{\theta \in (0,1)} u(f_x) = \begin{cases} u(1-x) & \text{if } x \leq \frac{1}{2} \\ u(x) & \text{if } x \geq \frac{1}{2} \end{cases} \quad (41)$$

The best case scenario (40) corresponds to a benevolent nature which always picks a state that coincides with the agent's probability judgment thereby maximizing his utility. In contrast, the worst case scenario (41) corresponds to a malevolent nature which always picks a state that minimizes the agent's utility for his probability judgment. The ambiguity parameter δ_I in (39) determines how much the agent cares about the best and worst case scenario as compared to his additive (=non-ambiguous) expectation. The parameter λ in (39) determines the degree by which the agent resolves his ambiguity by focusing on the best case versus the worst case scenario.

Collecting terms gives us the following characterization of the agent's CEU function (39) for any probability judgment x.

Observation 3. *Let*

$$U_1(x) \equiv \delta_I (1-\lambda) u(1-x) + (1-\delta_I) E \left[u(f_x), \mu\left(\tilde{\theta} \mid I\right) \right], \quad (42)$$

$$U_2(x) \equiv \delta_I (1-\lambda) u(x) + (1-\delta_I) E \left[u(f_x), \mu\left(\tilde{\theta} \mid I\right) \right] \quad (43)$$

with δ_I given by (35). The agent's objective function is then given as

$$E^C \left[u(f_x), \nu\left(\tilde{\theta} \mid I\right) \right] = \begin{cases} U_1(x) & \text{if } x \in \left(0, \frac{1}{2}\right] \\ U_2(x) & \text{if } x \in \left[\frac{1}{2}, 1\right) \end{cases} \quad (44)$$

Observe that the term for the best case scenario (40) vanishes from the agent's maximization problem. The reason is straightforward: Whatever probability judgment the agent chooses, in the best case scenario a benevolent nature plays along to ensure that the agent achieves his maximal utility. In contrast, the worst case scenario (41), in which a malevolent nature plays against the agent, enters the agent's maximization problem.[5] Moreover, this worst case scenario results in a kink of the agent's objective function at $x = \frac{1}{2}$ to the effect that a simple FOC argument is not sufficient to characterize the solution to (44). To exclude the trivial case where (42) and (43) are constantly zero, I henceforth assume that

$$\text{either } \delta_I < 1 \text{ or } \delta_I = 1,\ \lambda < 1. \tag{45}$$

Lemma. *Define x_1 and x_2 implicitly through the following FOCs*

$$\frac{d}{dx}\left(U_i\left(x_i\right)\right) = 0,\ i = 1, 2. \tag{46}$$

The solution x_I^C of the maximization problem (44), i.e.,

$$x_I^C = \arg \sup_{x \in (0,1)} E^C \left[u\left(f_x\right), \nu\left(\tilde{\theta} \mid I\right)\right], \tag{47}$$

is characterized as follows.

(a) *If*

$$x_1 < \frac{1}{2} \text{ and } x_2 > \frac{1}{2} \tag{48}$$

then

$$x_I^C = \arg \max_{\{x_1, x_2\}} E^C \left[u\left(f_x\right), \nu\left(\tilde{\theta} \mid I\right)\right]. \tag{49}$$

(b) *If*

$$x_1 \leq \frac{1}{2} \text{ and } x_2 \leq \frac{1}{2} \tag{50}$$

then

$$x_I^C = x_1. \tag{51}$$

[5] But of course, the parameter λ, which measures how much the agent cares about the best case scenario, still plays an important role in the agent's optimization problem if there is ambiguity. Therefore to say that "the term for the best case scenario vanishes from the agent's optimization problem" does not mean that "the agent's ambiguity attitudes towards the best case scenario have no impact on the agent's optimization problem".

(c) *If*

$$x_1 \geq \frac{1}{2} \text{ and } x_2 \geq \frac{1}{2} \tag{52}$$

then

$$x_I^C = x_2. \tag{53}$$

(d) *If*

$$x_1 \geq \frac{1}{2} \text{ and } x_2 \leq \frac{1}{2} \tag{54}$$

then

$$x_I^C = \frac{1}{2}. \tag{55}$$

To illustrate the findings of the Lemma let us consider the special case of a negative quadratic utility function, i.e., $u(z) = -z^2$. In that case, (42) and (43) become

$$U_1(x) \equiv \delta_I (1-\lambda)(-)(1-x)^2 + (1-\delta_I) E\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right], \tag{56}$$

$$U_2(x) \equiv \delta_I (1-\lambda)(-)(x)^2 + (1-\delta_I) E\left[-\left(x-\tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \tag{57}$$

and we can derive explicit analytical expressions for x_1 and x_2. Namely, by (46),

$$\frac{d}{dx}\left(U_1(x_1)\right) = 0 \tag{58}$$

$$\Leftrightarrow$$

$$0 = \delta_I (1-\lambda) \cdot 2(1-x_1) + (1-\delta_I) \int_{\theta \in (0,1)} -(2x_1 - 2\theta)\, d\mu\left(\tilde{\theta} \mid I\right) \tag{59}$$

$$\Leftrightarrow$$

$$x_1 = \frac{\delta_I - \delta_I \lambda + (1-\delta_I) E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]}{\delta_I - \delta_I \lambda + 1 - \delta_I}; \tag{60}$$

as well as

$$\frac{d}{dx}\left(U_2(x_2)\right) = 0 \tag{61}$$

$$\Leftrightarrow$$

$$0 = -2\delta_I (1-\lambda) \cdot x_2 + (1-\delta_I) \int_{\theta \in (0,1)} -(2x_2 - 2\theta)\, d\mu\left(\tilde{\theta} \mid I\right) \tag{62}$$

$$\Leftrightarrow$$

$$x_2 = \frac{(1-\delta_I)}{(1-\delta_I \lambda)} E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{63}$$

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

Corollary. *Suppose that u is given as the negative quadratic utility function, i.e., $u(z) = -z^2$.*

(a) *If $\delta = 0$, the analytical solution (47) becomes*

$$x_I^C = E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{64}$$

(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 (47) becomes (64).*

4 Choquet Bayesian Learning of Probability Judgments

This section combines the CEU estimation problem (47) with a standard model of Bayesian learning according to which the agent observes sample information that is generated by some i.i.d. process. More specifically, for a given measurable space $(\Omega', \mathcal{A})$ consider the additive probability space $(\Omega, \mathcal{F}, \mu)$ with $(\Omega, \mathcal{F})$ defined in Subsection 1.1. Any possible sample information after n trials is given as some event

$$I_n = \Theta \times \{s_1\} \times ... \times \{s_n\} \times S_{n+1} \times ... \in \mathcal{F} \tag{65}$$

with $s_j \in \{A, \neg A\}$, $j = 1, ..., n$. Suppose that the outcomes of the statistical trials are, conditional on θ, i.i.d. such that A occurs in every trial with true probability θ. That is,

$$\mu\left(I_n \mid \tilde{\theta} = \theta\right) = \prod_{j=1}^{n} \pi_\theta(s_j) \tag{66}$$

such that

$$\pi_\theta(s_j) = \begin{cases} \theta & \text{if } s_j = A \\ 1 - \theta & \text{if } s_j = \neg A \end{cases} \tag{67}$$

By Bayes' rule, we obtain the posterior $\mu\left(\tilde{\theta} \mid I_n\right)$ such that, for any $B \in \{B' \times \mathbb{I} \mid B' \in \mathcal{B}\}$,

$$\mu(B \mid I_n) = \frac{\int_{\theta\in B} \mu\left(I_n \mid \tilde{\theta}\right) d\mu\left(\tilde{\theta}\right)}{\mu(I_n)} \tag{68}$$

$$= \frac{\int_{\theta\in B} \prod_{j=1}^{n} \pi_\theta(s_j)\, d\mu\left(\tilde{\theta}\right)}{\int_{\theta\in(0,1)} \prod_{j=1}^{n} \pi_\theta(s_j)\, d\mu\left(\tilde{\theta}\right)}. \tag{69}$$

Define

$$\Omega_\theta = \left\{\omega \in \Omega \mid \tilde{\theta}(\omega) = \theta\right\}. \tag{70}$$

Recall that Doob's (1949) consistency theorem[6] implies that, for almost all true parameter values θ belonging to the support of μ, the posterior distribution $\mu\left(\tilde{\theta} \mid I_n\right)$ concentrates Ω_θ-almost everywhere (i.e., with $\mu(\cdot \mid \theta)$-probability one) at the true value θ as n gets large; that is,

$$\mu(B \mid I_n) \to \mathbf{1}_{B'}\theta,\ \mu(\cdot \mid \theta)\text{-a.s.} \tag{71}$$

for all $B \in \{B' \times \mathbb{I} \mid B' \in \mathcal{B}\}$ where $\mathbf{1}_{B'}$ denotes the indicator function of the Borel set B'. Or, simply expressed,

$$\mu\left(\tilde{\theta} \mid I_\infty\right) = \begin{cases} 1 & \text{if } \tilde{\theta} = \theta \\ 0 & \text{if } \tilde{\theta} \neq \theta \end{cases},\ \mu(\cdot \mid \theta)\text{-a.s.} \tag{72}$$

Applied to the standard Bayesian estimate of an EU agent (47), we therefore obtain in the limit that

$$x^*_{I_\infty} = \arg\sup_{x\in(0,1)} E\left[u\left(\left|x - \tilde{\theta}\right|\right), \mu\left(\tilde{\theta} \mid I_\infty\right)\right] \tag{73}$$

$$= \arg\sup_{x\in(0,1)} u(|x-\theta|) \tag{74}$$

$$= \theta \tag{75}$$

with $\mu(\cdot \mid \theta)$-probability one. Consequently, Doob's theorem immediately implies the following convergence result.

Observation 4. *Let $\mu\left(\tilde{\theta}\right)$ have full support on $(0,1)$. The EU agent's probability judgment about any essential event $A \in \mathcal{A}$ will, for almost all θ, almost surely converge to A's true probability if the sample size n gets large, i.e.,*

$$x^*_{I_n} \equiv \arg\sup_{x\in(0,1)} E\left[u(f_x), \mu\left(\tilde{\theta} \mid I_n\right)\right] \to \theta,\ \mu(\cdot \mid \theta)\text{-a.s.} \tag{76}$$

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

By Observation 4, the probability judgment of an EU decision maker will converge through Bayesian learning towards the true probability of any given event A. Whenever (76) holds, I speak of a (statistically) sophisticated agent.

The situation is different for CEU decision makers. As in the EU benchmark case of Observation 4, I assume that the i.i.d. sample information I_n is generated by θ, i.e., with true probability $\mu(I_n \mid \theta)$. However, in contrast to the conditional additive probability measure $\mu\left(\tilde{\theta} \mid I_n\right)$ given by (69), the CEU agent's Bayesian updating now results in the conditional neo-additive capacity

$$\nu\left(\tilde{\theta} \mid I_n\right) = \delta_{I_n}\lambda + (1 - \delta_{I_n})\,\mu\left(\tilde{\theta} \mid I_n\right). \tag{77}$$

Consequently, the sophisticated agent of Observation 4 corresponds to the non-generic special case of an CEU agent for whom $\delta = 0$.

The following Theorem states this paper's main result according to which the generic CEU agent becomes, already after finitely many observations, a (statistically) ignorant agent whose announced probability judgment is given as a "fifty-fifty" judgment.

Theorem. *For any $\delta \in (0,1]$ and $\lambda \in [0,1)$, the agent's announced probability judgment about any complementary essential events $A, \neg A \in \mathcal{A}$ becomes a fifty-fifty judgment for all $\omega \in \Omega_\theta$ if the sample size n gets sufficiently large. That is, for any fixed $\delta \in (0,1]$ and $\lambda \in [0,1)$ there exists some finite number n' such that*

$$x^C_{I_n} \equiv \arg\sup_{x\in(0,1)} E^C\left[u(f_x), \nu\left(\tilde{\theta} \mid I_n\right)\right] = \frac{1}{2}, \ \Omega_\theta\text{-everywhere} \tag{78}$$

for all $n \geq n'$.

The formal proof of the Theorem is driven by the fact that the conditional ambiguity parameter converges towards one, i.e., $\lim_{n\to\infty} \delta_{I_n} = 1$, because the (additive) prior probability attached to this sample information converges towards zero, i.e., $\lim_{n\to\infty} \mu(I_n) = 0$. As a consequence, we obtain for the limit of the objective function

$$\lim_{n\to\infty} E^C\left[u(f_x), \nu\left(\tilde{\theta} \mid I_n\right)\right] = (1-\lambda) \inf_{\theta\in(0,1)} u(f_x), \ \Omega_\theta\text{-everywhere} \tag{79}$$

with

$$\inf_{\theta\in(0,1)} u(f_x) = \begin{cases} u(1-x) & \text{if } x \leq \frac{1}{2} \\ u(x) & \text{if } x \geq \frac{1}{2} \end{cases} \tag{80}$$

That is, in the limit the agent only cares about the worst case scenario in which a malevolent nature picks a true parameter value θ that minimizes his utility for any x he is going to choose. Since u is strictly decreasing on $(0,1)$, $\inf_{\theta\in(0,1)} u(f_x)$ takes on a maximum at the kink $x=\frac{1}{2}$. Observe that this limit argument implies that $x^C_{I_\infty}=\frac{1}{2}$ amounts to the best probability judgment against a malevolent nature regardless of the shape of u, i.e., regardless of the agent's risk attitudes, because there remains no (subjective) uncertainty in the limit.

Beyond this limit argument, the formal proof of the Theorem uses the Lemma to show that $x^C_{I_n}=\frac{1}{2}$ becomes already the announced probability judgment of strictly risk averse agents (i.e., u is strictly concave) if the number of observations n is sufficiently large.

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 judgments have been formally described as the solution to a CEU maximization problem subject to Bayesian learning of neo-additive capacities.

The main result establishes that all announced probability judgments become after finitely many observations "fifty-fifty" judgments 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" judgments are the predominant empirical phenomenon of likelihood insensitivity. Interestingly, the CEU agent of our model announces his "fifty-fifty" probability judgment not due to a lack of statistical information but rather because he has received a large amount of information which increases his ambiguity parameter, i.e., his likelihood insensitivity.

Appendix: Formal Proofs

Proof of the Lemma. Step 1. Observe at first that

$$E\left[u\left(f_x\right),\mu\left(\tilde{\theta}\mid I\right)\right] \tag{81}$$

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

$$\frac{d}{dx}\left(u\left(f_x\right)\right) \tag{82}$$

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), (81) is continuously differentiable in x whereby

$$\frac{d}{dx}\left(E\left[u\left(f_x\right),\mu\left(\tilde{\theta}\mid I\right)\right]\right)=E\left[\frac{d}{dx}\left(u\left(f_x\right)\right),\mu\left(\tilde{\theta}\mid I\right)\right]. \tag{83}$$

Step 2. Focus on the function (42) and observe that it is, by assumption (45), strictly concave. Furthermore, (42) is, by (83), continuously differentiable with

$$\frac{d}{dx}\left(U_1\left(x\right)\right)=\delta_I\left(1-\lambda\right)\left(-\right)u'\left(1-x\right)+\left(1-\delta_I\right)E\left[\frac{d}{dx}\left(u\left(f_x\right)\right),\mu\left(\tilde{\theta}\mid I\right)\right]. \tag{84}$$

Evaluate (84) at $x=0$, i.e.,

$$\frac{d}{dx}\left(U_1\left(0\right)\right)=\delta_I\left(1-\lambda\right)\left(-\right)u'\left(1\right)+\left(1-\delta_I\right)E\left[\frac{d}{dx}\left(u\left(f_0\right)\right),\mu\left(\tilde{\theta}\mid I\right)\right]. \tag{85}$$

Since $u'\left(1\right)<0$, we have, by assumption (45),

$$\delta_I\left(1-\lambda\right)\left(-\right)u'\left(1\right)>0. \tag{86}$$

Furthermore, our assumptions on u and u', in particular $u'\left(0\right)=0$, imply

$$E\left[\frac{d}{dx}\left(u\left(f_x\right)\right),\mu\left(\tilde{\theta}\mid I\right)\right]=\int_{\theta\in(0,x)}u'\left(x-\theta\right)d\mu\left(\tilde{\theta}\mid I\right)+\int_{\theta\in(x,1)}\left(-\right)u'\left(\theta-x\right)d\mu\left(\tilde{\theta}\mid I\right). \tag{87}$$

so that

$$E\left[\frac{d}{dx}\left(u\left(f_0\right)\right),\mu\left(\tilde{\theta}\mid I\right)\right]=\int_{\theta\in(0,1)}\left(-\right)u'\left(\theta\right)d\mu\left(\tilde{\theta}\mid I\right)>0 \tag{88}$$

because $u'\left(\theta\right)<0$ for all $\theta\in(0,1)$. Combining (86) and (88) shows that

$$\frac{d}{dx}\left(U_1\left(0\right)\right)>0 \tag{89}$$

so that U_1 is strictly increasing at $x = 0$. Consequently,

$$0 \neq \arg \sup_{x \in (0,1)} E^C \left[u(f_x), \nu\left(\tilde{\theta} \mid I\right)\right] \tag{90}$$

implying that there exists a maximum x_1 of function (42) on the interval $\left(0, \frac{1}{2}\right]$ which is either $x_1 \leq \frac{1}{2}$ characterized by the FOC (46) for $i = 1$ or given as the boundary solution $x_1 = \frac{1}{2}$. That is, whenever $x_1 \in \left(0, \frac{1}{2}\right]$, x_1 is a local maximizer of the objective function (44).

Step 3. Turn now to the function (43), which is also strictly concave. (43) is continuously differentiable with

$$\frac{d}{dx}\left(U_2(x)\right) = \delta_I (1-\lambda) u'(x) + (1-\delta_I) E\left[\frac{d}{dx}\left(u(f_x)\right), \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{91}$$

Evaluated at $x = 1$, we have

$$\begin{aligned} \frac{d}{dx}\left(U_2(1)\right) &= \delta_I (1-\lambda) u'(1) + (1-\delta_I) \int_{\theta \in (0,1)} u'(1-\theta)\, d\mu\left(\tilde{\theta} \mid I\right) \\ &< 0 \end{aligned} \tag{92}$$

because of $u'(1) < 0$ and $u'(1-\theta) < 0$ for all $\theta \in (0,1)$. That is, $U_2(x)$ is strictly decreasing at $x = 1$, implying

$$1 \neq \arg \sup_{x \in (0,1)} E^C \left[u(f_x), \nu\left(\tilde{\theta} \mid I\right)\right]. \tag{93}$$

Consequently, there exists a maximum x_2 of function (43) on the interval $\left[\frac{1}{2}, 1\right)$ which is either $x_2 \geq \frac{1}{2}$ characterized by the FOC (46) for $i = 2$ or given as the boundary solution $x_2 = \frac{1}{2}$. That is, x_2 is a local maximizer of (44) iff $x_2 \in \left[\frac{1}{2}, 1\right)$.

Step 4. If condition (48) holds, we have thus two local maximizers, x_1 and x_2, of (44), characterized by FOCs (46), and whichever is greater is the global maximizer for (44). This proves (a).

Step 5. If condition (50) holds, we have one local maximizer x_1 characterized by the FOC (46) for $i = 1$. Since $x_1 \geq \frac{1}{2}$, with $\frac{1}{2}$ being the boundary maximum of (44) on the interval $\left[\frac{1}{2}, 1\right)$, x_1 is also the global maximizer for (44). This proves (b).

Step 6. If condition (52) holds, we have one local maximizer $x_2 \geq \frac{1}{2}$ characterized by the FOC (46) for $i = 2$, which is also the global maximizer for (44). This proves (c).

Step 7. If condition (54) holds, there is no local maximizer characterized by any FOC. Instead (44) takes on its maximum at the kink $x = \frac{1}{2}$. This proves (d).□

Proof of the Corollary. Case (a). If $\delta = 0$, then $\delta_I = 0$. By (60) and (63),

$$x_1 = x_2 = E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]$$

so that (64) is the global maximizer.

Case (b). Suppose, on the contrary, that

$$x_I^C = E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right] \neq \frac{1}{2}. \tag{94}$$

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

$$E\left[-\left(x - \tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \neq x_1 \text{ and } E\left[-\left(x - \tilde{\theta}\right)^2, \mu\left(\tilde{\theta} \mid I\right)\right] \neq x_2, \tag{95}$$

by (45), implying

$$x_I^C \neq E\left[\tilde{\theta}, \mu\left(\tilde{\theta} \mid I\right)\right]. \tag{96}$$

□

Proof of the Theorem. Step 1. Let

$$\theta_{\max} = \max\{\theta, 1 - \theta\} \tag{97}$$

and observe that, for any $I_n = \Theta \times \{s_1\} \times ... \times \{s_n\} \times \Omega' \times ...$ with $s_j \in \{A, \neg A\}$, $j = 1, ..., n$,

$$\begin{aligned} \mu(I_n) &= \int_{\theta \in (0,1)} \pi_\theta(s_1, ..., s_n)\, d\mu\left(\tilde{\theta}\right) & (98) \\ &= \int_{\theta \in (0,1)} \pi_\theta(s_1) \cdot ... \cdot \pi_\theta(s_n)\, d\mu\left(\tilde{\theta}\right) & (99) \\ &\leq \int_{\theta \in (0,1)} (\theta_{\max})^n\, d\mu\left(\tilde{\theta}\right), & (100) \end{aligned}$$

because the sample observations are θ-conditionally independent. Since the $\{w_n(\theta) \equiv (\theta_{\max})^n\}_{n \in \mathbb{N}}$ are non-negative measurable functions on $(0, 1)$ decreasing in n pointwise to the constant function 0 whereby $\int_{\theta \in (0,1)} w_1(\theta)\, d\mu\left(\tilde{\theta}\right) < \infty$, an application of the monotone convergence theorem (cf. Lemma 19.36 in Aliprantis and Border 2006) gives

$$\begin{aligned} \lim_{n \to \infty} \int_{\theta \in (0,1)} (\theta_{\max})^n\, d\mu\left(\tilde{\theta}\right) &= \int_{\theta \in (0,1)} \lim_{n \to \infty} w_n(\theta)\, d\mu\left(\tilde{\theta}\right) & (101) \\ &= 0 & (102) \end{aligned}$$

implying, for any I_n,

$$\lim_{n \to \infty} \mu(I_n) = 0.^7 \tag{103}$$

[7] Here the assumption that $\Theta = (0, 1)$ instead of $\Theta = [0, 1]$ is crucial. Otherwise pointwise convergence of $w_n(\theta)$ to the zero function would break down at $\theta = 0$ and $\theta = 1$. For degenerate probability measures μ that put probability mass one on $\theta = 0$ or $\theta = 1$, we would then obtain that $\lim_{n \to \infty} \mu(I_n) > 0$.

Step 2. Notice that (103) together with (35) implies

$$\lim_{n\to\infty} \delta_{I_n} \to 1 \tag{104}$$

for any I_n.

Step 3. Consider at first the FOC (46) for $i = 1$ implying

$$\begin{aligned} \frac{d}{dx}\left(U_1\left(x_{1,I_n}\right)\right) &= 0 \quad (105)\\ &\Leftrightarrow \\ \delta_{I_n}(1-\lambda)(-)\,u'\left(1-x_{1,I_n}\right) + \left(1-\delta_{I_n}\right) E\left[\frac{d}{dx}\left(u\left(f_{x_{1,I_n}}\right)\right), \mu\left(\tilde{\theta} \mid I_n\right)\right] &= 0 \quad (106)\end{aligned}$$

for all n. Taking the limit $n \to \infty$ gives, by (104),

$$u'\left(1 - x_{1,I_\infty}\right) = 0,\ \Omega_\theta\text{-everywhere} \tag{107}$$

because $\lambda < 1$ and $E\left[\frac{d}{dx}\left(u\left(f_{x_{1,I_n}}\right)\right), \mu\left(\tilde{\theta} \mid I_n\right)\right]$ is bounded for all n. By (38), we therefore have

$$x_{1,I_\infty} = 1. \tag{108}$$

Step 4. Turn to the FOC (46) for $i = 2$ implying

$$\begin{aligned} \frac{d}{dx}\left(U_2\left(x_{2,I_n}\right)\right) &= 0 \quad (109)\\ &\Leftrightarrow \\ \delta_{I_n}(1-\lambda)\,u'\left(x_{2,I_n}\right) + \left(1-\delta_{I_n}\right) E\left[\frac{d}{dx}\left(u\left(f_{x_{2,I_n}}\right)\right), \mu\left(\tilde{\theta} \mid I_n\right)\right] &= 0 \quad (110)\end{aligned}$$

for all n. Taking the limit $n \to \infty$ gives, by (104),

$$u'\left(x_{2,I_\infty}\right) = 0,\ \Omega_\theta\text{-everywhere} \tag{111}$$

because $\lambda < 1$ and $E\left[\frac{d}{dx}\left(u\left(f_{x_{2,I_n}}\right)\right), \mu\left(\tilde{\theta} \mid I_n\right)\right]$ is bounded for all n. Consequently,

$$x_{2,I_\infty} = 0. \tag{112}$$

Step 5. Collecting (108) and (112) establishes that x_{1,I_n} converges to one whereas x_{2,I_n} converges to zero. By the definition of convergence in the Euclidean distance, there must thus exist some finite n' such that for all $n \geq n'$

$$\left|1 - x_{1,I_n}\right| < \frac{1}{2} \text{ and } \left|x_{2,I_n} - 0\right| < \frac{1}{2},\ \Omega_\theta\text{-everywhere.} \tag{113}$$

Consequently, for all $n \geq n'$,

$$x_{1,I_n} > \frac{1}{2} \text{ and } x_{2,I_n} < \frac{1}{2},\ \Omega_\theta\text{-everywhere} \tag{114}$$

so that part (d) of the Lemma implies the desired result (78).□

References

Aliprantis, D. C., and K. Border (2006) *Infinite Dimensional Analysis* (2nd edn.) Springer: Berlin.

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.

Berger, J.O. (1980) *Statistical Decision Theory and Bayesian Analysis* (2nd edn.) Springer: New York.

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 (2007) "Updating Choquet beliefs," *Journal of Mathematical Economics* **43**, 888-899.

Eichberger, J., Grant, S., and D. Kelsey (2010) "Comparing three ways to update Choquet beliefs," *Economics Letters* **107**, 91-94.

Eichberger, J., Grant, S., and J.P. Lefort (2009) "Neo-additive capacities and updating," *Discussion paper* **No 490**, University of Heidelberg.

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.

Zimper, A. (2011) "Do Bayesians learn their way out of ambiguity?," *Decision Analysis* **8**, 269-285.

Zimper, A. (2012) "Asset pricing in a Lucas fruittree economy with the best and worst in mind," *Journal of Economic Dynamics and Control* **36**, 610-628.

Zimper, A., and A. Ludwig (2009) "On attitude polarization under Bayesian learning with non-additive beliefs," *Journal of Risk and Uncertainty* **39**, 181-212.