

Eliciting Subjective Probabilities Through Exchangeable Events: An Advantage and a Limitation

Aurélien Baillon

Department of Applied Economics, Erasmus University, NL-3000 DR Rotterdam, The Netherlands,
baillon@few.eur.nl

Two events are exchangeable for an agent when she is indifferent to permutations of their outcomes. Such events are thus revealed to be equally likely. If they are complementary, the subjective probability associated with each event should be $1/2$ (assuming the additivity of the probability measure). This paper reports the results of an experiment that elicits probabilities through exchangeable events. The experiment shows that this method does not suffer from source dependence, i.e., the preference for betting on some events based on knowledge about the mechanism that generates them. However, it also highlights how additivity might be violated. This paper deduces the practical implications of these results.

Key words: subjective probability; elicitation; source of uncertainty; support theory; event-splitting effect
History: Received on November 16, 2007. Accepted on April 27, 2008, after 2 revisions.

1. Introduction

When sufficient data about a potential risk is not available, decision makers are often required to seek advice from experts who are asked to assess the probability of the risk. Such assessments may be required by an insurance company before insuring a new risk or a firm before investing in a new technology. This is why probability elicitation has been extensively studied in decision analysis. Even if subjective probability assessment may be done through a simple direct judgment, a particular credibility is often assigned to the probabilities derived from choices. This latter viewpoint is known as the revealed preference approach and is prevalent in fields such as economics. Through choices, agents are supposed to have incentives to tell the truth. This is why several choice-based methods have been proposed to elicit beliefs. For instance, the *lottery method* consists in finding a probability p and an event E such that the expert is indifferent between a $p\%$ chance of winning $\$x$ and winning the same amount if E occurs. The probability that E occurs can be inferred based on this choice: Assuming Savage's (1954) subjective expected utility, it should be p . This

method can be implemented with p or E remaining constant (see Abbas et al. 2007).

This paper focuses on a choice-based implementation of another particular family of methods that are based on a bisection process, i.e., the subsequent partition of the state space into two equally likely subevents. The intuition behind this method was introduced by Ramsey (1931) and Fellner (1961). The method itself is described by Raiffa (1968) using judgments and, alternatively, by Spetzler and Staël von Holstein (1975) in terms of judgments and choices. Chesley (1978) and Wright (1988) provided tests of the judgments-based version and comparisons with other methods. Abdellaoui et al. (2008) implemented a choice-based version of this general technique. This specific choice-based implementation will be referred to as the *exchangeability method* because it is based on Ramsey (1931) and de Finetti's (1937) basic idea of *exchangeable events*. Two events are exchangeable for an assessor if she is indifferent to permutations of their outcomes. Chew and Sagi (2006) formally derived the existence of probabilistic beliefs from this concept. Thus, the exchangeability method refers to the subsequent splitting of the state space into equally

likely events that are revealed through binary choices between binary prospects.

In practice, this method consists of several series of choices. Assume that I want to determine your subjective probabilities concerning the temperature tomorrow (at noon) in Paris. Would you prefer to win \$100 if the temperature is higher than 50°F (and nothing otherwise) or the same amount if the temperature is lower than 50°F? According to your preference, I must increase or decrease the common boundary of the two events (50°F in the example) and ask you for new preferences up to the point of indifference. Two equally likely events are thus obtained; further, assuming the existence of an (additive) subjective probability measure, probability of the events must be 1/2. In a second series of choices, you would have to bet on the subevents of one of these events to determine two new equally likely events. Two equally likely disjoint subevents—the union of which has probability 1/2—should have probability 1/4. It is worth noting that the exchangeability method requires neither reference to the concept of probability nor a direct judgment, but only simple choices between binary prospects. Moreover, this method is cognitively easier for the person whose beliefs are under consideration (an expert or a subject in an experiment) than a direct matching method, despite it being more time consuming.

This paper focuses on two main aspects of this method. First, this method only deals with events from the same variable. In the above example, I did not refer to anything other than the temperature in Paris. With the lottery method, one would have had to propose two different kinds of events: the unknown variable under consideration (e.g., the temperature in Paris) and an external device (e.g., a probability wheel). In this case, you may have preferred a bet on an objective probability rather than one on an event with an unknown probability. This phenomenon was pointed out by Ellsberg (1961). An experiment will be designed to show how the exchangeability method avoids this bias and how the lottery method would not. Secondly, this method is based on the assumption that beliefs are additive. Indeed, additivity is needed to infer that indifference between betting on two complementary events implies that their probability is one half. It is, however, well known that

violations of additivity may occur in probability elicitation (Tversky and Koehler 1994). This is why the experiment presented in this paper will also test this hypothesis. Therefore, this paper will introduce new insights on this elicitation technique, highlighting one of its strengths and testing one of its limitations.

In §2, a review of the literature will enable us to clearly express what needs to be tested. Section 3 will present the experimental method, whereas the results will be described in §4. Section 5 discusses some limitations of this work, and §6 concludes the paper.

2. Review of the Literature and Predictions

2.1. Source Dependence

Biases (and how they can be avoided) have formed an important part of the literature on probability elicitation since the 1960s. Raiffa (1968, p. 111) pointed out the following situation: A subject was asked to choose between betting on one of two baseball teams winning a match and betting on the color of a ball drawn in an urn with 50 orange and 50 blue balls (outcomes of the different bets being the same). The subject responded “I would naturally prefer the drawing. [...] Because with the urn I know the probability of winning.” Raiffa put forth convincing normative arguments that this should not be the case. However, many descriptive studies show that this subject is not an outlier and that decision makers do not weight known probabilities and uncertain events equally in choice (e.g., Einhorn and Hogarth 1985, 1986; Hogarth and Einhorn 1990).

Furthermore, there do not exist only two cases—known and unknown probabilities. When probabilities are unknown, very different situations might occur in terms of knowledge. Betting on the temperature in a city one knows very little about is different from betting on the temperature in the city one lives in, even if the exact probabilities are not known in both cases. This dependency on information implies the necessity of defining a *source of uncertainty*, a set of events that are generated by a common mechanism of uncertainty.

This concept was developed by Heath and Tversky (1991), Tversky and Kahneman (1992), Tversky and Fox (1995), and Tversky and Wakker (1995). For

instance, football games correspond to events that belong to a first source of uncertainty, whereas basketball games belong to another source. Football fans would rather bet on a football team than on a basketball team even if they think that both teams have the same chance of winning, because they have considerable knowledge about the chances of winning a given football team has, but very little knowledge about basketball teams. Many experimental findings are consistent with source-dependent attitudes (Abdellaoui et al. 2008, Dolan and Jones 2004, Fox and See 2003, Fox and Tversky 1998, Kilka and Weber 2001). Consequently, *source dependence*, i.e., the impact of the sources of uncertainty on choice, may generate biases in any elicitation technique that mixes sources. It is worth noting that the lottery method implies a direct comparison of events from two different sources (e.g., the temperature and an external “objective” device), whereas the exchangeability method does not (the introductory example only dealt with temperature).

Let us express this argument in a more formal manner. Denote S_A as a set of possible states that pertain to a similar mechanism of uncertainty (called A) and such that one and only one of these states is true. (E, x) is a *binary prospect* that yields the strictly positive outcome ϵx if an event (a subset of S_A) $E \subset S_A$ occurs, and nothing otherwise. When the probability of E is known and equal to p , such a prospect will be denoted as (p, x) . Assuming $u(0) = 0$, a binary prospect (E, x) is represented by $P(E)u(x)$ if subjective expected utility holds (Savage 1954). Source dependence is typically modeled through a *source (dependent weighting) function*, i.e., an increasing function w_A from $[0, 1]$ to $[0, 1]$ that need not be additive and such that $(E, x) \mapsto w_A(P(E))u(x)$.

The exchangeability method only refers to events from the same source. If two events E and F are revealed to be exchangeable, this implies that $w_A(P(E)) = w_A(P(F))$, and thus $P(E) = P(F)$. The lottery method needs to refer to two sources of uncertainty, one of these being a source with objective probabilities. Assume that E pertains to source A , and let R represent the source with objective probabilities. $(E, x) \sim (p, x)$ does not imply $P(E) = p$ but $w_A(P(E)) = w_R(p)$. Several studies support models with source functions (see the list above). Therefore, they suggest

that the exchangeability technique is more robust than the lottery method.

PREDICTION 1 (SOURCE DEPENDENCE). *The exchangeability method is not biased by source dependence, whereas the lottery method is.*

2.2. Additivity of the Probability Measure

Probability judgments may be subadditive (e.g., Tversky and Fox 1995) even if they are enunciated by experts (Fox et al. 1996, Redelmeier et al. 1995), and can even violate set inclusion (Fischhoff et al. 1978, Tversky and Kahneman 1983). Binary additivity—the probabilities of two complementary events summing to one—is sometimes satisfied (e.g., Wallsten et al. 1993) or only weakly violated (e.g., Ariely et al. 2000). However, additivity remains a matter of concern in probability elicitation. Tversky and Koehler (1994) suggested that the subadditivity of subjective probabilities arises from the description of each event, and they proposed their *support theory* in which subjective probabilities are modeled as a ratio of supports in favor of and against an event. They assume that the support of an implicit disjunction is always lower than that of an explicit disjunction. They present the following example: “Ann majors in a natural science” is an implicit disjunction, whereas “Ann majors in either a biological or a physical science” is an explicit one. An explicit disjunction might influence judgments because it provides greater intuition and insights as to when the event can occur than does the implicit disjunction. This type of phenomenon may also appear with choices. A similar phenomenon is indeed observed in the evaluation of uncertain prospects (Johnson et al. 1993).

In the same vein, Starmer and Sugden (1993) and Humphrey (1995) demonstrated the so-called *event-splitting effect (ESE)*, i.e., the fact that two incompatible events appear more attractive to bet on than their union. The ESE can be explained by support theory because splitting an event can be viewed as giving an explicit disjunction of this event. The abovementioned studies highlighted this effect using lottery tickets on which a number is written and a table displaying payoffs as a function of the randomly drawn number. Hence, the ESE cannot really imply that decomposing an event helps subjects understand when the event might occur. Rather, it suggests that violations

of additivity (of the probability measure) are strongly embedded in decisions. If subjective probabilities are not robust to the ESE, additivity will be violated. As a consequence, using additivity to derive subjective probabilities from choices might constitute a strong assumption on the exchangeability method and the ESE can challenge it.

Assume that S_A is an interval of the real line that is split into four equally likely intervals $E, F, G,$ and H such that any element of E ($F, G,$ respectively) is lower than any element of F ($G, H,$ respectively). $E \cup F$ and $G \cup H$ can be represented as convex intervals, and they should be revealed to be equally likely. However, $E \cup H$ must be written as a nonconvex union of two intervals, whereas $F \cup G$ can be written as a single convex interval. The nonconvex union of intervals may be viewed as an explicit disjunction, whereas the convex union is likely to be viewed as an implicit disjunction. Hence, it is quite likely that the ESE will occur: $F \cup G$ will appear less attractive to bet on than $E \cup H$.

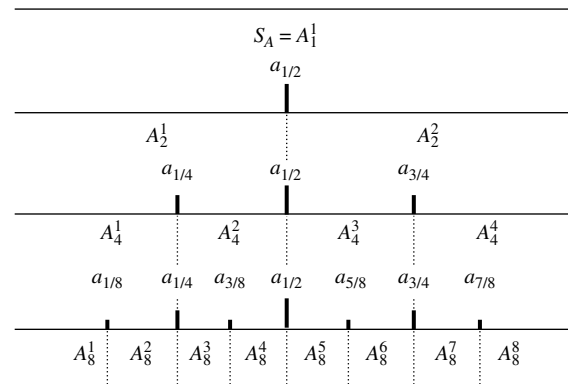
PREDICTION 2 (ADDITIVITY). *Through the ESE, the nonadditivity of beliefs may generate inconsistencies in probability measurements based on the exchangeability method.*

3. Experimental Design

3.1. Subjects

Fifty-two subjects (25 females and 27 males) participated in the experiment conducted during March–May 2005. All the participants were students of economics and management (27 subjects) or the social sciences (25 subjects) at Ecole Normale Supérieure de Cachan, France. They were recruited through posters and presentations at the beginning of their courses. None of them were aware of the true goal of the experiment. They were only told that the experimenter wanted to collect their choices in an uncertain framework. The computer-based experiment was conducted through individual interviews, using software specifically developed for the experiment. Each participant was seated in front of a screen in the presence of the experimenter, who entered the participant’s statement into the computer and submitted it after obtaining a clear confirmation.

Figure 1 Decomposition of the State Space



3.2. Elicitation Technique

First, two complementary events A_2^1 and A_2^2 from S_A —a source of uncertainty—are determined such that for some x , $(A_2^1, x) \sim (A_2^2, x)$. From this twofold partition of S , a fourfold one can be generated by splitting each of these two events into two equally likely subevents, i.e., by finding $A_4^1, A_4^2, A_4^3,$ and A_4^4 satisfying $A_4^1 \cap A_4^2 = \emptyset, A_4^3 \cap A_4^4 = \emptyset, (A_4^1, x) \sim (A_4^2, x)$, and $(A_4^3, x) \sim (A_4^4, x)$. An eightfold partition of S_A can be done by splitting $A_4^1, A_4^2, A_4^3,$ and A_4^4 in the same manner. For a given j , the set of the A_j^i s is an exchangeable partition of the state space, i.e., a partition into j exchangeable events. The subjective probability distribution over S_A can be inferred in this manner. Indeed, the events of an n -fold exchangeable partition have the same subjective probability, $1/n$. Figure 1 describes the process and the notations.

Let us explain the intuition underlying the notation A_j^i . The lower index j indicates in how many exchangeable events the state space is split. The upper index i is the number of the specific event, read from left to right. Each A_j^i is then divided into A_{2j}^{2i-1} and A_{2j}^{2i} . The ratio i/j yields the cumulative probability of the right-hand boundary, which is called $a_{i/j}$. Therefore, $A_j^i = (a_{(i-1)/j}, a_{i/j}]$, with the exception of $A_j^1 = (-\infty, a_{1/j}]$ and $A_j^j = (a_{(j-1)/j}, +\infty]$.

3.3. Implementation

Three sources of uncertainty are used: the temperature in Paris (this source will be referred to as S_T , and the corresponding events as $T_j^i = (t_{(i-1)/j}, t_{i/j}]$), the euro/dollar exchange rate expressed as the amount in dollars for one euro (with events $E_j^i = (e_{(i-1)/j}, e_{i/j}] \subset$

S_E), and the daily variation of the French stock index CAC 40 (with $C_j^i = (c_{(i-1)/j}, C_{i/j}] \subset S_C$). Note that the generic notations S_A and A_j^i are reserved for general comments. For each source, uncertainty resolution occurs exactly four weeks after the experiment.

The experiment begins with a 15-minute preparatory task (presentation of the sources, calibration, and training). For each variable, calibration involves asking the subjects for b_0 and b_1 such that they think that there is “almost no chance” that the value of the variable will lie outside the interval $[b_0, b_1]$. The result of the calibration task has strictly no impact on the state space, which remains¹ $(-\infty, +\infty)$, but it is necessary for graphical reasons and to avoid the participant’s belief being influenced. Indeed, $[b_0, b_1]$ is the part of the state space that is displayed as a graduated ruler on the screen. After this preparatory task, the main part of the experiment begins; this part has no time limit and lasts 50 minutes on the average.

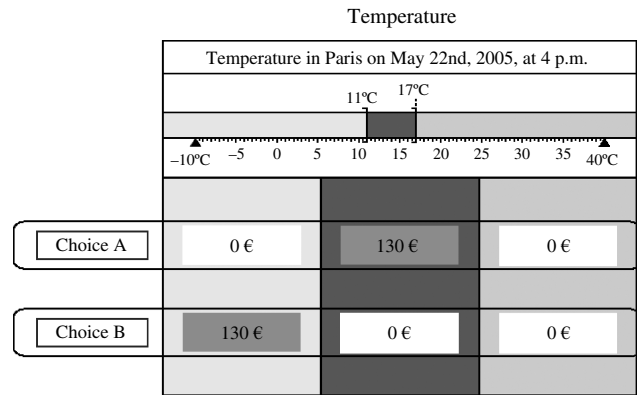
First, $t_{1/2}$ is determined in order to obtain T_1^2 and T_2^2 . The first question is based on the b_0 and b_1 that are previously determined for this particular source, and the subject is offered to bet either on $(-\infty, b_0 + (b_1 - b_0)/2]$ or on $(b_0 + (b_1 - b_0)/2, +\infty)$. Then, the determination of indifferences is conducted through a bisection process until a precision of 0.5°C , $\$0.01$, and 0.1% (for the temperature, exchange rate, and stock index, respectively) is reached. An example and the explanation of the algorithm can be found in the appendix. After this first step, the other $a_{i/j}$ s are elicited, switching between sources and probability levels, until an eightfold partition is obtained for the three sources (the entire list of questions is presented in the appendix). To introduce diversity, the high consequence x of each prospect is randomly drawn between $\text{€}130$, $\text{€}140$, and $\text{€}150$ (but the low consequence is always $\text{€}0$).²

On the screen (see Figure 2), the position of bets on the right-hand and the left-hand events is randomly mixed across questions.

¹ In this experiment, the real minima are certainly not $-\infty$ but -273.15°C , $0 \$$, and -100% for the temperature in Paris, the euro/dollar, and CAC 40, respectively.

² In March–May 2005, the exchange rate was about $\text{€}1 = \$1.30$.

Figure 2 Screenshot of a Choice Related to the Elicitation of $t_{1/4}$



3.4. Consistency

The first test of consistency is conducted by repeating the fourth choice that was made during the bisection process of $t_{1/2}$, $t_{1/4}$, $t_{3/4}$, $e_{1/2}$, $e_{1/4}$, $e_{3/4}$, $c_{1/2}$, $c_{1/4}$, and $c_{3/4}$ and by computing the rate of identical answers across questions. Then, a second method of testing the consistency of the elicited probabilities involves checking whether some qualitative predictions are satisfied: Assume that $a' \in A_8^6 = (a_{5/8}, a_{3/4}]$, that is, $5/8 < P((-\infty, a']) \leq 3/4$; further, assume that $(-\infty, a']$ is exchangeable with $(a'', a']$. Hence, $5/16 < P((-\infty, a'']) \leq 3/8$ and a'' should be in $A_8^3 = (a_{1/4}, a_{3/8}]$ and is likely to be closer to $a_{3/8}$ than to $a_{1/4}$. Unlike the previous test, which only consists in repeating one choice for each of the nine cases $(t_{1/2}, \dots, c_{3/4})$, this test is implemented by determining a new indifference through the same bisection process as the one that estimates the main boundaries.

3.5. Testing Prediction 1 (Source Dependence)

Prediction 1 is tested through preferences across sources, i.e., an event from one source can be strictly preferred to an equally likely event from another source. The participants are thus asked for their preferences between bets, either on an event with a known probability $p = 1/4$ or on T_4^1 , E_4^1 , or C_4^1 ; i.e., they have to rank $(p, 140)$, $(T_4^1, 140)$, $(E_4^1, 140)$, and $(C_4^1, 140)$ in the order of preference. This question is replicated with $p = 1/2$, T_2^1 , E_2^1 , and C_2^1 , and then with $p = 7/8$, $S_T - T_8^8$, $S_E - E_8^8$, and $S_C - C_8^8$. It is noteworthy that if there is no strict preference between sources, the lottery method will be equivalent to the exchangeability method. Indeed, this test can be viewed as a

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

means to verify whether an event with probability p according to the exchangeability method would have the same probability according to the lottery method. If this is not the case, numerous studies suggest that the difference arises from source dependence, implying that the lottery method is biased.

3.6. Testing Prediction 2 (Additivity)

The test of Prediction 2 concerns whether $T_4^1 \cup T_4^4$ and $T_4^2 \cup T_4^3$ are exchangeable. This test compares a convex event ($T_4^2 \cup T_4^3$) with a nonconvex one ($T_4^1 \cup T_4^4$), which is a weak point of the exchangeability method according to the ESE. Indeed, the first event will be represented on the ruler as a unique interval and will therefore be an implicit disjunction, whereas the second event will correspond to an explicit disjunction with two subintervals of the ruler. Prediction 2 says that $T_4^1 \cup T_4^4$ should be preferred. This test will be implemented by eliciting two exchangeable events, $(-\infty, t_{1/4}] \cup (t'_{3/4}, +\infty)$ and $(t_{1/4}, t'_{3/4}]$. Then, $t'_{3/4}$ should be compared with the original value $t_{3/4}$. If they differ, two probability measurements with the same method will imply two different results. Similar tests are implemented with the euro/dollar exchange rate and with CAC 40.

3.7. Real Incentives

There is no flat payment in this experiment. During the presentation, the participants are told that some questions will be played for real. First, a subject will be randomly selected, and then, one of his/her answers (except those for the ranking between bets on the different sources) will be played for real. The same process will then be applied to another subject. The subjects are told that at most one question per subject will be drawn and that the process will stop as soon as four subjects have won. This implies that at most €600 would be distributed to the subjects. Four weeks after each experiment, the values taken by the variables are recorded. After the last record is made, the payments are determined.³

It could be argued that even if real incentives are implemented, the process may not be incentive compatible owing to the chaining between the questions

and the use of a bisection process (Harrison 1986). However, there exist several strong arguments suggesting that these problems do not appear. First, the randomization of questions between the sources of uncertainty makes the chaining unclear. Second, in our experiment, there exists no simple alternative strategy that clearly dominates telling the truth, even for a participant that is completely acquainted with the elicitation process. Moreover, telling the truth is the simplest strategy for subjects who decide to maximize their gains and minimize the cognitive cost that the experiment entails. Eventually, repeated choices, which are not chained with subsequent questions, prove the consistency of the data (see the next subsection).

4. Results

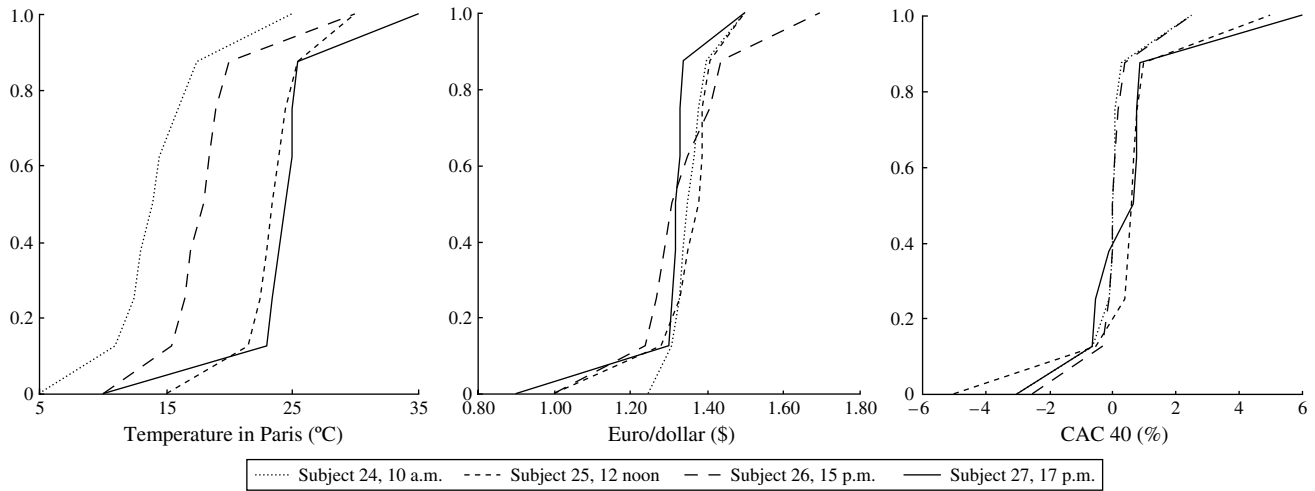
4.1. Subjective Probabilities and Consistency

In the experiment, 156 probability distributions are elicited. Because each distribution concerns a particular day, mean or median results are irrelevant. To present an idea of the result obtained, Figure 3 displays the probability distributions of four subjects (data are in the appendix), who had to bet on the values taken by the variables on May 3rd (each at a different time).

First, the consistency of the method needs to be checked. Some binary choices are therefore repeated at the end of the experiment. For 70.51% of the repeated questions, the preferred bet is identical to the first one. These questions concern the three sources and the determination of $a_{1/2}$, $a_{1/4}$, and $a_{3/4}$. According to Cochran tests, the consistency is not significantly different across question types $a_{1/2}$, $a_{1/4}$, or $a_{3/4}$ ($\chi_2^2 = 2.16$, $p = 0.34$) nor across sources ($\chi = 0.53$, $p = 0.77$). To further analyze consistency, the rate of identical second preferences should be related to the distance from indifference: When participants are indifferent between two bets, they are supposed to randomly choose the events they bet on. Thus, if the repeated questions only concern these indifferent bets, the consistency rate will be close to 50%. This is why the distance from the indifference is measured as the absolute value of the difference between the common boundary of the two events on which the participant can bet and the previously obtained value

³ In June 2005, participants 6 and 17 received €140 and participants 25 and 49 received €150.

Figure 3 Subjective Probability Distributions Obtained for May 3rd



$t_{1/2}, \dots, c_{3/4}$. Figure 4 displays the proportion of consistent second choices; the abscises represent the distance from indifference. Because any departure from indifference dramatically increases the rate of identical second answers, our results strongly suggest that the exchangeability method is able to catch true indifferences and, therefore, subjective probabilities.

The second way to evaluate the exchangeability method involves predicting the qualitative features of beliefs and then testing for consistency of the agent's behavior with respect to these predictions. Recall that $a' \in A_8^6 = (a_{5/8}, a_{3/4}]$ was first chosen arbitrarily, implying that $5/8 < P((-\infty, a']) \leq 3/4$; then, we determine a'' such that $(-\infty, a'')$ is exchangeable with $(a'', a']$. Eventually, a'' should be in $A_8^3 = (a_{1/4}, a_{3/8}]$ and should be closer to $a_{3/8}$ than to $a_{1/4}$. Three predictions follow:

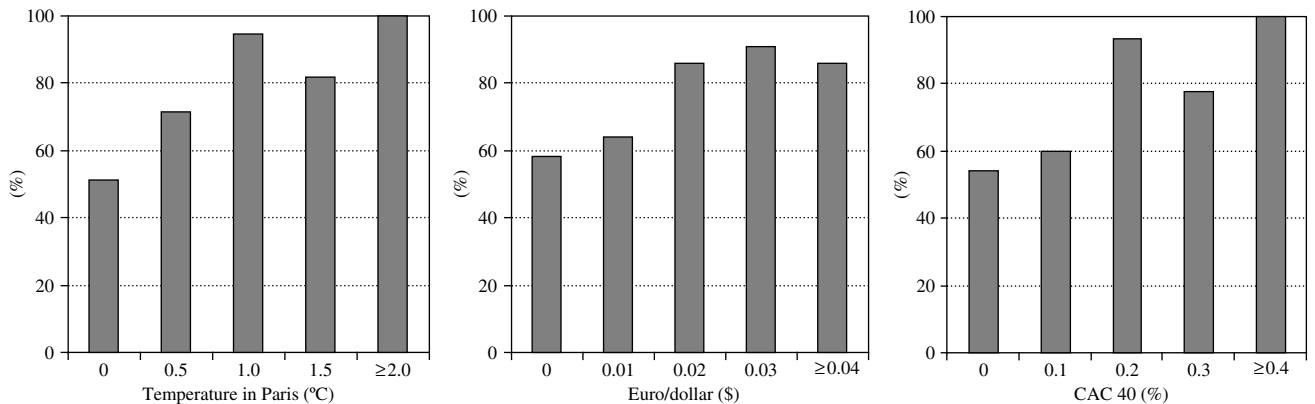
$a_{1/4} < a', a' < a_{3/8}$, and $|a_{3/8} - a'| < |a' - a_{1/4}|$. Table 1 displays the statistics.

Most of the predictions are verified. a' is indeed higher than $a_{1/4}$ for all sources, and lower than $a_{3/8}$ for the exchange rate and the stock index. Furthermore, it is closer to $a_{3/8}$ than to $a_{1/4}$ for the temperature (and CAC 40 at the significance level of 10%). The low identity coefficient and the high p -values of the exchange rate indicate that the answers are noisier or less consistent for this specific variable. As a conclusion of this test, the participants' behavior does not appear inconsistent with respect to the elicited probability distribution.

4.2. Prediction 1: Source Dependence

The interest of this method (and the nonrobustness of the lottery method) is grounded on the preference

Figure 4 Proportion of Consistent Second Answers with Respect to the Distance From Indifference



INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Table 1 Consistency Tests

	Temperature	Euro/dollar	CAC 40
$H_1: a_{1/4} < a'$			
Mean of $a_{1/4}$	17.8 °C	1.26 \$	−0.25%
Mean of a'	18.6 °C	1.27 \$	−0.12%
Identity coefficient*	0.90	0.94	0.70
p -values (paired t -test)	0.00	0.00	0.00
N	52	52	52
$H_1: a' < a_{3/8}$			
Mean of a'	18.6 °C	1.27 \$	−0.12%
Mean of $a_{3/8}$	18.6 °C	1.28 \$	−0.02%
Identity coefficient*	0.94	0.94	0.80
p -values (paired t -test)	0.48	0.01	0.00
N	52	52	52
$H_1: a_{3/8} - a' < a' - a_{1/4} $			
Mean of $ a_{3/8} - a' $	0.8 °C	0.02 \$	0.19%
Mean of $ a' - a_{1/4} $	1.1 °C	0.02 \$	0.23%
Identity Coefficient*	0.81	0.24	0.38
p -values (paired t -test)	0.00	0.43	0.09
N	52	52	52

*Zegers' (1986) chance-corrected identity coefficient.

Table 2 Comparison Between Each Source and the Known Probabilities

	1/4	1/2	7/8
Risk			
Median rank	3	2	1
Mean rank	3.12	2.12	1.54
Temperature			
Median rank	2	3	3
Mean rank	2.35	2.92	3.04
p -value (Wilcoxon)	0.00	0.00	0.00
Ambiguity attitude	AS	AA	AA
Euro/dollar			
Median rank	2	2	3
Mean rank	2.35	2.29	2.62
p -value (Wilcoxon)	0.00	0.34	0.00
Ambiguity attitude	AS	AN	AA
CAC 40			
Median rank	2	2	3
Mean rank	2.06	2.38	2.79
p -value (Wilcoxon)	0.00	0.12	0.00
Ambiguity attitude	AS	AN	AA

for betting on a given event instead of another even if their occurrence is equally likely. A test that aims to compare events having the same subjective probability is implemented in this section. This test is built on four bets, one on an event from each of the three sources, and one with an explicit probability of winning. Each bet displays the same (subjective) probability distribution over outcomes and the participants are asked to rank them. A Friedman test on ranks shows the significantly different ranks between the three sources and the known probability source for each of the three probabilities under consideration ($p = 0.02$ at probability $1/2$, $p = 0.00$ at probabilities $1/4$ and $7/8$). Removing the known probabilities from the analysis, the subjects are still significantly influenced by the source at probability $1/2$ ($p = 0.04$), but not at the other probabilities.

Table 2 displays the results of a signed rank test (Wilcoxon test) that compares the rank of each source with that of the known probability source at the three probability levels under consideration. Assuming that the three sources are more *ambiguous* than the source with objective probability, such a comparison highlights the ambiguity attitude. If a subject bets on an ambiguous event rather than on a known probability, he/she is ambiguity seeking (AS). The opposite preference is characteristic of ambiguity aversion (AA).

Indifference between these bets suggests ambiguity neutrality (AN). Our results clearly show that AA increases with probability, which is consistent with the previous literature on ambiguity (e.g., Hogarth and Einhorn 1990).

These results confirm that the lottery method would not be equivalent to the exchangeability method because of source dependence. For instance, if an event has a probability of $7/8$ according to the exchangeability method, the lottery method will induce a lower probability, only because the subjects prefer known probabilities. This clearly constitutes a strong advantage of the exchangeability method over the lottery method.

4.3. Prediction 2: The Event-Splitting Effect

Let us challenge our method with a weakness of subjective probability elicitation, namely, the nonadditivity of beliefs and sensitivity to the description of the events. The test deals with two events, one being a convex union of disjoint exchangeable subevents, and the other a nonconvex union of such events. Does the ESE occur and induce a violation of exchangeability? For each source, two events $(-\infty, a_{1/4}] \cup (a'_{3/4}, +\infty)$ and $(a_{1/4}, a_{1/2}] \cup (a_{1/2}, a'_{3/4}] = (a_{1/4}, a'_{3/4}]$ are elicited such that they are revealed to be equally likely. Then, $a'_{3/4}$ and the original value $a_{3/4}$ are compared. Table 3 displays the results.

Table 3 Exchangeability vs. ESE

	Temperature	Euro/dollar	CAC 40
$H_1: a'_{3/4} \neq a_{3/4}$			
Mean of $a'_{3/4}$	21.1°C	1.35 \$	0.55%
Mean of $a_{3/4}$	20.8°C	1.33 \$	0.50%
Identity coefficient*	0.92	0.88	0.57
p -values (paired t -test)	0.12	0.02	0.52
N	50	46	42

*Zegers' (1986) chance-corrected identity coefficient.

Let us first explain why there are fewer than 52 observations. A few participants change their minds during the experiment, especially when they do not have sufficient knowledge about the source. Those who think they have overestimated $a_{1/4}$ (such that $P((-\infty, a_{1/4}]) \geq 1/2)$ clearly always preferred $(-\infty, a_{1/4}] \cup (a'_{3/4}, +\infty)$ even when $a'_{3/4}$ tends to infinity. This is the case when the new belief coincides with a subjective probability of $(-\infty, a_{1/4}]$ that is higher than $1/2$. Whereas this problem appears once for the temperature in Paris, it appears five times for the stock index, which with the subjects were less familiar. To compensate for the bias generated by the absence of these observations, the same number of participants exhibiting the opposite behavior (i.e., participants who think they have underestimated $a_{1/4}$) are eliminated. This is why 10 subjects are missing for the stock index.

Note that $a'_{3/4} - a_{3/4} > 0$ indicates that $A_4^1 \cup A_4^4$ is preferred to $A_4^2 \cup A_4^3$ and, consequently, that the nonconvex events appears more attractive. This corresponds to the ESE, and it acquires significance only for the exchange rate. One can conclude that even with the exchangeability method, belief can be manipulated through the description of the events: An explicit disjunction may appear to have more support than an implicit one. However, the original elicitation process does not suffer from such limitation because it entails comparing similar convex events.

5. Discussion

5.1. Scoring Rules

One could argue that there are more elicitation techniques available apart from the lottery method or bisection-based techniques. For instance, *scoring rules*

are often applied in practice. According to Winkler (1969, p. 1073), a scoring rule is “a payoff function which depends on the assessor’s stated probabilities and on the event which actually occurs.” It is used “to keep the assessor honest” and “to evaluate assessors and to help them to become ‘better’ assessors.” The assessors have to choose between the prospects generated by the payoff function, and they have to report the probability that yields their preferred prospect.

However, scoring rules are defined such that an expected value maximizer has incentive to tell the truth, but an expected utility maximizer will report higher (lower) probabilities than what (s)he thinks when the real probability is strictly lower (higher) than $1/2$ (Murphy and Winkler 1970, Kadane and Winkler 1988). The methods under consideration in this paper do not suffer from such a limitation. There exist some attempts to correct scoring rules (e.g., Offerman et al. 2007), but the results of these experiments are equivalent to those obtained using the lottery method (assuming the model introduced in §2.1).

5.2. Measuring the Impact of Source Dependence

The results of the experiment are in favor of the exchangeability method because it is not influenced by the sources of uncertainty. However, even if the lottery method is biased, one could argue that the extent of this bias should be studied. Assuming that subjective probabilities are additive, a simple technique to observe the bias of the lottery method would be to sum the probabilities of an n -fold partition of the state space. The more this sum differs from one, the more biased are the elicited probabilities. Nevertheless, Prediction 2 highlights the limitations of such reasoning. Nonadditivity may be intrinsic in subjective probability and might therefore not be a good tool with which to discriminate between biases and beliefs.

Budescu et al. (2008) argue that in most cases, source dependence is compensated for by the tendency of reporting probabilities that are multiples of 0.05 or 0.10. In this case, this precision (or rather, imprecision) is sufficient for the source bias to be nonsignificant. In other words, the lottery method may suffer from even stronger biases (the tendency of

reporting only multiples of 0.05 or 0.10) than source dependence.

5.3. Additivity

A major argument can be put forward in favor of additive beliefs:

“Objective rules of coherence (the axioms and theorems of probability theory) must be strictly obeyed in any subjective probability evaluation. Coherence is necessary to prevent substantial contradictions, such as the possibility of incurring sure losses as a result of an action” (de Finetti 1974, p. 2).

During the elicitation process, the exchangeability method enforces a built-in additivity that may be useful to prevent actions possibly based on this evaluation from violating rationality rules.

However, the ESE and nonadditivity can pose a problem; this was confirmed by the experiment. The exchangeability method ensures additivity so long as only simple, convex events will be compared. A practical implication follows. The implementation of the exchangeability method should not compare any implicit disjunction with any explicit one. In the elicitation process proposed in the first part of the experiment (before testing the predictions), no such comparisons were used. Moreover, further research is required to verify whether other effects related to support theory or, more generally, to belief nonadditivity, may generate biases in probability measurement through the exchangeability method.

5.4. State-Dependent Preferences

The first limitation of most choice-based elicitation techniques is that they assume Savage’s (1954) separation between utilities and consequences. Thus, they are ineffective when utilities are state dependent, i.e., when the agent associates an intrinsic utility to the states of the world. For instance, even if you think that the probability of raining tomorrow is one half, you may not be indifferent between winning an umbrella if it rains and winning one if it does not rain. This issue also occurs when the agent has “stakes” in an event. Assume that an ice cream seller must choose between winning \$1,000 if the temperature is higher than 20°C and the same amount if the temperature is lower than 20°C. He might prefer the second gamble—not because he thinks the event is more likely, but because he wants to cover a potential loss.

Kadane and Winkler (1988) discussed the relevance of what they term the “no-stakes condition,” which is necessary to prevent biases in any choice-based technique. They pointed out various situations in which it cannot hold. In the experiment presented in this paper, subjects might have had stakes in the event: for instance, if they have a stock portfolio. However, unless they were managing their (large) portfolio every day, it is quite unlikely that the consequences implied by a daily variation in the stock index were significant with respect to what they could earn in the experiment (approximately €140).

Dreze (1988) and Karni (2008) suggested a solution for eliciting subjective probabilities even with state-dependent preferences. They theoretically proved the existence of a unique subjective probability measure as soon as there is moral hazard, i.e., as soon as the person whose beliefs are under consideration can act in a manner that will impact the events’ likelihood. Nonetheless, a practical implementation of this result is yet to be developed.

6. Conclusion

This paper dealt with the exchangeability method, i.e., the choice-based implementation of the general elicitation technique entailing splitting a state space into equally likely events. It demonstrated how this technique may benefit from the absence of an external device like objective probabilities. By avoiding a comparison with a situation with more information, the exchangeability method is more robust than, for instance, the lottery method. However, it needs to assume the additivity of the probability measure although violations of this property can be found. As a conclusion, the exchangeability method is preferable to the lottery method if source dependence is strong; however, the implementation should preclude any comparison between explicit and implicit disjunctions.

Acknowledgments

The author thanks Mohammed Abdellaoui and Peter Wakker for their helpful advice for this study. Han Bleichrodt, Nathalie Etchart-Vincent, Olivier L’Haridon, and Laetitia Placido are gratefully acknowledged for their comments on earlier drafts of the paper. The experiment was conducted at GRID (ENSAM-ESTP, in Paris, France) and received financial support from the RiskAttitude project

(ANR05-BLAN-0345-01). This research was also made possible by a VICI grant from the Netherlands Organization for Scientific Research. This paper particularly benefitted from the comments of the editor, the associate editor, and two referees.

Appendix

A.1. Algorithm

Table A.1 presents an example of the implementation of the exchangeability method for the temperature in Paris and Participant 24 (see Figure 3 for the entire elicited probability distribution).

According to Subject 24, it was very unlikely that the temperature in Paris on May 3rd would be outside the interval [5°C, 25°C]. This is why the first choice is based on the midpoint 15°C. Because option A was preferred (that is, a lower temperature is more likely), the second question concerns the midpoint between 5°C and 15°C. The precision used to determine the midpoints was 0.5°C. It can be observed that the last two questions of the first part of the table suggest that $t_{1/2}$ lies between 13.5°C and 14°C. The maximum (14°C) was selected. In the second part of the table, $(-\infty, 14^\circ\text{C}]$ is split into two exchangeable subevents. The same precision and the same rules were used, yielding $t_{1/4} = 12.5^\circ\text{C}$. The precision for the exchange rate and the stock index were \$0.01 and 0.1%, respectively.

Table A.1 Choices of Subject 24 for the Elicitation of $a_{1/2}$ and $a_{1/4}$

Searched boundary	Option A	Option B	Choice
$t_{1/2}$	$(t \leq 15^\circ\text{C}, \text{€}140)$	$(t > 15^\circ\text{C}, \text{€}140)$	A
	$(t \leq 10^\circ\text{C}, \text{€}140)$	$(t > 10^\circ\text{C}, \text{€}140)$	B
	$(t \leq 12.5^\circ\text{C}, \text{€}130)$	$(t > 12.5^\circ\text{C}, \text{€}130)$	B
	$(t \leq 14^\circ\text{C}, \text{€}140)$	$(t > 14^\circ\text{C}, \text{€}140)$	A
	$(t \leq 13.5^\circ\text{C}, \text{€}150)$	$(t > 13.5^\circ\text{C}, \text{€}150)$	B
$t_{1/4}$	$(t \leq 9.5^\circ\text{C}, \text{€}150)$	$(9.5^\circ\text{C} < t \leq 14^\circ\text{C}, \text{€}150)$	B
	$(t \leq 12^\circ\text{C}, \text{€}140)$	$(12^\circ\text{C} < t \leq 14^\circ\text{C}, \text{€}140)$	B
	$(t \leq 13^\circ\text{C}, \text{€}150)$	$(13^\circ\text{C} < t \leq 14^\circ\text{C}, \text{€}150)$	A
	$(t \leq 12.5^\circ\text{C}, \text{€}130)$	$(12.5^\circ\text{C} < t \leq 14^\circ\text{C}, \text{€}130)$	A

Table A.2 Order of Elicitations

	$a_{1/2}$	$a_{1/4}$	$a_{3/4}$	$a_{1/8}$	$a_{3/8}$	$a_{5/8}$	$a_{7/8}$	a'	$a'_{3/4}$
Temperature in Paris	1	4	7	10	16	19	13	22	25
Exchange rate	2	8	5	14	20	11	17	26	23
Stock index CAC 40	3	6	9	18	12	15	21	24	27

A.3. Data for Figure 3 (Table A.3)

Table A.3 Elicited Points for the Graphs of Figure 3

	b_0	$a_{1/8}$	$a_{1/4}$	$a_{3/8}$	$a_{1/2}$	$a_{5/8}$	$a_{3/4}$	$a_{7/8}$	b_1
Temperature in Paris (°C)									
Subject 24	5	11	12.5	13	14	14.5	16	17.5	25
Subject 25	15	21.5	22.5	23	23.5	24	24.5	25.5	30
Subject 26	10	15.5	16.5	17	18	18.5	19	20	30
Subject 27	10	23	23.5	24	24.5	25	25	25.5	35
Exchange rate (\$)									
Subject 24	1.25	1.31	1.33	1.34	1.35	1.37	1.38	1.40	1.50
Subject 25	1.00	1.28	1.33	1.35	1.38	1.39	1.39	1.41	1.50
Subject 26	1.00	1.24	1.27	1.29	1.31	1.35	1.41	1.44	1.70
Subject 27	0.90	1.30	1.31	1.32	1.32	1.33	1.33	1.34	1.50
Stock index (CAC 40) (%)									
Subject 24	-3	-0.6	-0.1	0	0	0.1	0.1	0.3	2.5
Subject 25	-5	-0.5	0.4	0.5	0.6	0.7	0.8	1	5
Subject 26	-2.5	-0.3	-0.1	0	0	0.1	0.2	0.4	2.5
Subject 27	-3	-0.6	-0.5	-0.1	0.7	0.8	0.8	0.9	6

A.2. Elicitation Order

Table A.2 presents the order of the elicitation: The first elicited point is $t_{1/2}$, followed by $e_{1/2}$, $c_{1/2}$, $t_{1/4}$, $e_{3/4}$...

After this, the consistency test comprising nine repeated binary choices was applied. Eventually, the preferences (see §3.5.) across sources were elicited.

References

Abbas, A., D. V. Budescu, H.-T. Yu, R. Haggerty. 2007. A comparison of two probability encoding methods: Fixed probability wheel vs. fixed variable values. Working paper, University of Illinois, Urbana-Champaign.

Abdellaoui, M., A. Baillon, P. P. Wakker. 2008. Combining Bayesian beliefs and willingness to bet to analyze attitudes towards uncertainty. Working paper, Econometric Institute, Erasmus University, Rotterdam, The Netherlands.

Abdellaoui, M., A. Baillon, L. Placedo, P. P. Wakker. 2008. The rich domain of uncertainty. Working paper, Econometric Institute, Erasmus University, Rotterdam, The Netherlands.

Ariely, D., W. T. Au, R. H. Bender, D. V. Budescu, C. B. Dietz, H. Gu, T. S. Wallsten, G. Zauberman. 2000. The effects of averaging subjective probability estimates between and within judges. *J. Experiment. Psych.: Appl.* 6 130–147.

Budescu, D., A. Abbas, L. Wu. 2008. Does probability weighting matter in probability elicitation? Working paper, University of Illinois, Urbana-Champaign.

Chesley, G. R. 1978. Subjective probability elicitation techniques: A performance comparison. *J. Accounting Res.* 16 225–241.

Chew, S. H., J. Sagi. 2006. Event exchangeability: Probabilistic sophistication without continuity or monotonicity. *Econometrica* 74 771–786.

de Finetti, B. 1937. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'I.H.P.* 7 1–68.

de Finetti, B. 1974. The value of studying subjective evaluations of probability. C.-A. S. Staël von Holstein, ed. *The Concept of Probability in Psychological Experiments*. D. Reidel Publishing Company, Dordrecht-Holland, 2–14.

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

- Dolan, P., M. Jones. 2004. Explaining attitudes towards ambiguity: An experimental test of the comparative ignorance hypothesis. *Scottish J. Political Econom.* **51** 281–301.
- Drèze, J. H. 1987. *Essays on Economic Decision Under Uncertainty*. Cambridge University Press, Cambridge, UK.
- Einhorn, H. J., R. M. Hogarth. 1985. Ambiguity and uncertainty in probabilistic inference. *Psych. Rev.* **92** 433–461.
- Einhorn, H. J., R. M. Hogarth. 1986. Decision making under ambiguity. *J. Bus.* **59** S225–S250.
- Ellsberg, D. 1961. Risk, ambiguity and the Savage axioms. *Quart. J. Econom.* **75** 643–669.
- Fellner, W. 1961. Distortion of subjective probabilities as a reaction to uncertainty. *Quart. J. Econom.* **75** 670–689.
- Fischhoff, B., P. Slovic, S. Lichtenstein. 1978. Fault trees: Sensitivity of estimated failure probabilities to problem representation. *J. Experiment. Psych.: Human Perception Performance* **4** 330–344.
- Fox, C. R., B. A. Rogers, A. Tversky. 1996. Options traders exhibit subadditive decision weights. *J. Risk Uncertainty* **13** 5–17.
- Fox, C. R., K. E. See. 2003. Belief and preference in decision under uncertainty. D. Hardman, L. Macchi, eds. *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making*, Wiley, New York, 273–314.
- Fox, C. R., A. Tversky. 1998. A belief-based account of decision under uncertainty. *Management Sci.* **44** 879–895.
- Harrison, G. W. 1986. An experimental test for risk aversion. *Econom. Lett.* **21** 7–11.
- Heath, C., A. Tversky. 1991. Preference and belief: Ambiguity and competence in choice under uncertainty. *J. Risk Uncertainty* **4** 5–28.
- Hogarth, R. M., H. J. Einhorn. 1990. Venture theory: A model of decision weights. *Management Sci.* **36** 780–803.
- Humphrey, S. J. 1995. Regret aversion or event-splitting effects? More evidence under risk and uncertainty. *J. Risk Uncertainty* **11** 263–274.
- Johnson, E. J., J. C. Hershey, J. Meszaros, H. C. Kunreuther. 1993. Framing, probability distortions, and insurance decisions. *J. Risk Uncertainty* **7** 35–51.
- Kadane, J. B., R. L. Winkler. 1988. Separating probability elicitation from utilities. *J. Amer. Stat. Assoc.* **83** 357–363.
- Karni, E. 2007. Bayesian decision theory and the representation of beliefs. Working paper, Johns Hopkins University, Baltimore.
- Karni, E. 2008. A theory of Bayesian decision making. Working paper, Johns Hopkins University, Baltimore.
- Kilka, M., M. Weber. 2001. What determines the shape of the probability weighting function under uncertainty. *Management Sci.* **47** 1712–1726.
- Murphy, A. H., R. L. Winkler. 1970. Scoring rules in probability assessment and evaluation. *Acta Psychologica* **34** 917–924.
- Offerman, T., J. Sonnemans, G. van de Kuilen, P. P. Wakker. 2007. A truth-serum for non-Bayesians: Correcting proper scoring rules for risk attitudes. Working paper, CREED, University of Amsterdam, The Netherlands.
- Raiffa, H. 1968. *Decision Analysis*. Addison-Wesley, London.
- Ramsey, F. P. 1931. Truth and probability (1926). R. B. Braithwaite, ed. *The Foundations of Mathematics and Other Logical Essays*. Routledge, London, 156–198.
- Redelmeier, D. A., D. J. Koehler, V. Liberman, A. Tversky. 1995. Probability judgment in medicine: Discounting unspecified possibilities. *Medical Decision Making* **15** 227–230.
- Savage, L. J. 1954. *The Foundations of Statistics*. Wiley, New York.
- Spetzler, C. S., C.-A. S. Stael Von Holstein. 1975. Probability encoding in decision analysis. *Management Sci.* **22** 340–358.
- Starmer, C., R. Sugden. 1993. Testing for juxtaposition and event-splitting effects. *J. Risk Uncertainty* **6** 235–254.
- Tversky, A., C. R. Fox. 1995. Weighing risk and uncertainty. *Psych. Rev.* **102** 269–283.
- Tversky, A., D. Kahneman. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psych. Rev.* **90** 293–315.
- Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* **5** 297–323.
- Tversky, A., D. J. Koehler. 1994. Support theory: A nonextensional representation of subjective probability. *Psych. Rev.* **101** 547–567.
- Tversky, A., P. P. Wakker. 1995. Risk attitudes and decision weights. *Econometrica* **63** 1255–1280.
- Wallsten, T. S., D. V. Budescu, R. Zwick. 1993. Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Sci.* **39** 176–190.
- Winkler, R. L. 1969. Scoring rules and the evaluation of probability assessors. *J. Amer. Statist. Assoc.* **64** 1073–1078.
- Wright, W. F. 1988. Empirical comparison of subjective probability elicitation methods. *Contemporary Accounting Res.* **5** 47–57.
- Zegers, F. E. 1986. A family of chance-corrected association coefficients for metric scales. *Psychometrika* **51** 559–569.