Journal of Mathematical Psychology 55 (2011) 320–327

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Notes and comment

Does probability weighting matter in probability elicitation?

David Budescu[a,*], Ali Abbas[b], Lijuan Wu[c]

[a] Fordham University, United States
[b] University of Illinois at Urbana-Champaign, United States
[c] University of Kentucky, United States

ARTICLE INFO

Article history:
Received 12 February 2008
Received in revised form
1 April 2011
Available online 31 May 2011

Keywords:
Subjective probability
Probability elicitation
Probability weighting
Prospect theory
Decision weights

ABSTRACT

One of the most widely used methods for probability encoding in decision analysis uses binary comparisons (choices) between two lotteries: one that depends on the values of the random variable of interest and another that is contingent on an external reference chance device (typically a probability wheel). This note investigates the degree to which differences in probability weighting functions between the two types of events could affect the practice of subjective probability encoding. We develop a general methodology to investigate this question and illustrate it with two popular probability weighting functions over the range of parameters reported in the literature. We use this methodology to (a) alert decision analysts and researchers to the possibility of reversals, (b) identify the circumstances under which overt preferences for one lottery over the other are not affected by the weighting function, (c) document the magnitude of the differences between choices based on probabilities and their corresponding weighting functions, and (d) offer practical recommendations for probability elicitation.

The elicitation of representative probability distributions of continuous variables is a fundamental step in decision analysis and, as such, has engendered a substantial literature (see O'Hagan et al., 2006; Von Winterfeldt & Edwards, 1986; Wallsten & Budescu, 1983, for partial reviews). Spetzler and Staël von Holstein (1975) distinguished between direct and indirect elicitation methods. Direct methods ask the judges to state, or identify by some other means (such as graphs or sliders), the probability distribution of the random variable of interest, X. Indirect methods, the focus of this note, infer the correspondence between values of X and their probabilities from comparative judgments and/or choices made by the Decision Makers (DMs). For example, in a meteorological context, one could ask: "Is it more likely that the average temperature next year at a given location will be (a) between 20° and 23° or (b) between 24° and 25°?" By asking a series of questions of this format, it is possible to trace the probability distribution of the variable of interest. In general, indirect methods are considered more natural, and less demanding cognitively.

Often the comparisons involve lotteries based on external chance devices with well calibrated probabilities. The most popular device is a probability wheel with two sectors of different colors (say, Orange and Blue) whose relative size is not explicitly stated but can be inferred from their relative area (see discussion of the probability wheel and the "reference experiment" in Chapter 5 of French, 1988). The judge asked "Is it more likely that (a) the average temperature next year at a given location will be between 20° and 23°, or (b) that the wheel will land on the orange sector when it is spun?" When eliciting distributions of a continuous variable, the target event is often specified in the context of its cumulative probability distribution. For example, in the process of eliciting a dose–response curve a judge could be asked "Is it more likely that (a) less than 10% of the at-risk population will be adversely affected if the concentration of a pollutant exceeds C particles/mm^3, or (b) that the wheel will land on the orange sector when it is spun?" (e.g. Wallsten, Forsyth, & Budescu, 1983; Whitfield & Wallsten, 1989). All our arguments in this note apply equally to both cases so, without any loss of generality, we will refer exclusively to cumulative probabilities. These judgments are overt realizations of the binary relation *(weakly) more likely than* and, if the judgments obey certain behavioral axioms (see e.g. French, 1988, Chapter 6), they can be mapped into (unique) subjective probabilities.

Alternatively, these comparisons can be framed as binary choices involving a positive outcome, O. This has the advantage of adding a proper incentive for the judge as his/her choices determine the eventual payoffs. Thus, the judge could be asked "Would you prefer to bet on a lottery that pays $100 if (a) the total amount of precipitation in your town this month is less than 6 cm, or (b) the wheel will land in the orange sector when it is spun?"

These choices are overt realizations of the binary relation *is (weakly) preferred over* and, if the judge's preferences obey the

* Corresponding address: Department of Psychology, Fordham University, 441 East Fordham Road, Bronx, NY 10017, United States.
E-mail address: budescu@fordham.edu (D. Budescu).

0022-2496/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2011.04.002

behavioral axioms of the Subjective Expected Utility model (Sarin & Wakker, 1997; Savage, 1954), they can be mapped into utilities and (unique) probabilities. More precisely, we infer that:

Lottery on $Event$ is ($weakly$) preferred to Lottery on Wheel

iff

SEU (Lottery on $Event$) $\geq$ SEU(Lottery on Wheel).

Kadane and Winkler (1988) discuss the effects of non-linear utility functions on several probability elicitation procedures and the corresponding separation of beliefs from expected utility formulations. They show that when the judge has a "stake" in the outcome of the variable of interest – his/her distribution of wealth is correlated with the variable being assessed – non-linear utility functions do not allow the separation. Here the two lotteries are independent of the actual wealth distribution, and involve the same outcome, *O*. It is reasonable to assume that the utility is invariant over (independent of) the source of the prize (wheel or event), and the inference can now be further simplified (for an excellent review of the composition of risk preference and belief, see Wakker (2004)). The relative (subjective) likelihood of the two events can be inferred from the revealed preferences between the two gambles as,

Lottery on $Event$ is ($weakly$) preferred to Lottery on Wheel

iff

SP ($Event$) $\geq$ SP (Wheel).

This logic applies to any pair-wise comparison (i.e., between any setting on the wheel and any value of the target variable), and it provides the theoretical underpinning for various sequences of choices used to determine the full distribution of the random variable, X. The two most popular sequences (encoding methods) are fixed probability and fixed variable value (see Spetzler & Staël von Holstein, 1975; Abbas, Budescu, Yu, & Haggerty, 2008). The Fixed Probability method uses a fixed setting on the probability wheel and asks judges for the value of the variable whose cumulative probability corresponds to the wheel setting. The second approach assesses the fractiles using a fixed value of the variable (hence, the label Fixed Value) and searches for the probability wheel setting that corresponds to the cumulative probability of that variable value.

Subjective probabilities and decision weights

Many empirical findings caused behavioral decision theorists to question the descriptive validity of the SEU model. Kahneman and Tversky (1979) summarized many of these reservations in their seminal paper on Prospect Theory (PT). Of special relevance for our purpose is their critique of the "expectation principle". Kahneman and Tversky (1979) showed that it is violated systematically, and suggested that when making choices among risky prospects DMs transform probabilities, p, into decision weights, $W(p)$, that are monotonic transformations of the probabilities "which reflect the total impact of p on the overall value of the prospect" (Kahneman & Tversky, 1979). Originally, $W(p)$ was characterized by its key properties—sub-additivity, over-(under-)weighting of low(high) probabilities, sub-certainty, and sub-proportionality. Subsequently, various parametric functional representations have been proposed. The function proposed by Tversky and Kahneman (1992) is:

$$W_1(p) = \frac{p^\beta}{(p^\beta + (1-p)^\beta)^{1/\beta}}. \quad (1)$$

The single parameter, β, determines the nature and magnitude of the discrepancy between $W_1(p)$ and p, by capturing features such as the curvature of the function and its elevation—the location of the cross-over point (where $W_1(p) = p$). The parameter of the function may vary slightly across domains (Tversky & Kahneman, 1992, report median $\beta = 0.61$ for gains and median $\beta = 0.69$ for losses) but there are large variations in the estimates across studies (see Booij, van Praag, & van de Kuilen, 2010; Stott, 2006), as well as between judges (see discussion by Gonzalez & Wu, 1999).

Although probability weighting has received a lot of attention in behavioral decision theory, it has not affected the practice of decision analysis to the same degree (see Abdellaoui, 2000; Bleichrodt, Pinto, & Wakker, 2001, for exceptions). If DMs' comparisons of the lotteries are driven by the principle underlying PT, we can infer that:

Lottery on $Event$ is ($weakly$) preferred to Lottery on Wheel $Event$

iff

$W[SP(Event)] \geq W[SP(Wheel)]$.

Since $W(p)$ is a strictly monotonic function of p, when the two subjective probabilities are converted into decision weights by the same instantiation of the function (i.e., the same value of β), decision weights do not affect choices. However, if the two relevant events are transformed at different rates, this is not necessarily true. For example, consider two events, E_1 and E_2, and assume that a certain person judges $SP(E_1) = 0.50$ and $SP(E_2) = 0.60$. If the decision weight of E_1 is determined by $\beta_1 = 0.8$, but the weight of E_2 is determined by a more extreme $\beta_2 = 0.6$, we obtain $W(SP(E_1)) = 0.48$ and $W(SP(E_2)) = 0.47$. This would reverse the judge's overt preference, so our inferences about the judge's subjective probabilities would be inaccurate.

An obvious case where this could happen is when one of the events is not transformed, i.e., $W[SP(Event)] = SP(Event)$, and the other is, but there could be other circumstances where the probabilities of the two events would be transformed in different ways. Essentially, the probability wheel is a well defined chance event subject to purely aleatory uncertainty that is external to the judge, whereas the target event is described verbally and is governed by different sources of epistemic uncertainty that are internal to the judge. Differential sensitivity to various types of events was documented originally in a series of studies by Fox and Tversky (1998) and Tversky and Fox (1995) and, more recently by Abdellaoui, Baillon, and Wakker (2007) and Kilka and Weber (2001). Wakker (2004) offered a general characterization of such cases by suggesting that DMs are *less sensitive to uncertainty than to risk* (Wakker, 2004, page 238).

The recent interest in the differential pattern of choices in decisions from description and from experience offer new instances where probabilities are weighted differentially. Papers by Hau, Pleskac, Kieffer, and Hertwig (2008) and Ungenmach, Chater, and Stewart (2009) suggest that the probability weighing functions are different for decisions from experience and from description, and Wu, Delgado, and Maloney (2009) have shown such differences between various framings and presentation modes of, essentially, equivalent decision tasks.

In this note we seek to determine the degree to which decision weights could affect the practice of subjective probability encoding. We present a general methodology for addressing this question and illustrate it with two popular weighting functions over the range of parameter values reported in the literature. We use these analyses to (a) identify the circumstances under which overt preferences for one lottery over the other are not affected by the weighting function; (b) document the magnitude of the differences between choices based on probabilities and their corresponding weights in order to determine the severity of distortions in these cases; and (c) draw some practical recommendation for probability elicitations.

We perform this analysis with both the one-parameter probability weighting function used by Tversky and Kahneman (1992),

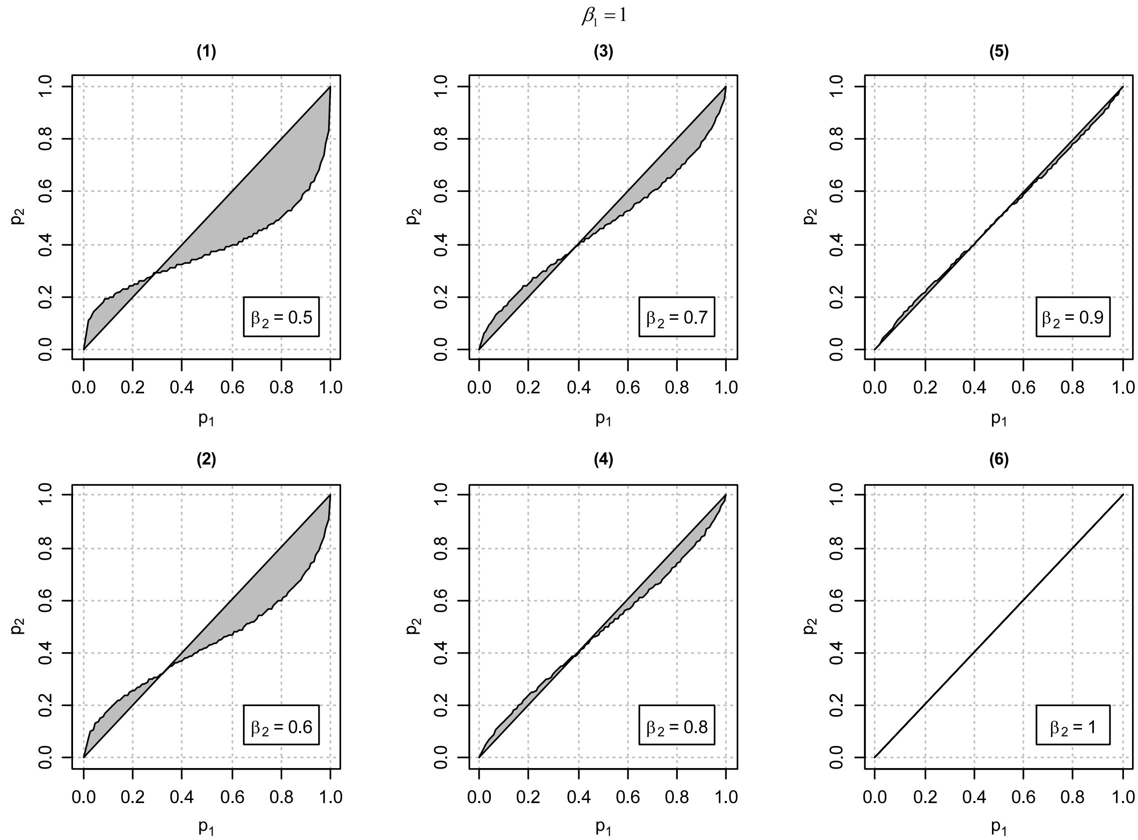

Fig. 1. Reversals with the one-parameter probability weighting function: $\beta_1 = 1.0$ and $\beta_2 = 0.5$ to 1 by 0.1. The shaded areas represent cases where applying the decision weights reverse preferences.

and the two-parameter function proposed by Prelec (1998), where δ controls its elevation and γ its curvature:

$$W_2(p) = \exp(-\delta(-\log(p))^{\gamma}). \tag{2}$$

We do not necessarily claim any general superiority for these functions (but see Stott, 2006), and use them only as illustrations. The methodology used here can be applied to any other function (e.g. Gonzalez & Wu, 1999; Rieger & Wang, 2006).

Analysis of the one-parameter weighting function

Let $W_1(p, \beta_\alpha)$ denote the value of the weighting function of probability p under parameterization, β_α, which can vary as a function of the source of the uncertainty. We can identify the range of values of β_1, β_2 that preserves the original order of the untransformed probabilities, p_1, p_2. We define reversals as cases where sign $(p_1 - p_2) \neq \text{sign}(W_1(p_1, \beta_1) - W_1(p_2, \beta_2))$. The rate of reversals for each relevant β_1, β_2 combination is the area of the unit square where the inequality holds. In practice we estimate this quantity as the fraction of (p_1, p_2) pairs (rounded to the nearest 0.01) for which we observe reversals.

Recent reviews by Booij et al. (2010) and Stott (2006) identified 8 papers in which estimates of the β parameter were reported in the domain of gains. The median estimates from these papers are between 0.56 and 0.96, and when the two most extreme estimates are eliminated the range is reduced to $0.60 \leq \beta \leq 0.91$. Consequently, we focus our analysis on the range $0.50 \leq \beta_1, \beta_2 \leq 1.0$.[1] In support of this choice, note that 37 of the 44 *individual* estimates (84%) reported by Budescu, Kuhn, Kramer, and Johnson (2002) are also in this range.

[1] Rieger and Wang (2006) point out that for low values of β this function is not always monotonic (when $\beta \leq 0.28$, $W_1(p)$ can be a decreasing function of p).

Fig. 1 displays results for the benchmark case where $\beta_1 = 1.0$ (i.e., the first variable is not weighted), while β_2 varies from 0.5 to 1.0 in increments of 0.1, for all pairs of p_1 and p_2 (the plane of the plot). The background area of the plots represents the consistent cases where the overt preferences are not affected by the differential weighting. The shaded area, in the vicinity of the diagonal, designates the cases where differential decision weights induce reversals. This figure illustrates vividly (a) the dependence of the rate of reversals and the discrepancy between the two βs – as long as they are close (e.g., $\beta_2 > 0.7$) reversals are negligible, but they increase for larger discrepancies; and (b) their asymmetric pattern – as long as $\beta_1 > \beta_2$ most reversals occur below the diagonal where $p_1 < p_2$.

Fig. 2 displays similar results for the case where $\beta_1 = 0.75$. The figure shows that in the presence of a moderate degree of distortion of one of the events ($\beta_1 = 0.75$) the effect of the weights is small, and shows up especially when the two probabilities are close to each other (more on this later), and it increases monotonically as a function of the difference $|\beta_2 - \beta_1|$. The order reversals are not distributed symmetrically: Reversals are most prevalent and more pronounced when $\beta_1 > \beta_2$, and occur mostly when $p_1 < p_2$, just as in Fig. 1. The highest rate of violations (14.4% of the various p_1, p_2 pairs) is observed when $\beta_1 = 0.75$ and $\beta_2 = 0.50$. In this case, over 95% of the order reversals are below the diagonal where $p_1 < p_2$. Conversely, when $\beta_1 < \beta_2$ more reversals occur above the diagonal, where $p_1 > p_2$.

Analysis of the two-parameter weighting function

We now consider the two-parameter weighting function (Eq. (2)) proposed by Prelec (1998). We distinguish between reversals – cases where $sign(p_1 - p_2) \neq sign(W_2(p_1; \delta_1, \gamma_1) - W_2(p_2; \delta_2, \gamma_2))$ – and circumstances where $sign(p_1 - p_2) = sign(W_2(p_1; \delta_1, \gamma_1) - W_2(p_2; \delta_2, \gamma_2))$, preserving the preference order implied by the original probabilities. More specifically we seek the constraints

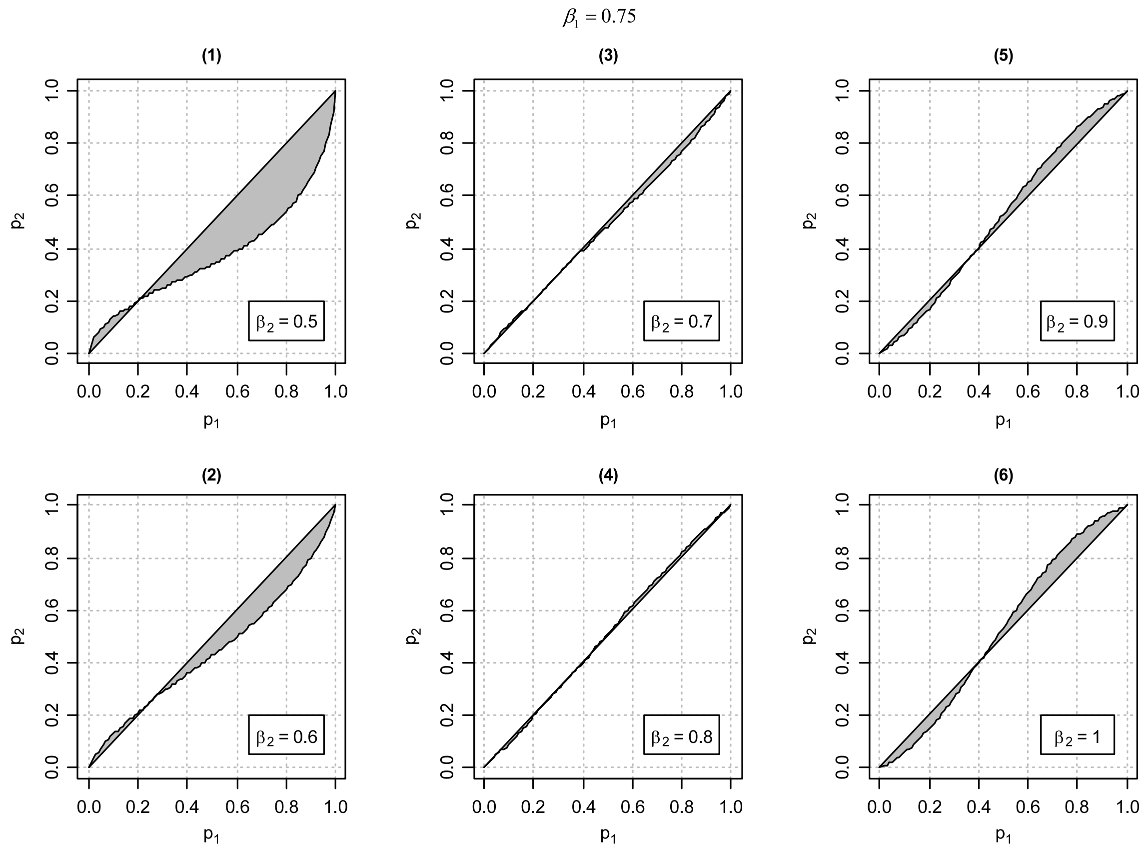

Fig. 2. Reversals with the one-parameter probability weighting function $\beta_1 = 0.75$ and $\beta_2 = 0.5$ to 1 by 0.1. The shaded areas represent cases where applying the decision weights reverse preferences.

on the δs, that control the functions' elevations, and on the γs, that control their curvatures, to satisfy this constraint. If $p_1 > p_2$ invariance under differential weighting is achieved when:

$$\delta_1(-\log(p_1))^{\gamma_1} > \delta_2(-\log(p_2))^{\gamma_2}. \tag{3}$$

This inequality can be mapped into a relationship between δ_1 and δ_2 (assuming fixed γ's):

$$\frac{\delta_1}{\delta_2} > \frac{(-\log(p_2))^{\gamma_2}}{(-\log(p_1))^{\gamma_1}}. \tag{4}$$

Conversely, if one keeps δ_1 and δ_2 constant, it is possible to derive the relationship between γ_1 and γ_2. When $1/e > p_1 > 0$, $\log(-\log(p_1)) > 0$, so

$$\gamma_1 > \frac{\log(\delta_2) - \log(\delta_1) + \gamma_2 \log(-\log(p_2))}{\log(-\log(p_1))}. \tag{5}$$

On the other hand, when $1 > p_1 > 1/e$, and $\log(-\log(p_1)) < 0$, we have:

$$\gamma_1 < \frac{\log(\delta_2) - \log(\delta_1) + \gamma_2 \log(-\log(p_2))}{\log(-\log(p_1))}. \tag{6}$$

The reviews by Booij et al. (2010) and Stott (2006) identified 4 papers reporting estimates of the two parameters of Prelec's function. The median estimates in these papers are such that $0.53 \le \gamma \le 1.05$ and $1.0 \le \delta \le 2.12$. Based on these results, we focus our attention to cases where $0.50 \le \gamma \le 1.0$ and $1.0 \le \delta \le 2.0$. Fig. 3 examines the case where $\delta_1 = \delta_2 = \delta = 1$ where both functions have identical elevations and illustrates the effects of the differences in the curvatures of the two functions. To illustrate the results we consider in Fig. 3 the case where $\gamma_1 = 0.75$, and γ_2 ranges from 0.5 to 1.0 in increments of 0.1. The effects follow a pattern similar to the one-parameter function but are smaller (The highest rate of violations for the case where $\delta = 1$ is observed in the upper left corner plot when $\gamma_1 = 0.75$ and $\gamma_2 = 0.50$).

Next we examine the case where both functions have identical curvatures ($\gamma_1 = \gamma_2 = \gamma = 0.8$) and one of the elevations is fixed at $\delta_1 = 1.0$. Fig. 4 displays the effects of the differences in elevation by changing systematically the value of δ_2. We consider the case $\delta_2 = 0.8$ (upper left corner) and values of δ_2 from 1.2 to 2.0 in increments of 0.2. These effects are much stronger than those associated with the curvatures, and are easy to summarize: their direction is determined by sign($\delta_2 - \delta_1$) and their magnitude increases as a function of $|\delta_2 - \delta_1|$: When $|\delta_2 - \delta_1| = 0.2$ (the left column) we observe only 6.7% violations, but when we consider extreme cases, such as the lower right corner where $|\delta_2 - \delta_1| = 1$, order is reversed in 20.44% of the p_1, p_2 combinations.

Fig. 5 displays the joint effects of the two parameters. We use as a reference a "typical" judge with $\gamma_1 = 0.9$ and $\delta_1 = 1.2$. The effects of different elevations are reflected in the changes across the three rows (corresponding to $\delta_2 = 1.0$, 1.50 and 2.0), and the effects of curvature are displayed in the changes cross the three columns (corresponding to $\gamma_2 = 0.4$, 0.7 and 1.0). As the overall distance between the reference case and the alternative weighting function increases (moving from the upper right to the lower left panels), the rate of reversals increases. The pattern of reversals is consistent with the special cases illustrated in Figs. 3 and 4.

When does probability weighting matter? A sensitivity analysis

The analysis in the previous section suggests that in most cases preferences between pairs of events whose probabilities are weighted differentially would be insensitive to the parameters of the weighting functions. However, there are clear instances of reversals, especially when the two probabilities are close to each other, so it is natural to ask whether these cases are of practical relevance and importance. Given the many possible functions (e.g. Stott, 2006), and their high non-linearity it is impossible to offer a simple unequivocal answer to this question.

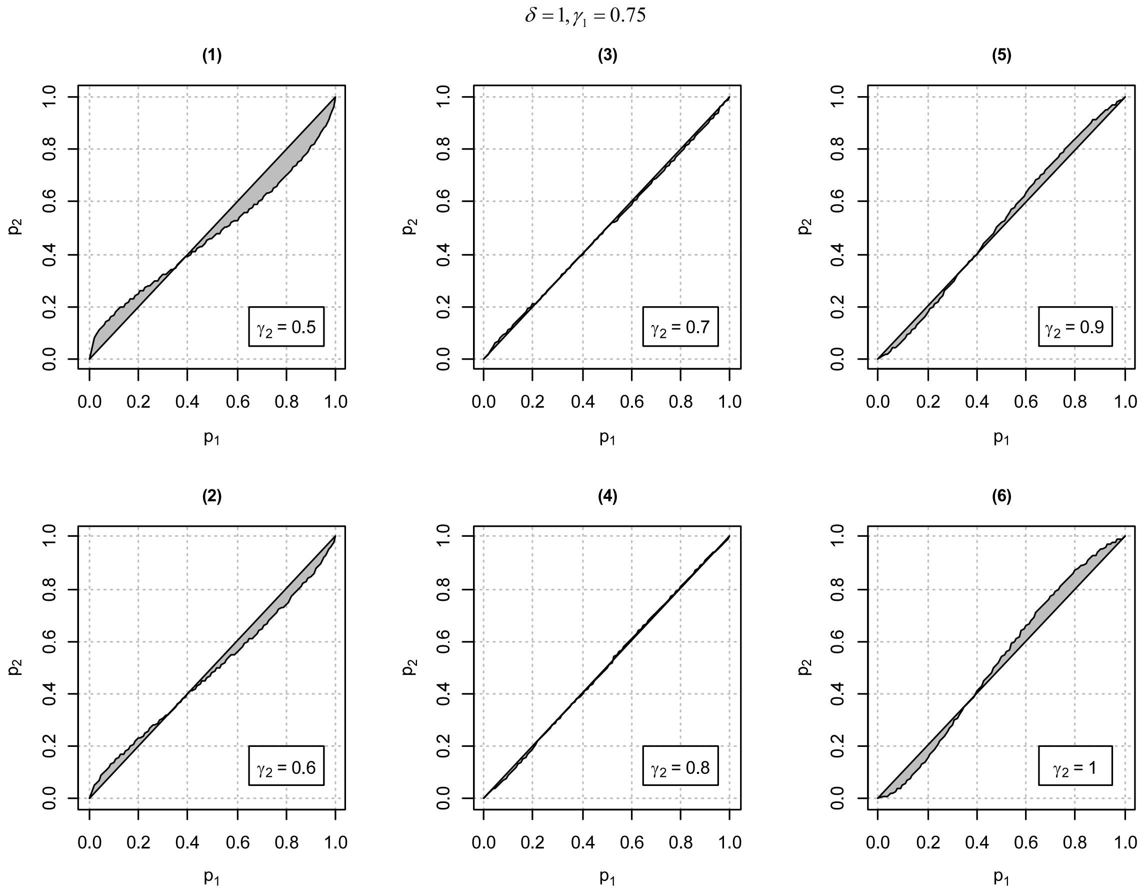

Fig. 3. Effects of curvature for the two-parameter probability weighting function $\delta = 1$, $\gamma_1 = 0.75$, $\gamma_2 = 0.5$ to 1 by 0.1 The shaded areas represent cases where applying the decision weights reverse preferences.

Table 1
Distribution of reversals for the case $|\beta_1 - \beta_2| = 0.10$, as a function of the parameters of the weighting functions (β_1, β_2).

(β_1, β_2) or (β_2, β_1)	Overall reversal rate	Relative reversal rate	Proportion of cases where $\lvert p_1 - p_2\rvert \leq$		
			0.05	0.10	0.15
0.50, 0.60	0.079	0.282	0.48	0.84	0.98
0.55, 0.65	0.057	0.204	0.61	0.97	1.00
0.60, 0.70	0.041	0.147	0.76	1.00	1.00
0.65, 0.75	0.031	0.110	0.90	1.00	1.00
0.70, 0.80	0.023	0.083	1.00	1.00	1.00
0.75, 0.85	0.018	0.063	1.00	1.00	1.00
0.80, 0.90	0.013	0.048	1.00	1.00	1.00
0.85, 0.95	0.010	0.035	1.00	1.00	1.00
0.90, 1.00	0.008	0.029	1.00	1.00	1.00
Overall		$N = 5440$	0.73	0.95	1.00

The answer depends on the interplay between the desired, or expected, resolution of the probability elicitation process, and the dissimilarity between the two weighting functions. In this section we offer two sensitivity analyses driven by these factors. Our main goal is to characterize circumstances where differential weighting can make a difference.

There is ample empirical evidence that when judges provide numerical probability judgments pertaining to graphical displays (e.g. Budescu, Weinberg, & Wallsten, 1988), or general knowledge of historical or geographical facts (e.g. Wallsten, Budescu, & Zwick, 1993) most subjective probabilities reported are multiples of 0.10 or 0.05. In other words, the resolution of probability judgments involves at most 21 levels and for many judges as few as 11 categories. Thus, events whose probabilities are within 0.05 of each other are, essentially, indistinguishable, but it makes sense to expect good differentiation between events whose probabilities are 0.10 apart. We consider reversals involving probabilities that are 0.15 (or more) apart as unacceptable. Unfortunately, we do not have enough good quality data to quantify the "typical" difference between the parameters of various weighting functions, so we examine a wide range of values.

We use the one-parameter weighting function (see Eq. (1)) in this illustration. We examined all pairs of probabilities, $p_1 p_2$, in the unit square [0, 1] and used all parameter pairs (rounded to the nearest 0.05) in the relevant range $\beta_1, \beta_2 \in [0.5, 1]$ to identify all the cases where differential weighting induced reversals.[2] Tables 1 and 2 summarize the results for cases where $|\beta_2 - \beta_1| = 0.1$ and 0.2, respectively. In each case we examine separately all relevant pairs of β_1 and β_2. The first data column in each table

[2] For the purpose of these analyses all the probabilities (p_1, p_2), the weights $(w(p_1), w(p_2))$, and the differences between them are rounded to 2 significant digits.

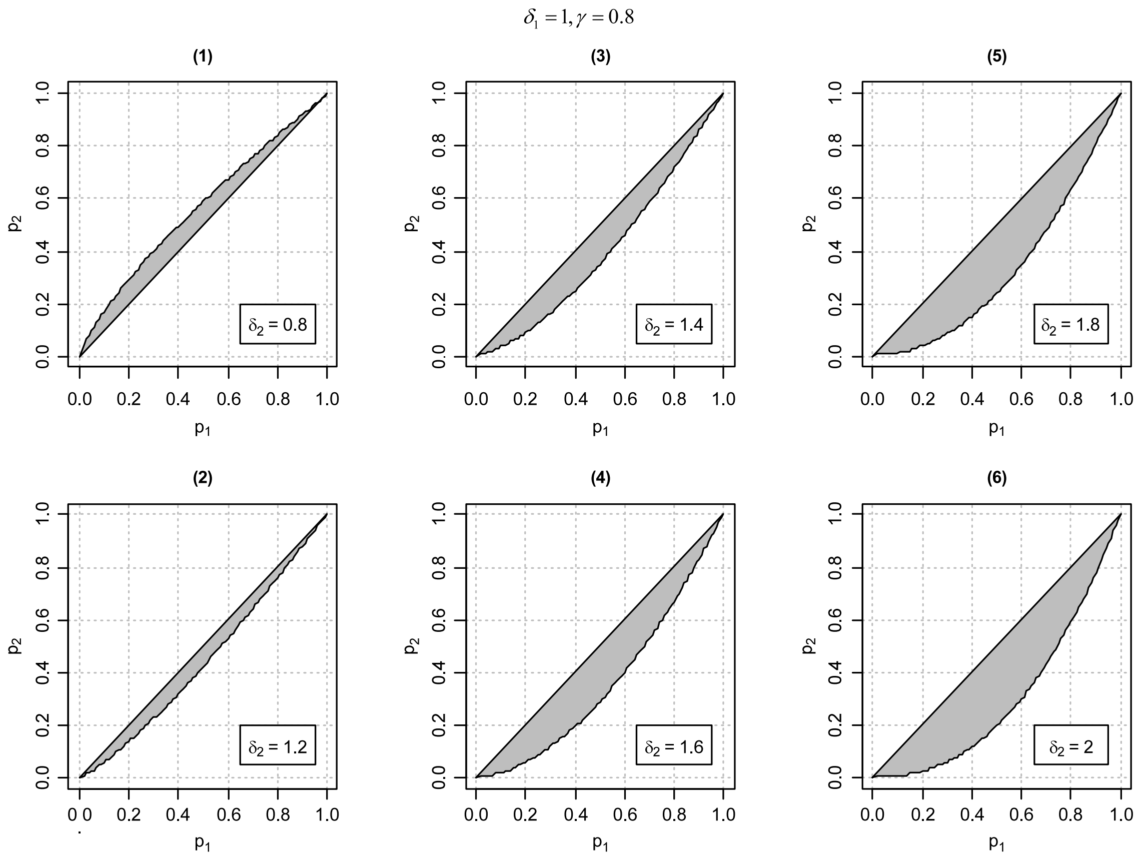

Fig. 4. Effects of elevation for the two-parameter probability weighting function $\gamma = 0.8$, $\delta_1 = 1$, $\delta_2 = 0.8$ and 1.2 to 2.0 by 0.2. The shaded areas represent cases where applying the decision weights reverse preferences.

Table 2
Distribution of Reversals for the case $|\beta_1 - \beta_2| = 0.20$, as a function of the parameters of the weighting functions (β_1, β_2).

(β_1, β_2) or (β_2, β_1)	Overall reversal rate	Relative reversal rate	Proportion of cases where $\lvert p_1 - p_2\rvert \le$		
			0.05	0.10	0.15
0.50, 0.70	0.123	0.270	0.33	0.60	0.81
0.55, 0.75	0.093	0.205	0.42	0.73	0.94
0.60, 0.80	0.071	0.157	0.52	0.86	1.00
0.65, 0.85	0.056	0.123	0.64	0.97	1.00
0.70, 0.90	0.045	0.099	0.75	1.00	1.00
0.75, 0.95	0.036	0.080	0.87	1.00	1.00
0.80, 1.00	0.030	0.065	0.96	1.00	1.00
Overall		$N = 8830$	0.54	0.81	0.94

presents the global rate of reversals for each β_1, β_2 pair. The second data column, labeled relative reversal rate, list what proportion of all the observed reversals (out of the 5400 reversal for $|\beta_2 - \beta_1| = 0.1$ and the 8830 cases for $|\beta_2 - \beta_1| = 0.2$) are associated with each combination of parameter values. The next three columns show the cumulative distributions of reversals, conditional of the given β_1 and β_2, at $|p_2 - p_1| \le 0.05$, 0.10 and 0.15, respectively.

Both tables show that both the global and the relative reversal rates decrease as the parameters (β_1, β_2) approach the $\beta = 1$ benchmark. In fact they are small, and negligible, as long as min $(\beta_1, \beta_2) > 0.75$. Table 1 shows that with a 10% tolerance for reversals due to weighting (which maps into a Kendall τ rank-order correlation of 0.80) (a) it is always possible to obtain resolution for cases where $|p_2 - p_1| = 0.15$, (b) almost always possible to achieve resolution when $|p_2 - p_1| = 0.10$ (the only exception being the case where $\beta_1 = 0.5$ and $\beta_2 = 0.6$), and (c) obtain very high resolution if $|p_2 - p_1| = 0.05$ if Max $(\beta_1, \beta_2) = 0.75$. Table 2 shows that with more dissimilar functions, where $|\beta_2 - \beta_1| = 0.2$, resolutions of $|p_2 - p_1| = 0.15, 0.10$ and 0.05 are obtainable, as long as Max $(\beta_2, \beta_1) = 0.70, 0.80,$ and 0.95, respectively.

An alternative sensitivity analysis, conditioned of the difference between the two probabilities, is presented in Figs. 6 and 7. We used, again, the one-parameter function (Eq. (1)) as an illustrative tool, and we calculated the rate of reversals for all pairs of probabilities, (p_1, p_2), in the unit square [0, 1] for a wider range of parameters and summarized them as a function of the difference between the two probabilities, $(p_1 - p_2)$. In each figure we fixed β_1 ($\beta_1 = 1$ in Fig. 6 and $\beta_1 = 0.75$ in Fig. 7) and examined all values of $\beta_2 \in [0.40, 1.60]$. Each plot shows 4 curves corresponding to $(p_1 - p_2) = -0.10, -0.05, 0.05,$ and 0.10, as a function of β_2.

Consider first Fig. 6 that pertains to the case where p_1 is not transformed ($\beta_1 = 1$). Note that even for probabilities that are highly similar there are large ranges of β_2 (in the vicinity of β_1) without any reversals. For example, the original ordering is preserved for $(p_1 - p_2) = -0.05$ as long as $0.86 \le \beta_2 \le 1.20$, for $(p_1 - p_2) = -0.10$ as long as $0.74 \le \beta_2 \le 1.46$, and for $(p_1 - p_2) = 0.05$ as long as $0.68 \le \beta_2$. If one is willing to tolerate a

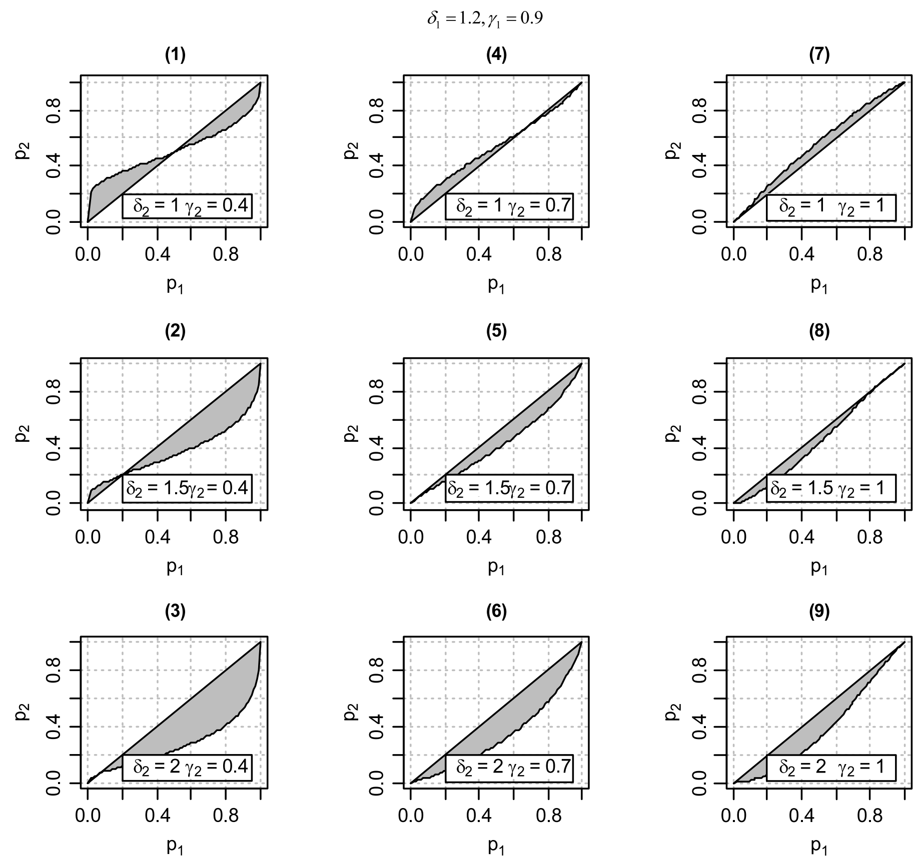

Fig. 5. Joint effects of elevation (rows) and curvature (columns) for the two-parameter probability weighting function with $\gamma_1 = 0.9$, $\delta_1 = 1.2$, $\gamma_2 = 0.4$ to 1.0 by 0.3, and $\delta_2 = 1.0$ to 2.0 by 0.5. The shaded areas represent cases where applying the decision weights reverse preferences.

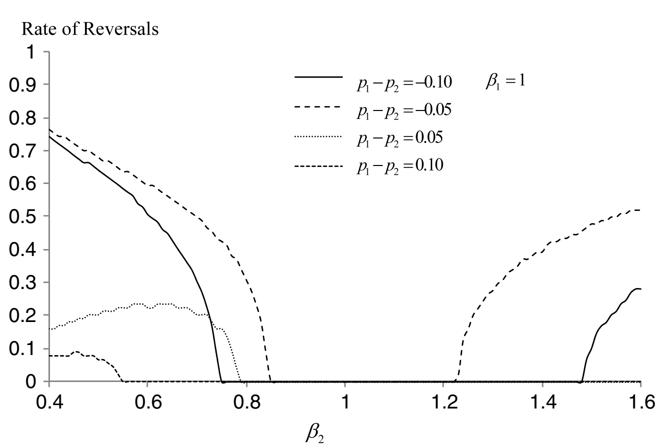

Fig. 6. Rates of reversals with the one-parameter probability weighting function when $\beta_1 = 1$ as a function of β_2, for selected values of $(p_1 - p_2)$.

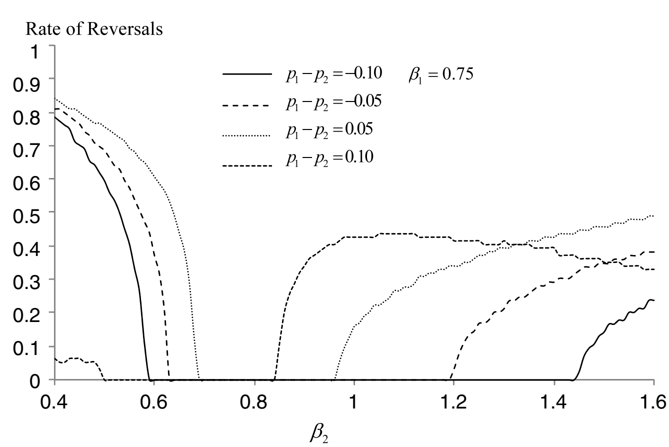

Fig. 7. Rates of reversals with the one-parameter probability weighting function when $\beta_1 = 0.75$ as a function of β_2, for selected values of $(p_1 - p_2)$.

relatively low rate of reversals (e.g. 10%) these ranges of invariance are slightly wider, but not by much.

Fig. 7 focuses on the case where $\beta_1 = 0.75$, so there is a considerable degree of distortion in p_1. Although the rate of change in the rate of reversals is steeper and the tolerance ranges are more asymmetric (compared to Fig. 6), the basic pattern is similar. For example, the original ordering is preserved for $(p_1 - p_2) = -0.05$ as long as $0.60 \leq \beta_2 \leq 0.94$, and for $(p_1 - p_2) = 0.05$ when $0.54 \leq \beta_2 \leq 0.82$.

Final remarks

The standard methods for eliciting subjective probabilities in decision analysis (see Spetzler & Staël von Holstein, 1975; Wallsten & Budescu, 1983) predate the development of the probability weighting function. When finding the point of indifference between the setting of the generic probability wheel and a value of the variable of interest (temperature, currency exchange rate, etc.) we infer that the probability on the wheel represents the cumulative probability of the value of the variable. This analysis implicitly assumes that the two probabilities – the one on the wheel and the one from the distribution of the variable of interest – affect choices identically.

The purpose of this paper was to provide a general methodology for analyzing the potential magnitude of the effects induced by the differential weighting transformations.

Using two popular functional forms of the weighting function as illustrative examples, we examined the behavior over all possible values of probabilities and over reasonable subsets of their parameter space, as implied by the two recent review papers (Booij et al., 2010; Stott, 2006). The general method developed in this paper can be easily applied with other weighting functions (e.g. Gonzalez & Wu, 1999; Rieger & Wang, 2006) and on other regions of their respective parameters' space.

Our results suggest that in most cases the choices based on differential weighting (within the ranges considered) are consistent with choices assuming identical weighting functions. It follows that the elicited subjective probability functions would be quite insensitive to the shape and parameter of the weighting functions. We were also able to show regions where inconsistencies may occur— mainly in a narrow region around the diagonal of each plot, corresponding to probabilities that are quite similar. Our sensitivity analysis confirms that the vast majority of reversals involve extreme non-linear weighting of similar probabilities.

We asked whether probability weighting functions affect probability elicitation. At the end of the day, the answer depends on one's expected resolution, tolerance for errors due to reversals, and the exact nature of the actual weighting transformations. In our view it is sensible to rely on the judges' stated preferences and it is possible to estimate their subjective probabilities with reasonable resolution (e.g., $|p_2 - p_1| \leq 0.10$) and tolerable levels of error ($\leq$10%) without worrying about the effects of the weighting functions, as long as these transformations are not too extreme. However, we recognize that different people under various circumstances may wish to apply different criteria and may not accept this interpretation. The general method proposed in this paper can be applied with different evaluation criteria (e.g., different expectations about the judgments' resolution and other (stricter or looser) levels of tolerance for error).

This analysis of the effects of probability weighting during the encoding process highlights the need for more empirical research designed to examine and quantify the degree to which different sources and types of uncertainty, as well as various devices, settings, elicitation procedures, instructions, incentives, etc. can affect the shape (parameters) of the weighting functions over the relevant range of values and probabilities. The work by Gonzalez and Wu (1999), Hau et al. (2008) and Wakker (2004) provides promising starting points for such future investigations.

Acknowledgments

This work was supported by the National Science Foundation under Award Number SES 06-20008.

We wish to thank Aurelien Baillon for identifying this problem, and Craig Fox, Richard Gonzalez, Peter Wakker, Thomas Wallsten, the Associate Editor and the anonymous referees for useful comments on an earlier version.

References

Abbas, A. E., Budescu, D. V., Yu, H., & Haggerty, R. (2008). A comparison of two probability encoding methods: fixed probability vs. fixed variable values. *Decision Analysis*, *5*, 190–202.

Abdellaoui, M. (2000). Parameter free elicitation of utility and probability weighting functions. *Management Science*, *46*, 1497–1512.

Abdellaoui, M., Baillon, A., & Wakker, P. P. (2007). Combining Bayesian beliefs and willingness to bet to analyze attitudes towards uncertainty. Rotterdam, the Netherlands: Econometric Institute, Erasmus University. http://people.few.eur.nl/wakker/pdf/sources.pdf.

Bleichrodt, H., Pinto, J. L., & Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science*, *47*, 1498–1514.

Booij, A. S., van Praag, B. M. S., & van de Kuilen, G. (2010). A parametric analysis of risk attitudes for the general population under prospect theory. *Theory and Decision*, *68*, 115–148.

Budescu, D. V., Kuhn, K. M., Kramer, K. M., & Johnson, T. (2002). Modeling certainty equivalents for imprecise gambles. *Organizational Behavior and Human Decision Processes*, *88*, 748–768.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 281–294.

Fox, C., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science*, *44*, 879–895.

French, S. (1988). *Decision theory: an introduction to the mathematics of rationality*. England: Ellis Horwood: Chichester.

Gonzalez, R., & Wu, G. (1999). On the form of the probability weighting function. *Cognitive Psychology*, *38*, 129–166.

Hau, R., Pleskac, T. J., Kieffer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: the role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*, 493–518.

Kadane, J. B., & Winkler, R. L. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association*, *83*, 357–363.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kilka, M., & Weber, M. (2001). What determines the shape of the probability weighting function under uncertainty. *Management Science*, *47*, 1712–1726.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain judgments: eliciting experts' probabilities*. West Sussex, England: Wiley.

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*, 497–527.

Rieger, M. O., & Wang, M. (2006). Cumulative prospect theory and the St. Petersburg paradox. *Economic Theory*, *28*, 665–679.

Sarin, R. K., & Wakker, P. P. (1997). A single-stage approach to Anscombe and Aumann's expected utility. *Review of Economic Studies*, *64*, 399–409.

Savage, J. L. (1954). *The foundations of statistics*. New York: Wiley.

Spetzler, C. S., & Staël von Holstein, C. A. (1975). Probability encoding in decision analysis. *Management Science*, *22*, 340–358.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, *32*, 101–130.

Tversky, A., & Fox, C. (1995). Weighing risk and uncertainty. *Psychological Review*, *102*, 269–283.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *26*, 297–323.

Ungenmach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcoems are experiences (rarely)? *Psychological Science*, *20*, 473–479.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

Wakker, P. (2004). On the composition of risk preference and belief. *Psychological Review*, *11*, 236–241.

Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, *29*, 151–173.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, *39*, 176–190.

Wallsten, T. S., Forsyth, B., & Budescu, D. V. (1983). Stability and coherence of health experts upper and lower subjective probabilities about dose-response curves. *Organizational Behavior and Human Decision Processes*, *31*, 277–302.

Whitfield, R. G., & Wallsten, T. S. (1989). A risk assessment for selected lead-induced health effects: an example of a general methodology. *Risk Analysis*, *9*, 197–207.

Wu, S. W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences*, *106*, 6088–6093.