Theory and Decision (2006) 60: 283–313
DOI 10.1007/s11238-005-4590-0

HELGA FEHR-DUDA, MANUELE DE GENNARO
and RENATE SCHUBERT

GENDER, FINANCIAL RISK, AND PROBABILITY WEIGHTS

ABSTRACT. Women are commonly stereotyped as more risk averse than men in financial decision making. In this paper we examine whether this stereotype reflects gender differences in actual risk-taking behavior by means of a laboratory experiment with monetary incentives. Gender differences in risk taking may be due to differences in valuations of outcomes or in probability weights. The results of our experiment indicate that value functions do not differ significantly between men and women. Men and women differ in their probability weighting schemes, however. In general, women tend to be less sensitive to probability changes. They also tend to underestimate large probabilities of gains more strongly than do men. This effect is particularly pronounced when the decisions are framed in investment terms. As a result, women appear to be more risk averse than men in specific circumstances.

KEY WORDS: gender differences, risk aversion, financial decision making, prospect theory, probability weighting function

1. INTRODUCTION

Women are often assumed to be more risk averse than men, and numerous questionnaire studies have confirmed this common stereotype. Psychologists and sociologists find strong gender-specific differences in responses to non-financial risks which are particularly pronounced when it comes to physical or life-threatening risks (Byrnes et al., 1999). Surprisingly, little work has been done on gender-specific differences in financial decision making. Only a few laboratory experiments using real monetary incentives and some studies based on field data have investigated whether women are more risk averse than men when financial risks are concerned. The studies based on field data conclude that women are relatively more risk averse than men, whereas the laboratory experiments render inconclusive results (see the survey by Eckel and Grossman, 2005).

The experimental studies yield a diversity of findings: When gains are at stake, two outcomes prevail: either the women exhibit relatively higher risk aversion or there is no clear difference between the sexes. "Everything seems possible" when losses may be incurred, depending on the study considered: either the female, neither sex, or even the male sex is relatively more risk averse. Since the studies differ in decision context, levels of probabilities, potential payoffs, variance of payoffs, choice task, and other dimensions, one might argue that they lack comparability and, therefore, it is not surprising that results do not coincide. However, even in studies with quite similar experimental designs we find divergent results. How can this puzzle be resolved?

Subjects' risk-taking behavior is driven by their valuations of outcomes and assessments of probability information. We surmise that the diversity in the experimental findings may be caused by the way men and women weight probabilities in their decisions. If there are systematic gender differences in the way probabilities are weighted, laboratory results may depend on the mix of lotteries used in the experiment. Suppose, for the sake of argument, that women place a much higher weight on small probabilities than do men, and that their behavior does not differ otherwise. If the experimental design comprises winning lotteries with mostly medium and large probabilities (of the best outcome) the researcher will not find any gender differences. If, on the other hand, the design relies heavily on gambles with small probabilities, she will most likely find women to be the relatively more risk-seeking gender. We conjecture, therefore, that some of the contradictory results of the experiments done so far may be caused by the differing ranges of probabilities used in the experiments.

To our knowledge, hardly any work has been done on the question of whether women assess probabilities differently than do men. In order to explore the issue of gender-specific probability weighting, we conducted a laboratory experiment based on a wide range of probabilities. To be able to infer gender-specific average behavior, we recruited a large number of male and female subjects, allowing us to generate data on certainty equivalents for winning and losing gambles in an abstract and in a contextual environment. In the abstract environment the decision problems were framed as abstract gambles, whereas the very same gambles were presented as investment and insurance decisions in the contextual environment. In both environments decisions entailed gains and losses. The elicited certainty equivalents were used to estimate the parameters of Prospect Theory, enabling us to check value and probability weighting functions for systematic gender differences.

On average, female probability weighting functions differ from male ones in a specific way. Women's curves are more curved irrespective of the treatment condition. Moreover, in the domain of gains, women tend to underestimate larger probabilities much more strongly than do men. This gender difference is particularly pronounced when the decisions are framed in investment terms rather than in abstract terms. Since the majority of subjects values outcomes linearly, risk taking behavior is mostly reflected by probability weighting. And indeed, we find women to be more risk averse than men when facing investment choices.

The paper is organized as follows. In Section 2 we describe the design and procedure of our experiment. In Section 3 we briefly introduce the decision model and specify the functional forms for the value and probability weighting functions. We present the results of our experiment and their implications for gender-specific risk taking behavior in Section 4. Finally, our findings are discussed in Section 5. Section 6 concludes the paper.

2. THE EXPERIMENT

We designed a computerized experiment to elicit subjects' certainty equivalents which serve as base for estimating value and probability weighting functions. The experiment is characterized by a 2 (environments) $\times 2$ (domains) design. The environmental treatment conditions differ in the way the lotteries are framed: subjects are confronted with abstract gamble choices in the abstract environment, while the same lotteries are framed as investment and insurance decisions in the contextual environment. Each subject participates in only one of the environments.[1]

The second dimension of the design concerns the domains of gains and losses. In both environmental treatments, each subject has to consider the same 50 two-outcome lotteries. Of these, 25 of the lotteries offer potential gains and the remaining 25 lotteries are framed as losses. Each one of the losing lotteries is equivalent to a corresponding winning lottery as it is assigned a lottery-specific initial endowment such that total payoff is equal to the corresponding winning lottery's. These initial endowments compensate subjects for potential losses they might incur.

The experimental design comprises lotteries with probabilities of 5, 10, 25, 50, 75, 90, and 95%. Outcomes range from zero to 150 Swiss Francs.[2] The lotteries are summarized in the first three columns of Table I panel a, b. The expected payoff per participant amounts to 31 Swiss Francs.

The experiment is programmed using "Z-Tree", a special software package for conducting economic experiments (Fischbacher, 1999). The 50 lotteries appear in random order for each subject. The participants fill out a separate decision form for each one of these lotteries. The computer screen displays the respective lottery (option A) and a list of 20 guaranteed amounts (options B; see Figure 1). These guaranteed amounts are arranged in algebraically descending order,[3] starting with the larger gamble outcome and descending in equal steps towards the smaller gamble outcome. Going down the list, on each line of the decision sheet the subjects have to decide whether they prefer the (fixed) lottery (option A) or the respective guaranteed payment (option B) by clicking on the box next to the preferred option. If subjects change from preferring guaranteed payments to preferring the lottery and

TABLE I

(a) Lotteries—Gain Domain: (b) Lotteries—Loss Domain, Median Relative Risk Premiums (RRP) by Sex and Environment. Shaded Areas Indicate Risk Seeking

Lottery Design			Median RRP		Median RRP	
Probability	Outcome x_1	Outcome x_2	Abstract Environment		Context Environment	
p of x_1 (%)	CHF	CHF	Women	Men	Women	Men
(a) Lotteries—Gain Domain						
5	20	0	−3.50	−3.50	−2.50	−2.50
5	40	10	−0.46	−0.33	−0.39	−0.33
5	50	20	−0.17	−0.17	−0.17	−0.21
5	150	50	−0.23	−0.14	−0.23	−0.23
10	10	0	−2.25	−1.75	−1.75	−1.25
10	20	10	−0.16	−0.16	−0.16	−0.14
10	50	0	−0.75	−1.25	−0.75	−0.75
25	20	0	−0.50	−0.50	−0.10	−0.10
25	40	10	−0.13	−0.13	−0.13	−0.04
25	50	20	−0.08	−0.08	−0.08	−0.08
50	10	0	0.05	−0.05	0.05	0.05
50	20	10	0.02	0.02	0.02	0.00

TABLE I
Continued

Lottery Design			Median RRP		Median RRP	
Probability	Outcome	Outcome	Abstract Environment		Context Environment	
p of x_1(%)	x_1 CHF	x_2 CHF	Women	Men	Women	Men
(a) Lotteries—Gain Domain						
50	40	10	0.03	0.03	0.15	0.03
50	50	0	0.05	0.05	0.10	0.05
50	50	20	0.11	0.04	0.09	0.02
50	150	0	0.25	0.25	0.25	0.25
75	20	0	0.17	0.03	0.37	0.10
75	40	10	0.21	0.09	0.14	0.09
75	50	20	0.16	0.07	0.12	0.07
90	10	0	0.08	0.08	0.19	0.08
90	20	10	0.12	0.07	0.14	0.07
90	50	0	0.19	0.14	0.31	0.14
95	20	0	0.18	0.08	0.24	0.14
95	40	10	0.14	0.10	0.18	0.08
95	50	20	0.11	0.11	0.17	0.11

(b) Lotteries—Loss Domain						
5	20	0	3.50	1.50	3.50	2.50
5	40	10	0.33	0.26	0.59	0.33
5	50	20	0.24	0.17	0.24	0.17
10	10	0	2.25	1.25	2.25	0.75
10	20	10	0.23	0.16	0.16	0.16
10	50	0	2.25	1.25	2.00	1.25
25	20	0	0.50	0.30	0.80	0.30
25	40	10	0.21	0.17	0.39	0.13
25	50	20	0.25	0.08	0.14	0.14
50	10	0	0.05	−0.05	0.05	0.05
50	20	10	0.23	0.16	0.16	0.16
50	40	10	0.03	0.03	0.09	0.03
50	50	0	0.15	0.05	0.10	0.05
50	50	20	0.02	0.00	0.00	0.02
50	150	0	0.35	0.20	0.45	0.05
75	20	0	−0.17	−0.10	−0.07	−0.03
75	40	10	−0.07	−0.07	−0.07	−0.07
75	50	20	−0.05	−0.05	−0.02	−0.05

TABLE I
Continued

Lottery Design			Median RRP		Median RRP	
Probability	Outcome	Outcome	Abstract Environment		Context Environment	
p of x_1(%)	x_1 CHF	x_2 CHF	Women	Men	Women	Men
(b) Lotteries—Loss Domain						
90	10	0	−0.14	−0.14	−0.07	−0.14
90	20	10	−0.09	−0.09	−0.09	−0.07
90	50	0	−0.14	−0.11	−0.08	−0.11
95	20	0	−0.13	−0.08	−0.13	−0.13
95	40	10	−0.09	−0.09	−0.09	−0.09
95	50	20	−0.08	−0.08	−0.06	−0.06
95	150	50	−0.09	−0.09	−0.17	−0.05

Decision situation: 22						
	Option A	Your Choice:				Option B Guaranteed payoff amounting to:
1		A		o	B	20
2		A		o	B	19
3		A		o	B	18
4		A		o	B	17
5		A		o	B	16
6		A		o	B	15
7	Profit of CHF 20 with	A		o	B	14
8		A	o		B	13
9	probability 25%	A	o		B	12
10		A	o		B	11
11	and profit of CHF 0 with	A	o		B	10
12		A	o		B	9
13	probability 75%	A	o		B	8
14		A	o		B	7
15		A	o		B	6
16		A	o		B	5
17		A	o		B	4
18		A	o		B	3
19		A	o		B	2
20		A	o		B	1
OK						

Figure 1. Design of computer screen (gain domain). For each of the 20 lines on the screen the subject has to decide whether she prefers option A, the lottery, or option B, the guaranteed payoff in the respective line. Preference is indicated by marking the box next to A or B in each line. Suppose the subject chooses the guaranteed option for payoffs from CHF 20 to 14 and then switches to the lottery. In this case the certainty equivalent amounts to CHF 13.50. Participants are informed in the experimental instructions that option B is taken to be their choice throughout if they do not make any entries.

switch back again, a message appears on the computer screen informing them that they have switched between A and B more than once. In this case, subjects can either reconsider their choices or stick to their previous entries. A lottery's certainty equivalent is determined as the arithmetic mean of the minimum guaranteed payment which is preferred to the lottery and the following smaller guaranteed payment on the list (see Figure 1).

At the end of the experiment, subjects are asked to fill out a questionnaire eliciting information on a number of socioeconomic variables, such as gender, age, and income. When the subjects have completed the questionnaire, one of their lottery choices is randomly selected for payment by rolling dice. Total payment includes a show-up fee of 10 Swiss Francs. Average actual payments amounted to 40.61 Swiss Francs. The experimental sessions lasted about one hour in total. Financial incentives are salient considering the local student assistant's hourly wage of 22.50 Swiss Francs.

The experiments took place at the computer lab of the Institute of Empirical Economic Research, University of Zurich, in June and August 2003. We recruited 204 students of various faculties of the Swiss Federal Institute of Technology and the University of Zurich. We conducted five sessions with abstract choices and five sessions with contextual choices. Since the calculation of certainty equivalents requires subjects to switch from option B to option A (or vice versa) just once, we can only use decisions that meet this condition. On the other hand, since we estimate the parameters of the value and the probability weighting functions for each individual, we should not exclude too many decisions from a person's data set. Therefore, we use the following rule: if a subject exhibits inconsistent choices, i.e. if she switches back and forth between guaranteed and risky outcomes for more than two lotteries, all her decisions are excluded from the data set. Subjects' data with one or two inconsistent decision sheets are still considered; the erroneous decision sheets are deleted. This procedure leaves 181 subjects' data for analysis. An overview of the number of subjects according to sex and environment is presented in Table II.

TABLE II

Number of Participants by Sex and Environment

	Abstract Environment	Context Environment	Pooled Environments
Female	37	40	77
Male	54	50	104
Both sexes	91	90	181

3. THE MODEL

Observed risk-taking behavior depends on two factors: how outcomes are valued and how probabilities are weighted. Since we surmise that men and women assess probabilities differently, we need to specify a model that allows us to estimate individual value and probability weighting functions to test our hypothesis. For this purpose we invoke the concepts of Prospect Theory (Tversky and Kahneman, 1992).

For a two-outcome lottery

$$L[x_1, p; x_2], \quad |x_1| > |x_2|, \quad x_1, x_2 \geqslant 0 \text{ or } x_1, x_2 \leqslant 0, \tag{1}$$

where x_i $(i = 1, 2)$ denotes outcomes, and p denotes the probability of x_1 occurring, the certainty equivalent CE is defined by the equation

$$v(\mathrm{CE}) = \pi_1 v(x_1) + \pi_2 v(x_2). \tag{2}$$

Decision weights are denoted by π_i $(i = 1, 2)$, v is the value function defined on the monetary outcome x. Both value function and decision weights are assumed to depend on the sign of the outcomes. The decision weights depend on the subject's domain-specific probability weighting function $w(p)$. In our two-outcome case, the decision weights can be represented as

$$\pi_1 = w(p), \tag{3}$$

$$\pi_2 = 1 - w(p). \tag{4}$$

We have to choose functional forms for v and w to make the model operational. One of the functional forms most frequently used for the probability weighting function w is a one-parameter version introduced by Quiggin (1982) as well as Tversky and Kahneman (1992). Lattimore et al. (1992) propose the following two-parameter functional:

$$w(p) = \delta p^{\gamma} / [\delta p^{\gamma} + (1 - p)^{\gamma}]; \quad \delta \geqslant 0, \gamma \geqslant 0, \tag{5}$$

which we tested against the Quiggin version on the basis of the Akaike information criterion. It turns out that the Akaike

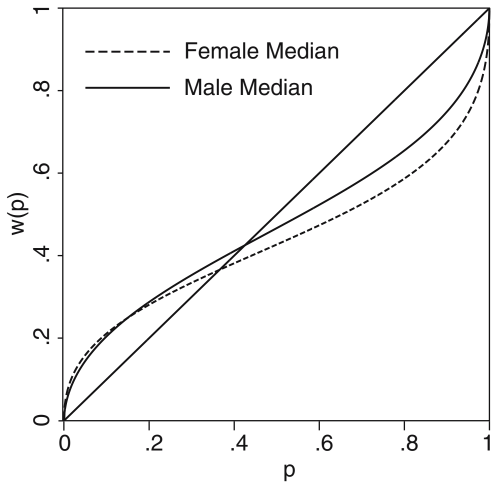

Figure 2. Gender-specific median probability weighting functions: Abstract gains.

criterion favors the Lattimore functional for the majority of our participants. This result holds across both environments and domains. Therefore, we present parameter estimates for the Lattimore version only. Note that the parameters γ and δ are domain specific, i.e. they may take on different values for winning and losing gambles.

Elevation and slope of this functional cannot be varied totally independently from each other since $w(p)$ is fixed "at both ends" ($w(0)=0, w(1)=1$; for typical specimens of probability weighting functions refer to Figure 2). However, parameter δ largely governs the elevation of the curve, while γ largely determines its slope: the smaller the value of γ, the more curved (flatter in the range of medium probabilities and steeper near the ends) the $w(p)$ curve; and the greater the value of δ, the more elevated the curve, ceteris paribus. Linear weighting is characterized by $\gamma=\delta=1$.

The parameters have a neat psychological interpretation (Gonzalez and Wu, 1999): γ reflects a subject's responsiveness to changes in probability: the smaller γ, the less responsive. The interpretation of δ depends on the domain considered. For winning gambles δ can be viewed as a gamble's attractiveness. The more elevated the probability weighting curve, the greater are the weights placed on the probabilities. In this sense, a person finds a gamble more attractive than does another person if she puts more weight on the (best outcome's) probability than does the other person. For a given γ, the elevation parameter also determines where the curve intersects the diagonal, i.e. the linear probability weighting line in $(p, w(p))$ space. Therefore, for the typical inverted S-shape of the curve, the higher the point of intersection with the diagonal, the larger the range of probabilities where the subject displays optimism $(w(p) > p)$. It works the other way round for losing gambles: the more elevated the curve, the less attractive a gamble is judged to be, and the more pessimistic the person views the probabilities.

After discussing the weighting function we now turn to the second component of the model, the value function v. We assume the valuation of outcomes to be represented by the following power functional (Tversky and Kahneman, 1992):

$$v(x) = \begin{cases} x^{\alpha} & x \geqslant 0 \\ -\lambda(-x)^{\beta} & x < 0 \end{cases} \tag{6}$$

The parameters α and β of the power functional are assumed to be greater than zero. Depending on their size, domain-specific value functions are either concave $(\alpha < 1, \beta > 1)$, convex $(\alpha > 1, \beta < 1)$ or linear $(\alpha = \beta = 1)$. Since the experimental design includes lotteries with two non-zero outcomes, we are able to estimate α and β. The parameter for loss aversion λ is not identifiable, since we only consider single-domain gambles, and no mixed gambles, i.e. gambles with both positive and negative outcomes.

Based on the aforementioned assumptions, we estimate the parameters (α, γ, δ) for the gain domain and (β, γ, δ) for the loss domain for each single individual using the maximum likelihood method.[4] Moreover, we also estimate the parameters for the female and male median person. The latter estimates are based on the respective median certainty equivalents for each lottery. For the model to be economically meaningful, the value function parameters α and β need to differ significantly from zero. This does not necessarily apply to γ and δ, however. When $\gamma = 0$, the subject exhibits a flat probability weighting curve, i.e. probabilities are totally neglected in the decision. When $\delta = 0$, the decision model reduces to x_2 being the only relevant input. Judged by the significance of α and β, our model works well for 93% of the subjects. If we require all six parameters to be significantly different from zero, 80% of our participants meet this condition.

Median standard errors of the estimates for the individual parameters are in the range of 0.07 for γ across all the treatment conditions. The standard errors of the value function parameters α and β as well as the respective elevation parameters δ lie, on average, between 0.20 and 0.26. They tend to be higher in the loss domain than in the gain domain. The highest median standard errors are reported for females in the insurance treatment (0.35 for both β and δ).

4. RESULTS

The data elicited by our experiment are analyzed at two levels. First, we report subjects' observed risk-taking behavior. Second, our main findings on the estimated value and probability weighting functions are presented.

Risk-taking behavior is measured by relative risk premiums. The relative risk premium RRP is defined as

$$\mathrm{RRP} = (\text{expected payoff} - \text{certainty equivalent}) / |\text{expected payoff}|. \tag{7}$$

$\mathrm{RRP} > 0$ indicates risk aversion, $\mathrm{RRP} = 0$ risk neutrality, and $\mathrm{RRP} < 0$ risk seeking. A descriptive overview of female and male risk-taking behavior is presented in Table I panel a, b (see above). These tables comprise information on the lottery design and relative risk premiums by sex and environmental condition.

Table I, panel a, b reveal a fourfold pattern of risk attitudes in our data: men and women in both environments are risk averse for medium and large probabilities of a gain, and risk seeking for small probabilities. They are also risk seeking for large probabilities of a loss, and risk averse for small and medium probabilities. This pattern of behavior has been found in many empirical investigations and fostered the idea that subjects weight probabilities nonlinearly. We also detect a regularity in gender differences: women are either equally risk averse or even more risk averse than men for the range of lotteries for which both sexes tend to be risk averse. The opposite holds for the lotteries for which subjects exhibit predominantly risk seeking behavior: women tend, with a few exceptions, to be either as risk seeking as men or even more so. These observations imply that women seem to react more extremely to risk in the sense that the female fourfold pattern seems more pronounced. Regarding statistically significant differences in the gender-specific risk premiums, we find women to be more risk averse particularly in the investment environment where they exhibit greater risk aversion than men in 56% of the decisions with large probabilities of a gain.[5]

We now turn to our estimates of probability weights and value function parameters. As far as the individual parameter estimates are concerned, we find that subjects' probability weighting schemes differ vastly. Some of our participants weight probabilities linearly, but the majority under- and overweighs them to some extent, with some people being optimistic or pessimistic over practically the whole range of probabilities. We find subjects with flat weighting functions, i.e. with γ close to zero, as well. Median behavior, however, is characterized by the typical inverted S-shape of the probability weighting function (for both sexes and all treatment conditions), consistent with earlier estimates (Gonzalez and Wu, 1999; Tversky and Kahneman, 1992, among others).

Individual variety in probability weights contrasts starkly with the homogeneity in value function parameters: the estimated individual values for α and β do not differ significantly from one for 76% of the subjects (70% of the women and 81% of the men; at the 5% level of significance), i.e. the value functions are practically linear for the vast majority of subjects. The linearity of the value functions is not surprising: the range of outcomes typically used in experiments with real financial incentives is rather limited. Therefore, decreasing marginal utility may not play a role.

We now address the focal point of our paper: are female probability weights and outcome valuations different from male ones?

4.1. *Result 1*

4.1.1. *Women are significantly less responsive to changes in probability in all four treatments, i.e. women's probability weighting functions are, on average, more curved than are men's*

Support. Two groups of findings back gender differences in responsiveness to probabilities, expressed by the parameter γ: the distributions of the individual parameter estimates and the estimates of the median parameters. A bootstrapped[6] Mann–Whitney test assesses differences in the distributions of the individual parameter estimates. The female sensitivity to changes in probability tends to be significantly smaller than the male one (at the 1%-level of significance) across both domains and environments. The same characteristic can be seen in the estimates for the median probability weighting functions. Table III summarizes the estimated median parameters for all treatment conditions. Focusing on the estimates for γ, one can discern the following regularities: the estimate for the female responsiveness to probabilities is consistently smaller (by 0.09 or more) across all treatment conditions. Moreover, the parameter estimates for both the females and the males are strikingly stable across treatments. The gender differences in responsiveness are also highly significant as Table IV reveals. The numbers in Table IV are calculated as female coefficients relative to the male coefficients. These significant differences in γ across all treatment conditions imply that the median female curve is significantly more curved.

4.2. *Result 2*

4.2.1. *Women are on average more pessimistic than men in the gain domain*

Support. Table III reveals that the median parameter estimates for δ, the elevation of the probability weighting curve, differ between the sexes. This difference appears to be more pronounced in the gain domain. Whether these differences are statistically significant can be inferred from Table IV. Whereas we do not find any gender effects for losing gambles, the parameter estimates for the gain domain differ significantly between the sexes. The women's average probability weighting curve is less elevated than the men's for gains in both environments, signifying a higher degree of female pessimism.

Significant differences in single parameters, however, do not necessarily imply significant differences in the probability weights since the shape of the probability weighting curve simultaneously depends on both γ and δ. Therefore, we need to examine the combined effect of these parameters on the probability weights.

4.3. *Result 3*

4.3.1. *In general, women's probability weighting functions are different from men's. This difference, however, is significant only for investment decisions where the gender differences are greatest both in responsiveness (γ) and attractiveness (δ)*

Support. The main body of evidence for Result 3 is depicted in the graphs of the median probability weighting functions. The respective graph of the female median weighting function is plotted against the male one for each of the treatment conditions in Figures 2–5. The female function is more curved than the male function in all four figures and, in the gain domain, it tends to be more depressed. The largest difference between the gender-specific probability weighting curves can be observed for contextual gains, i.e. investment decisions (Figure 3).

TABLE III
Estimates of Median Parameters and Standard Errors

Parameter Estimates	α, β		γ		δ	
	Women	Men	Women	Men	Women	Men
	Abstract Environment					
Gains	1.16 (0.105)	1.13 (0.096)	0.47 (0.028)	0.56 (0.028)	0.74 (0.077)	0.88 (0.081)
Losses	1.09 (0.089)	1.03 (0.064)	0.47 (0.031)	0.57 (0.021)	1.10 (0.089)	1.00 (0.064)
	Context Environment					
Gains	0.94 (0.066)	0.91 (0.065)	0.41 (0.021)	0.56 (0.025)	0.79 (0.059)	1.00 (0.075)
Losses	1.20 (0.131)	0.94 (0.044)	0.47 (0.031)	0.57 (0.016)	1.06 (0.130)	1.14 (0.044)

All parameters estimates of the Lattimore functional are significant at least at the 5%-level (t-test). The F-test for the respective vector of parameters yields significance at 1%.

TABLE IV

Relative Gender Differences in Median Parameters* and Values of t-test

Parameter Estimates	γ		δ	
Environment	Abstract	Context	Abstract	Context
Gains	−17% (4.16)	−27% (6.94)	−16% (2.10)	−21% (3.28)
Losses	−18% (5.31)	−18% (4.68)	10% (1.49)	−7% (1.01)

*(Female coefficient minus male coefficient)/male coefficient.

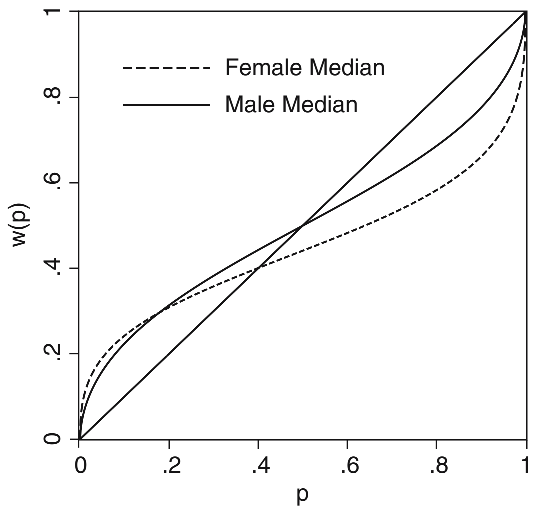

Figure 3. Gender-specific median probability weighting functions: Contextual gains (investment).

As argued above, the combined effect of the parameters drives the shape of the probability weighting function. Table IV reveals that the gender difference in the estimates for γ (indicating responsiveness) is highest for contextual gains ($\Delta = -27\%$), whereas the differences in the other treatment conditions are approximately of the same order of magnitude. There are also significant effects for δ (indicating attractiveness) in the gain domain, which, again, are much more pronounced for contextual decisions ($\Delta = -21\%$). Thus, the total

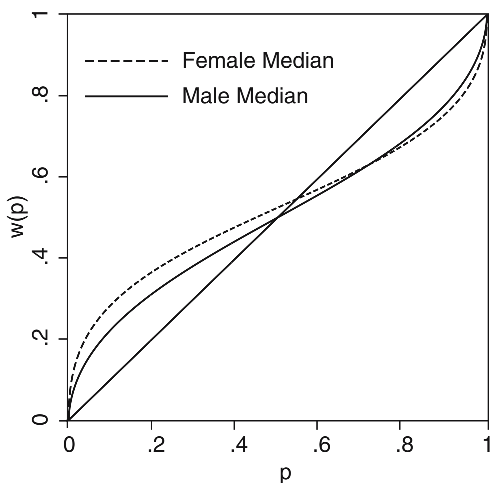

Figure 4. Gender-specific median probability weighting functions: Abstract losses.

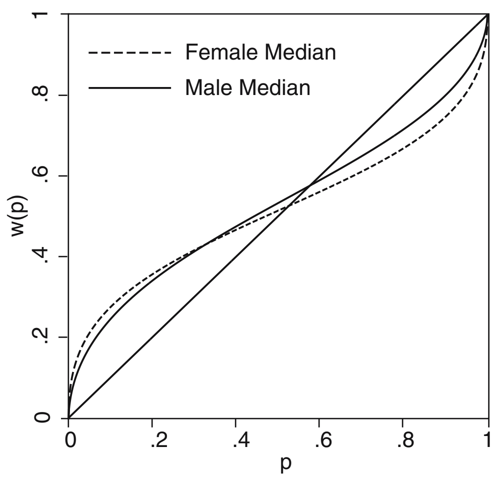

Figure 5. Gender-specific median probability weighting functions: Contextual losses (Insurance).

gender difference is largest for winning gambles in the context environment, i.e. for investment decisions.

Finally, we turn to the statistical evidence for Result 3, i.e. the 95% confidence bands[7] for the median probability weighting functions. The gender specific confidence bands diverge only for contextual gains when probabilities exceed approx-

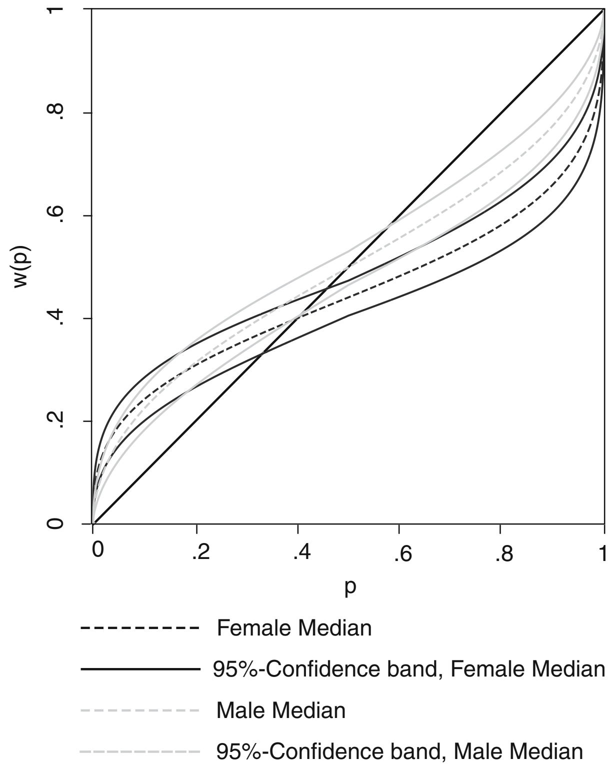

Figure 6. 95%-Confidence bands of median probability weighting functions: Contextual gains.

imately 0.5 (Figure 6). The confidence bands for men's and women's median probability weighting functions overlap over the whole range of probabilities in the other treatments.

4.4. *Result 4*

4.4.1. *For both sexes, the probability weighting functions tend to be more elevated for losses than for gains. However, women's domain-specific curves differ significantly whereas men's do not*

Support. Individual probability weighting functions exhibit a significant domain effect. The hypothesis of equal domain-specific distributions of the parameter estimates for δ can be rejected at the 1% level (Mann–Whitney test, bootstrapped). The elevation of the probability weighting function tends to be higher for losses than for gains. This finding is mirrored in the median probability weights. We constructed the 95% confidence bands of the median probability weighting functions to analyze the combined impact of both parameters γ and δ on the domain-specific curves. The confidence bands of the female curves do not overlap for considerable portions of the probability range in either environment. The male confidence bands do not exhibit such a pattern which means that the observed gender effects are, at least partly, caused by women's strong discriminative reaction to the presentation of lotteries in terms of gains or losses.

4.5. *Result 5*

4.5.1. *The vast majority of women and men value monetary outcomes linearly. Hence, there are no significant gender differences in the parameters of the individuals' value functions*

Support. As already mentioned above, we cannot reject the hypothesis of linear value functions for 76% of our subjects. Furthermore, a bootstrapped Mann–Whitney test yields no significant gender differences in the distributions of α and β at the 5% level of significance.

5. DISCUSSION

Are women relatively more risk averse than men in financial decision making? It depends. We can safely conclude that women and men do not strongly differ in their valuations of monetary outcomes, but there is convincing evidence that the sexes weight probabilities differently. Our experiment yields three major conclusions:

(1) *Universal feature:* In general, women are less sensitive to changes in probability than are men. Female probability weighting curves are more curved than male ones, irrespective of context and domain.

(2) *Domain specificity:* Women's reaction to gains differs strongly from their reaction to losses. Underweighting of larger probabilities is much more pronounced in the gain domain than in the loss domain where probability weights lie much closer to the identity line.
(3) *Context dependence:* Relative to men, women tend to be especially pessimistic when winning gambles are framed in investment terms.

These findings imply that we expect women to be the more risk averse gender mostly in the domain of investment decisions when the probability of a gain is of medium or large size. The same picture emerges at the level of observed risk premiums.

How do our findings relate to the experimental literature? Our results are confirmed by Harbaugh et al. (2002), the only other work addressing gender-specific probability weights. The experimental design in their study includes only abstract gambles, with probabilities evenly spread out over the (0,1) interval. Consistent with our findings on abstract gains, neither probability weights nor risk-taking behavior differ by sex in the Harbaugh experiment. As we would expect on the basis of our results, most of the laboratory experiments done in a contextual gains framework (Eckel and Grossman, 2002; Moore and Eckel, 2003; Schubert et al., 2000 among others) conclude that women are the relatively more risk averse sex. In contrast, Gysler et al. (2002) do not find a significant gender effect in their investment experiment. This seemingly contradictory result can be explained in the light of our findings: Gysler et al. use predominantly small probabilities for which our estimates suggest no significant differences in the gender-specific weights.

The Gysler et al. study also illustrates the observed context dependence of female behavior. Whereas in their experiment gender turns out to be insignificant overall, it is highly significant when interaction terms with overconfidence and financial market knowledge are taken into account. With increasing objectively measured knowledge men are relatively more risk averse, while women tend to be relatively more risk prone. It might be the case that women with a low level of knowledge and experience in financial matters feel less competent when lotteries are framed in investment terms and, therefore, behave particularly risk averse in this environment. The questionnaire administered at the end of our experiment indeed indicates a low level of experience with investment decisions among our women subjects.

As regards the reasons underlying women's more strongly curved probability weighting curves we can only speculate. One possible explanation could lie in women's stronger emotionality. The model of affective utility by Walther (2003) provides a link between a person's shape of the probability weighting curve and her sensitivity to emotions. At the moment of decision the decision maker anticipates that the resolution of the lottery's uncertainty will have two separate effects: besides utility from wealth, it will generate utility from emotional reactions to the resolution of uncertainty. In this model, the decision maker will have feelings of disappointment or elation if ex post events fall outside a "normal" range of deviations from the utility of the lottery's certainty equivalent. Disappointment and elation are assumed to decay over time. The model yields a nonlinear transformation of probabilities which depends, among others, on the parameters of emotional sensitivities and their rates of decay. The probability weighting function becomes more curved with increasing sensitivities to disappointment and elation and with decreasing rates of decay. Consequently, we could infer from the model that women, ceteris paribus, react much more intensely to feelings of disappointment and elation than do men.

6. CONCLUSION

Our analysis demonstrates that gender differences in risk-taking behavior crucially depend on probabilities. Studies that do not take this relationship into account may yield flawed conclusions.

Our experiment is limited to single-domain gambles. Mixed gambles allow estimation of λ, the parameter of loss aversion. Gender-specific estimates of loss aversion are an obvious candidate for future research. Aside from estimating loss aversion, another interesting feature of observed behavior can be investigated when mixed gambles are included in the experimental design: Subjects responding to mixed gambles tend to have relatively more curved probability weighting functions than when responding to gains and losses in isolation (Wu and Markle, 2004).

In our view, a particularly interesting venue for future endeavors concerns explanations for the curvature of the probability weighting function. Why women's probability weighting functions are more curved than men's is an open question. The link between risk preferences and emotions seems to be a promising candidate for this endeavor.

APPENDIX A

The Wilcoxon–Mann–Whitney Test with Bootstrap

The Wilcoxon–Mann–Whitney statistic tests the hypothesis that two independent samples X, of size m, and Y, of size n, are from populations with the same distribution. The null hypothesis, that the two samples were drawn from the same population, is tested against the alternative hypothesis, that they were not. If the distributions of the two samples do not differ in their higher moments, rejection of the null hypothesis implies that their first moments are different, i.e. that the two distributions differ in central tendency.

In the context of our paper, we are interested in whether the parameters of the value and probability weighting functions for female and male individuals are identically distributed. Let θ represent any one of these parameters. The null hypothesis H_0 asserts equality of the gender-specific distributions G(θ):

$$H_0: G(\theta)^{\text{male}} \overset{\text{dist.}}{=} G(\theta)^{\text{female}} \tag{8}$$

where θ denotes the true parameter value. Since we only have estimates of θ, denoted by $\hat{\theta}$, at our disposal, we cannot test H_0 directly. Instead we can test the null,

$$H_0 : \tilde{G}(\hat{\theta})^{\text{male}} \stackrel{\text{dist.}}{=} \tilde{G}(\hat{\theta})^{\text{female}} \tag{9}$$

where $H_0 : \tilde{G}(\hat{\theta})^{\text{male}} \stackrel{\text{dist.}}{\equiv} \hat{\theta}^{\text{male}} \equiv \{\hat{\theta}_1,^{\text{male}} \hat{\theta}_2^{\text{male}}, \dots, \hat{\theta}_I^{\text{male}}\}$ with the subscript denoting the number of the individual concerned. The distribution of the parameter estimates for the females is defined accordingly.

For the true parameter values, the Wilcoxon–Mann–Whitney statistic is defined by

$$Z = \frac{U(\theta^{\text{male}}\theta^{\text{female}}) - n/2}{\sqrt{n(m+n+1)/12m}}. \tag{10}$$

The placement of element θ_i^{female} in sample θ^{female} is defined as the number of lower-valued observations in θ^{male}—the other sample—and is denoted by $\mathrm{U}(\theta^{\text{male}}\theta_i^{\text{female}})$. The mean placement $\mathrm{U}(\theta^{\text{male}}\theta^{\text{female}})$ is the arithmetic mean of the $\mathrm{U}(\theta^{\text{male}}\theta_i^{\text{female}})$'s. For ties of θ^{male} values with θ^{female} values, a correction is employed (Siegel and Castellan, 1988). For large sample sizes, the statistic is approximately N(0,1).

Since θ is unknown, it is not obvious how good a proxy $\tilde{G}(\hat{\theta})$ is for $G(\theta)$ nor how safe we are in assuming that the Wilcoxon–Mann–Whitney test statistic based on $\hat{\theta}$ is approximately normally distributed.

Instead of relying on the normal approximation, we apply a non parametric bootstrap procedure to estimate the empirical distribution of the Wilcoxon–Mann–Whitney statistic. By resampling $\hat{\theta}$ we directly bootstrap the Wilcoxon–Mann–Whitney statistic.

APPENDIX B

Construction of the Confidence Bands of the Median Probability Weighting Functions by the Bootstrap Percentile Method

In the following we present the procedure for obtaining confidence bands for the median probability weighting functions.

First, we define the median person's certainty equivalents which are used to estimate the parameters of the median functions. Then we elaborate on the bootstrap method which enables us to calculate the confidence intervals of the parameter estimates. Finally, these confidence intervals have to be combined in a specific way to render confidence bands for the probability weighting curves.

The estimates for the (gender, environment, domain) specific median value and probability weighting functions are based on the respective median certainty equivalents $\mathrm{CE}_{\mathrm{med}}$. This vector of medians comprises $\mathrm{CE}^{l}_{\mathrm{med}}$ for each single lottery $l = 1, 2, \ldots, 25$ in the gain and loss domains, respectively, and is calculated as follows:

$$\mathbf{CE}^{l}_{\mathrm{med}} \equiv \mathrm{median}\{CE^{l}_{1}, CE^{l}_{2}, \ldots, CE^{l}_{i}\}, l \in \{1, 2, \ldots, 25\} \text{ and } i \in \{1, 2, \ldots\}, \tag{11}$$

i counts the number of individuals the median is calculated for (depending on gender, environment and domain). The median person's vector of certainty equivalents $\mathrm{CE}_{\mathrm{med}}$ is defined as:

$$\mathbf{CE}_{\mathrm{med}} \equiv \{CE^{1}_{\mathrm{med}}, CE^{2}_{\mathrm{med}}, \ldots, CE^{25}_{\mathrm{med}}\}. \tag{12}$$

Based on these $\mathbf{CE}_{\mathrm{med}}$, the domain-specific point estimator $\hat{\xi}$ for the parameters of the value and probability weighting functions is calculated using maximum likelihood.

We estimate the distribution of the point estimator $F_{\hat{\xi}}$ by a non parametric bootstrap (Efron, 1979). The bootstrap samples $\mathbf{CE}^{*}_{\mathrm{med}}$ are obtained by sampling $\{CE^{1}_{\mathrm{med}}, CE^{2}_{\mathrm{med}}, \ldots, CE^{25}_{\mathrm{med}}\}$ with replacement. We run the bootstrap procedure with B = 9999 repetitions. Analogous to $\hat{\xi}, \hat{\xi}^{*}$ is based on the bootstrap samples $\mathbf{CE}^{*}_{\mathrm{med}1}, \ldots, .\mathbf{CE}_{\mathrm{med}B}$. The estimator for $F_{\hat{\xi}}$ is defined as:

$$\widehat{F}_{\hat{\xi}}(x) \equiv P^{*}(\xi^{*} \leqslant x) \tag{13}$$

The bootstrap α-percentiles define the confidence intervals for the two parameters of the Lattimore functional:

$$\left[\widehat{F}^{-1}_{\hat{\delta}}\left(\frac{\alpha}{2}\right); \widehat{F}^{-1}_{\hat{\delta}}\left(1-\frac{\alpha}{2}\right)\right], \left[\widehat{F}^{-1}_{\hat{\gamma}}\left(\frac{\alpha}{2}\right); \widehat{F}^{-1}_{\hat{\gamma}}\left(1-\frac{\alpha}{2}\right)\right]. \tag{14}$$

To construct confidence bands for the probability weighting curve, we need to examine the combined effects of both parameters. We have to minimize $w(p)$ over γ and δ for the lower confidence bound ($\underline{w(p)}$) and we have to maximize $w(p)$ over these two parameters for the upper bound ($\overline{w(p)}$).

$$\underline{w(p)} \equiv \left\{ \min_{\delta,\gamma} w(p) | p \in [0,1], \delta \in \left[\widehat{F}_{\hat{\delta}}^{-1}\left(\frac{\alpha}{2}\right); \widehat{F}_{\hat{\delta}}^{-1}\left(1-\frac{\alpha}{2}\right)\right], \gamma \in \left[\widehat{F}_{\hat{\gamma}}^{-1}\left(\frac{\alpha}{2}\right); \widehat{F}_{\hat{\gamma}}^{-1}(1-\frac{\alpha}{2})\right]\right\}. \tag{15}$$

$$\overline{w(p)} \equiv \left\{ \max_{\delta,\gamma} w(p) | p \in [0,1], \delta \in \left[\widehat{F}_{\hat{\delta}}^{-1}\left(\frac{\alpha}{2}\right); \widehat{F}_{\hat{\delta}}^{-1}\left(1-\frac{\alpha}{2}\right)\right], \gamma \in \left[\widehat{F}_{\hat{\gamma}}^{-1}\left(\frac{\alpha}{2}\right); \widehat{F}_{\hat{\gamma}}^{-1}\left(1-\frac{\alpha}{2}\right)\right]\right\}. \tag{16}$$

The partial derivatives of $w(p)$ with respect to γ and δ are calculated as follows:

$$\frac{\partial w(p)}{\partial \gamma} = \frac{\delta p^\gamma (1-p)^\gamma [\ln(p) - \ln(1-p)]}{[\delta p^\gamma (1-p)^\gamma]^2} \tag{17}$$

and

$$\frac{\partial w(p)}{\partial \delta} = \frac{p^\gamma (1-p)^\gamma}{\delta p^\gamma + (1-p)^\gamma}. \tag{18}$$

Examining the signs of the partial derivatives, we find that $w(p)$ is monotone in δ, but not monotone in γ:

$$\frac{\partial w(p)}{\partial \delta} > 0 \quad \text{for } 0 < p < 1$$

but

$$\frac{\partial w(p)}{\partial \gamma} \begin{cases} > 0 & \text{for } p > 0.5 \\ = 0 & \text{for } p = 0.5 \\ < 0 & \text{for } p < 0.5 \end{cases}$$

Therefore, the confidence bands have to be constructed in the following way:

For $p < 0.5$ follows:

$$\underline{w(p)} \equiv \left\{ w(p) | p \in [0, 0.5[, \delta = \widehat{F}_{\hat{\delta}}^{-1}\left(\frac{\alpha}{2}\right), \gamma = \widehat{F}_{\hat{\gamma}}^{-1}\left(1-\frac{\alpha}{2}\right)\right\}, \tag{19}$$

$$\overline{w(p)} \equiv \left\{ w(p) | p \in [0, 0.5[, \delta = \widehat{F}_{\hat{\delta}}^{-1}\left(1 - \frac{\alpha}{2}\right), \gamma = \widehat{F}_{\hat{\gamma}}^{-1}\left(\frac{\alpha}{2}\right) \right\}. \quad (20)$$

and for $p \geqslant 0.5$:

$$\underline{w(p)} \equiv \left\{ w(p) | p \in [0.5, 1], \delta = \widehat{F}_{\hat{\delta}}^{-1}\left(\frac{\alpha}{2}\right), \gamma = \widehat{F}_{\hat{\gamma}}^{-1}\left(\frac{\alpha}{2}\right) \right\}, \quad (21)$$

$$\overline{w(p)} \equiv \left\{ w(p) | p \in [0.5, 1], \delta = \widehat{F}_{\hat{\delta}}^{-1}\left(1 - \frac{\alpha}{2}\right), \gamma = \widehat{F}_{\hat{\gamma}}^{-1}\left(1 - \frac{\alpha}{2}\right) \right\}, \quad (22)$$

which combine into the respective confidence bands.

ACKNOWLEDGEMENTS

We thank Ernst Fehr, Michael Kosfeld, George Wu, and an anonymous referee for their helpful comments.

NOTES

1. Instructions are available in our working paper version at www.wif.ethz.ch/research/workingpapers.
2. One Swiss Franc equals about 0.80 U.S. Dollars.
3. Guaranteed losses (in absolute terms) are arranged in ascending order.
4. To correct for heteroscedasticity the intervals $[x_2, x_1]$ are transformed to uniform length.
5. For a detailed analysis of relative risk premiums refer to our working paper version at www.wif.ethz.ch/research/workingpapers.
6. Normally, a standard Mann–Whitney test is sufficient for discerning among distributions of two random variables. In our case, we would like to test the distributions of the true parameter values for equality. Unfortunately, these are unobservable and we only have estimates at our disposal. Therefore, we construct the empirical distributions of the Mann–Whitney test statistic by a non parametric bootstrapping procedure which then serves as a basis for our tests (see Appendix A).
7. Due to the nonlinearity of the model, standard deviations estimated by the usual delta method may not be trustworthy. Instead, we estimated combined empirical confidence intervals for γ and δ based on a non parametric bootstrap procedure with 9999 repetitions (see Appendix B for details).

REFERENCES

Byrnes, J. P., Miller, D. C. and Schafer, W. D. (1999), Gender differences in risk taking: A meta-analysis, *Psychological Bulletin* 125, 367–383.

Eckel, C. C. and Grossman, P. J. (2002), Sex differences and statistical stereotyping in attitudes toward financial risk, *Evolution and Human Behavior* 23, 281–295.

Eckel, C. C. and Grossman, P. J. (2005), The difference in the economic decisions of men and women: Experimental evidence, In Plott, C. and Smith, V. L. (eds.), *Handbook of Experimental Economics Results*, Vol. 1, Amsterdam: North-Holland.

Efron, B. (1979), Bootstrap methods: Another look at the Jackknife, *Annals of Statistics* 7, 1–26.

Fischbacher, U. (1999), Z-Tree. Zurich Toolbox for Readymade Economic Experiments. *Institute of Empirical Economic Research, University of Zurich*, Working Paper No. 21.

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

Gysler, M., Brown Kruse, J. and Schubert, R. (2002), Ambiguity and gender differences in financial decision making: An experimental examination of competence and confidence effects. *Center for Economic Research, Swiss Federal Institute of Technology*, Working Paper.

Harbaugh, W. T., Krause, K. and Vesterlund, L. (2002), Risk attitudes of children and adults: Choices over small and large probability gains and losses, *Experimental Economics* 5, 53–84.

Lattimore, P. K., Baker, J. K. and Witte, A. D. (1992), The influence of probability on risky choice: A parametric examination, *Journal of Economic Behavior and Organization* 17, 377–400.

Moore, E. and Eckel, C. (2003), Measuring Ambiguity Aversion, *mimeo*.

Quiggin, J. (1982), A theory of anticipated utility, *Journal of Economic Behavior and Organization* 3, 323–343.

Schubert, R., Gysler, M., Brown, M. and Brachinger, H. W. (2000), Gender-specific attitudes towards risk and ambiguity: An experimental investigation. *Center for Economic Research, Swiss Federal Institute of Technology*, Working Paper.

Siegel, S. and Castellan, N. J. -Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York.

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

Walther, H. (2003), Normal-randomness expected utility, time preferences and emotional distortions, *Journal of Economic Behavior & Organization* 52, 253–266.

Wu, G. and Markle, A. B. (2004), An Empirical Test of Gain–Loss Separability in Prospect Theory, *mimeo*.

Addresses for correspondence: Helga Fehr-Duda, Institute of Economic Research, Swiss Federal Institute of Technology, Weinbergstrasse 35, CH-8092 Zurich, Switzerland. E-mail: fehr@wif.gess.ethz.ch. Tel: +41-44-632-46-25; Fax: +41-44-632-10-42.
Manuele de Genaro, Institute of Economic Research, Swiss Federal Institute of Technology, Weinbergstrasse 35, CH-8092 Zurich, Switzerland. Tel: +41-44-632-46-25; Fax: +41-44-632-10-42.
Renate Schubert, Institute of Economic Research, Swiss Federal Institute of Technology, Weinbergstrasse 35, CH-8092 Zurich, Switzerland Tel: +41-44-632-47-17; Fax: +41-44-632-10-42.