ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 78, No. 4 (July, 2010), 1375–1412

RISK AND RATIONALITY: UNCOVERING HETEROGENEITY IN PROBABILITY DISTORTION

ADRIAN BRUHIN
Institute for Empirical Research in Economics, University of Zurich, 8006 Zurich, Switzerland

HELGA FEHR-DUDA
ETH Zurich, 8092 Zurich, Switzerland

THOMAS EPPER
ETH Zurich, 8092 Zurich, Switzerland

Econometrica, Vol. 78, No. 4 (July, 2010), 1375–1412

RISK AND RATIONALITY: UNCOVERING HETEROGENEITY IN PROBABILITY DISTORTION

BY ADRIAN BRUHIN, HELGA FEHR-DUDA, AND THOMAS EPPER[1]

It has long been recognized that there is considerable heterogeneity in individual risk taking behavior, but little is known about the distribution of risk taking types. We present a parsimonious characterization of risk taking behavior by estimating a finite mixture model for three different experimental data sets, two Swiss and one Chinese, over a large number of real gains and losses. We find two major types of individuals: In all three data sets, the choices of roughly 80% of the subjects exhibit significant deviations from linear probability weighting of varying strength, consistent with prospect theory. Twenty percent of the subjects weight probabilities near linearly and behave essentially as expected value maximizers. Moreover, individuals are cleanly assigned to one type with probabilities close to unity. The reliability and robustness of our classification suggest using a mix of preference theories in applied economic modeling.

KEYWORDS: Individual risk taking behavior, latent heterogeneity, finite mixture models, prospect theory.

1. INTRODUCTION

RISK IS A UBIQUITOUS FEATURE of social and economic life. Many of our everyday choices, and often the most important ones, such as what trade to learn and where to live, involve risky consequences. While it has long been recognized that individuals differ in their risk taking attitudes, comparatively little is known about the distribution of risk preferences in the population.[2] Since preferences are one of the ultimate drivers of behavior, knowledge of the composition of risk attitudes is paramount to predicting economic behavior. Economic models often allow for heterogeneity, but this heterogeneity is usually defined by the boundaries of the standard model of preferences, expected utility theory (EUT). The empirical evidence, however, reveals that heterogeneity in risk taking behavior is of a substantive kind, that is, some people evaluate risky prospects consistently with EUT, whereas other people depart substantially from expected utility maximization (Hey and Orme (1994)). Moreover, it seems to be the case that rational decision makers, revealing EUT preferences, constitute only a minority of the population (Lattimore, Baker, and Witte (1992)).

[1]We thank Ernst Fehr, David Levine, Rainer Winkelmann, Michael Wolf, three anonymous referees, and the participants of the ESA World Meeting 2007 and the EEA-ESEM Meeting 2008 for helpful comments. The usual disclaimer applies. This research was supported by the Swiss National Science Foundation (Grant 100012-109907).

[2]Exceptions include Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2005), Eckel, Johnson, and Montmarquette (2005), Harrison, Lau, Rutström, and Sullivan (2005), and Harrison, Lau, and Rutström (2007).

 DOI: 10.3982/ECTA7139

To improve descriptive performance, a plethora of alternative theories have been developed. Unfortunately, no single best fitting model has been identified so far (Harless and Camerer (1994), Starmer (2000)) and, depending on the individual, one or the other model fits better. This finding poses a serious problem for applied economics. What the modeler needs is a *parsimonious* representation of risk preferences that is empirically well grounded and robust, and not a host of different functionals. Providing such a parsimonious characterization of heterogeneity in risk taking behavior is the objective of this paper.

Our method is based on a literature on classifying individuals which has been recently adopted by the social sciences. On the basis of statistical classification procedures, such as finite mixture models, investigators have tried to discover which decision rules people actually apply when playing games or dealing with complex decision situations (El-Gamal and Grether (1995), Stahl and Wilson (1995), Houser, Keane, and McCabe (2004), Houser and Winter (2004)). The finite mixture approach does not require fitting a model for each individual, which is—given the usual quality of choice data—frequently impossible and often not desirable in the first place. Instead, our method reveals latent heterogeneity by estimating the proportions of distinct behavioral types in the population and assigning each individual to one endogenously defined behavioral type, characterized by a unique set of parameter values.

We apply such a finite mixture model to choice data from three different experiments, two of which were conducted in Zurich, Switzerland. The third experiment took place in Beijing, People's Republic of China. We analyze 448 subjects' decisions over real monetary gains and losses, which comprise a total of nearly 18,000 observations. All three experiments were designed in a similar manner and served to elicit certainty equivalents for binary lotteries. Using a flexible sign-dependent functional as the basic behavioral model, we show the following main results.

First, the estimation procedure renders a robust classification of risk taking behavior across all three data sets. Moreover, the proportions of these distinct types in their respective populations are very similar.

Second, almost all the experimental subjects are unambiguously assigned to one distinct type. Measuring the quality of classification by the normalized entropy criterion (Celeux and Soromenho (1996)), ambiguity of assignments turns out to be extremely low. Thus, we observe hardly any mixed types, that is, individuals with a high probability (of say 0.4) of being one type *and* a high probability (of say 0.6) of being another type. This clean segregation suggests that the classification procedure is able to capture the distinctive characteristics of each behavioral type.

Third, without restricting parameter values a priori, we find that in all three data sets, the minority type, which constitutes about 20% of the population, weights probabilities and values monetary outcomes near linearly. Consequently, this group of individuals can essentially be characterized as expected value maximizers. This result is particularly interesting in the light of Rabin's (2000) calibration theorem, which shows that expected utility maximizers should be approximately risk neutral for small stakes, which typically are encountered in laboratory experiments, if behavior under high stakes is to remain within a plausible range of risk aversion. Therefore, we label subjects belonging to this group of nearly risk neutral people as EUT types. Moreover, the EUT group remains robust to increasing the number of types in the mixture.

Fourth, the majority of individuals, labeled cumulative prospect theory (CPT) types, are characterized by significant departures from linear probability weighting, consistent with prospect theory. As three-group classifications show, this group's behavior can be characterized as a mixture of two different types: In all three data sets a proportion of approximately 30% of the subjects display pronounced departures from linear probability weighting, whereas the relative majority of 50% differ less radically from linear probability weighting.

Finally, within the class of CPT types, we find major differences between Swiss and Chinese behavior. Sensitivity to changes in probabilities is generally lower for the Chinese subjects than for the Swiss. While the majority CPT groups' probability weighting curves do not differ dramatically between countries, the minority groups display diametrically opposed patterns of probability weighting. In particular, the minority Chinese CPT group weights probabilities extremely favorably, rendering them risk seeking over a considerable range of probabilities. The minority Swiss CPT group, however, is characterized by the opposite behavior. Thus, our analysis provides a deeper understanding for the finding that, on average, the Chinese tend to be more risk seeking than westerners (Kachelmeier and Shehata (1992)).

Our results show that the classification procedure successfully uncovers latent heterogeneity in the population. If there is heterogeneity of a substantive kind, as the data suggest, basing predictions on a single preference theory is inappropriate and may lead to biased results (Wilcox (2006)). EUT preferences should be taken into account alongside prospect theory preferences, even if rational EUT individuals constitute only a minority in the population. As the literature on the role of bounded rationality under strategic complementarity and substitutability shows, the mix of rational and irrational actors may be decisive for aggregate outcomes (Haltiwanger and Waldman (1985, 1989), Fehr and Tyran (2005), Camerer and Fehr (2006), Fehr and Tyran (2008)). Depending on the nature of strategic interdependence, the behavior of even a minority of players may drive the aggregate outcome. Therefore, the mix of types in the population is a crucial variable in predicting market outcomes. Since the finite mixture model provides a robust and reliable classification of individuals, the resulting estimates of group sizes and group-specific parameters may serve as valuable inputs for applied economics.

The finite mixture method has been used by others in the context of modeling risk taking. However, to the best of our knowledge, there is no previous study showing a nearly identical classification of risk preference types for three independent data sets. Additionally, our analysis breaks new ground by showing that EUT types emerge endogenously and by extending classification to three groups. Related work by Harrison and colleagues (Andersen, Harrison, and Rutström (2006), Harrison, Humphrey, and Verschoor (2010), Harrison and Rutström (2009)) applies finite mixture models as well, but differs from our approach. Their estimation procedure is based on the a priori assumption that *choices*, irrespective of by whom they were taken, are either EUT consistent or CPT consistent, that is, it sorts choices by a predefined decision model. In contrast, we aim to classify *individuals* by endogenously defined type. Therefore, if there is a group of people whose behavior can best be described by EUT, they should get identified by the classification procedure. Furthermore, in certain decision situations, choices of EUT individuals and CPT individuals do not differ substantially from one another and, therefore, both decision models fit equally well. Consequently, depending on the data available, classification by EUT- and CPT-consistent *decisions* may differ markedly from classification by *decision makers' types*.

A recent study by Conte, Hey, and Moffatt (2010) is also dedicated to finite mixture modeling of risk taking behavior. Their results for British subjects corroborate our conclusions: Even though their work differs from ours in set of lotteries, elicitation method, and estimation procedure, and restricts one behavioral type to be EUT a priori, they also find that in the domain of gains, 80% of the individuals exhibit nonlinear probability weighting, whereas 20% are assigned to EUT.

The paper is structured as follows. Section 2 describes the experimental design and procedures of the three experiments. The functional specification of the behavioral model and the finite mixture model are discussed in Section 3. Section 4 presents descriptive statistics of the data and the results of the classification procedure. Section 5 concludes.

2. EXPERIMENTAL DESIGN

In the following section we describe the experimental setup and procedures. The experiments took place in Zurich in 2003 and 2006 as well as in Beijing in 2005. In Zurich, all subjects were recruited from the subject pool of the Institute for Empirical Research in Economics, which consists of students of all fields of the University of Zurich and the Swiss Federal Institute of Technology Zurich. In Beijing, subjects were recruited by flier distributed at the campuses of Peking University and Tsinhua University. Since all three experiments are based on the same design principles, we will present the prototype experiment Zurich 2003 in detail and describe the extent to which the other two experiments deviate. The main distinguishing features of the different experiments are summarized in Table I.

We elicited certainty equivalents for a large number of two-outcome lotteries. One-half of the lotteries were framed as choices between risky and certain

TABLE I

DIFFERENCES IN EXPERIMENTAL DESIGN

	Zurich 03	Zurich 06	Beijing 05
Number of			
Subjects	179	118	151
Lotteries	50	40	28
Observations	8906	4669	4225
Procedure	Computerized	Computerized	Paper and pencil
Framing	Abstract and contextual	Contextual	Abstract and contextual

gains ("gain domain"); the other half were presented as choices between risky and certain losses ("loss domain").[3] For each decision in the loss domain, subjects were endowed with a specific monetary amount, which served to cover potential losses and equalized expected payoffs of corresponding gain and loss lotteries. In the Zurich 2003 and the Beijing experiments, 50% of the subjects were confronted with decisions framed in the standard gamble format. The other 50% of the subjects had to make choices framed in contextual terms, that is, gains were represented as risky or sure investment gains, and losses were represented as repair costs and insurance premiums, respectively. The Zurich 2006 experiment was based on contextually framed lotteries only. In Zurich, outcomes x_1 and x_2 ranged from 0 to 150 Swiss francs.[4] The payoffs in the Beijing 2005 experiment were commensurate with the compensation in Zurich and varied between 4 and 55 Chinese yuan.[5] Expected payoffs per subject amounted to approximately 31 Swiss francs and 20 Chinese yuan, respectively, which was considerably more than a local student assistant's hourly compensation, plus a show up fee of 10 Swiss francs and 20 Chinese yuan, thus generating salient incentives. Probabilities p of the lotteries' higher gain or loss x_1 varied from 5% to 95%. The gain lotteries for Zurich 2003 are presented in Table II. The other two experiments essentially included a subset of these. The lotteries appeared in random order on a computer screen[6] in the Swiss experiments and on paper in Beijing.

In the computerized experiments, the screen displayed a decision sheet containing the specifics of the lottery under consideration and a list of 20 equally spaced certain outcomes, ranging from the lottery's maximum payoff to the

[3]There were no mixed lotteries involving both gains and losses.

[4]At the time of the experiments, 1 Swiss franc equaled about 0.76 and 0.84 U.S. dollars, respectively.

[5]At the time of the experiment, 1 Chinese yuan equaled about 0.12 U.S. dollars.

[6]The experiment was programmed and conducted with the software z-Tree (Fischbacher (2007)).

TABLE II

GAIN LOTTERIES (x_1, p; x_2), ZURICH 2003[a]

p	x_1	x_2	p	x_1	x_2	p	x_1	x_2
0.05	20	0	0.25	50	20	0.75	50	20
0.05	40	10	0.50	10	0	0.90	10	0
0.05	50	20	0.50	20	10	0.90	20	10
0.05	150	50	0.50	40	10	0.90	50	0
0.10	10	0	0.50	50	0	0.95	20	0
0.10	20	10	0.50	50	20	0.95	40	10
0.10	50	0	0.50	150	0	0.95	50	20
0.25	20	0	0.75	20	0			
0.25	40	10	0.75	40	10			

[a]Outcomes x_1 and x_2 are denominated in Swiss francs.

lottery's minimum payoff, as shown in Figure 1.[7] The subjects had to indicate in each row of the decision sheet whether they preferred the lottery or the certain payoff. The lottery's certainty equivalent was calculated as the arithmetic mean of the smallest certain amount the subject preferred to the lottery and the subsequent certain amount on the list, when the subject had, for the first time, reported preference for the lottery. For example, if the subject had decided as indicated by the small circles in Figure 1, her certainty equivalent would amount to 13.5 Swiss francs.

Before subjects were permitted to start working on the real decisions, they had to correctly calculate the payoffs for two hypothetical choices. In the computerized experiments, there were two trial rounds to familiarize the subjects with the procedure. At the end of the experiment, one row number of one decision sheet was randomly selected for each subject, and the subject's choice in that row determined her payment. Subjects were paid in private afterward. The subjects could work at their own speed; the vast majority of them needed less than an hour to complete the experimental tasks as well as a socio-economic questionnaire.

3. ECONOMETRIC MODEL

This section discusses the specification of the finite mixture model, which allows controlling for latent heterogeneity in risk taking behavior in a parsimonious way. For the purpose of classifying subjects according to risk taking type, we need to specify three ingredients of the mixture model: the basic theory of decision under risk, the functional form of the decision model, and the specification of the error term.

[7]The format of the decision sheet for the Beijing experiment was identical to the Zurich one.

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	A profit of CHF 20 with	A		o	B	14
8		A	o		B	13
9	probability 75%	A	o		B	12
10		A	o		B	11
11	and a profit of CHF 0 with	A	o		B	10
12		A	o		B	9
13	probability 25%	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 the decision sheet.

The underlying theory of decision under risk should be able to accommodate a wide range of different behaviors. Sign- and rank-dependent models capture reference dependence and nonlinear probability weighting. Therefore, a flexible approach in the spirit of cumulative prospect theory (CPT) lends itself to describing risk taking behavior. Moreover, CPT nests EUT as special case.[8] If there is a group of people whose behavior can best be described by EUT, these individuals should be identified by the finite mixture estimation as a unique group exhibiting the predicted behavior.

Suppose that there are C different types of individuals in the population. According to CPT, an individual belonging to a certain group $c \in \{1, \dots, C\}$ values any binary lottery $\mathcal{G}_g = (x_{1g}, p_g; x_{2g}),\ g \in \{1, \dots, G\}$, where $|x_{1g}| > |x_{2g}|$, by

$$v(\mathcal{G}_g) = v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g)).$$

[8]The bulk of previous research has been conducted under the tacit assumption that utility is defined over lottery outcomes rather than lottery outcomes integrated with total wealth. In Section 4.8.1, we extend the model to accommodate the possibility of integration.

The function $v(x)$ describes how monetary outcomes x are valued, whereas the function $w(p)$ assigns a subjective weight to every outcome probability p. The lottery's certainty equivalent $\hat{\text{ce}}_g$ can then be written as

$$\hat{\text{ce}}_g = v^{-1}\big[v(x_{1g})w(p_g) + v(x_{2g})(1 - w(p_g))\big].$$

To make CPT operational, we have to assume specific functional forms for the value function $v(x)$ and the probability weighting function $w(p)$. A natural candidate for $v(x)$ is a sign-dependent power function

$$v(x) = \begin{cases} x^\alpha, & \text{if } x \geq 0, \\ -(-x)^\beta, & \text{otherwise,} \end{cases}$$

which can be conveniently interpreted and has turned out to be the best compromise between parsimony and goodness of fit in the context of prospect theory (Stott (2006)). Our specification of the value function seems to lack a prominent feature of prospect theory, loss aversion, capturing that "[...] most people find symmetric bets of the form $(x, 0.5; -x, 0.5)$ distinctly unattractive" (Kahneman and Tversky (1979, p. 279)). In this interpretation, loss aversion measures a decision maker's *attitude toward mixed lotteries*, encompassing both gains and losses.[9] Our lottery design does not contain any mixed lotteries, however. When there are only single-domain lotteries and loss aversion is introduced into our model in the conventional way, that is, by assuming $v(x) = -\lambda(-x)^\beta$ for $x < 0$ and $\lambda > 0$ (Tversky and Kahneman (1992)), the parameter of loss aversion λ is not identifiable: λ cancels out in the definition of the certainty equivalent ce of a loss lottery $(x_1, p; x_2)$ with $x_1 < x_2 \leq 0$, as $\lambda(-\text{ce})^\beta = \lambda(-x_1)^\beta w(p) + \lambda(-x_2)^\beta(1 - w(p))$ holds for *any* value of λ. Consequently, when there are no mixed lotteries available, estimating such a parameter is neither feasible nor meaningful.

Obviously, this argument rests on the assumption that subjects' reference point with respect to which gains and losses are defined is equal to zero. However, subjects might encode positive payments as gains only if they exceed a certain threshold, which would turn some of the objectively given gain lotteries into mixed ones, containing both subjective gains and losses. While in principle this is possible, estimating this reference point is questionable when there are no mixed lotteries from the onset, which would provide valuable additional information for locating the reference point reliably. To complicate matters, near linear value functions, as is predominantly the case for our data, pose severe

[9]Köbberling and Wakker (2005, p. 125) viewed loss aversion as a component of risk attitudes which is logically independent from basic utility: "Prospects [...] will exhibit considerably less risk aversion if [...] they are nonmixed than if [...] they are mixed. [...] [T]he difference in risk aversion between them is due to loss aversion."

identification problems.[10] For these reasons, we stick to common practice and assume a zero reference point.

Turning to the second component of the model, a variety of functional forms for modeling probability weights $w(p)$ have been proposed in the literature (Quiggin (1982), Tversky and Kahneman (1992), Prelec (1998)). We use the two-parameter specification suggested by Goldstein and Einhorn (1987) and Lattimore, Baker, and Witte (1992):

$$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}, \quad \delta \geq 0, \gamma \geq 0.$$

We favor this specification because it has proven to account well for individual heterogeneity (Wu, Zhang, and Gonzalez (2004)) and the parameters are nicely interpretable. The parameter $\gamma < 1$ largely governs the slope of the curve and measures sensitivity toward changes in probability. The smaller the value of γ is, the more strongly the probability weighting function departs from linear weighting.[11] The parameter δ largely governs curve elevation and measures the relative degree of optimism. The larger is the value of δ for gains, the more elevated is the curve, the higher is the weight placed on every probability, and, consequently, the more optimistically the prospect is valued, ceteris paribus. For losses, the opposite holds. Linear weighting is characterized by $\gamma = \delta = 1$. In a sign-dependent model, the parameters may take on different values for gains and for losses.

We now turn to the third step of model specification. In the course of the experiments, we measured risk taking behavior of individual $i \in \{1, \dots, N\}$ by her certainty equivalents ce_{ig} for a set of different lotteries. Since CPT explains *deterministic* choice, we have to add an error term ε_{ig} so as to estimate the parameters of the model based on the elicited certainty equivalents. The observed certainty equivalent ce_{ig} can then be written as $\mathrm{ce}_{ig} = \hat{\mathrm{ce}}_g + \varepsilon_{ig}$. There may be different sources of error, such as carelessness, hurrying, or inattentiveness, that result in accidentally wrong answers (Hey and Orme (1994)). The central limit theorem supports our assumption that the errors are normally distributed and simply add white noise.

Furthermore, we allow for three different sources of heteroskedasticity in the error variance. First, for each lottery, subjects had to consider 20 certain outcomes, which are equally spaced throughout the lottery's range $|x_{1g} - x_{2g}|$. Since the observed certainty equivalent ce_{ig} is calculated as the arithmetic mean of the smallest certain amount preferred to the lottery and the subsequent amount on the list, the error is proportional to the lottery range.[12]

[10]Previous attempts to estimate model parameters simultaneously with the reference point are extremely scarce and suggest that the reference point is of negligible magnitude (Harrison, List, and Towe (2007)); their experimental design included mixed lotteries, however.

[11]If linear probability weighting is accepted as a standard of rationality, $\gamma < 1$ can be interpreted as an index of departure from rationality (Tversky and Wakker (1995)).

[12]See Wilcox (2010) for a similar approach in the context of discrete choice under risk.

Second, as the subjects may be heterogeneous with respect to their previous knowledge, their attention span, and their ability to find the correct certainty equivalent, we expect the error variance to differ by individual. Third, lotteries in the gain domain may be evaluated differently from those in the loss domain. Therefore, we allow for domain-specific variance in the error term. This yields the form $\sigma_{ig} = \xi_i|x_{1g} - x_{2g}|$ for the standard deviation of the error distribution, where ξ_i denotes an individual domain-specific parameter. Note that the model allows us to test for both individual-specific and domain-specific heteroskedasticity either by imposing the restriction $\xi_i = \xi$ or by forcing all the ξ_i to be equal in both decision domains. Both types of restrictions are rejected by their corresponding likelihood ratio tests in all three samples with p-values close to zero. Therefore, we control for all three types of heteroskedasticity in the estimation procedure.

Having discussed all the necessary ingredients, we now turn to the specification of the finite mixture model. The basic idea of the mixture model is assigning an individual's risk taking choices to one of C types of behavior, each characterized by a distinct vector of parameters $\theta_c = (\alpha_c, \beta_c, \gamma_c', \delta_c')'$.[13] When estimating the model parameters, the number of types C is held fixed. The optimum number of classes is determined by estimating mixture models with varying C and applying some suitable test to decide among them (see Section 4.2). We denote the proportions of the C different types in the population by π_c. Given our assumptions on the distribution of the error term, the density of type c for the ith individual can be expressed as

$$f(\text{ce}_i, \mathcal{G}; \theta_c, \xi_i) = \prod_{g=1}^{G} \frac{1}{\sigma_{ig}} \phi\left(\frac{\text{ce}_{ig} - \hat{\text{ce}}_g}{\sigma_{ig}}\right),$$

where ϕ denotes the density of the standard normal distribution. Since we do not know a priori to which group a certain individual belongs, the proportions π_c are interpreted as probabilities of group membership. Therefore, each individual density of type c has to be weighted by its respective mixing proportion π_c, which, of course, is unknown and has to be estimated as well. Summing over all C components yields the individual's contribution to the model's likelihood L. The log likelihood of the finite mixture model is then given by

$$\ln L(\Psi; \text{ce}, \mathcal{G}) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c f(\text{ce}_i, \mathcal{G}; \theta_c, \xi_i),$$

where the vector $\Psi = (\theta_1', \ldots, \theta_C', \pi_1, \ldots, \pi_{C-1}, \xi_1, \ldots, \xi_N)'$ summarizes all the parameters of the model.

[13]The vectors γ_c and δ_c contain the domain-specific parameters for the slope and the elevation of the probability weighting functions.

The parameters are estimated by the iterative expectation maximization (EM) algorithm (Dempster, Laird, and Rubin (1977)),[14] which provides an additional feature: In each iteration, the algorithm calculates by Bayesian updating an individual's posterior probability τ_{ic} of belonging to group c. The final posterior probabilities represent a particularly valuable result of the estimation procedure. Not only do we obtain the probabilities of individual group membership, but we also have a method of judging the quality of classification at our disposal. If all the τ_{ic} are either close to 0 or 1, all the individuals are unambiguously assigned to one specific group. The τ_{ic} can be used to calculate a suitable measure of entropy, such as the normalized entropy criterion (Celeux and Soromenho (1996)), to gauge the extent of ambiguous assignments. If classification has been successful, that is, if genuinely distinct types have been identified, we should observe a low measure of entropy.

4. RESULTS

In this section we present descriptive statistics of the raw data and the results of the finite mixture estimations.

4.1. *Descriptive Statistics*

At the level of observed data, risk taking behavior can be conveniently summarized by relative risk premia $\text{RRP} = (\text{ev} - \text{ce})/|\text{ev}|$, where ev denotes the expected value of a lottery's payoff and ce stands for its certainty equivalent. $\text{RRP} > 0$ indicates risk aversion, $\text{RRP} < 0$ risk seeking, and $\text{RRP} = 0$ risk neutrality. In the context of EUT, risk preferences are captured solely by the curvature of the utility function, which in turn determines the sign of relative risk premia. Hence, the sign of RRP should be independent of p, the probability of the more extreme lottery outcome. In Figures 2–4, median risk premia sorted by p show a systematic relationship between RRP and p, however: In all three data sets subjects' choices display a fourfold pattern, that is, they are risk averse for low-probability losses and high-probability gains, and they are risk seeking for low-probability gains and high-probability losses. Therefore, at a first glance, average behavior is adequately described by a model such as CPT rather than EUT. As the following sections show, the median RRPs gloss over an important feature of the data as there is substantial latent heterogeneity in risk taking behavior.

[14]Various problems may be encountered when maximizing the likelihood function of a finite mixture model and, therefore, a customized estimation procedure was used that can adequately deal with these problems. Details of the estimation procedure, written in the R environment (R Development Core Team (2006)), are discussed in the Supplemental Material (Bruhin, Fehr-Duda, and Epper (2010)) available online.

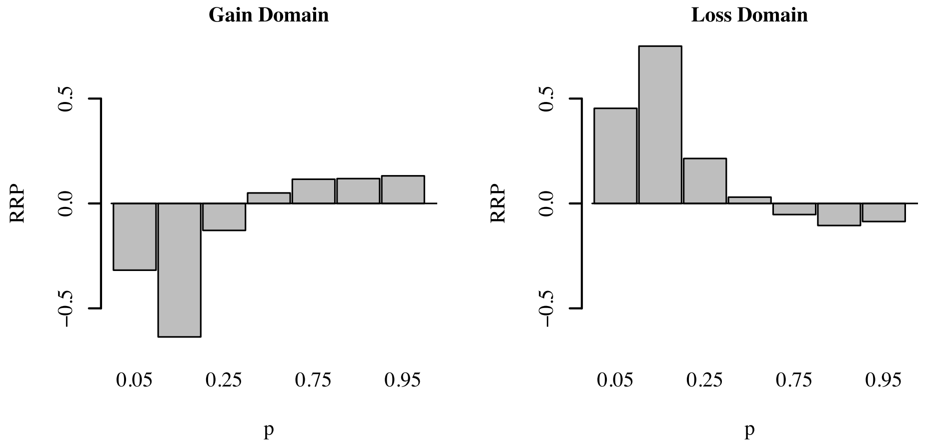

FIGURE 2.—Median relative risk premia, Zurich 2003.

4.2. *Model Selection*

So far we have not addressed the issues of whether a finite mixture model is actually to be preferred over a single-component model in the first place, and of what the number of groups C in the mixture model, often termed *model size*, should be. To deal with these questions, the researcher needs a criterion for assessing the correct number of mixture components. The literature on model selection in the context of mixture models is quite controversial, however, and

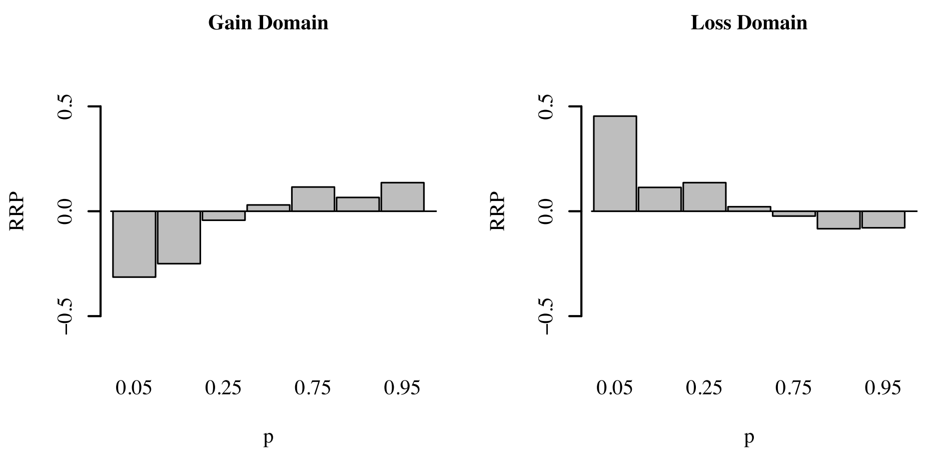

FIGURE 3.—Median relative risk premia, Zurich 2006.

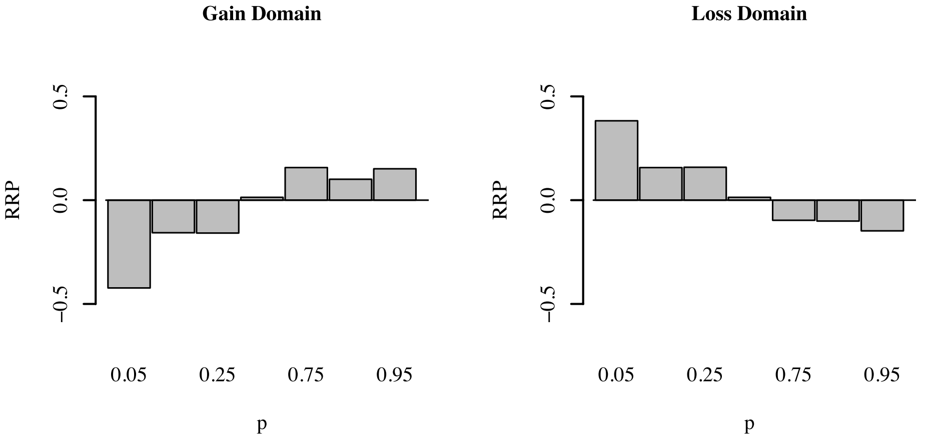

FIGURE 4.—Median relative risk premia, Beijing 2005.

there is no best solution.[15] For this reason, rather than relying on a single measure, we examine several criteria with differing characteristics to get a handle on the problem of model selection.

Obviously, the classical information criteria, the *Akaike information criterion* (AIC) and the *Bayesian information criterion* (BIC), are a natural starting point for our analysis. Unfortunately, the AIC is order inconsistent, that is, the probability that it is minimized at the true model size does not approach unity with increasing sample size, and it tends to overfit models (Atkinson (1981), Geweke and Meese (1981), Celeux and Soromenho (1996)). The BIC, on the other hand, has been proved to be consistent under suitable regularity conditions, but may suffer from over- or underestimating the number of mixture components (Biernacki, Celeux, and Govaert (2000)).

Aside from these problems, both classical criteria share the principle of trading off model parsimony against goodness of fit, but do not directly measure the ability of the mixture to provide well separated and nonoverlapping components, which, ultimately, is the objective of estimating mixture models. Therefore, Celeux and Soromenho (1996) proposed the *normalized entropy criterion* (NEC), which is based on the posterior probability of group membership τ_{ic}. Biernacki, Celeux, and Govaert (1999) argued that the NEC criterion appears to be less sensitive than AIC and BIC. However, the NEC focuses solely on the quality of classification and does not take model fit into account.

Ideally, what the researcher would like to have at her disposal is a criterion that delivers an assessment of both model fit, making allowance for parsimony,

[15]"The problem of identifying the number of classes is one of the issues in mixture modeling with the least satisfactory treatment" (Wedel (2002, p. 364)). For example, a standard likelihood ratio test is not appropriate here (Cameron and Trivedi (2005, p. 624)).

and the quality of classification. Biernacki, Celeux, and Govaert (2000) therefore suggested modifying the BIC criterion by factoring in a penalty for mean entropy. When the mixture components are well separated, mean entropy is close to zero and its weight in their proposed *integrated completed likelihood criterion* (ICL) is negligible. In the one-component case, there is no entropy by definition, and therefore ICL coincides with BIC. While there is no theoretical justification for this approach, simulations seem to show a superior performance compared to other heuristic criteria, such as NEC (Biernacki, Celeux, and Govaert (2000)), as well as compared to AIC and BIC (McLachlan and Peel (2000)).

As different criteria may come up with conflicting results concerning the correct number of mixture components, model selection is a difficult problem. One way to deal with this issue is to use one's central research question as a guideline.[16] Our concern here is twofold: First, given the vast heterogeneity in individual risk taking behavior, it is doubtful whether a single-component model is adequate. Therefore, the crucial question is whether $C > 1$ should be preferred to $C = 1$.[17] Second, considering the heated dispute about the "right" model of choice under risk, another objective of our study is to identify relative group sizes of EUT and non-EUT types. Bearing these objectives in mind, we calculated values for four different criteria, AIC, BIC, NEC, and ICL, and three different model sizes, $C \in \{1, 2, 3\}$, which are presented in Table III. According to these criteria, the model size which minimizes the respective criterion value should be preferred.

TABLE III

MODEL SELECTION CRITERIA

	AIC	BIC	NEC	ICL
Zurich 03				
$C = 1$	−38,398	−35,815	n.a.	−35,815
$C = 2$	−39,629	−36,997	**0.0099**	−36,991
$C = 3$	**−40,504**	**−37,822**	0.0131	**−37,807**
Zurich 06				
$C = 1$	−20,858	−19,297	n.a.	−19,297
$C = 2$	−22,173	−20,568	**0.0041**	−20,566
$C = 3$	**−22,622**	**−20,971**	0.0049	**−20,968**
Beijing 05				
$C = 1$	−18,485	−16,529	n.a.	−16,529
$C = 2$	−19,585	−17,585	**0.0061**	−17,582
$C = 3$	**−19,965**	**−17,920**	0.0114	**−17,912**

[16]Cameron and Trivedi (2005, p. 622) argued in this context: "Therefore, it is very helpful in empirical application if the components have a natural interpretation."

[17]Parameter estimates for $C = 1$ are presented in the Supplemental Material.

As AIC, BIC, and therefore also ICL, are highest at $C = 1$ for all three data sets, $C > 1$ is clearly favored over $C = 1$. As the NEC criterion is not defined for $C = 1$, Biernacki, Celeux, and Govaert (1999) argued in favor of a multicomponent model if there is a $C > 1$ with $\text{NEC}(C) \leq 1$, which is the case here. We therefore conclude that a finite mixture model is superior to a single-component model, given the unanimous recommendation by all four criteria.

With regard to the choice between $C = 2$ and $C = 3$, the three-group classifications seem to be favored by all criteria but NEC. Given the minimum level of NEC at $C = 2$, a two-group classification is preferable if the central issue is a parsimonious representation of risk taking types rather than best model fit. As entropy is generally extremely low for both the two-group and three-group classifications, both model sizes seem quite sensible, however. Before we infer from these results that we should choose $C = 3$, we take a closer look at the difference between the two-group and the three-group classifications.[18] What is of special interest here is whether one group remains fairly stable and the other group gets subdivided into two new ones when model size is increased, or whether the individuals get reshuffled to three new types. If the latter were the case, a two-group specification would clearly be misleading. To answer this question, we examine relative group sizes and transition patterns of individuals' type assignment.

Table IV displays the estimated relative group sizes of the behavioral types for model sizes $C = 2$ and $C = 3$. As the percentages reveal, all the Type I groups remain stable with respect to relative group size. Moreover, with a few exceptions, Type I individuals remain Type I when model size is increased: Only a total of 2% of the individuals move into or out of Type I when an additional component is introduced into the finite mixture model.[19] Increasing model size results in a decomposition of the original Type II groups into two subtypes, Type IIa and Type IIb, as there is still considerable heterogeneity within these groups. Thus, from the point of view of identifying Type I individuals, the

[18]Since there is quite some heterogeneity within the majority group, it is to be expected that even finer segmentations improve model fit. However, when we extend the number of groups beyond three, multimodality of the log likelihood function becomes a severe problem as, depending on the randomly drawn start values, even a stochastic extension of the EM algorithm tends to converge to local maxima. For "poorly drawn" start values, the estimation algorithm diverges, with one group getting smaller in each iteration, which might indicate that the likelihood is unbounded (McLachlan and Peel (2000, p. 54)). Therefore, estimating larger models may ask too much of our data. See also the discussion of overparametrization in Cameron and Trivedi (2005, p. 625)). Nevertheless, in the case of Zurich 06 we were able to estimate four- and five-group models: In both cases, the relative size of the minority group declines only slightly. This finding supports our conjecture that heterogeneity is particularly pronounced within the majority group, whereas the minority group is fairly homogeneous and robust to model size. Since we are not able to present results for all three data sets, we do not discuss these findings here.

[19]Across all three data sets, only two individuals are newly assigned to Type I and seven individuals leave Type I when C is increased from 2 to 3.

TABLE IV

RELATIVE GROUP SIZES

	Type I	Type II/IIa	Type IIb
Zurich 03			
$C = 2$	17.1%	82.9%	
$C = 3$	16.7%	27.3%	56.0%
Zurich 06			
$C = 2$	22.3%	77.7%	
$C = 3$	22.0%	29.8%	48.2%
Beijing 05			
$C = 2$	20.3%	79.7%	
$C = 3$	19.9%	29.3%	50.8%

two-group classifications are informative by themselves and provide the most parsimonious classification, whereas three groups render a more detailed description of Type II individuals. To keep interpretation of graphs manageable, we will present results for $C = 2$ when contrasting Type I with Type II, and for $C = 3$ when discussing subtypes of Type II behavior.[20]

4.3. *Clean and Robust Segregation of Behavioral Types*

To be of value to applied economics, a classification of risk taking behavior should meet two conditions. First, it should be clean, that is, all the individuals should be clearly associated with one specific risk taking type. Second, the classification should be robust across different experiments based on the same design principles. Regarding the first condition, entropy criteria, based on the posterior probabilities of group membership, can be used to evaluate the quality of classification. One such measure is the *normalized entropy criterion* introduced in the previous section. If all the individuals can be clearly assigned to one of the different behavioral groups, the posterior probabilities of group membership τ_{ic} are close to 0 or 1, and NEC ≈ 0. A τ_{ic} distinctly different from 0 or 1 indicates that the individual is classified as a "mixed" type belonging to group c with probability τ_{ic} and to the other group(s) with probability $1 - \tau_{ic}$. As Table III shows, NEC always lies in the vicinity of 0, irrespective of model size C being $1, 2$, or 3, that is, there are hardly any mixed types with ambiguous group affiliation.

The high quality of classification can also be inferred directly from the distributions of the individuals' posterior probabilities of group membership. In Figure 5, based on $C = 2$, τ_{EUT} denotes the posterior probability of belonging

[20]The interested reader is referred to Bruhin, Fehr-Duda, and Epper (2007) for an extensive discussion of $C = 2$.

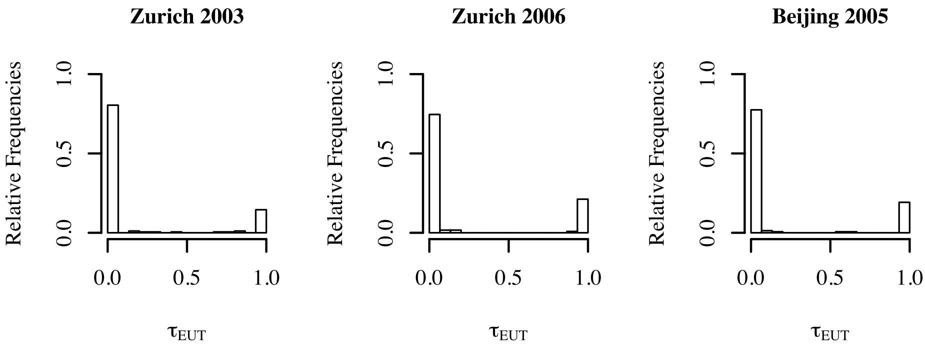

FIGURE 5.—Distribution of posterior probability of assignment to EUT, τ_{EUT} ($C = 2$).

to the first type, which can indeed be characterized, as we will demonstrate below, as expected utility maximizers.[21] As the distributions of τ_{EUT} show, the individuals' posterior probabilities of behaving consistently with EUT are either close to 1 or close to 0 for practically all the individuals in all three data sets, indicating an extremely clean segregation of subjects to types. Our result is quite remarkable as it substantiates that there are distinct types in the population—be they Swiss or Chinese—and it also shows that the underlying behavioral model provides a sound basis of discriminating between them.

With respect to the second criterion, robustness of classification, Figure 5 illustrates the probably most striking result of our study, namely similar distributions of types across all three data sets. In all three histograms of Figure 5, there are about four times as many individuals with τ_{EUT} close to 0, compared to individuals with τ_{EUT} close to 1. This finding is mirrored by the estimates of the relative group sizes, displayed in Table IV, which show a stable proportion of Type I of about 20%, irrespective of model size C. Moreover, it can be shown that the hypothesis of the same distribution of types prevailing in all three data sets cannot be rejected. Similarly, when model size is increased to $C = 3$, relative group sizes turn out to be of equal magnitudes in all three data sets and are statistically indistinguishable from one another. Therefore, classification is not only unambiguous, but also results in roughly equal mixing proportions, demonstrating that classification is robust across experiments.

This finding leads us to the next question. Do the respective types identified in each data set also exhibit similar patterns of behavior? This question will be addressed in the following sections, dedicated to the characterization of the endogenously defined types of behavior.

[21]As group membership is stable, histograms of τ_{EUT} for $C = 3$ are qualitatively the same.

4.4. *Characterization of the Minority Type*

Irrespective of model size, the first type of individuals encompasses about 20% of the subjects in all three data sets, thus constituting the minority type. To characterize risk taking behavior, we examine the parameter estimates of the value functions and probability weighting functions. Table V displays, for $C = 2$, the type-specific parameter estimates of the finite mixture model and their standard errors, obtained by the bootstrap method with 4000 replications (Efron and Tibshirani (1993)).[22] When model size is increased to three groups, parameter estimates, presented in Tables VI–IX, remain unchanged for the minority type, as group membership does not change substantially. Therefore, from the point of view of identifying this type of individuals, model size is not

TABLE V

CLASSIFICATION OF BEHAVIOR ($C = 2$)[a]

	EUT Types				CPT Types			
Parameters	ZH 03	ZH 06	BJ 05	Pooled	ZH 03	ZH 06	BJ 05	Pooled
π	0.171	0.223	0.203	0.193	0.829	0.777	0.797	0.807
	(0.026)	(0.025)	(0.020)	(0.013)				
Gains								
α	0.978	0.988	1.083	0.981	1.054	0.901	0.389	0.941
	(0.014)	(0.018)	(0.102)	(0.011)	(0.021)	(0.026)	(0.107)	(0.013)
γ	0.954	0.945	0.911	0.943	0.415	0.425	0.245	0.377
	(0.022)	(0.020)	(0.033)	(0.021)	(0.015)	(0.015)	(0.014)	(0.009)
δ	0.910	0.909	0.889	0.911	0.845	0.862	1.315	0.926
	(0.015)	(0.019)	(0.052)	(0.012)	(0.022)	(0.028)	(0.074)	(0.013)
Losses								
β	1.007	1.013	1.023	1.015	1.107	1.122	1.144	1.139
	(0.018)	(0.023)	(0.084)	(0.013)	(0.028)	(0.047)	(0.107)	(0.019)
γ	0.871	0.953	0.949	0.950	0.417	0.451	0.309	0.397
	(0.043)	(0.020)	(0.040)	(0.023)	(0.017)	(0.014)	(0.013)	(0.009)
δ	0.967	1.049	1.066	1.072	1.025	1.059	0.937	0.991
	(0.062)	(0.033)	(0.065)	(0.026)	(0.028)	(0.044)	(0.053)	(0.016)
$\ln L$	20,185	11,336	10,108	41,385				
Parameters	371	249	315	909				
Individuals	179	118	151	448				
Observations	8906	4669	4225	17,800				

[a]Standard errors (in parentheses) are based on the bootstrap method with 4000 replications. Parameters include additional estimates for ξ_i for domain- and individual-specific error variances. ZH stands for Zurich; BJ stands for Beijing.

[22]"[U]nless the sample size is very large, the standard errors found by an information-based approach may be too unstable to be recommended" (McLachlan and Peel (2000, p. 68)).

TABLE VI

CLASSIFICATION OF BEHAVIOR WITH $C = 3$, ZURICH 2003[a]

	Gains				Losses		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.167 (0.016)	0.273 (0.015)	0.560 (0.022)				
α	0.954 (0.013)	1.007 (0.016)	1.075 (0.015)	β	1.006 (0.020)	1.237 (0.044)	1.091 (0.015)
γ	0.944 (0.041)	0.302 (0.031)	0.467 (0.013)	γ	0.885 (0.042)	0.304 (0.029)	0.459 (0.015)
δ	0.930 (0.020)	0.622 (0.023)	0.944 (0.017)	δ	1.024 (0.043)	1.371 (0.075)	0.897 (0.016)
$\ln L$			20,630				
Parameters			378				
Individuals			179				
Observations			8906				

[a]Standard errors (in parentheses) are based on the bootstrap method with 4000 replications. Parameters include estimates of ξ_i for domain- and individual-specific error variances.

a crucial issue and the two-group classifications nicely contrast the distinctive characteristics of the minority type with the majority type.

Concerning probability weighting, Table V displays almost identical parameter estimates across all three data sets as well as the pooled data. Without any restrictions imposed on the parameters, we find that the minority types' probability weighting functions are roughly linear, as the parameter estimates for both γ and δ are close to 1. Since the probability weights are a nonlinear combination of these parameters, inference needs to be based on γ and δ jointly. Therefore, we constructed the 95%-confidence bands for the probability weighting curves by the bootstrap method. Figures 6, 7, and 8 contain the graphs of the type-specific probability weighting functions for each decision domain. The gray solid lines correspond to the estimated curves for the first type, referred to as EUT type, and the gray dashed lines delimit their respective confidence bands. For both gains and losses, the confidence bands for the first type in fact include the diagonal over a wide range of probabilities, demonstrating high congruence with linear probability weighting. Where the confidence bands do not include the diagonal, the curves still lie extremely close to linear weighting. In sum, in all three data sets, we find the first behavioral type to exhibit near linear probability weighting.

With respect to the valuation of monetary outcomes, the second element of the decision model, the estimated parameters α and β also display a high degree of conformity. As can be inferred from the standard errors in Table V, the 95%-confidence intervals of each single curvature estimate contain unity,

TABLE VII

CLASSIFICATION OF BEHAVIOR WITH $C = 3$, ZURICH 2006[a]

	Gains				Losses		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.220 (0.020)	0.298 (0.025)	0.482 (0.030)				
α	0.990 (0.024)	0.884 (0.042)	0.908 (0.031)	β	1.012 (0.029)	1.100 (0.083)	1.141 (0.049)
γ	0.946 (0.084)	0.362 (0.081)	0.465 (0.022)	γ	0.952 (0.081)	0.393 (0.078)	0.491 (0.023)
δ	0.905 (0.042)	0.658 (0.054)	1.012 (0.043)	δ	1.054 (0.074)	1.460 (0.122)	0.878 (0.054)
$\ln L$			11,567				
Parameters			256				
Individuals			118				
Observations			4669				

[a]Standard errors (in parentheses) are based on the bootstrap method with 4000 replications. Parameters include estimates of ξ_i for domain- and individual-specific error variances.

TABLE VIII

CLASSIFICATION OF BEHAVIOR WITH $C = 3$, BEIJING 2005[a]

	Gains				Losses		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.199 (0.017)	0.293 (0.026)	0.508 (0.027)				
α	1.083 (0.098)	0.032 (0.155)	0.489 (0.113)	β	1.023 (0.070)	1.348 (0.149)	1.111 (0.102)
γ	0.911 (0.051)	0.244 (0.049)	0.254 (0.023)	γ	0.948 (0.053)	0.263 (0.046)	0.332 (0.019)
δ	0.889 (0.094)	2.194 (0.241)	1.085 (0.113)	δ	1.062 (0.057)	0.600 (0.093)	1.106 (0.075)
$\ln L$			10,304				
Parameters			322				
Individuals			151				
Observations			4225				

[a]Standard errors (in parentheses) are based on the bootstrap method with 4000 replications. Parameters include estimates of ξ_i for domain- and individual-specific error variances. Estimates for CPT-I α statistically not distinguishable from logarithmic utility.

TABLE IX

CLASSIFICATION OF BEHAVIOR WITH $C = 3$, POOLED[a]

	Gains				Losses		
	EUT	CPT-I	CPT-II		EUT	CPT-I	CPT-II
π	0.198 (0.010)	0.316 (0.011)	0.486 (0.013)				
α	0.960 (0.009)	0.901 (0.009)	0.957 (0.010)	β	1.019 (0.008)	1.250 (0.010)	1.139 (0.009)
γ	0.915 (0.032)	0.309 (0.015)	0.451 (0.010)	γ	0.935 (0.027)	0.339 (0.013)	0.444 (0.011)
δ	0.935 (0.009)	0.726 (0.012)	1.063 (0.010)	δ	1.055 (0.013)	1.230 (0.013)	0.878 (0.011)
$\ln L$			42,105				
Parameters			916				
Individuals			448				
Observations			17,800				

[a]Standard errors (in parentheses) are based on the bootstrap method with 4000 replications. Parameters include estimates of ξ_i for domain- and individual-specific error variances.

implying that the hypothesis of linear value functions cannot be rejected. Together with near linear probability weighting, this result justifies regarding the first type of individuals as largely consistent with expected value maximization, and therefore EUT.

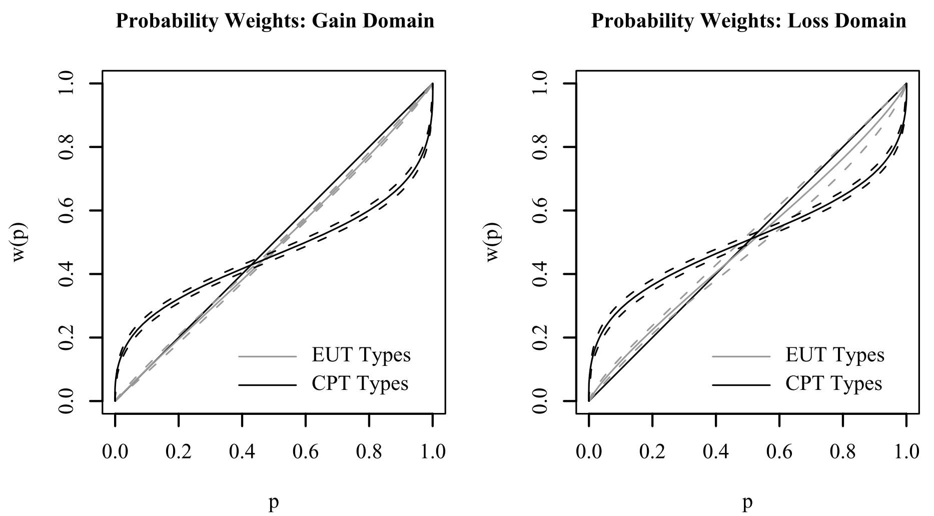

FIGURE 6.—Type-specific probability weighting functions, Zurich 2003.

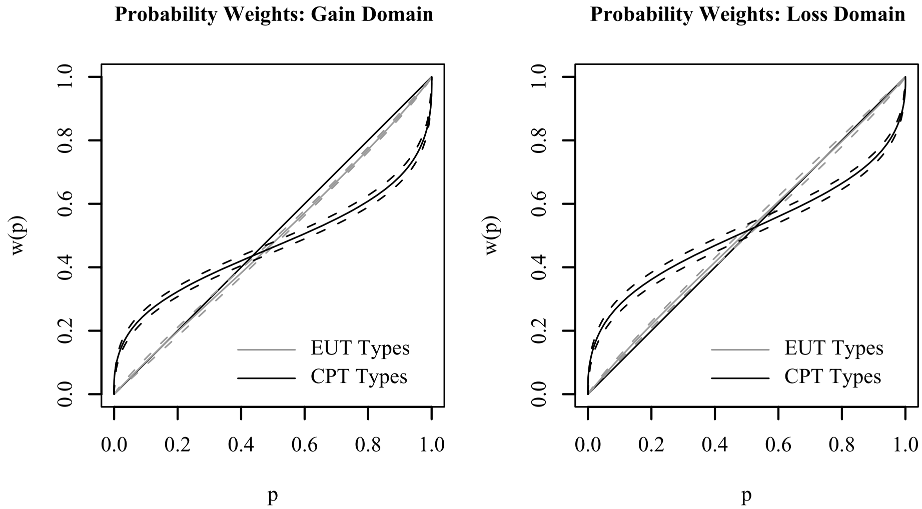

FIGURE 7.—Type-specific probability weighting functions, Zurich 2006.

4.5. *Characterization of the Majority Types*

As the discussion on model selection revealed, model size makes a difference when characterizing the majority types. Due to the stability of the minority EUT groups in all three data sets, the behavior of the majority groups can be described by a mixture of two different subtypes. As the majority groups exhibit

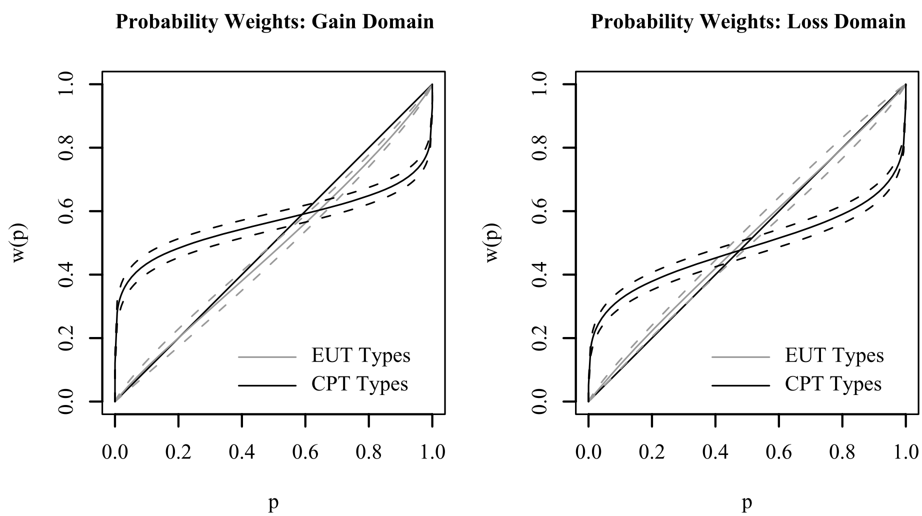

FIGURE 8.—Type-specific probability weighting functions, Beijing 2005.

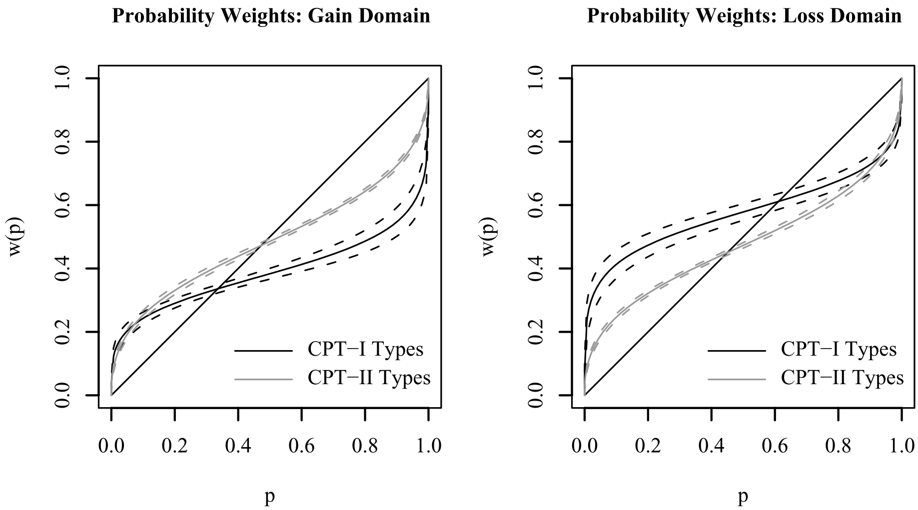

FIGURE 9.—Probability weights CPT-I versus CPT-II, Zurich 2003.

inverted S-shaped probability weighting curves, apparent in Figures 6, 7, and 8, we label them CPT types and label their corresponding subtypes CPT-I and CPT-II.

CPT-I and CPT-II groups are characterized by specific varieties of nonlinear probability weighting as Figures 9–11 show. The difference between CPT-I

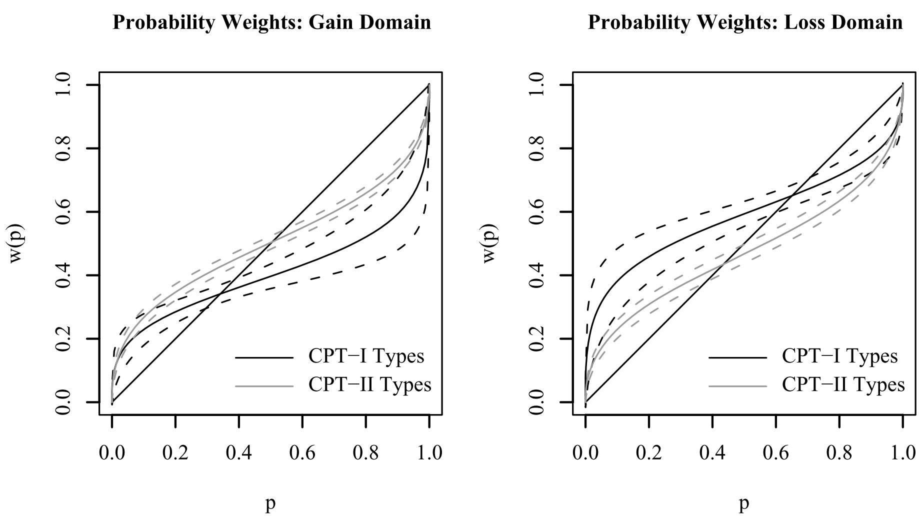

FIGURE 10.—Probability weights CPT-I versus CPT-II, Zurich 2006.

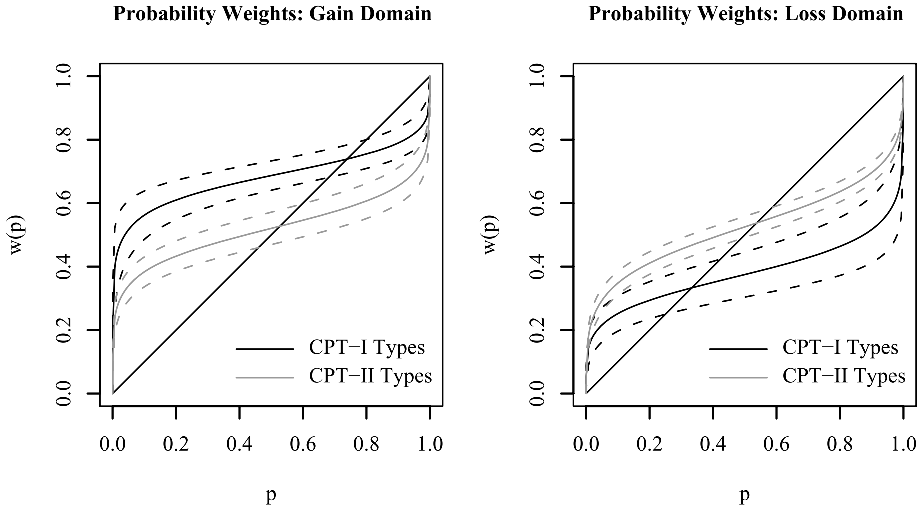

FIGURE 11.—Probability weights CPT-I versus CPT-II, Beijing 2005.

and CPT-II types manifests itself predominantly in relative strength of optimism: the elevation of the probability weighting curves, measured by δ,[23] differs substantially between CPT-I and CPT-II. CPT-II individuals, who constitute the relative majority of approximately 50% in all three data sets, exhibit moderately S-shaped probability weighting curves with δ in the vicinity of 1. The remaining 30% of the individuals, however, are characterized by differing patterns of behavior. Swiss CPT-I individuals are systematically less optimistic than Swiss CPT-II types, whereas the Chinese CPT-I group encompasses highly optimistic individuals, overweighting gain probabilities and underweighting loss probabilities over a wide range of probabilities. This specific feature of Chinese CPT-I types might explain the prevalence of high risk tolerance in the Chinese population, documented by previous research (Kachelmeier and Shehata (1992)).

The three-group classifications constitute a valuable piece of information when more disaggregate estimates of risk taking behavior are called for. When the focus of research lies on a parsimonious characterization of risk taking types, juxtaposing rational decision makers, not prone to probability distortions, with nonrational ones, two-group classifications are sufficiently informative due to the stability of EUT group membership.

[23]Parameter estimates are presented in Tables VI–IX.

4.6. *Observed Behavior by Type*

So far we have characterized the different behavioral types by their estimated parameter values. The obvious question that arises is whether the discriminatory power of the classification procedure can also be traced at the behavioral level. After assigning the subjects to one of the types, EUT, CPT-I, or CPT-II, based on their posterior probability of group membership τ_{ic}, the observed relative risk premia can be broken down by type as depicted in Figure 12, aggregated for the pooled data set. As can be seen, median RRP of the EUT types are close to 0, reflecting near risk neutral behavior in accordance with expected value maximization.

When we trace the behavior of the CPT-I and CPT-II types at the level of observed RRP in Figure 12, we find a fourfold pattern of risk attitudes, with distinctive departures from risk neutrality. Consistent with the characterizations before, CPT-I types exhibit more pronounced deviations. These findings document that individuals' type assignment is highly congruent with observed behavioral differences.

Obviously, the type-specific median relative risk premia do not differ greatly at $p = 0.5$. In decision situations when the more extreme reward materializes with a 50% chance, the typical CPT individual will not over- or underweight its probability significantly, and therefore her behavior will often not be distinguishable from a typical EUT type's behavior. This consideration can be illustrated by means of Figure 13, which displays the departures of average CPT behavior, aggregated over both subtypes CPT-I and CPT-II, from EUT, measured by the type-specific differences in median normalized certainty equivalents. Each circle in Figure 13 corresponds to one specific lottery played in any of the three experiments, encompassing a total of 59 gain and 59 loss lotteries, ordered by the probability of the more extreme lottery outcome. At a gain probability of 25%, for instance, CPT lottery evaluations, on average, exceed EUT by up to 30% of their corresponding expected values. The dashed lines in the graphs represent the case when median CPT behavior does not differ from median EUT behavior. Positive values in the graphs indicate that, on average, CPT types are relatively more risk seeking than EUT types. The opposite holds for negative values. As the graphs show, zero differences occur solely at the 0.5 probability level, where, in some cases, average CPT behavior is totally indistinguishable from EUT behavior. The bulk of type-specific differences in lottery evaluations lie in the range of about ±20% of expected values, but there are also a few observations with up to ±300% of expected value, where the more extreme outcomes materialize with a low probability. In these cases, CPT types tend to overreact pronouncedly to stated probabilities. To provide an overall measure, we conducted two-sided Mann–Whitney tests which indicate significant differences (at the 5% level) in the type-specific distributions of the certainty equivalents for 75% of the lotteries.

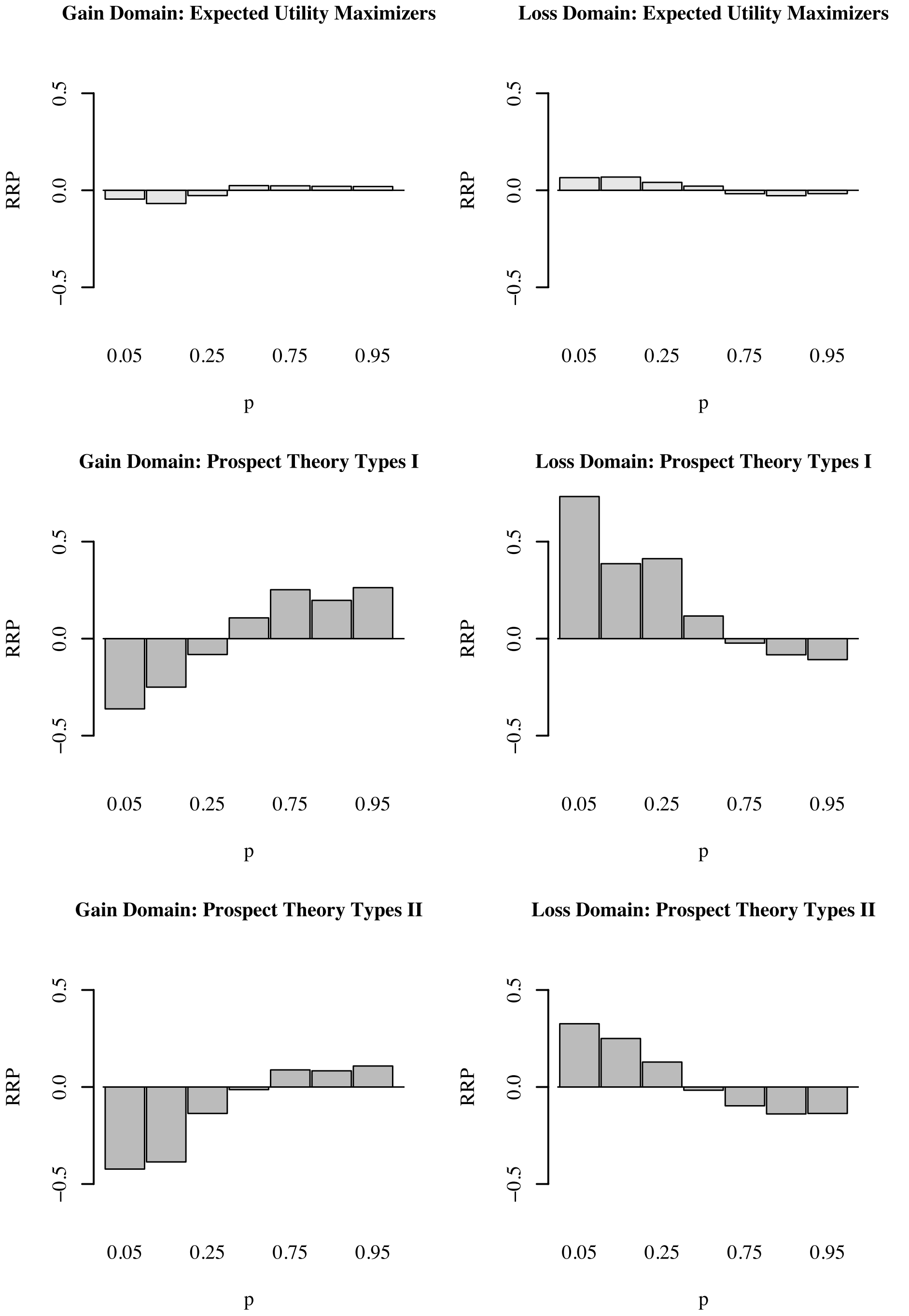

FIGURE 12.—Median relative risk premia by type, pooled.

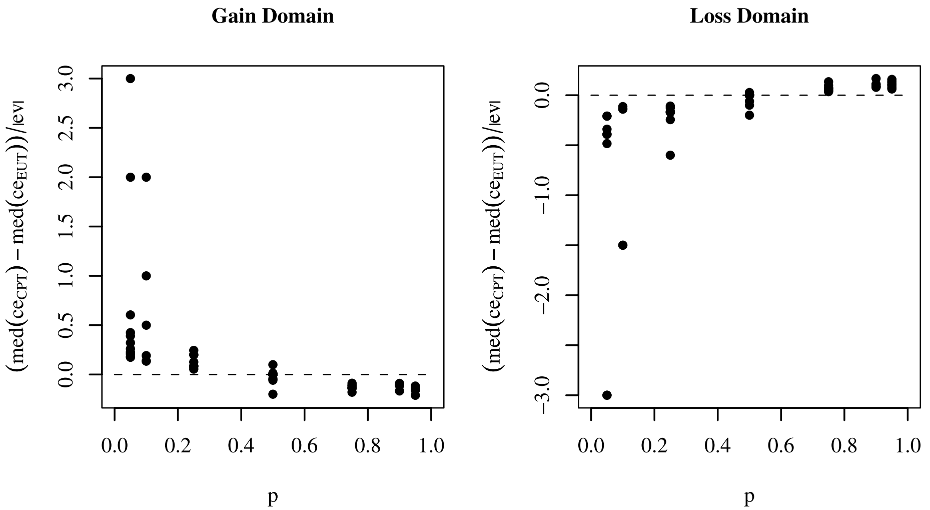

FIGURE 13.—Differences in median normalized certainty equivalents, pooled.

4.7. *Demographics and Group Membership*

The finite mixture model is a powerful tool to uncover latent heterogeneity in behavior. Given our clean and robust classification of types, an interesting question is whether we can characterize the composition of the different groups by demographic variables. In particular, can we explain who the EUT types are? To answer this question, we conducted two kinds of analyses. First, we estimated a single-component model with demographic variables as covariates. This procedure uncovers systematic behavioral differences among groups defined by observable socio-economic characteristics. We included the following variables: a gender dummy *female*, the number of semesters enrolled at university *semester*, and a binary variable *highincome* for incomes above a certain threshold. Summary statistics for these variables are included in the Supplemental Material. It turns out that the only variable that consistently affects behavioral parameters across experiments is *female*[24]: Being female is associated with a substantially lower value of γ, the slope of the probability weighting function. This finding implies that women tend to be less sensitive to changes in probability than men, in line with the evidence in Fehr-Duda, de Gennaro, and Schubert (2006).[25]

[24]Note that the percentage of females is approximately 50% in all three data sets. Parameter estimates for the single-component model are available in the Supplemental Material.

[25]In our experience, in student subject pools we generally do not find socio-economic characteristics, other than gender, that are systematically correlated with the curvature of the probability weighting function. Factors other than demographics may be more important here, but this question is still underresearched.

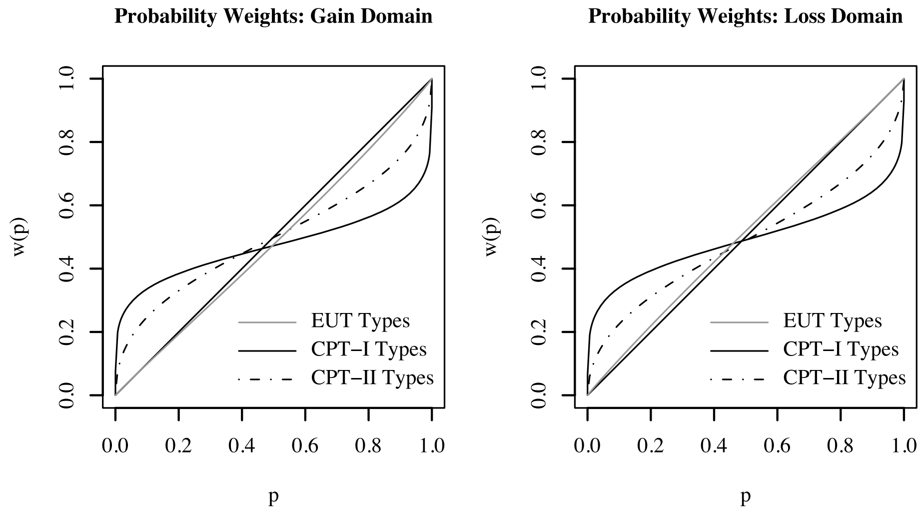

FIGURE 14.—Type-specific probability weighting functions: men.

Second, given that only gender systematically influences parameter values, we estimated the finite mixture model separately for men and women. In the following text we limit discussion to the results for the pooled model with $C = 3$. The gender-specific probability weighting functions classified by types are presented in Figures 14 and 15.[26] Whereas the distributions of types are quite similar, the probability weights display a striking gender difference. The men's groups differ essentially by their degree of rationality, characterized by the magnitude of the slope parameter γ. As in the overall data, the EUT group's probability weights lie very close to the diagonal. The male CPT-I types deviate quite strongly from linear weighting, whereas the CPT-II types, who constitute the relative majority of 49% of the men, lie somewhere in between these two more extreme groups. The women's minority group, however, departs more strongly from linear weighting than does the men's. One may conclude from these findings that the overall EUT group is dominated by the behavior of male individuals exhibiting near rational probability weighting.

The female CPT-I and CPT-II curves differ predominantly in the size of the elevation parameter δ. Compared with its male counterpart, the female CPT-I type also exhibits quite pronounced probability distortions, albeit with a larger fraction of optimistically weighted probabilities. The largest gender difference is displayed by the CPT-II types. Women in this group are characterized by the widest region of pessimistically weighted probabilities. This group's behavior seems to have a decisive influence on women's greater average risk aversion, usually found in empirical studies. While previous research has typically

[26]Parameter estimates are available in the Supplemental Material.

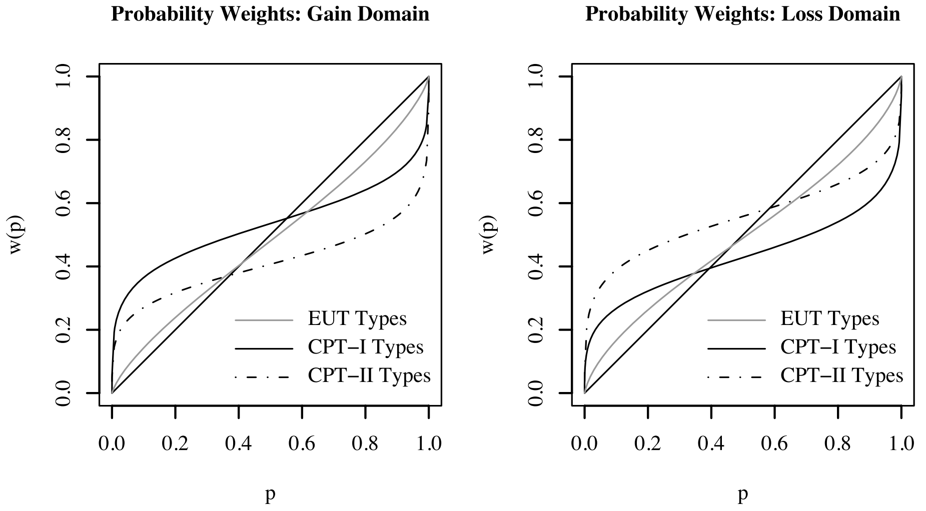

FIGURE 15.—Type-specific probability weighting functions: women.

centered on comparative risk aversion, our finite mixture estimations provide new, much more detailed, insights in gender-specific differences in risk taking behavior.

4.8. *Extensions*

4.8.1. *Robustness to Model Specification*

An additional part of our analysis concerns the robustness of classification results with respect to alternative specifications of the value function. For instance, people may not evaluate lotteries in isolation, but integrate prospective outcomes with their wealth or consumption spending. To account for the possibility that subjects integrate prospective outcomes with some background variable, we reestimated the model with the value function being defined over the sum of the prospective lottery outcome and an additional type-specific background parameter k, to be estimated, such that $v(x) = (x + k)^\alpha$ over gains and *mutatis mutandis* over losses, that is, $v(x) = (x + \omega + k)^\beta$, where ω stands for the initial endowment.[27]

[27]Estimating such an additional parameter comes at a cost, however. As Wakker (2008, p. 1338) noted, k represents an "anti-index of concavity" and therefore serves a similar function as the exponents of the value function α and β. For this reason, their respective contributions to utility curvature cannot be reliably separated unless one has observations over two distinct sets of lotteries (e.g., over low stakes and high stakes) at one's disposal (Heinemann (2008)). Moreover, k is not identifiable when functions v are near linear. Previous research suggests that under EUT, people generally do not integrate their wealth in their choices over risky lotteries (Binswanger

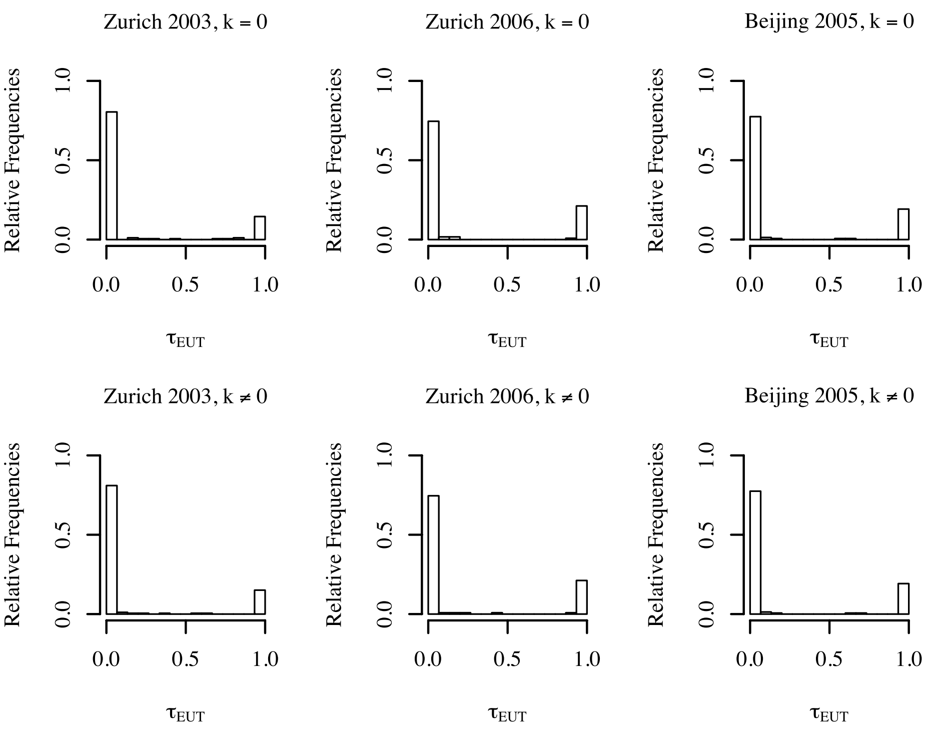

FIGURE 16.—Distribution of posterior probability of assignment to EUT τ_{EUT}.

Extending our model in such a manner yields the following insights. First, the stability of classification is not affected by the alternative model specification: For all three data sets, the distribution of the posterior probability of belonging to EUT is almost unaltered when background consumption is introduced into the model as Figure 16 shows. The stability of group assignment is also reflected in the estimated relative group sizes π_{EUT}. Table X clearly shows

TABLE X

ESTIMATED MODEL-SPECIFIC PROPORTIONS OF EUT TYPES, π_{EUT}

	Zurich 03	Zurich 06	Beijing 05
$k = 0$	0.171	0.223	0.203
k endogenous	0.163	0.227	0.203

(1981), Harrison, List, and Towe (2007), Heinemann (2008)). Since group affiliation of EUT types remains stable, we limit discussion to $C = 2$ here.

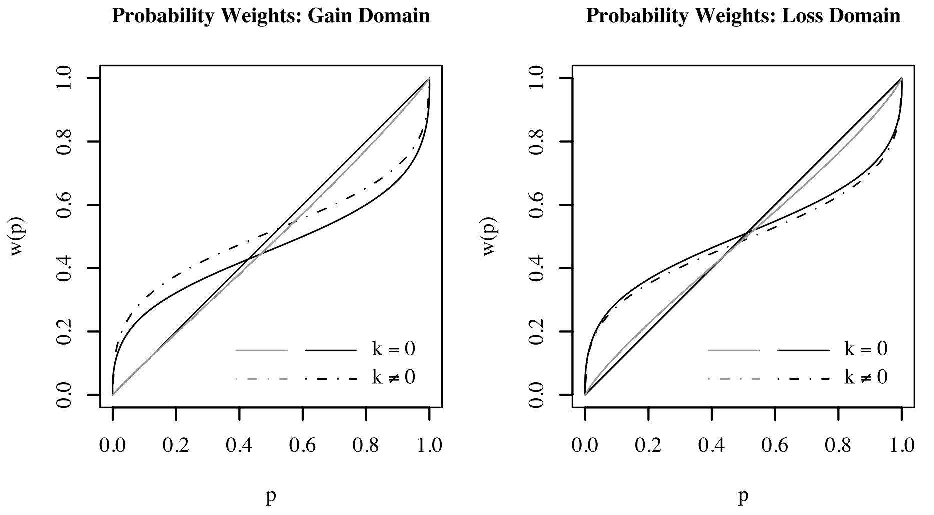

FIGURE 17.—Probability weights Zurich 2003, k endogenous.

that these values practically do not change. Moreover, not a single subject out of 448 is assigned to a different group, defined by $\tau_{ic} \geq 0.5$, after allowing for integration with background consumption. Finally, the estimated probability weighting functions for both the EUT types and the CPT types remain stable as well, as Figures 17–19, confirm. In sum, our analysis attests that the distribution of types, individuals' type affiliations, and the estimated probability weighting functions are robust to inclusion of background consumption. This robustness result represents further evidence that decision makers' tendency to weight probabilities nonlinearly is the driving force of classification.

4.8.2. *Heterogeneity in Error Variance*

The finite mixture model supplied us not only with estimates of type-specific behavioral parameters, but also with estimates of the error parameters, ξ_i—the normalized standard deviations of the error distributions. These parameters are idiosyncratic to the individual and, thus, capture some of the heterogeneity across subjects. A high error variance does not necessarily stem from random behavior, however. In an aggregate model such as ours, individual errors also reflect the degree of congruence between individual and group behavior. The question then arises of how well average behavioral group parameters describe subjects with differing degrees of departure from average behavior.[28] To investigate this matter, we classified individuals as either low- or high-variance

[28]We are grateful to an anonymous referee who called our attention to this issue.

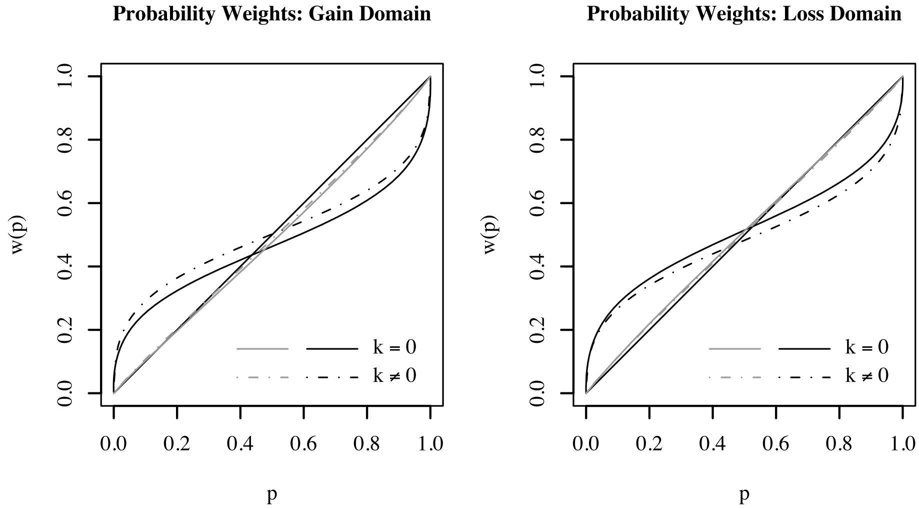

FIGURE 18.—Probability weights Zurich 2006, k endogenous.

type, depending on their estimated ξ_i being below or above the respective median value, and reestimated the behavioral parameters for each of the resulting six types (two types of variance × three types of behavior), pooled over all three data sets. The upper panel in Figure 20 displays the average probability weighting curves for the aggregate types estimated from the pooled data.

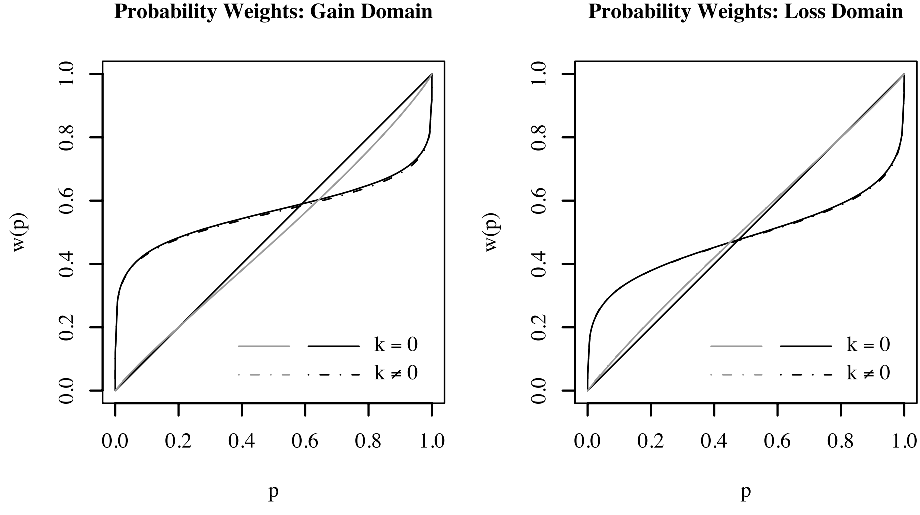

FIGURE 19.—Probability weights Beijing 2005, k endogenous.

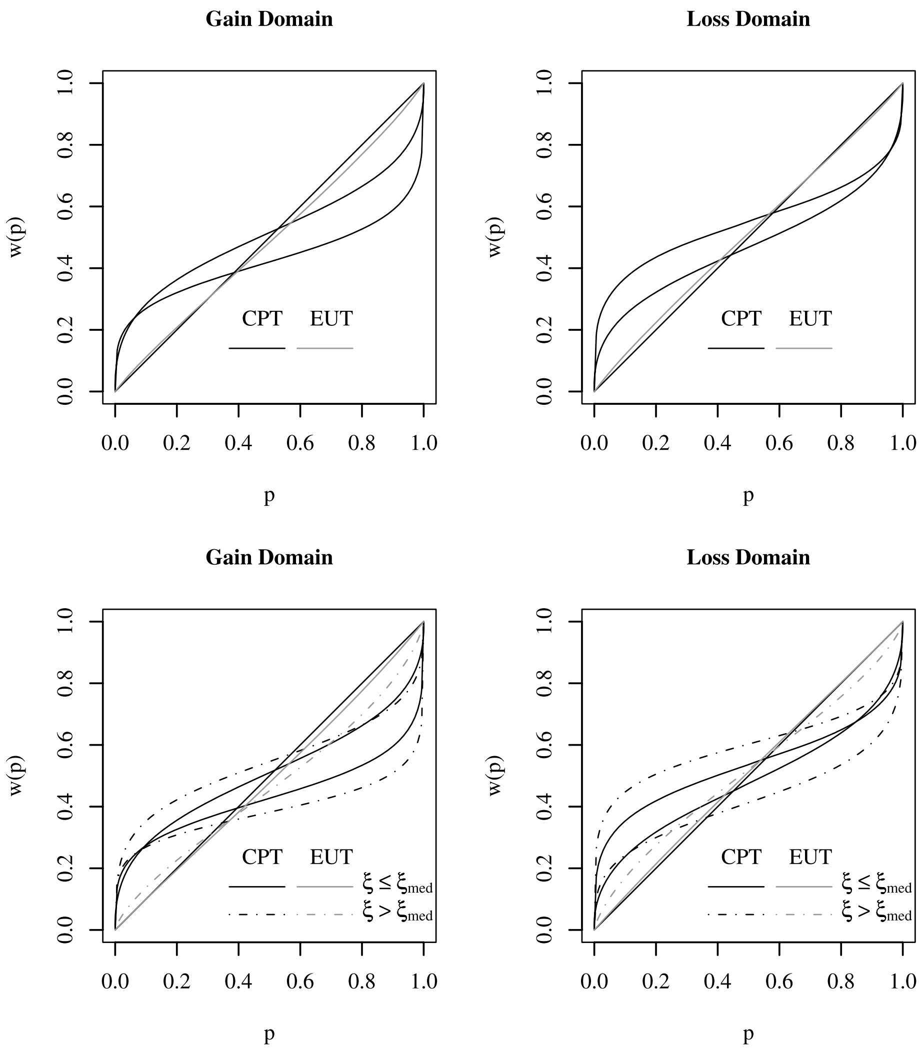

FIGURE 20.—Probability weights by error variance, pooled.

The lower panel contains the curves for the variance-specific types where the solid lines mark the low-variance people's curves and the dot-dashed lines denote the respective high-variance ones. Comparing the variance-specific curves with the overall averages reveals that low- and high-variance EUT probability weighting functions generally differ somewhat in degree of rationality, but largely remain within a comparatively narrow band around linear weighting. CPT individuals, however, exhibit a wide range of degrees of optimism. Typical high-variance individuals are either distinctly less optimistic (CPT-I) or more optimistic (CPT-II) than their low-variance colleagues.[29] Not surprisingly perhaps, decomposing behavior according to error variance widens the spectrum of emerging probability weighting types. These findings underscore that EUT types are a fairly homogeneous group, whereas CPT types display a much wider range of behaviors.

5. CONCLUDING REMARKS

We conducted three experiments based on the same design principles and applied a finite mixture model to the choice data. Our results provide novel insights: In all three data sets, the procedure renders a parsimonious characterization of risk taking behavior. Across experiments, we find an equal mix of distinct types, each characterized by a specific pattern of probability distortion. Almost every single individual is identified as one specific type, rendering segregation extremely clean. By and large, 20% of the population adhere to linear probability weighting and behave essentially as expected value maximizers, whereas majority preferences are more suitably represented by a model such as prospect theory, which can accommodate nonlinear probability weighting. In each data set, the overall CPT group is composed of a smaller group of 30% of the subjects who display substantial departures from linear probability weighting, and a relative majority of 50% who depart less radically from linear probability weighting. Moreover, classification is robust to an alternative model specification.

Whereas the distribution of types is the same in the Swiss and the Chinese data sets, there are substantial cultural differences in CPT-type behavior, the most prominent being the existence of a pronouncedly optimistic group of Chinese subjects who distort small- and medium-sized probabilities much more strongly than do the Swiss. This prevalence of Chinese optimism in lottery valuation may explain previous findings that, on average, Chinese respondents are relatively more risk seeking than westerners (Kachelmeier and Shehata (1992), Hsee and Weber (1999)). We also identify a gender difference in risk taking behavior: Women generally depart more strongly from linear probability weighting than men. This finding corroborates previous research (Fehr-Duda, de Gennaro, and Schubert (2006), Harrison and Rutström (2009)). Moreover, on average, female probability distortions vary predominantly in degree of optimism, whereas male probability distortions vary in degree of rationality.

Our findings demonstrate that the finite mixture approach is a powerful tool to identify and to characterize the distribution of risk taking types in the population. In this study, the individual is the unit of classification, that is, the *entirety* of an individual's choices governs group affiliation. As the low measures of entropy demonstrate, almost every individual got unambiguously assigned to one

[29] In the upper panel of Figure 20, comparatively more optimistic probability weighting represents CPT-II and comparatively less optimistic weighting represents CPT-I.

endogenously defined behavioral type. Previous work by Harrison and collaborators tried to accomplish a different goal: They estimated the probability that any one lottery choice, irrespective of the identity of the decision maker, was consistent with EUT or CPT, respectively, and found that "each [specification] is equally likely for these data" (Harrison and Rutström (2009, p. 144)). Clearly, in certain decision situations CPT-consistent choices are indistinguishable from EUT-consistent ones. Our findings indicate, for example, that this is the case for outcome probabilities in the neighborhood of 0.5. Since a CPT individual's choices in this region are interlinked with all her other choices, the respective observations are categorized as CPT by our method, but may be interpreted as either CPT or EUT in the choice-based approach. Therefore, classification results may differ depending on the unit of classification and the type of data available.

When we started this project, we were quite confident that we would find a considerable percentage of expected utility maximizers. What really surprised us is the robust percentage of EUT individuals, even across two so different cultures as the Swiss and Chinese. Our findings were recently corroborated by Conte, Hey, and Moffatt (2010), who found a similar distribution of behavioral types for British subjects.

Since the subject pools in all three of our experiments consisted of students, further research is needed to see whether the proportion of near rational EUT types changes significantly in a representative sample and whether the complexity of decision tasks greatly alters type assignment. If it can be ascertained that near rational actors constitute a nonnegligible proportion of the population, their behavior, depending on the nature of the strategic environment, may be decisive for aggregate outcomes. The existence of a robust share of near rational actors suggests using a mix of preference theories for modeling behavior rather than a single theory, which would yield systematically biased results. In our data, prospect theory adequately describes behavior of the majority of subjects, but the parameter estimates exhibit culture- as well as type-specific values. Researchers should take this evidence into account when constructing, estimating, and applying models of choice under risk.

REFERENCES

ANDERSEN, S., G. W. HARRISON, AND E. E. RUTSTRÖM (2006): "Choice Behavior, Asset Integration and Natural Reference Points," Working Paper 06-07, Department of Economics, College of Business Administration, University of Central Florida. [1378]

ATKINSON, A. (1981): "Likelihood Ratios, Posterior Odds and Information Criteria," *Journal of Econometrics*, 16, 15–20. [1387]

BIERNACKI, C., G. CELEUX, AND G. GOVAERT (1999): "An Improvement of the NEC Criterion for Assessing the Number of Clusters in a Mixture Model," *Pattern Recognition Letters*, 20, 267–272. [1387,1389]

——— (2000): "Assesing a Mixture Model for Clustering With the Integrated Completed Likelihood," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719–725. [1387,1388]

BINSWANGER, H. P. (1981): "Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal*, 91, 867–890. [1403,1404]

BRUHIN, A., H. FEHR-DUDA, AND T. F. EPPER (2007): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," Working Paper 0705, Socioeconomic Institute, University of Zurich. [1390]

——— (2010): "Supplement to 'Risk and Rationality: Uncovering Heterogeneity in Probability Discortion'," *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/7139_tables.pdf; http://www.econometricsociety.org/ecta/Supmat/7139_data and programs.zip. [1385]

CAMERER, C. F., AND E. FEHR (2006): "When Does Economic Man Dominate Social Behavior?" *Science*, 311, 47–52. [1377]

CAMERON, A., AND P. TRIVEDI (2005): *Microeconometrics. Methods and Applications.* Cambridge: Cambridge University Press. [1387-1389]

CELEUX, G., AND G. SOROMENHO (1996): "An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model," *Journal of Classification*, 13, 195–212. [1376,1385,1387]

CONTE, A., J. D. HEY, AND P. G. MOFFATT (2010): "Mixture Models of Choice Under Risk," *Journal of Econometrics* (forthcoming). [1378,1409]

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1–38. [1385]

DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP, AND G. G. WAGNER (2005): "Individual Risk Attitudes: New Evidence From a Large, Representative, Experimentally-Validated Survey," Discussion Paper 1730, Institute for the Study of Labor (IZA), Bonn, Germany. [1375]

ECKEL, C., C. JOHNSON, AND C. MONTMARQUETTE (2005): "Savings Decisions of the Working Poor: Short and Long-Term Horizons," in *Field Experiments in Econometrics*. Research in Experimental Economics, Vol. 10, ed. by J. Carpenter, G. W. Harrison, and J. A. List. Greenwich, CT: JAI Press, 219–260. [1375]

EFRON, B., AND R. J. TIBSHIRANI (1993): *An Introduction to the Bootstrap*. New York: Chapman & Hall. [1392]

EL-GAMAL, M. A., AND D. M. GRETHER (1995): "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 1137–1145. [1376]

FEHR, E., AND J. R. TYRAN (2005): "Individual Irrationality and Aggregate Outcomes," *Journal of Economic Perspectives*, 19, 43–66. [1377]

——— (2008): "Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia," *Econometrica*, 76, 353–394. [1377]

FEHR-DUDA, H., M. DE GENNARO, AND R. SCHUBERT (2006): "Gender, Financial Risk, and Probability Weights," *Theory and Decision*, 60, 283–313. [1401,1408]

FISCHBACHER, U. (2007): "z-Tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*, 10, 171–178. [1379]

GEWEKE, J., AND R. MEESE (1981): "Estimating Regression Models of Finite but Unknown Order," *International Economic Review*, 22, 55–70. [1387]

GOLDSTEIN, W., AND H. EINHORN (1987): "Expression Theory and the Preference Reversal Phenomena," *Psychological Review*, 94, 236–254. [1383]

HALTIWANGER, J. C., AND M. WALDMAN (1985): "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity," *American Economic Review*, 75, 326–340. [1377]

——— (1989): "Limited Rationality and Strategic Complements: The Implications for Macroeconomics," *Quarterly Journal of Economics*, 104, 463–483. [1377]

HARLESS, D. W., AND C. F. CAMERER (1994): "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62, 1251–1290. [1376]

HARRISON, G. W., AND E. E. RUTSTRÖM (2009): "Representative Agents in Lottery Choice Experiments: One Wedding and a Decent Funeral," *Experimental Economics*, 12, 133–158. [1378,1408,1409]

HARRISON, G. W., S. J. HUMPHREY, AND A. VERSCHOOR (2010): "Choice Under Uncertainty: Evidence From Ethiopia, India and Uganda," *The Economic Journal*, 120, 80–104. [1378]

HARRISON, G. W., M. I. LAU, AND E. RUTSTRÖM (2007): "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 109, 341–368. [1375]

HARRISON, G. W., M. I. LAU, E. E. RUTSTRÖM, AND M. B. SULLIVAN (2005): "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues," in *Field Experiments in Econometrics*. Research in Experimental Economics, Vol. 10, ed. by J. Carpenter, G. W. Harrison, and J. A. List. Greenwich, CT: JAI Press, 125–218. [1375]

HARRISON, G. W., J. A. LIST, AND C. TOWE (2007): "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," *Econometrica*, 75, 433–458. [1383,1404]

HEINEMANN, F. (2008): "Measuring Risk Aversion and the Wealth Effect," in *Risk Aversion in Experiments*. Research in Experimental Economics, Vol. 12, ed. by J. Cox and G. W. Harrison. Bingley: Emerald Group Publishing Limited, 293–313. [1403,1404]

HEY, J. D., AND C. ORME (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291–1326. [1375,1383]

HOUSER, D., AND J. WINTER (2004): "How Do Behavioral Assumptions Affect Structural Inference?" *Journal of Business and Economic Statistics*, 22, 64–79. [1376]

HOUSER, D., M. KEANE, AND K. MCCABE (2004): "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm," *Econometrica*, 72, 781–822. [1376]

HSEE, C. K., AND E. U. WEBER (1999): "Cross-National Differences in Risk Preferences and Lay Predictions," *Journal of Behavioral Decision Making*, 12, 165–179. [1408]

KACHELMEIER, S. J., AND M. SHEHATA (1992): "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence From the People's Republic of China," *American Economic Review*, 82, 1120–1141. [1377,1398,1408]

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–292. [1382]

KÖBBERLING, V., AND P. WAKKER (2005): "An Index of Loss Aversion," *Journal of Economic Theory*, 122, 119–131. [1382]

LATTIMORE, P. K., J. R. BAKER, AND A. D. WITTE (1992): "The Influence of Probability on Risky Choice," *Journal of Economic Behavior and Organization*, 17, 377–400. [1375,1383]

MCLACHLAN, G., AND D. PEEL (2000): *Finite Mixture Models*. Wiley Series in Probabilities and Statistics. New York: Wiley. [1388,1389,1392]

PRELEC, D. (1998): "The Probability Weighting Function," *Econometrica*, 66, 497–527. [1383]

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343. [1383]

R DEVELOPMENT CORE TEAM (2006): *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. [1385]

RABIN, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292. [1377]

STAHL, D. O., AND P. W. WILSON (1995): "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–254. [1376]

STARMER, C. (2000): "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk," *Journal of Economic Literature*, 38, 332–382. [1376]

STOTT, H. P. (2006): "Cumulative Prospect Theory's Functional Menagerie," *Journal of Risk and Uncertainty*, 32, 101–130. [1382]

TVERSKY, A., AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323. [1382,1383]

TVERSKY, A., AND P. WAKKER (1995): "Risk Attitudes and Decision Weights," *Econometrica*, 63, 1255–1280. [1383]

WAKKER, P. P. (2008): "Explaining the Characteristics of the Power (CRRA) Utility Family," *Health Economics*, 17, 1329–1344. [1403]

WEDEL, M. (2002): "Concomitant Variables in Finite Mixture Models," *Statistica Neerlandica*, 56, 362–375. [1387]

WILCOX, N. (2006): "Theories of Learning in Games and Heterogeneity Bias," *Econometrica*, 74, 1271–1292. [1377]

——— (2010): "Stochastically More Risk Averse: A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics* (forthcoming). [1383]

WU, G., J. ZHANG, AND R. GONZALEZ (2004): "Decision Under Risk," in *The Blackwell Handbook of Judgment and Decision Making*, ed. by D. Koehler and N. Harvey. Oxford: Oxford University Press, 399–423. [1383]

Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstrasse 10, 8006 Zurich, Switzerland; bruhin@iew.uzh.ch,

Chair of Economics, ETH Zurich, Weinbergstrasse 35, 8092 Zurich, Switzerland; fehr@econ.gess.ethz.ch,

and

Chair of Economics, ETH Zurich, Weinbergstrasse 35, 8092 Zurich, Switzerland; epper@econ.gess.ethz.ch.

Manuscript received May, 2007; final revision received February, 2010.