`

Upside Versus Downside Risk: Gender, Stakes, and Skewness *

Irene Comeig, Charles Holt, and Ainhoa Jaramillo-Gutiérrez

October 16, 2021

Abstract:

Risky choices often involve a tradeoff between expected payoff and payoff variability. Subjects in a simple experiment, however, exhibit more aversion to "downside risk" (with a small probability of a low payoff) and more attraction to "upside risk" (with a small probability of a high payoff). Women tend to be more averse than men for downside risk, but not for upside risk. These patterns are evaluated in terms of the utility curvature and probability weighting components of risk preferences. Gender differences in downside risk are relevant for the design of appropriately gender-tailored policies and algorithms for saving, financing, and entrepreneurship.

Keywords: risk aversion, skewness, payoff scale, probability weighting, rank-dependent utility, gender differences, experiments

JEL Codes: C92, G20

---

* Irene Comeig, Department of Finance, University of Valencia, Campus Tarongers 46022 Valencia, Spain (irene.comeig@uv.es); Charles Holt, Department of Economics, University of Virginia, Charlottesville, VA 22904-4182, USA (cah2k@virginia); Ainhoa Jaramillo, Universitat Jaume I, Castellón de la Plana, Spain, (jaramill@uji.es).
We wish to thank Sam Brott, Courtney Carpenter, Jim Cox, Andre de Palma, Emel Feliz-Ozbay, Glenn Harrison, Maggie Isaacson, Sophia Lochner, Erkut Ozbay, Mike Schreck, Emily Snow, Erica Sprott, Sean Sullivan, Sarah Tulman, Elin Wolf, and Max Zilvitis for comments on an earlier draft. This research was funded in part by the University of Virginia Bankard Fund, a 4-VA grant, the Spanish Ministry of Economy PID2019-110790RB-I00, RTI2018-096927-B-I00, and ECO2016-75575-R projects, the Generalitat Valenciana PROMETEO/2019/095 and AICO/2021/005, the University Jaume I UJI-B2018-76, and the U.S. National Science Foundation (SBR 1459918).

`

I. Introduction

Major financial decisions regarding loans, investments, insurance, or pension plans are typically characterized by a "safer" option with low variance, and a "riskier" option with a higher spread between best and worst outcomes. Preferences over such alternatives may be influenced by a bundle of emotions, motivations, and perceptions, which vary from person to person, and over time for the same person. Despite these fluctuations, previous research has provided important insights about demographic factors, such as gender, that seem to have persistent effects on risk preferences.[1] However, it is difficult to reconcile conclusions of studies of gender differences that examine a wide variety of risk and payoff structures.

In order to improve theoretical insights and models of gender differences in risk preferences, an essential first step is to clarify whether and under what circumstances women are more risk averse than men. It is important to distinguish different types of financial risks. Some risks involve a boom, or "upside risk," with salient payoffs that are above reference points, e.g. speculation in booming housing markets, or the investments of "angels" in startup business projects with high growth potential. Another example involves savings accounts with small chances of a winning a high lottery-based prize (in lieu of interest). This type of upside risk generated broad appeal in a laboratory study (Feliz-Ozbay, et al., 2015). Other risks involve a bust, or downside risk: with salient payoffs that are much lower than a reference point, e.g. accepting a loan with high collateral. The experiment with financial professionals of Cohn et al. (2015) showed that professionals primed with a graphics for a financial asset market boom were substantially less risk averse than those who were primed with a bust. The gender factor was not considered in the analysis of data collected in those experiments.

There is prior work on gender effects for specific settings that can be classified as either downside or upside risk. Harrison and Mason (2007) studied the behavior of investors in the business angel market, an upside risk environment. They found that women did not differ significantly from their male counterparts. However, strong gender differences have been found in a downside risk environment involving loans for projects with specified failure probabilities that would trigger a payment of the collateral amount. Comeig et al. (2022) report significant

[1] See Eckel and Grossman (2008), Harrison and Rutström (2008), Croson and Gneezy (2009), Charness and Gneezy (2012), Charness et al. (2013), and Holt and Laury (2014) for surveys of that cover the literature on gender and risk. Gender differences are prevalent but not uniform across measurement methods (Crosetto and Filippin, 2017).

`

gender differences in an experiment where borrowers (men and women) with a 90% probability of repaying the loan had to choose between (a) low interest payments and high collateral (only paid in case of no loan repayment) or (b) high interest rate and low collateral. Women tended to avoid the high collateral, while men did not, even knowing that the probability of transferring the required collateral was only 10%.

This paper reports results for a carefully selected set of risky decisions in a context-free setting with a range of payoff magnitudes and risk structures. The primary treatment difference is based on whether there is upside risk, a small probability of a good outcome, or downside risk, a small probability of a bad outcome. The objective is to differentiate settings in which gender differences in risk preferences are prevalent from those in which there are no systematic differences. The main experiment is based on a balanced design that varies gender, the nature of the risk ("upside" or "downside"), and the magnitude of the payoff scale (low or high). Subjects only make a single decision, with payoffs in cash. A second experiment involves a within-subjects design in which each person makes 20 decisions using various risk, probability, and payoff scale features, with payoff-relevant decisions selected randomly ex post. Random selection can be problematic as discussed below, but it permits examination of a broader range of decisions in order to disentangle probability-weighting and utility-curvature components of risk attitudes. The results indicate that women are more risk averse then men for downside risk, and that high payoffs tend to amplify risk aversion for both genders. In contrast, there is no gender difference for upside risk, and the same subjects who are risk averse for downside risk tend to exhibit risk preference for upside risk.

The observed gender difference for downside risk is important because standard financial planning algorithms are often based on assumptions of risk tolerance or aversion. These algorithms can be used to specify guidelines and/or constraints for someone requesting a collateralized loan or setting up a retirement plan. Our results also underscore the importance of tailoring the advice to the particular risk structure, upside or downside, since we find no gender difference with upside risk.

The upside and downside risk settings used for the main experiment are explained in the next section, which is followed by a description of experiment procedures in Section III. The results of the main single-choice experiment are presented in Section IV, with a discussion of the effects of risk structure, gender, and payoff scale. Then V summarizes comparable results for a different (within-subjects) design in which each person made multiple decisions, with random selection used to determine the decision to be used for payoffs. Section VI concludes.

II. Upside Versus Downside Risk

One commonly cited experimental result is that women are more risk averse than men. Such differences have been observed in decisions for which the riskier choice (e.g. not purchasing insurance) is characterized by a low probability of a very low payoff. In this case, the safer choice has less variance, but it also has a lower expected money payoff. In effect, the safer option provides less "downside risk." For example, consider the choice between two options, A and B shown in the top row of Table 1. Option A has higher payoff variability, but the expected payoff of $6 for A is $0.80 higher than the expected payoff for Option B. In this case, those who are risk neutral or slightly risk averse would select the riskier option A, and those who are sufficiently risk averse would select B. Prior experiments indicate that women often exhibit more risk aversion in such situations.[2] There is also some evidence that gender differences would be even stronger if the "safe" option B offers a sure payoff instead of a lower-variance pair of payoffs ($6.18 and $3.25 in this case).[3]

Table 1. Decisions, Expected Payoff difference, and Standard Deviation difference

Decision	Option A	Option B	Expected Payoff Difference	Standard Deviation Difference
Downside Risk (skewness[a] = −0.72)	0.67 of $8.38, 0.33 of $1.25	0.67 of $6.18, 0.33 of $3.25	$0.80	1.98
Upside Risk (skewness= +0.72)	0.33 of $9.50, 0.67 of $4.25	0.33 of $5.90, 0.67 of $4.85	$0.80	1.98

[a] The Fisher-Pearson skewness measure is the same for options A and B, and reverses sign for the case of upside risk.

One avenue for explaining gender differences is to specify a utility function for women with more curvature than for men. The amount of curvature required to explain the observed choice frequencies, however, is often quite high. Alternatively, some reluctance to take risks could

[2] Fehr-Duda et al. (2006) find a similar pattern of gender differences in decisions characterized by a small probability of a low payoff.

[3] Crosetto, and Filippin (2017) have noted that the types of gender differences observed in investment task menus (Eckel/Grossman) are not generally observed in BRET (bomb) or structured choice menus (Holt/Laury). They provide evidence that these differences can be explained by the presence of a safe option in these investment tasks.

`

be due to an overweighting of the low probability of a low payoff, which would tend to skew decisions away from the option A that has the extreme low payoff of $1.25. Therefore, it is difficult to distinguish between curvature and probability weighting, using decision problems in which the risky choice involves a low probability of a low payoff, as with the Downside Risk choice in the top row of the table. But if the low probability outcome involves relatively *high* payoffs, then an overweighting of that low probability could imply *preference* for risky choices instead of aversion. For example, consider the Upside Risk decision in the second row of Table 1, for which the expected payoff difference for A over B is again $0.80, but the high-variance Option A offers an "upside risk" of a relatively high payoff of $9.50.

The idea behind the comparison of the two decision problems in Table 1 is that if probability weighting is important, then the one-third probability would tend to be overweighted in each case. This would generate a pattern of risk avoidance with downside risk, but it would result in an attraction to upside risk with a low probability of a high payoff. This observation would be consistent with some prior work and with casual observation.[4] Probability weighting is a key component of prospect theory, and experimental tests have provided some mixed evidence that low probabilities are overweighted.[5]

On an intuitive level, downside risk indicates the extent to which a payoff may fall below an anticipated value, in the same manner that a stock might fall in value, unless a put option to sell at a specified price is used to mitigate risk. Conversely, upside risk indicates the extent to which a payoff might rise above a reference point or normal level. This difference is illustrated in Figure 1, where the riskier option in each panel has a greater payoff spread, as indicated by the dark bars that bracket the light bars for the safer option. Bar heights indicate probabilities, so the riskier option in the left panel offers a 0.33 chance of a low payoff of $1.25, whereas the riskier option on the right offers a 0.33 chance of a high payoff of $9.50, which represents greater upside risk.

[4] Cohen, Jaffray, and Said (1985) used a choice menu with a lottery on the left side, e.g. 1000 FF. if a "Diamond" is drawn from a deck of cards, and various sure money amounts on the right side ranging from 0 to 1000 F.F. The crossover "price" in such a menu determines a certainty equivalent. Risk aversion is indicated if the certainty equivalent is less than the expected payoff for the lottery. Payoffs were either positive or 0 for all choice menus in a "gain domain," with the probability of a gain being changed from one menu to another. Subjects switched from being risk averse to risk seeking as the probability of a high payoff decreased. Gender differences were not discussed.

[5] Harbaugh et al. (2010) summarize this literature, noting that tests supporting the predictions of prospect theory often involve using hypothetical payoffs as in Kahneman and Tversky (1979), or giving subjects hundreds of decision problems with one to be selected at random, as in Hey and Orme (1994). In contrast, Harbaugh et al. (2010) find no support for the notion that subjects are more risk seeking for small probability gains, a finding that is roughly consistent with their 2002 paper that considered behavior of both children and adults.

The risky option (dark bars) in each panel of Figure 1 has a higher standard deviation, and the difference in standard deviations (1.98) between risky and safe options is the same in each panel. Also, each option in Figure 1 has a skewed distribution, with negative skewness (a "tail" to the left) in the downside risk panel and with positive skewness (a "tail" to the right) in the upside risk panel. In fact, the standard Fisher-Pearson measure of skewness is negative (−0.72) for both safer and riskier options in the downside risk, and is positive (0.72) for both safer and riskier options in the upside risk.[6] This figure suggests that any tendency to overweight low probabilities might produce risk aversion for downside risk and risk seeking for upside risk.

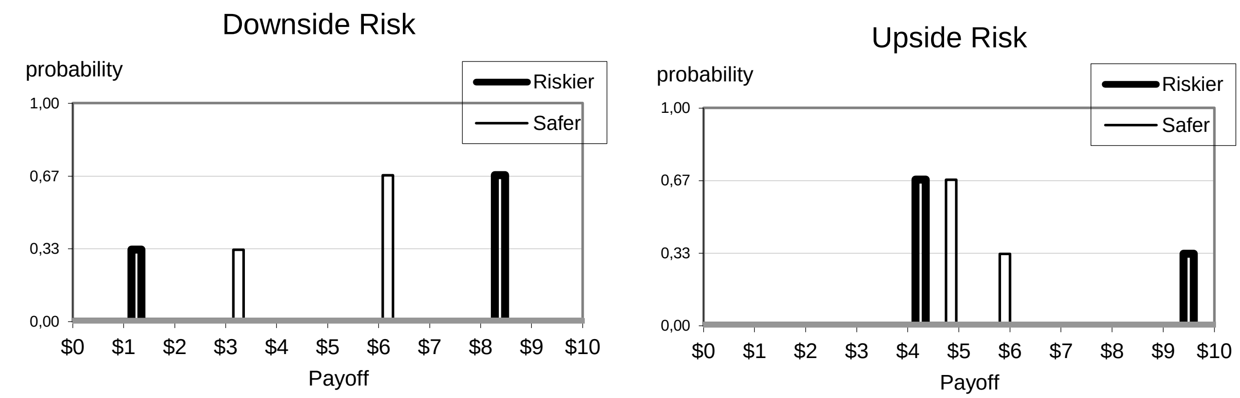

Figure 1. Downside Risk, Negative Skewness (Left), Upside Risk, Positive Skewness (Right)
Note: Each panel provides exactly the same expected payoff and standard deviation difference between the safe and risky choices. The equal negative skewness measures for both choices on the left exactly match the absolute value of the equal positive skewness for both choices on the right.

In contrast, the classic Friedman and Savage (1948) explanation for the simultaneous purchase of insurance and lottery tickets is based on risk seeking (convex utility) for very high payoffs. In other words, it is possible that any risk seeking observed in the upside risk setting might be due to utility curvature rather than to a tendency to overweight the low probability of a very high payoff with upside risk. Therefore, we decided to include a set of decisions for which all payoffs (for both upside and downside risk treatments) were scaled up by a factor of 5. Since probabilities are unaffected by such payoff scaling, any observed behavioral changes would have to be due to curvature aspects of utility instead of nonlinear probability perceptions. So the idea behind a switch from downside to upside risk with low payoff scales is to isolate effects of probability weighting that cannot be explained by utility curvature. Then a scaling up of all payoffs

[6] The formula for this skewness measure is $\gamma = \frac{E(X-\mu)^3}{\sigma^3}$, where μ is the expected payoff, and σ is the standard deviation.

`

by a factor of 5 in a second treatment variation (holding probabilities fixed) is used to isolate the effect of utility curvature that cannot be explained by probability perceptions.

III. Procedures for the Single-Choice Experiment

The 2x2x2 experimental design involved 256 subjects making a single decision, with variations in gender (half men), payoff scale (half with 5X payoffs), and risk skewness (half with downside risk and half with upside risk). The two decisions used are shown in Table 1, which provides expected payoff, dispersion, and skewness measures. Note that both decisions have been carefully chosen to have the same expected payoff difference between options ($0.80) and the same difference in standard deviation (1.98), so that any differences in risk taking behavior cannot be attributed to dispersion or expected payoff differences. On the other hand, the 5X scale would raise the low payoff in the Downside Risk decision from $1.25 to $6.25, and would raise the high payoff in the Upside Risk decision from $9.50 to $47.50. The 2x2x2 design has 8 "bins" with 32 subjects in each, for a total of 256 subjects (128 men and 128 women).[7]

The experiment was run with web-based software, using instructions presented in the online appendix.[8] The instructions are the same for all treatments, with common examples used "for illustrative purposes only." The results page had a link to a demographic survey that included age, gender, major, etc. Subjects were recruited from the University of Virginia. The experiment lasted 30-45 minutes, and earnings ranged from $10-$54, including a $6 show-up payment.

IV. Single-Choice Results

Each subject made a choice between safe and risky options shown in one of the rows of Table 1, with differences in expected payoffs and variability being the same for each row. The key difference between these settings is that the sign of the skewness measure has switched from

---

[7] The use of 0.67 and 0.33 probabilities in the experiment instead of 2/3 and 1/3 caused some of the expected payoff differences to be slightly different from the difference of $0.80 for the other decisions. This change was made because of the way that probabilities were explained to subjects in terms of "chances in 100" instead in terms of probabilities. The truncated 0.67 and 0.33 probabilities that were used in the experiment are used for the econometric estimation in Appendix A. Since we wanted subjects to be fully aware of payoff scale, the payoffs were labeled in dollars and cents, instead of using higher numbers of points that could have been converted to cash subsequently.

[8] The Pairwise Lottery Choice program that was used can be found on the Decisions menu of the V*e*conlab site: http://veconlab.econ.virginia.edu/admin.php This web-based software is written and maintained by one of the coauthors (Holt) and is freely available for instructional and research use.

`

being negative for each option in the downside case (a "tail" to the left) to being positive for each option in the upside case (a "tail" to the right).

Gender Effects

The choice counts and percentages shown in Table 2 support several insights about choice patterns in terms of gender, payoff scale, and risk structure. First, consider the top row (Downside, 1X scale). The 32 men chose the risky option 25 times, whereas the women only chose risky 19 times. Thus the risky choice percentage was 78% for men and 59% for women, for a treatment difference of 19 percentage points.

Table 2. Choices by Treatment for the Between Subjects (Single Choice) Experiment

		Men Choosing Risky (A)	Women Choosing Risky (A)
Downside 1X *	32 men, 32 women	25 78%	19 59%
Downside 5X ***	32 men, 32 women	17 53%	4 13%

		Men Choosing Risky (A)	Women Choosing Risky (A)
Upside 1X	32 men, 32 women	30 94%	29 91%
Upside 5X	32 men, 32 women	28 88%	27 84%

*, *** Significant gender differences at 10% (1-tail) and 1% (2 tails) respectively, using an Exact Probability Test.

Gender Effects: *Women are more risk averse than men in the downside risk treatment, especially with high (5X) payoff stakes, but there is no gender difference in the upside risk treatment, irrespective of payoff scale.*

Support: A Fisher permutation test is based on the idea that there is no gender effect under the null hypothesis, so permutations in gender labels should not matter. In other words, each possible assignment of the gender labels to observed decisions would be equally likely ex ante. The observation of an extreme outcome (e.g. men making most of the observed risky choices) suggests that the null can be rejected. For each payoff scale, 1X or 5X, the test statistic can be constructed by randomly permuting the gender labels assigned to each of the 64 observed A or B decisions, keeping equal numbers of male and female labels. The p value for such a test is the proportion of such permutations that yield a treatment difference that is greater than or equal to the observed difference in choice proportions (19% for 1X scale, or 41% for 5X scale). Since there are "64 take

`

32" ways that gender labels can be permuted (a number in the thousands of trillions), we used 100,000 simulations in which the gender labels (32 men and 32 women) are randomly reassigned to decisions in each simulation. For the 1X scale, the *p* value (proportion of simulations that yield a treatment difference exceeding 19 percent) is only marginally significant, even if prior expectations are used to justify a 1-tailed test ($p = 0.09$). With high (5X) payoff stakes, which the *p* value is 0.0012 for a 2-tailed test.[9] One way to avoid the dangers of running multiple tests is to use a *joint* test of the gender effect for downside risk. Such a joint test can be constructed by *simultaneously* permuting the gender labels in each of the top two rows of Table 2, while holding the total number of risky decisions constant in each row. This *stratified permutation test* controls for payoff-scale effects by permuting the 64 gender labels *separately* for each payoff-scale row or "strata."[10] This joint test provides strong evidence that women are more risk averse in the downside risk treatment, with a *p* value of 0.0002 (128 observations, 2-tailed test, 2 strata). In contrast, there is no significant gender difference with upside risk, as would be expected from a review of the bottom two rows of Table 2. The main insight here is that gender differences in risk taking are context specific.

Risk Structure and Skewness Effects

The most salient feature of the results in Table 2 is that subjects are much more willing to take the high-variance upside risks than is the case for the equivalent high-variance downside risks. This risk effect can be seen by comparing the proportions of risky choices for each combination of gender and payoff stakes, as shown in Figure 2.

Downside versus Upside Risk: *Subjects are more likely to select the risky option when it involves the upside risk than when it involves the downside risk, despite the fact that the expected payoff and standard deviation differences between the safe and risky options are the same in each case.*

[9] More precisely, 61 of 100,000 simulations yielded a treatment difference greater than +40.06, and 121 of 100,000 simulations yielded a treatment difference greater than 40.06 in absolute value ($p = 0.0012$).

[10] The Fisher permutation test is equivalent to a stratified permutation test with a single cluster or "strata." A stratified permutation test (with 2 or more strata) allows one to control for "nuisance variables," in this case payoff scale. This test is discussed in more detail in Holt and Sullivan (2021), and examples from laboratory experiments are provided in the methodology chapter 13 of Holt (2019).

`

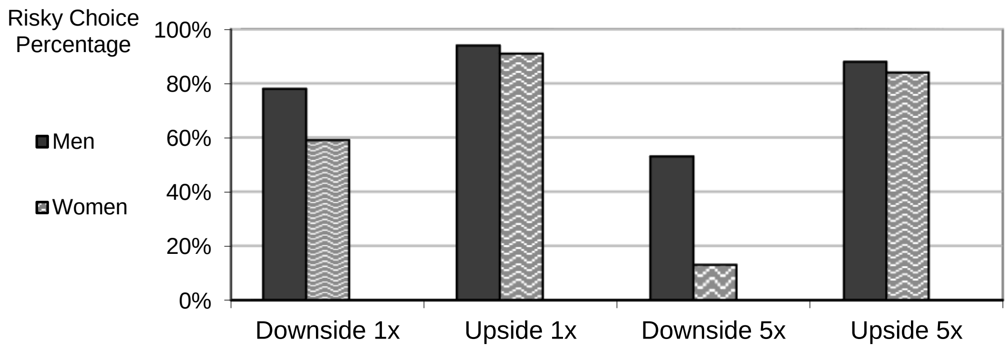

Figure 2. Single Decision Data: Men Versus Women Percentage of Risky Choices by Risk Type (Upside or Downside) Payoff Scale (1x or 5x)

Support: The proportion of upside risky choices is higher for all four categories (men/1X, men/5X, women/1X, and women/5X). Under the null hypothesis that the type of risk (upside or downside) has no effect, the risk labels do not matter and could be randomly switched for each person. The stratified permutation test involves randomly switching the risk label for each of the 256 choices of A or B, but keeping permutations separate for men and women and for 1X scale and 5X scale. Thus the test "holds constant" both gender and payoff scale dimensions in order to calculate the proportion of risky decisions for each risk treatment for each permutation. Then the *p* value is determined from the tails of the distribution of the difference between the proportion of risky choices under upside risk and that under downside risk. This *p* value is the proportion of differences (from random reassignments of upside/downside labels) that are at least as extreme as the difference observed in Table 2. As before, we use 100,000 simulations to determine the proportions of "as or more extreme" treatment differences, and the resulting *p* value is less than 0.001 (256 observations, 2-tailed test, with 4 data clusters based on gender and payoff scale).

Payoff Scale

The payoff scale treatment can be evaluated by comparing the heights of the risky choice bars on the right (5X) side of Figure 2 with the corresponding bars on the left (1X):

Payoff Scale: *An increase in payoff scale increases risk aversion, especially with downside risk.*

Support: Restricting attention to downside risk, the null of no scale effect can be rejected with p = 0.006 (2-tailed test, 128 subjects for downside risk, stratified by gender). Running the same test

`

with all 256 subjects and stratifying by gender and risk structure yields the same *p* value. Similarly, it can be verified that the effect of payoff scale is not significant for the upside risk data alone.

Other Possible Explanations

An alternative perspective might be based on threshold effects, which imply that risky choice is influenced by target or desired payoff levels. For example, consider the Upside Risk decision in the bottom row of Table 1, where all payoffs for both exceed $4, which participants might view as meeting earnings expectations, which might cause subjects to be attracted to the risky option with the highest payoff. In contrast, with downside risk, the low payoff of $1.25 for the risky option might not be perceived as meeting expectations. In this manner, threshold effects could explain the tendency for subjects to take more risk in the upside risk setting. But if the payoffs for downside risk in the top row were scaled up by a factor of 5, then the lowest payoff of $6.25 would exceed the $4 threshold mentioned above. Thus the same threshold effect that could explain an aversion to downside risk and attraction to upside risk with low stakes might not explain a similar pattern with scaled-up payoffs. Of course, thresholds could be context specific, a possibility that is not addressed with the experiment in this paper.

An alternative approach would be to consider a different theoretical framework, such as range-dependent utility, in which the utility is normalized to fit into the range (from low to high) of lottery outcomes. Kontek and Lewandowski (2018) show that this approach can explain perplexing data patterns in laboratory experiments, without resorting to probability weighting. We conjecture, however, that range-dependent would not provide an explanation for the reduced tendency to make risky choices in our high-stakes downside risk treatment. A scaling up of all payoffs will also scale up the range from the highest to the lowest payoff by 5X, and it will scale up the certainty equivalent that is used to assess utility within this range. As a result, the location of the certainty equivalent within the *normalized* payoff range will not be affected by scaling up the payoffs, which is inconsistent with one of the most salient results of our experiment, the payoff scale effect with downside risk.[11]

[11] Another alternative is to model the decision process in stages. For example, Bayrak and Hey (2020) assume that a person first identifies best and worst utility estimates for a lottery and then uses a weighted average to evaluate the lottery, with a skewness-dependent weight. But scaling up payoffs by a constant (5X) will not alter skewness, so we are unsure about whether this approach will explain the payoff scale effects that we observe in our high-stakes downside risk treatment.

`

Finally, we consider whether the treatment effects are due to secondary differences, e.g. a tendency for men to choose technical majors that predispose them to select A on the basis of expected value calculations. Equation (1) shows marginal values for a logistic regression with independent variables for gender (man = 1), risk (upside = 1), payoff scale (5X = 1) and dummy variables for categories of academic major (science, business, economics, etc.). The dependent variable is 1 if the risky option was chosen, and standard errors are shown in parentheses.

$$\text{(1)}\quad \begin{array}{l} \text{Logistic Regression for} \\ \text{Risky Choice Probability} \end{array} \quad \underset{(0.06)}{0.18}\ \text{Men} + \underset{(0.06)}{0.42}\ \text{Upside Risk} - \underset{(0.06)}{0.24}\ \text{Scale}$$

Notably, none of the coefficients of the dummy variables for various majors (not shown) were significant. All treatment variables (gender, risk type, and payoff scale) are significant at the 1 percent level. The next task is to consider structural estimates of probability weighting and utility curvature that shed light on factors that are driving observed patterns of risk taking and avoidance.

V. A Within-Subjects Experiment with Random Selection

Our results for upside risk are consistent with those of Eckel and Grossman (2015), who document the tendency for subjects to "love the long shot" of a high payoff. They offer a conjecture and some indirect evidence that this behavior is due to nonlinear probability weighting, which typically results in overweighting low probabilities and underweighting high probabilities. A similarly suggestive result is reported Filiz-Ozbay et al (2015), who show that laboratory subjects are more likely to save by deferring income receipts when those receipts are linked to random lotteries. This effect is strongest for men and self-reported lottery players, and the authors' explanation is based on estimates of a nonlinear probability weighting parameter.[12]

In order to evaluate probability weighting in our design, we ran a within-subjects experiment with a range of payoffs and probabilities that included 0.1, 0.33 0.5, 0.67, and 0.9. Subjects made 10 decisions with low (1X) scale and another ten decisions with a 5X scale. The decisions included four paired options with upside risk and four more with downside risk, all with an expected payoff difference of 80 for the risky option. These decisions are listed in Table A1 in the Appendix A, with decisions D1 and D3 in that table corresponding to the Downside and Upside Risk decisions in Table 1 (and shown in Figure 1). There are two decisions in Table A1 with

[12] Ebert (2015) does not consider probability weighting, but does report that subjects prefer lotteries with right-skewness to those with left skewness.

moderate (0.67, 0.33) downside risk, two with moderate (0.67, 0.33) upside risk, two with extreme (0.90, 0.10) downside risk, and two with extreme (0.90, 0.10) upside risk. The other two decisions had 0.5 probabilities, one taken from a basic Holt and Laury (2002) menu.

The range of treatment structures motivated the use of a within-subjects design. Each subject made all 10 decisions with low stakes and 10 decisions with high stakes, with the low stakes encountered first in half of the sessions. The order in which decisions were presented for each scale was randomized for each subject. The objective was to generate enough data to estimate probability weighting and stakes-sensitive risk aversion parameters, and to determine whether the *same* subjects exhibit risk aversion with downside risk and risk seeking with upside risk.

Experiment Results with Random Selection

This experiment involved 32 men and 32 women, as was the case for each of the rows in Table 2. Each person was paid for one randomly selected low scale (1X) decision and for one randomly selected (5X) decision. The within-subjects results are similar to those obtained in the single-choice experiment. In particular, an increase in payoff scale reduces the tendency to choose the risky option with downside risk, and subjects select the riskier choice about twice as often with upside risk than with downside risk, for both payoff scales and for risks that are moderate (0.33 probability) or extreme (0.10 probability). This pattern in the aggregate data also shows up clearly in Table 3 for the two decisions in (1) and (2) that used previously in the single-choice experiment:

Table 3. Choice Percentages for Selected Decisions in the Within-Subjects Experiment[a]

		Men Choosing Risky (A)		Women Choosing Risky (A)	
D1: Downside 1X	32 men, 32 women	19	59%	13	41%
D1: Downside 5X	32 men, 32 women	9	28%	9	28%

		Men Choosing Risky (A)		Women Choosing Risky (A)	
D3: Upside 1X	32 men, 32 women	30	94%	31	97%
D3: Upside 5X	32 men, 32 women	30	94%	28	88%

[a] The results are for the two decisions D1 and D3 shown in the Appendix table A1.

`

` The overall pattern for the two upside and downside risk decisions provided in Table 3 is similar to the single-choice data in Table 2 for the same decisions.[13] In addition, random selection data for experiment 2 permits a within-subjects comparison based on a wider range of decisions:

Downside versus Upside Risk (Between Subjects): *The same subjects who tend to select the safer option when faced with the downside risk also tend to select the riskier option when faced with the upside risk, despite the fact that the expected payoff differences between the options are identical.*

Support: Each subject made four choices with downside risk (decisions 1, 2, 5, and 6 in Table A1) and we summed the number of riskier choices to obtain individual measures of downside risk taking. Similarly, each subject made four choices with upside risk (decisions 3, 4, 7, and 8), and we constructed a measure of upside risk taking in the same manner. Most subjects (51 out of 64) chose the risky option more often with upside risk. This result is highly significant ($p < 0.001$) for both the low payoff scale and the high payoff scale using a Wilcoxon matched-pairs test.

The justification often given for using random selection is that subjects tend to "isolate" attention on the current choice and make a decision that is unaffected by the range of other possible decisions that might be used to determine payoffs. Random selection is the most common procedure for controlling for wealth effects. Although it has been defended by Starmer and Sugden (1991), Hey and Lee (2005), and others, some researchers have suggested caution (e.g. Holt, 1986, Davis and Holt, 1993, and more forcefully Cox et al. 2015 and Brown and Healy, 2018). Random selection may not work properly if subjects fail to view each decision in isolation.

Specifically, Harrison and Swarthout (2012) find that experiments with a single decision provide different econometric estimates of probability weighting and risk preference parameters than are obtained by using the random selection with multiple decisions per subject. Brown and Healy (2018) also report a significant difference between a single choice and random selection, when the decision is embedded in a structured table of options. For their single choice treatment, the other options in the table are visible but not available. The options used in our second

[13] It is important to note that both men and women were consistent with their choices across stake levels for the upside risk. With upside risk, all 28 women who chose option A with 1X stakes also chose option A in the 5X treatment. For men with upside risk, there were 30 who chose option A with 1X, and 28 of those also chose A with 5X. However, subjects reduced consistency across stake levels for the downside risk treatments. Of the 13 women who chose A for 1X stakes with downside risk, only 5 of these keep the A choice with 5X. There were 19 men chose A with 1X stakes with downside risk, only 6 of these keep the A choice with 5X stakes.

`

experiment were *not* embedded in a table, but rather were presented in random order, which is the procedure recommended by Brown and Healy.

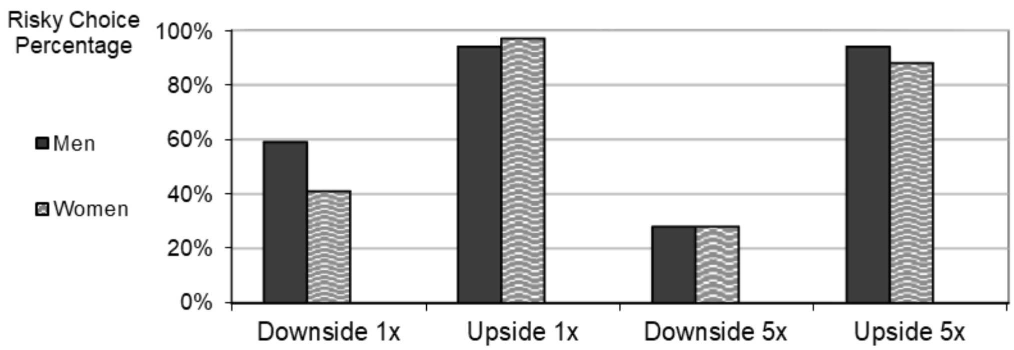

Figure 3. Multiple Decision Data with Random Selection:
Men Versus Women Percentage of Risky Choices

The patterns shown in Table 3 and Figure 3 with random selection are similar to those in Table 2 and Figure 2 for single decisions. In addition to the strong increase in risk taking with upside risk documented above, there is evidence that women are more risk averse than men, with downside risk only, and there are strong payoff-scale effects. With downside risk D1, for example, a scale effect is indicated by the lower numbers of risky choices in the second row of Table 3 than in the top row. A permutation test of payoff scale (stratified by gender) supports a significant effect (p =0.006, 128 observations, 2-tailed test). Despite similarities based on comparisons of patterns within a table, a cross-comparison of the top rows of Tables 2 with those in Table 3 suggests that men and women take more downside risk with 1X scale in the single-choice setting, as compared with the random selection setting. This observation is supported by a permutation test p = 0.043 (128 observations, stratified by gender, 2-tailed test). Therefore, the comparison of the results for single choice and random selection provides a mixed message: both procedures produce the same overall patterns of treatment effects for payoff scale and risk structure, but there is more risk taking with random selection in one of four dimensions (downside risk, 1X scale).

Estimation of Risk Aversion and Probability Weighting Parameters

Recall that the decision problems were structured so that the expected payoff advantage in favor of the riskier option is \$0.80 for the low payoff scale in all cases (with the exception of decision 10 in the within-subjects experiment, which was taken from a standard risk assessment

`

choice menu to serve as a benchmark). Instead of using expected payoff maximization as the basis for the design, we could have used the parametric model, e.g. expected utility with constant relative risk aversion (CRRA). But observed payoff scale effects cannot be explained by with constant relative risk aversion, which is insensitive to payoff scale. The estimation reported in the Appendix provides a more general "expo-power" utility function, $U(x) = \frac{1- \exp(-ax^{1-r})}{a}$, for $r \neq 1$, which exhibits increasing relative risk aversion (more aversion at high payoff scales). In addition, we use a nonlinear probability weighting function, $w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}$ *for* $\gamma > 0$. This function, popularized by Kahneman and Tversky (1979), reduces to $w(p) = p$ with no weighting as the parameter γ goes to 1. This function will overweight low probabilities when $0 < \gamma < 1$. A number of prior studies have reported probability weighting parameters estimates between 0.5 and 1.0, with an estimate of 0.7 being typical.

To summarize, we used a two-parameter expo-power utility function to allow for payoff scale effects, and a nonlinear probability weighting function to permit overweighting of low probabilities (of a low payoff with downside risk and of a high payoff with upside risk). The maximum likelihood estimates (and robust standard errors) of the weighting parameters are 0.66 (0.04) for men and 0.67 (0.03) for women. Both weighting parameters are significantly different from 1, which implies overweighting of low probabilities. The risk aversion parameter estimates are significantly different from zero, suggesting risk aversion for both genders, with slightly more risk aversion for women. Details are provided in the appendix, where the parameter estimates are used to predict choice proportions that follow the general patterns observed in the data in Table 2: more risk seeking with upside risk, clear payoff scale effects, and modest gender effects with downside risk.

VI. Conclusion

Gender differences documented in previous studies turn out to be present in our downside risk setting, and are amplified with high stakes. Moreover, these gender differences are greatly diminished in our upside risk treatments. This distinction is important because it helps us understand gender differences found in the literature. For example, the previous work was often motivated by insurance or bankruptcy, which tends to generate a downside risk (low probability

`

of a low payoff). The main insight is that any analysis of risk should distinguish between upside and downside risk structures, taking gender into consideration in the downside risk environments:

1) Men and women tend to be more risk averse for downside risk than for upside risk. The same people choose the riskier option about twice as often when it involves upside risk instead of downside risk, even though the expected payoff advantage of the riskier option is the same for all decision pairs.
2) Men exhibit less risk aversion than women with downside risk, but there is no significant gender difference for upside risk, irrespective of payoff scale.

Laboratory experiments are well suited for deconstructing components of complex decision theories, since key aspects of incentives and probabilities can be varied independently. The second experiment was structured to evaluate the utility curvature and probability weighting components of risk aversion, using a within-subjects design to reproduce risk aversion tendencies and gender effects for the downside risk setting. Notably the same subjects tended to choose the riskier option much more frequently in the upside risk setting. Maximum likelihood estimates reveal a significant and similar curvature in the probability weighting function, but with greater utility curvature for women. Finally, a scaling up of payoffs resulted in more risk aversion, especially with downside risk. The estimated probabilistic choice model (with probability weighting and utility curvature) explains salient features of the data in terms of gender, scale, and skewness effects. Another possible extension would be to convert the safe options for both downside and upside risk decisions into sure payoffs, to determine whether certainty effects might alter risk-taking behavior and gender effects. Indeed, it would be surprising if strong behavioral attractors due to loss aversion or certainty bias would not mitigate the effects we see in some contexts. Another extension (in progress) involves using different subject pools found in Amazon Turk to evaluate behavior over a wider range of cultural and demographic factors.

The distinction between upside and downside risk is important in understanding gender differences in risk preferences in various domains, e.g. financing with collateral, purchases of insurance and lottery tickets, or saving for security and enrollment in prize-linked savings accounts. In models of "directed search," for example, a worker who decides to concentrate efforts on obtaining a more highly paid position must keep in mind the possibility that many others may also direct searches to that position. Here there is an element of upside risk associated with the

`

higher wage. But there is also a downside risk associated with the chance of not getting hired at all, because the congestion associated with excess applications for the high paying job can result in an unsuccessful search yielding a very low payoff. Similarly, in the decision of becoming an entrepreneur, there is a downside risk due to the small probability of bankruptcy, and the effect might be for more women to avoid it and take the safe route of not becoming an entrepreneur, an outcome with a glass ceiling flavor.

`

References

Bayrak, O. and J. D. Hey (2020) "Decisions Under Risk, Dispersion and Skewness," *Journal of Risk and Uncertainty*, 61(1), 1-24.

Binswanger, H. P. (1981) "Attitudes toward Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal,* 91, 867-890.

Brown, A. L. and P. J. Healy (2018) "Separated Choices," *European Economic Review*, 101, 20-34.

Charness, G. and U. Gneezy (2012) "Strong Evidence for Gender Differences in Risk Taking," *Journal of Economic Behavior and Organization*, 83(1), 50-58.

Charness, G., U. Gneezy, and A. Imas (2013) "Experimental Methods: Eliciting Risk Preferences," *Journal of Economic Behavior and Organization*, 87, 43-51.

Cohen, M., J.Y. Jaffrey, and T. Said (1985) "Individual Behavior Under Risk and Under Uncertainty: An Experimental Study," *Theory and Decision*, 18(2), 203-228.

Cohn, A., J. Engelmann, E. Fehr, and M. A. Maréchal (2015) "Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals," *American Economic Review*, 105 (2), 860-85.

Comeig, I., A. Jaramillo-Gutierrez, and F. Ramirez (2022) "Are Credit-Screening Contracts Designed for Men?" *Service Business*. https://doi.org/10.1007/s11628-022-00485-w

Crosetto, P. and A. Filippin (2017) "Safe Options Induce Gender Differences in Risk Attitudes," IZA Institute of Labor Economics Discussion Paper 10793.

Croson, R. and U. Gneezy (2009) "Gender Differences in Preferences," *Journal of Economic Literature*, 47 (2), 448-474.

Cox, J. C., V. Sadiraj, and U. Schmidt (2015) "Paradoxes and Mechanisms for Choice under Risk," *Experimental Economics*, 18(2), 215–250.

Davis, D. D. and C. A. Holt (1993) *Experimental Economics*, Princeton: Princeton University Press.

Ebert, S. (2015) "On Skewed Risks in Economic Models and Experiments," *Journal of Economic Behavior and Organization*, 112, 88-97.

Eckel, C. C., and P. J. Grossman (2008) "Men, Women, and Risk Aversion: Summary and Analysis," in C. Plott and V. Smith, *Handbook of Experimental Economic Results, Vol. 1*, 1061-1073.

`

Eckel, C. C., and P. J. Grossman (2015) "Loving the Long Shot," *Journal of Risk and Uncertainty*, 51, 195-217.

Fehr-Duda, H., M. de Gennaro, and R. Schubert (2006) "Gender, Financial Risk, and Probability Weights," *Theory and Decision*, 60(2-3), 283-313.

Fehr, H., T. Epper, A. Bruhin, and R. Schubert (2011), "Risk and Rationality: The Effects of Mood and Decision Rules on Probability Weighting," *Journal of Economic Behavior and Organization*, 78(1), 14-24.

Filiz-Ozbay, E.; J. Guryan, K. Hyndman, M. Kearney, and E. Y. Ozbay (2015) "Do Lottery Payments Induce Savings Behavior: Evidence from the Lab," *Journal of Public Economics,* 126, 1-24.

Friedman, M. and L. Savage (1948) "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56(August), 279-304.

Handa, J. (1977) "Risk, Probabilities, and a New Theory of Cardinal Utility," *Journal of Political Economy,* 85, 97-122.

Harbaugh, W. T., K. Krause, and L. Vesterlund (2002) "Risk Attitudes of Children and Adults: Choices Over Small and Large Probability Gains and Losses," *Experimental Economics*, 5, 53-84.

Harbaugh, W. T., K. Krause, and L. Vesterlund (2010) "The Fourfold Pattern of Risk Attitudes in Choice and Pricing Tasks," *Economic Journal*, 120, 595-611.

Harrison, G. W. and E. Rutström (2008) "Risk Aversion in the Laboratory," in J. Cox and G. Harrison, eds., in *Research in Experimental Economics*, *12*, Emerald Group Publishing, Bingley: UK., 41-196.

Harrison, G. W. and J.T. Swarthout (2012) "The Independence Axiom and the Bipolar Behaviorist," CEAR Working Paper.

Harrison, R. T. and C. M. Mason (2007) "Does Gender Matter? Women Business Angels and the Supply of Entrepreneurial Finance," *Entrepreneurship Theory and Practice*, 31(3), 445-472.

Hey, J. D. and C. Orme, (1994) "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica,* 62, 1291-1326.

Hey, J. D. and J. Lee (2005) "Do Subjects Separate (or Are They Sophisticated?", *Experimental Economics,* 8, 233-265.

`

Holt, C. A. (1986) “Preference Reversals and the Independence Axiom,” *American Economic Review*, 76(3), 508-515.

Holt, C. A. (2019) *Markets, Games and Strategic Behavior: A First Course in Experimental Economics*, Princeton NJ: Princeton University Press.

Holt, C. A., and S. K. Laury (2002) "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), 1644-1655.

Holt, C. A. and S. K. Laury (2014) “Assessment and Estimation of Risk Preferences,” *Handbook of the Economics of Risk and Uncertainty, Vol. 1*, M. Machina and K. Viscusi, eds., Oxford: North Holland, 2014, Chapter 4, pp. 135-201.

Holt, C.A. and S.P. Sullivan (2021) “Permutation Tests for Experimental Data,” working paper, University of Iowa Law School.

Kahneman, D., and A. Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263-291.

Kontek, Krzysztof and Michal Lewandowski (2018) “Range Dependent Utility,” *Management Science*, 64(6), 2473-2972.

Laury, S. K., M. M. McInnes, and J. T. Swarthout (2009) “Insurance Decisions for Low-Probability Losses,” *Journal of Risk and Uncertainty*, 39, 17-44.

Laury, S. K. (2012) “Pay One or Pay All: Random Selection of One Choice for Payment,” Discussion Paper, Andrew Young School, Georgia State University.

Quiggin, J. (1982) “A Theory of Anticipated Utility,” *Journal of Economic Behavior and Organization*, 3(4), 323-343.

Saha, A. (1993) “Expo-Power Utility: A ‘Flexible’ Form for Absolute and Relative Risk Aversion,” *American Journal of Agricultural Economics*, 75, 905-913.

Starmer, C. and R. Sugden (1991) “Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation,” *American Economic Review*, 81, 971-997.

`

Online Appendix A: Parameter Estimation

The ten decisions used in the second, within-subjects experiment are shown in Table A1. The expected payoff for the risky option A is 80 cents higher for decisions, D1 – D9. Decision D10 was taken from a standard Holt and Laury menu as a reference point.

Table A1. Payoff Structure for Riskier and Safer Options (1x Payoffs)

Decision	Option A	Option B
D1: Moderate Downside Risk	0.67 of $**8.38**, 0.33 of $1.25	0.67 of $**6.18**, 0.33 of $3.25
D2: Moderate Downside Risk	0.67 of $**8.13**, 0.33 of $1.75	0.67 of $**5.93**, 0.33 of $3.75
D3: Moderate Upside Risk:	0.33 of $**9.50**, 0.67 of $4.25	0.33 of $**5.90**, 0.67 of $4.85
D4: Moderate Upside Risk	0.33 of $**8.55**, 0.67 of $4.73	0.33 of $**5.75**, 0.67 of $4.93
D5: Extreme Downside Risk	0.9 of $**6.64**, 0.1 of $0.25	0.9 of $**5.47**, 0.1 of $2.75
D6: Extreme Downside Risk	0.9 of $**6.58**, 0.1 of $0.75	0.9 of $**5.53**, 0.1 of $2.25
D7: Extreme Upside Risk:	0.1 of $**25.00**, 0.9 of $3.89	0.1 of $**6.00**, 0.9 of $5.11
D8: Extreme Upside Risk	0.1 of $**21.00**, 0.9 of $4.33	0.1 of $**6.80**, 0.9 of $5.02
D9: Balanced Risk:	0.5 of $**8.25**, 0.5 of $3.75	0.5 of $**5.60**, 0.5 of $4.80
D10: Balanced Risk	0.5 of $**7.70**, 0.5 of $0.20	0.5 of $**4.00**, 0.5 of $3.20

Table A2 shows the proportions of risky choices for each of the 10 decisions, D1 – D10, organized by gender and payoff scale. The general pattern of results matches the pattern shown in Table 3 above for decisions D1 and D3, with more risk seeking for upside risk, and with gender differences and payoff scale effects for downside risk but not for upside risk.

Next, we evaluate the extent to which these rich data patterns can be explained by a unified theoretical perspective. To do this, we estimate a rank-dependent expected utility model, using a standard probability weighting function and a two-parameter utility function that is well adapted for examining effects of payoff scale. The third component is a probabilistic choice function, which generates choice probability predictions that are used in the estimation. The two-parameter "power-expo" utility function, $U(x) = \dfrac{1 - \exp(-ax^{1-r})}{a}$, reduces to constant absolute risk aversion in the limit as $r \to 0$ and to constant relative risk aversion as $\alpha \to 0$. When both parameters are

`

positive, the function exhibits increasing relative risk aversion (more risk aversion with scaled-up payoffs) and decreasing absolute risk aversion (less risk aversion as an additive constant is added to all payoffs).[14]

Table A2: Proportions of Risky Choices with Random Selection[15]

Decision Number	**Structure (Chances for High Payoff)**	**1x Payoffs Men**	**1x Payoffs Women**	**5x Payoffs Men**	**5x Payoffs Women**
D1	downside risk (33)	0.59	0.41	0.28	0.28
D2	downside risk (33)	0.59	0.41	0.38	0.44
D3	upside risk (33)	0.94	0.97	0.94	0.88
D4	upside risk (33)	0.91	0.88	0.97	0.91
D5	downside risk (10)	0.44	0.34	0.38	0.19
D6	downside risk (10)	0.44	0.31	0.38	0.34
D7	upside risk (10)	0.78	0.75	0.81	0.72
D8	upside risk (10)	0.94	0.75	0.84	0.78
D9	balanced risk (50)	0.81	0.75	0.69	0.78
D10	balanced risk (50)	0.19	0.09	0.09	0.06
	All Problems	**0.66**	**0.57**	**0.61**	**0.54**

The Kahneman and Tversky probability weighting function was applied to the inverse cumulative distribution in order to avoid violations of stochastic dominance.[16] With only two payoffs in each choice, with $L \leq H$, this means applying the weight to the probability of the high

[14] This function was proposed by Saha (1993). Holt and Laury (2002) provide maximum likelihood estimates of the two parameters of this function and use those parameters to show that the function offers a reasonably good explanation of the payoff scale effects, ranging from 1x to 90x, for the choice menus used in their experiments.

[15] There are 64 observed choices, safer or riskier, for each decision in each payoff treatment, half of which pertain to men and half to women. Note that the decision number refers to the number in Table A1 and does not indicate the order in which it was encountered, since the 10 decisions for each payoff treatment block were randomly shuffled.

[16] Applying the weighting directly to all probabilities is known to imply violations of stochastic dominance Handa (1977). In particular, for any nonlinear weighting function it is possible to find two probability distribution functions for which inverse cumulative function $(1 - F(x))$ of one is higher, but this dominating function has a lower weighted expected utility. The theoretical fix is to apply the weighting function directly to the inverse cumulative distribution (Quiggin, 1982).

`

payoff and using the residual for the low payoff: $w(p)U(H) + (1-w(p)U(L)$. This "weighted expected utility" expression is then used to derive choice predictions.

For any specific weighting and utility parameters, the implication would be that choice probabilities are either 0 or 1, so we introduce a probabilistic choice function to capture unobserved factors that produce "noise" in the data.[17] This "power" function is: $\Pr(A) = \dfrac{(U_A)^{1/\mu}}{(U_A)^{1/\mu} + (U_B)^{1/\mu}}$, where μ is a positive "noise" parameter and the weighted expected utilities for options A and B are represented by U_A and U_B respectively. As $\mu \to \infty$, the noise dominates and choice probabilities go to 0.5 as each of the utility terms are raised to the power 0 in the limit. In contrast, it can be shown that as $\mu \to 0$ the choice probability converges to 1 for the option with the higher weighted expected utility. For each decision, the utility and probability weighting functions are used to compute the weighted expected utilities of U_A and U_B for options A and B in the choice pair, and the associated choice probability prediction is determined from the probabilistic choice function.

The likelihood function to be maximized is the product of these probabilities, each raised to a power determined by the number of A choices for that problem. Thus maximum likelihood estimation essentially involves finding the parameters for utility (α and r), probability weighting (γ), and noise parameter (μ) that maximize the probability of seeing what was observed in the data. The maximum likelihood estimates are provided in Table A3, where standard errors have been adjusted for clustering of individual decisions. The parameter estimates in the table are significant and of plausible magnitudes.[18] Note that the estimated risk aversion parameters, r and α, are higher for women, indicating more curvature of the utility function that provides the best fit for their decisions. In contrast, the probability weighting parameter estimates are essentially the same

[17] The power function stochastic choice model can be derived by assuming that positive payoffs are perturbed by multiplicative errors, which implies that "noise" does not diminish as payoffs are scaled up. In contrast, scaling up payoffs in a standard logit model tends to diminish the effects of additive errors, and subjects' responses to incentives are predicted to be "sharper" at high payoff scales. Since we do not observe sharper response functions in experiments with high payoff scales, it is desirable to use a stochastic formulation for which noise does not diminish with high payoffs (Holt and Laury, 2002). The Luce power form is one way to accomplish this. To see the intuition, suppose for simplicity that the subject is risk neutral, and hence, the U_A and U_B terms in the equation are expected payoffs. In this case, a 5x increase in payoff scale would increase all expected payoffs by a factor of 5, and this multiplicative constant would factor out of both the numerator and the denominator, having no effect.

[18] For example, the coefficient estimates for r in Table A3 for men and women bracket the estimate of $r = 0.27$ reported by Holt and Laury (2002) for both genders combined, using data from a structured choice menu risk elicitation task.

`

for both genders.[19] The difference between the estimated probability weighting parameters and a value of 1 (no weighting) is statistically significant and in line with many prior estimates. Finally, note that the estimated error parameters in the bottom row of Table A3 are essentially the same for men and women.

Table A3. Maximum Likelihood Estimates

Parameter	Gender	Coefficient (Robust Standard Error)
Risk Aversion (r)	Men	0.18 (0.07)
	Women	0.39 (0.08)
Risk Aversion (α)	Men	0.02 (0.01)
	Women	0.04 (0.02)
Probability Weighting (γ)	Men	0.66 (0.04)
	Women	0.67 (0.03)
Error Parameter(μ)	Men	0.07 (0.01)
	Women	0.06 (0.00)

To summarize, the main estimation results reveal 1) significantly positive risk aversion parameters (α and r) that are higher for women than for men, and 2) estimated probability weighting parameters implying significant curvature that is comparable for men and women.[20]

The next step is to determine whether this model can explain the qualitative patterns observed in the data, e.g. risk aversion for downside risk and risk seeking for upside risk, or the presence of gender and scale effects in some settings and not in others. Figure A1 shows the percentages of riskier choices for each of the treatment combinations, with dark bars for men and light bars for women. For comparison, the gray bars indicate the predicted proportions calculated from the parameter estimates for each gender. The low stakes treatments are shown on the left side and the high stakes treatments are shown on the right side. The fitted predictions track the

[19] Gender differences in probability weighting have been reported by others. For example, Fehr et al. (2011) found that women were more likely to weight probabilities of good outcomes less optimistically when they are not in a good mood, and vice versa. Men seemed to be less responsive to moods, and more prone to rely on mechanical decision rules instead of intuition. In particular, about 40 per cent of the men in their study reported using expected value as a decision criterion, whereas only a handful of women in the study reported this behavior. As noted in the procedures section above, we used "odd numbers" for many of the payoffs listed in Table A1 in an attempt to diminish any tendency for subjects to rely on simple mathematical decision rules.

[20] These two features of the estimates do not change when we estimate a two-parameter version of the Kahneman and Tversky probability weighting function or a two-parameter version of the Prelec "expo-ln" weighting function. In each case, the two risk aversion parameters are significantly positive ($p < 0.03$ in all cases) and greater for women, whereas both probability weighting parameters are quite close in value for men and women, with a significant amount of curvature in the weighting function for both genders.

`

major difference between risk aversion for downside risk and risk seeking for upside risk, and the gender difference for low stakes downside risk that is diminished for the other treatments.

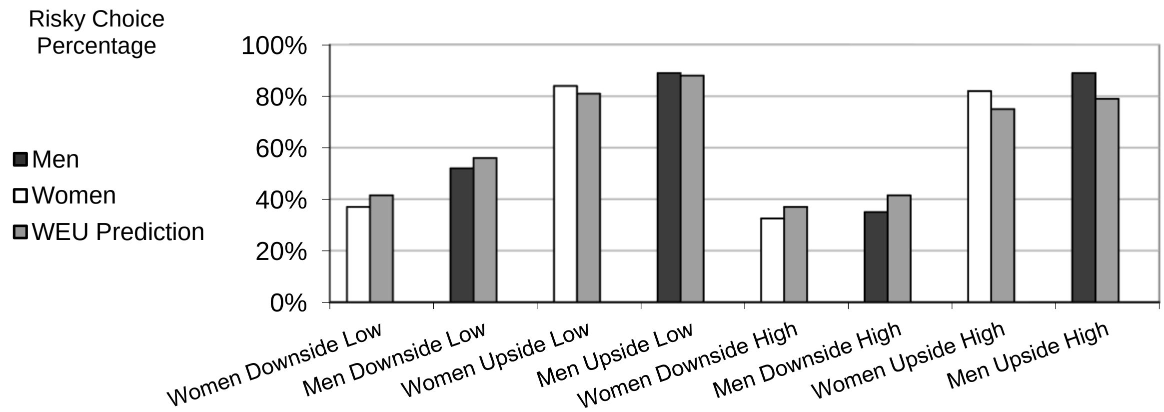

Figure A1. Percentages of Risky Choices for Men (Dark Bars) and Women (Light Bars), Together with Weighted Expected Utility Predictions (Gray Bars)

Although the main focus of this experiment is on gender differences in the structure of risk for salient low-probability events, we also included a couple of decisions (9 and 10) for balanced risk. Even though the probabilities for the high payoff are 0.5 in each of these two decisions, the expected payoff difference in favor of the risky choice is lower for problem 10 (only $0.35 instead of $0.80 as was the case for the other 9 problems). Hence it is not surprising that the theoretical predictions (using fitted parameter values for low stakes) are higher for decision 9 (67% for women and 74% for men), and lower for decision 10 (2% for women and 12% for men). This sharp difference in predicted proportions of risky choices is observed in the data; risky choice percentages were 75% for women and 81% for men in decision 9, and 9% for women and 19% for men in decision 10. In summary, the rank-dependent utility estimates provide a unified explanation of the main qualitative features of the choice data for a wide variety of settings in terms of payoff scales, risk intensity, and risk direction (upside or downside).

`

Online Appendix B: Experiment Instructions

(These instructions are for the case of a single decision.
Wording changes for sessions with random selection are indicated in italics.)

- **Options:** You will be making a **single choice** between alternative options, such as "Option A" and "Option B" below. Each option offers two (or more) possible money prizes. You must select one of these options, without knowing in advance which monetary amount will be obtained. (WITH RANDOM SELECTION: O***ptions:*** *In each part of this experiment, you will be making a* ***series of choices*** *between alternative options, such as "Option A" and "Option B" below. Each option offers two (or more) possible money prizes. You must select one of these options, without knowing in advance which monetary amount will be obtained.*)
- **Monetary Prizes:** The money prize that is relevant for the option you select is determined by the computer equivalent of throwing a ten-sided die or spinning a roulette wheel with ten equally-likely stops. In the example below, if you choose Option A, the wheel would have 5 stops labeled \$4.00 and 5 stops labeled \$6.00, and the wheel for option B would have 5 stops labeled \$0.00 and 5 stops labeled \$12.00. Thus if you choose Option A, you will have a **5 in 10** chance of earning **\$4.00** and a **5 in 10** chance of earning **\$6.00**. Similarly, Option B offers a **5 in 10** chance of earning **\$0.00** and a **5 in 10** chance of earning **\$12.00**. Please Note: The numbers used in the example below are for illustrative purposes only, the actual choice that you will consider will be different from those used in this example,
- **Choice:** You will register your choice by using the mouse to click on the small circle ("radio button") for the option you select. Then you must click on the gray Submit button at the bottom. Please go ahead and make a choice for this practice round to see how this process will work.

○ **Option A**	○ **Option B**
5 chances in 10 of **\$4.00**	**5** chances in 10 of **\$0.00**
5 chances in 10 of **\$6.00**	**5** chances in 10 of **\$12.00**

Submit Practice Decision

(new page)

5 chances in 10 of **\$12.00**

You selected Option B.
We will now use the computer to generate a random number,
which is equally likely to be any number between (and including) 1 and 10.

`

The payoff will be **$0.00** if the number is in the range **1 - 5**
The payoff will be **$12.00** if the number is in the range **6 - 10**.

Show Practice Results

(new page)

You selected Option B.
The payoff will be **$0.00** if the number is in the range **1 - 5**
The payoff will be **$12.00** if the number is in the range **6 - 10**.

Option B

5 chances in 10 of **$0.00**

5 chances in 10 of **$12.00**

Result: The random draw turned out to be **3**.
Thus the payoff would be **$0.00**.
Continue

(new page)

- **Additional Setup Details:** You will be making a single choice between alternative options. These options may be expressed in terms of **"chances in 100"** instead of "chances in 10" as in the previous example. (WITH RANDOM SELECTION: *A**dditional Setup Details:** You will be making a series of choices between alternative options. These options may be expressed in terms of* ***"chances in 100"*** *instead of "chances in 10" as in the previous example*.
- **Monetary Prizes:** In the example below, the money prize that is relevant for the option you select is determined by the computer equivalent of throwing a **100-sided die** or spinning a roulette wheel with **100 equally-likely stops**.
- **Options:** In particular, if you choose Option A, the wheel would have 25 stops labeled $4.00 and 75 stops labeled $6.00, and the wheel for option B would have 25 stops labeled $0.00 and 75 stops labeled $12.00.
- **Chances in 100:** Thus if you choose Option A, you will have a **25 in 100** chance of earning **$4.00** and a **75 in 100** chance of earning **$6.00**. Similarly, Option B offers a **25 in 100** chance of earning **$0.00** and a **75 in 100** chance of earning **$12.00**.

Option A	**Option B**
25 chances in 100 of **$4.00**	**25** chances in 100 of **$0.00**
75 chances in 100 of **$6.00**	**75** chances in 100 of **$12.00**

Continue with Instructions

Summary

- **Single Choice:** To summarize, you will begin by making a **single decision** that will determine your earnings. (WITH RANDOM SELECTION: *To summarize, you will begin by making a series of* ***10 decisions****.*)
- **Options:** For this decision problem, you must select one of the two options, A or B. (NEW PARAGRAPHS WITH RANDOM SELECTION: ***Relevant Decision:*** *After you have made all 10, decisions,* ***only one of these will be selected at random*** *to determine your total earnings for this part. Each of the 10 decisions has an equal chance of being selected, independently of the choices you made.* ***Possible Prizes:*** *After you have made all decisions, only one of the 10 choice problems will be used. The option that you have* ***already*** *selected for that choice problem will have 2 possible money earnings amounts, each associated with the chances that it will be the actual amount obtained.*)
- **Random Number:** Then the computer will generate a random number that determines which of the money prizes for the option you selected will be the amount of money that you earn.
- WITH RANDOM SELECTION*:* ***Subsequent Parts:*** *This whole process (making 10 decisions and having one selected at random to determine your earnings) will be repeated once, with some changes in the structure of the options themselves in the second part. Earnings for each decision will not be released until you finish the final part.*)
- **Earnings Record:** The computer keeps track of your earnings, i.e. the sum of the amounts earned in each part.