Investment decisions and time horizon: Risk perception and risk behavior in repeated gambles [*]

Alexander Klos[†]
University of Mannheim

Elke U. Weber[‡]
Columbia University

Martin Weber[§]
University of Mannheim and CEPR

March 2005

Abstract

To investigate the effect of time horizon on investment behavior, this paper reports the results of an experiment in which business graduate students provided certainty equivalents and judged various dimensions of the outcome distribution of simple gambles that were played either once or repeatedly for 5 or 50 times. Systematic mistakes in the ex-ante estimations of the distributions of outcomes after (independent) repeated plays were observed. Despite correctly realizing that outcome standard deviation increases with the number of plays, respondents showed evidence of Samuelson's (1963) fallacy of large numbers. Perceived risk judgments showed only low correlations with standard deviation estimates, but were instead related to the anticipated probability of a loss (which was overestimated), mean excess loss, and the coefficient of variation. Implications for future research and practical implications for financial advisors are discussed.

[†]Lehrstuhl für Bankbetriebslehre; L 5,2, Universität Mannheim, 68131 Mannheim, Germany, klos@bank.BWL.uni-mannheim.de

[‡]Center for the Decision Sciences and Graduate School of Business, Columbia University, 3022 Broadway, 716 Uris Hall, New York, NY 10027, euw2@columbia.edu

[§]Lehrstuhl für Bankbetriebslehre, L 5,2, Universität Mannheim, 68131 Mannheim, Germany and CEPR, London, weber@bank.BWL.uni-mannheim.de

1 Introduction

The role of investment horizon in asset allocation decisions has been widely discussed in finance and remains a controversial topic. Many proponents of the idea of time diversification, especially professional financial advisors, urge younger people to invest more heavily in stocks for long-term investment goals such as saving towards their pension. Kritzman (1994) for example argues that the probability of a loss declines with the investment horizon. Consider an investment option that offers equal chances to win $200 or to lose $100. The probability of losing money after 10 independent trials is 0.18, but goes down to 0.0077 for 50 independent trials. If risk is operationalized as the probability of a loss, this risky investment is less risky after more independent trials or, within the context of investment decisions, with a longer investment time horizon.

The time diversification debate evolves around the question of how "risk" ought to be defined. The fact that "risk has no universal definition; rather, like beauty, it is in the eye of the beholder" (Kritzman 1994, p. 67)[1], poses a fundamental problem. While everyone agrees that risk reduction is a desirable goal, not all definitions of risk are consistent with the observation that risk is reduced by repeated plays. Equating risk with the probability of a loss is only one possibility. Most people would agree that the magnitude of potential losses needs to be considered as well, as for example in a measure called mean excess loss, which is defined as the expected value of the loss, given that a loss has occurred. For our example investment option, mean excess loss increases from $216 for 10 independent trials to $407 for 50 independent trials. Investors who take mean excess loss as their index of investment risk thus should be less risk tolerant at longer investment horizons (Albrecht et al. 2001).

A third risk measure is the standard deviation of outcomes, probably the most common risk measure in finance since the work of Markowitz (1952). Clearly, if standard deviation is used as an index of risk, the risk of repeated gambles increases by a factor of $\sqrt{N}$, with N denoting the numbers of repeated plays.

A fourth measure of risk recently examined by E. U. Weber et al. (2004) is the coefficient of variation (CV). The CV, which is defined as the standard deviation of a lottery divided by its expected value and thus provides a standardized measure of risk per unit of return, was shown to predict risky choices in repeated decisions far better

---

[1]See also Brachinger and M. Weber (1997) and E. U. Weber (2001a).

than the standard deviation. Since the standard deviation increases only by a factor of $\sqrt{N}$ and the expected value increase by the factor N for N repeated plays, using the CV as an index of risk implies that the risk of a gamble decreases by a factor $\sqrt{N}$ as N, the number of plays, increases.

The fifth and final line of argument found in the debate about the effect of repeated plays on risk makes fallacious use of the law of large numbers. Samuelson (1963) reports a lunch-time conservation with a colleague who was not willing to play our example lottery once, but would have taken on 100 independent plays. The intuition behind the law of large numbers is that below-average and above-average realizations of a random variable (i.e., the return of a stock market investment) tend to cancel each other out, the more so with larger numbers of realizations (e.g., repeated plays), making it more certain that the expected return will be realized in the long run. Expressed mathematically (Samuelson 1989, page 292),

if each play of gamble brings positive expected return, $E\{X_j\} = \mu > 0$, then

$$Prob\{|(X_1 + ... + X_N) - N \cdot \mu| < \epsilon\} \to 1, \tag{1}$$

for $0 < \epsilon \ll 1$, as $N \to \infty$.

As Samuelson (1963) points out, this reasoning is based on a misinterpretation of the law of large numbers. The law of large numbers states that $1/N \cdot \sum_{i=1}^{N} X_i$ tends to μ as $N \to \infty$ and not that $\sum_{i=1}^{N} X_i$ tends to $N \cdot \mu$ as $N \to \infty$. A reduction in risk can only be achieved if a given risk is subdivided into many smaller independent risks and not simply by adding more independent risks. It follows that the probability to end up in a given small interval around the long-run expected value declines with the number of repeated plays and converges to zero instead of one as N goes to infinity.

This paper does not address the prescriptive issue of which measure of risk decision makers should be using when comparing the risks and returns of risky options that involve single or multiple independent realizations (but see Kritzman and Rich (2002)). Instead, our study offers insight into descriptive issues raised by the time horizon debate. In particular, we examine decision makers' ex-ante predictions for the different summary statistics of the outcome distribution of lotteries after single and repeated plays that have been suggested to serve as proxies for risk and thus as affecting choice, namely the probability of loss, the mean excess loss, the standard deviation of outcomes, the

coefficient of variation, or the probability that the realized return of the lottery falls into a small interval around the long-run expected value. We do so by asking decision makers to estimate these statistics for different numbers of repeated plays of a set of lotteries. This allows us to assess whether people correctly anticipate how these statistics change as the number of repeated plays increase. We then examine how people's estimates of these five measures are related to (a) their judgments of the riskiness of these lotteries, where risk is left intentionally undefined, and (b) their preference for these lotteries as inferred from certainty equivalents.[2]

Studies of repeated gambles have much practical relevance.[3] Many decisions are made between risky options that have multiple, repeated realizations, including decisions regarding retirement savings. Many health and safety decisions also have this characteristic, e.g., the decision to (not) use condoms or seatbelts. While there is some question about the degree to which insights gained from studying repeated lotteries can be generalized to other domains, including investment decisions[4], an investigation like the one reported here provides a baseline to which more content-laden applications can be compared. There is reason to believe that possible misperceptions of the statistics of multiple-play outcome distributions and their relationship to perceived risk and choice are (at least) similar for multiple-play gambling decisions and long-term asset allocation decisions. An understanding of how people perceive the risk of repeated gambles is therefore the basis for an understanding of more complex retirement saving decisions and repeated-exposure health and safety decisions.

Our study was partly inspired by Benartzi and Thaler (1999), who elicited the probability of loss for several repeated gambles in an experimental study and found that undergraduate economics students overestimate the probability of a loss by a wide margin. Our study examines whether people misestimate this as well as other statistics widely used as proxies for risk. Benartzi and Thaler (1999) also examine the fallacy

---

[2]Throughout the paper, the term "risk behavior" includes both choices and preferences inferred from certainty equivalents.

[3]Note that there are two different types of repeated gamble situations: In Type 1 situations, people make one decision that will hold for multiple plays, which is made before experiencing any of the realizations and their outcomes. This situation is considered in this study. Decisions of Type 2 are made sequentially and people can switch their preferences across repetitions. Those situations have been studied more widely before, especially for animals (E. U. Weber et al. 2004), but more recently also for humans (Hertwig et al. 2004; see Klos 2004 for some evidence in an asset allocation experiment). There is only little work on Situation 1. A recent contribution is Langer and M. Weber (2005).

[4]Goldstein and E. U. Weber (1995) and E. U. Weber et al. (2002), for example, find differences in risk attitude between investment and gambling decisions.

of large numbers, according to which respondents should underestimate the standard deviation of multiple-play gambles. Instead, their subjects overestimate the standard deviation of multiple-play gambles, a result more in line with a fallacy of small numbers, and putting into question Samuelson's (1963) interpretation that a misinterpretation of the law of large numbers underlied his colleague's willingness to play the offered lottery 100 times. Our study will allow us to examine which of the explanations offered above influence both risk judgments and preference for multiple-play lotteries. We will directly test whether people's estimates of the riskiness of repeatedly played lotteries follows a misinterpretation of the law of large numbers, by asking respondents for their estimates of the probability to end up in a given small interval around the long-run expected value.

Recent research has examined perceptions of risk as a psychological construct that mediates risky choice (see E. U. Weber 2001a,b and M. Weber 2001c for reviews). These studies show that perceived risk, i.e., the subjective feeling of riskiness associated with a specific action or choice alternative, often deviates strongly from conventional risk indices, such as the probability of a loss or the standard deviation of possible outcomes (E. U. Weber et al. 2004, Loewenstein et al. 2001). In one study where people assess both the standard deviation of possible outcomes and the perceived riskiness of possible risky investment alternatives, it is the latter and not the former measure that predicts choice (E. U. Weber et al. 2005). Our study allows us to examine how closely people's subjective perceptions of the risk of repeated gambles are related to the risk measures described above.

In sum, we address three research questions: (1) Are people able to correctly assess statistics of outcome distributions commonly suggested as indices of risk for repeated-play lotteries? (2) What constitutes perceived risk in repeated gambles? (3) Does perceived risk predict preference?

The remainder of the paper is structured as follows. Section 2 describes the details of our experimental design. The risk behavior of our subjects is the topic of Section 3. In Section 4, we analyze respondents' estimates of the risk measures described above. Section 5 examines the determinants of respondents' judgments of perceived risk. Section 6 discusses the relationship between risk behavior and the various risk measures. Section 7 concludes with a discussion of the implication of our results.

2 Experimental Design and Hypotheses

Our experiment used discrete two-outcome lotteries that all had an expected value (EV) of €50 for a single play, in addition to Samuelson's (1963) classic lottery (200,0.5;-100,0.5). This lottery and the other three lotteries employed are shown in Figure 1 in the decision tree format in which they were shown to subjects. Keeping the EV (and thus presumably perceived return) constant within each level of repeated plays allowed for a simple test of the relationship between perceived risk and preference. The primary motivation for the use of the four lottery profiles was to ensure the robustness of the results with respect to the exact lottery profile. We chose therefore a positively and a negatively skewed distribution as well as a lottery without any losses in addition to Samuelson's famous bet.[5] Each lottery was presented as either a single-play option or as an option that would be played 5 or 50 times, respectively, resulting in 12 different stimuli.
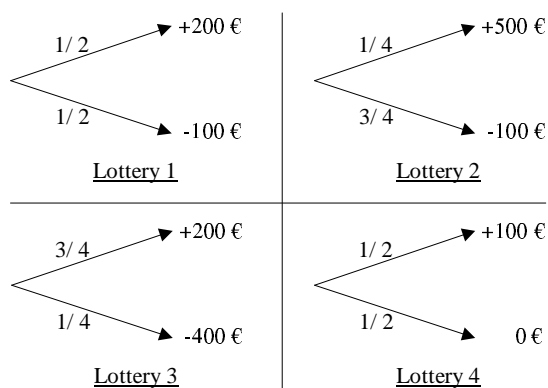

Figure 1: *Lotteries.* The four lotteries used in our experimental study. Note that all have the same expected values of €50.

2.1 Decision Tasks

Participants engaged in two PC-guided tasks in individual sessions in a research lab, conducted in the order described below. Task order was not randomized because of logical constraints. However, the order of the different stimuli within each task was randomized for each task and each respondent.

---

[5]See Lopes and Oden (1999) on the question how people respond to alternative distributions of complex gambles and Keller et al. (1986) and Shafir et al. (2003) on differences in risk perception with respect to skewness.

Task A: Certainty Equivalents for Repeated Gambles. Respondents saw each of the 12 lottery stimuli (4 lotteries shown as in Figure 1, with a number [1, 5, or 50] displayed in red letters next to it that indicated the number of independent times that the lottery would be played) and provided their certainty equivalents (CE) for each one as follows. They first indicated whether they were willing to play the specified lottery for the indicated number of repeated plays. For respondents who were so willing, the computer screen then displayed six different sure amounts in ascending order. Respondents were asked whether they preferred playing the indicated single or multiple-play lottery to each of the six stated sure payments. For decision makers who were not willingness to play the lottery, we applied the same procedure, except for using negative sure amounts. Instructions spelled out that negative numbers should be interpreted as the amount of money they were willing to pay in order not to have to play the displayed lottery for the indicated number of plays.[6]

Violations of monotonicity were not allowed, i.e., the program forced the decision maker to reassess her six preference statements via a message window. Based on stated preferences, the program inferred a lower and upper bound for the subject's CE. Subjects were then asked to state the exact amount that would make them indifferent between the lottery and the CE, within the range they specified before. The arithmetic mean of upper and lower bounds was displayed as a starting value and could be modified by using increment and decrement buttons or the number keys to type in a different amount.

Note, that subjects were first confronted with a choice situation (whether they were willing to take the gamble), then with a series of choices (between the sure amounts and the lottery), and finally had to state an exact certainty equivalent for the gamble.[7]

Task B: Assessment of Summary Statistics of Anticipated Lottery Outcome Distribution. Just as for Task A, respondents saw the graphical representation of one

---

[6]The six different sure amounts ranged from 0 to 10000 in steps of 2000 for 50 repeated plays, 0 to 1000 in steps of 200 for 5 repeated plays, and 0 to 200 in steps of 40 for the one-shot-gamble. If the subject declined the gamble in her initial choice, the sure amounts ranged from 0 to -5000 in steps of -1000 for 50 repeated plays, 0 to -500 in steps of -100 for 5 repeated plays, and 0 to -100 in steps of -20 for the one-shot-gamble.

[7]To our knowledge, the described elicitation method for certainty equivalents was first used by Kilka and M. Weber (2001). However, a common practice in decision analysis is to continue choices like the one described here until a "choice" certainty equivalent is elicited (see for example Tversky and Kahneman 1992 or Gonzales and Wu 1999, among others). We deviated from this common practice for the following reason. Given the high expected values, especially for 50 repeated plays, many choices would be needed to elicit one single certainty equivalent. Since we elicited many certainty equivalents, this procedure would have resulted in an unacceptable large series of questions, resulting in fatigue and boredom on part of the respondents, which introduce their own set of errors and biases.

of the four lottery, with the numbers of repeated plays to which they were committed indicated by a red numeral next to the lottery. They were asked for their assessment of each of the following statistics. For all but the final judgment (perceived risk), we did not ask respondents to provide estimates for single plays, of the shown lottery, under the assumption that everyone would correctly read the answer off the lottery display.

Probability of a loss. Following Benartzi and Thaler (1999), we asked for the probability of a loss in two ways, that is, either directly for their estimate of the probability of losing money, or for the positive complement, namely their estimate of the probability of coming out ahead (i.e., not losing any money) after playing the lottery for the indicated number of trials (5 or 50). The estimated probability of a loss was assumed to be one minus the estimate of coming out ahead. We randomly selected one of the two possible question frames for every subject. Subjects could enter their estimate into a text field via the keyboard or they could modify the starting value of zero[8] by using increments and decrement buttons.

Mean excess loss. For each multiple-play lottery, respondents answered the question: "What do you think the average loss would be, given that a loss has occurred, after playing the shown lottery for the indicated number of times?". Subjects could again enter their answer into a text field or could modify the starting value of zero by using increments and decrement buttons ($\pm 1000, \pm 100, \pm 50, \pm 10, \pm 1$).

Probability that the outcome of the lottery falls into a small interval around the long-run expected value. Subjects indicated their estimate of the probability that the outcome of the shown multiple-play lottery would fall in a € 100-interval around the long-run expected value. Thus the question for lotteries with five repeated plays was "What do you think is the probability that the final total outcome will lie between 150 and 350 (expected value: 250) after all 5 repeated plays?" and for lotteries with 50 repeated plays: "What do you think is the probability that the final total outcome will lie between 2400 and 2600 (expected value: 2500) after all 50 repeated plays?". We used the same input mechanism via the keyboard or using increment and decrement buttons for this

---

[8]Alternative starting values, for example 50% or a random number between 0 and 1, were possible. If we chose such a starting value, this may have resulted in an upward biased estimate. However, our ex-ante hypotheses were an overestimation of the elicited risk measures. We acknowledge the possibility that the starting value of zero could bias the estimated probability to the lower end, but the vast majority of the elicited risk measures are overestimated (see Section 4) and therefore this possibility does rather strengthen our results.

question as for the probability of a loss (including the starting value of zero). Figure 2, showing the aggregated outcome distribution of Lottery 1 for up to five trials, illustrates that the probability that the outcome of the lottery falls into a small interval around the long-run expected value declines with the number of repeated plays, i.e., from 1/2 for two repeated plays to 3/8 for four repeated plays. The core of the fallacy of large numbers is the erroneous belief that this probability increases.

Figure 2 also illustrates a problem arising from the fact that we used discrete distributions. The probability of the final outcomes falling into the interval of plus or minus €100 around the mean is zero for some numbers of repeated plays, i.e. for 5 repeated plays. However, we do not think that this was a serious problem since respondents probably had a continuous distribution of outcomes in mind and answered based on this approximation to our questions. Some subjects quoted such a heuristic explicitly in a short questionnaire administered after the experiment or in private conversations with the experimenter. Furthermore, if subjects recognized the problem, they should have answered 0 to one or more of the eight interval probability estimation questions. Almost 89% of all subjects answered with a strictly positive value to all eight questions and the results reported below remain qualitatively the same if we exclude the remaining 11%.

Standard Deviation. Subjects were asked for their estimates of the median and the 10th and 90th percentile of the aggregated distribution.[9] Given these three values, it is possible to approximate the estimated standard deviation. For the median, respondents were asked to state the amount of money for which they thought it equally likely that the final overall outcome would exceed it or remain under it. For the 10th (90th) percentile, they were asked to state the amount of money that they thought would be exceeded (remain under) by the final overall lottery outcome in only 10% of all cases. Responses were only accepted by the program if the answers satisfied the relationship 10th percentile < median < 90th percentile. This relationship was also mentioned in the instructions.

These questions also suffer from the shortcoming due to the discrete distributions. As illustrated by Figure 2, with the percentages in the framed boxes denoting the probability of a specific outcome, an unambiguous 10th percentile of this distribution does not exist. However, subjects probably had a continuous approximation to the questioned

---

[9]E. U. Weber et al. (2005) also elicit these quantiles in an experimental study using continuous distributions. Benartzi and Thaler (2002) use the same quantiles to communicate an approximate distribution of final wealth for an investment alternative in an asset allocation experiment.
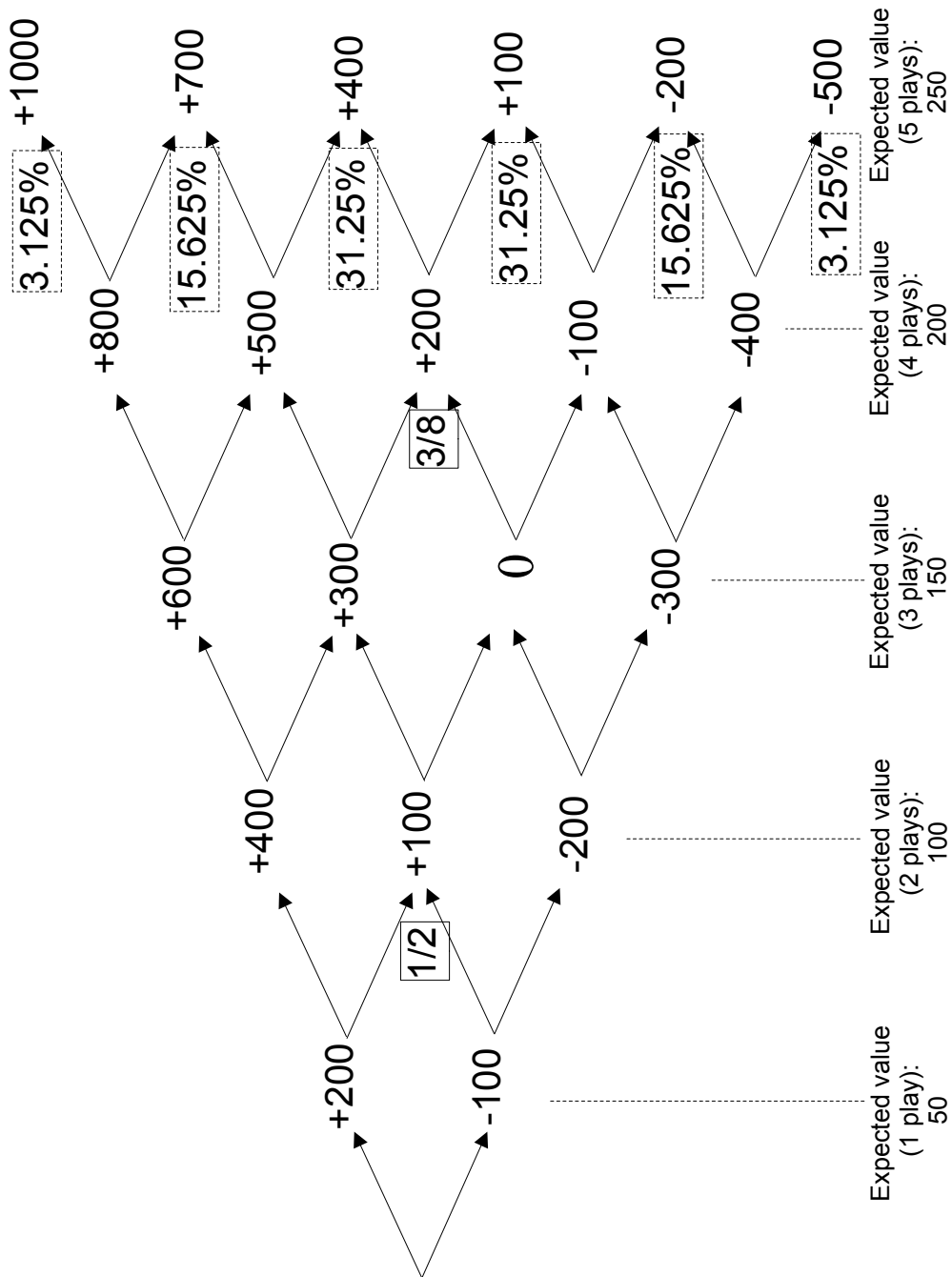
Figure 2: *Outcome Distribution of Lottery 1.* The graph shows the aggregated outcome distribution of Lottery 1 for up to five plays. The fractions in the framed boxes give the probability that the outcome of the repeated gamble will fall in a ±100-interval around the expected value. Also shown is the long-run EV for all possible numbers of repeated plays. The numbers in the fine dashed framed boxes denote the probability of every possible outcome after five independent repeated plays.

distribution in mind. Furthermore, we are mainly interested in the qualitative changes in their estimates from 5 to 50 repeated plays.

Perceived Risk. For all 12 lottery stimuli, respondents answered the questions: "How do you assess the risk of this lottery played the indicated number of times on a scale from 0 ('not at all risky') to 100 ('extremely risky')?" Subjects indicated their answer using a slider on a graphic response scale. The slider started out at the middle of the scale, i.e. at 50.

Correct Answers to the Estimations Elicited Under Task B. Table 2 in Section 4 shows the correct answers to the estimation questions. The "correct" value shown for the probability that the overall outcome lies in $\pm \text{\euro}\, 100$ interval around the long-run expected value is based on a normal approximation of the binomial distribution.

2.2 Hypotheses

Task A: Certainty Equivalents for Repeated Gambles. Generalizing from the behavior of Samuelson's colleague, we predicted that willingness-to-play for our positive EV lotteries would increase with the number of repeated plays (*Hypothesis A1: $p(play)_1 < p(play)_5 < p(play)_{50}$*).[10] The data to test this hypothesis comes from subjects initial choice whether they were willing to play the specified lottery for the indicated number of repeated plays. Assuming a monotonic relationship between willingness-to-play and willingness-to-pay (i.e. certainty equivalents), we also predicted that certainty equivalents would increase with numbers of play (*Hypothesis A2: $ce_1 < ce_5 < ce_{50}$*).

Task B: Assessment of Different Summary Statistics of Anticipated Lottery Outcome Distribution. *Probability of a loss.* Following Benartzi and Thaler (1999), we expected that subjects would overestimate the probability of a loss (*Hypothesis B1*). In addition, we hypothesized (based on introspection) that the decline in the probability of a loss in moving from 5 to 50 repeated plays would be correctly recognized (*Hypothesis B2: $Est_{Prob(Loss)}(5 trials) > Est_{Prob(Loss)}(50 trials)$*).

Mean excess loss. We are not aware of previous experimental results for this estimation task. However, if subjects overestimate the probability of a loss, one would expect that they also overestimate the mean excess loss, i.e. that they believe the aggregated dis-

---

[10]We will also use the term acceptance rate, defined as the percentage of subjects who accept the gamble, synonymously to $p(play)$ later in the paper.

tribution to have fatter tails than it really has (*Hypothesis B3*). Furthermore, we expected that the increase in the mean excess loss as a result of the increase in repeated plays would be recognized by our subjects (*Hypothesis B4 : $Est_{MEL}(5 trials) < Est_{MEL}(50 trials)$*).

Probability that the outcome of the lottery falls into a small interval around the long-run expected value. The fallacy of large numbers refers to the fact that people may believe that it is almost certain to receive the long-run expected value of an independently repeated gamble if the number of repetitions is high enough. If Samuelson's speculation about the reason for his colleague's inconsistent preferences is correct, we will predict that subjects state a higher probability of an outcome near the expected value for more iterations (*Hypothesis B5 : $F_1 := Prob\{|(X_1+...+X_5)-5\cdot\mu| < 100\} < F_2 := Prob\{|(X_1+ ... + X_{50}) - 50 \cdot \mu| < 100\}$*).

Standard Deviation. Given the results of Benartzi and Thaler (1999), we predicted that subjects would correctly realize that the standard deviation of lotteries increases with the number of repeated plays (*Hypothesis B6*). Note that Hypothesis B6 is normatively inconsistent with Hypothesis B5, though it is certainly possible that both an increase in standard deviation and the fallacy of large numbers could be simultaneously observed in the within-subject context of our experiment.

Perceived Risk. Although Samuelson's colleague makes no direct reference to how he perceives the risk of the repeated gamble compared to the risk of the single play, his behavior suggests that he perceives the risk of the lottery to decrease with the number of repeated plays, i.e., *Hypothesis B7 :* PR(1 trial)>PR(5 trials)>PR(50 trials). The simultaneous elicitation of judgments of perceived risk and of different summary statistics suggested to reflect risk gives us the opportunity to analyze the sources of the subjective perception of riskiness, especially in the context of repeated gambles, which has not been done before. Relating the responses in Task A to risk perception allows us to analyze the relationship between perceived risk and the elicited certainty equivalents. Based on the experimental work cited in the introduction, we expected that perceived risk predicts preference (i.e., CEs) better than outcome standard deviation (*Hypothesis B8*).

2.3   Incentives and Procedure

We paid subjects according to Becker, DeGroot, and Marschak's (BDM) (1963) incentive-compatible procedure for eliciting certainty equivalents. After the experiment we ran-

domly selected one of the lottery stimuli. For the selected lottery with its specified number of repeated plays, we randomly selected a bid that came from a uniform distribution that ranged from the minimal to the maximal value which could be realized in the selected repeated gamble. If this bid was higher than the CE the subject had provided for this lottery, then the subject received the amount of the randomly selected computer bid. If the bid was lower than the subject's CE, then we played out the lottery for the specified number of repetitions and the subject received the obtained payoff. The BDM mechanism was explained to subjects on-screen before the start of the experiment. We also presented an intuition why stating the true certainty equivalent is a dominant strategy. Respondents received detailed instructions before every decision task and could also call up instructions for each task and a description of the incentive scheme at any point of the experiment.

2.4 Respondents

The experiment was conducted at the University of Mannheim in June of 2002. The volunteer participants were 44 advanced business students, who took between 40 and 110 minutes to complete all tasks. Payments ranged from € 9 to € 14 with an average payment about € 11. The median age was 25, and 7 subjects were female. Participants were asked how clear they found the instructions after the experiment, on a scale from 1 (completely unclear) to 9 (completely clear). The median response was 7. Given the complexity of the tasks, a rating of this magnitude suggests quite strongly that respondents thought they understood the decision tasks.[11]

3 Results: Risk Behavior

Table 1 shows the arithmetic means and medians of the certainty equivalents provided by our respondents for the lottery stimuli in Task A. Testing hypothesis A1 (that $p(play)$

---

[11]Note that we used a within subject design. One question is what we would have gotten in a between subject design. We assessed the possibility of different results in a between subject design on the basis of Table 1 (see section 4). We compared the data of only the first judgment of each of the four lotteries made by each subject: for approximately 1/2 of them, this first judgment will be for a repeated play of 5, and for approximately 1/2 it is for a repeated play of 50. (Clearly, this is not a "clean" between subject design, but in our view it is a sufficient approximation to what we might have gotten.) Although some p-values increases (especially due to the lack of a sufficient number of subjects), the results of this analysis suggest that a between-subject design would not have changed the qualitative results.

increases with the number of trials) suffers from a ceiling effect as all acceptance rates are near 100%. However, in most cases acceptance rates differ in the predicted direction, but only 2 out the 12 pairwise comparisons were statistically significant, after correcting for multiple comparisons. We do not reported the details of this analysis due to space limitations.

We performed a nonparametric test for ordered hypothesis introduced by Page (1963) to address hypothesis A2. This statistical method tests the null hypothesis of no difference in certainty equivalents ($H_0 : CE_{1trial} = CE_{5trials} = CE_{50trials}$) against the ordered alternative $H_1 : CE_{1trial} > CE_{5trials} > CE_{50trials}$.[12]

A Page test strongly supports hypothesis A2 (p<0.0000*** for all four lottery profiles) that CEs increase with the number of repeated plays.[13] This should come as no surprise, since the expected value of the lotteries increases substantially with the number of repeated plays.

Certainty Equivalents						
	Lottery 1			*Lottery 2*		
Repeated Plays	1	5	50	1	5	50
Mean	55.09	274.05	2500.75	46.86	173.34	2058.18
Median	50	250	2500	40	207.5	2500
Acceptance Rates	88.6%	100%	97.7%	70.5%	79.5%	86.4%
within Wilcoxon z p-value	-5.721 (0.0000)***	-5.330 (0.0000)***	-5.254 (0.0000)***	-4.203 (0.0000)***	-4.838 (0.0000)***	-4.623 (0.0000)***
	Lottery 3			*Lottery 4*		
Repeated Plays	1	5	50	1	5	50
Mean	32.7	178.36	2259.32	49.84	237.59	2633.43
Median	60	51.5	2350	50	249.5	2500
Acceptance Rates	70.5%	79.5%	93.2%	100%	100%	100%
within Wilcoxon z p-value	-4.174 (0.0000)***	-5.195 (0.0000)***	-5.406 (0.0000)***	-5.731 (0.0000)***	-5.695 (0.0000)***	-5.695 (0.000)***

Table 1: *Mean and Median Certainty Equivalents.* Also shown are the z-value of a within Wilcoxon test for differences in the median and the corresponding p-values in parentheses. A positive z-value indicates that estimates are higher for 5 than for 50 trials. A negative z-value indicates the opposite. The upper left number belongs to a test between 1 and 5 plays and the upper right number to a test between 5 and 50 plays. The number in the middle gives the test statistic for a test between 1 and 50 plays.

---

[12]We report the value of the test statistic L and compute corresponding approximate p-values based on the χ^2_L-statistic given in Page (1963). Page (1963) also shows that this approximation works very well, especially in cases like ours with 44 subjects and 3 tasks. See Wellek (1989) for computing exact p-values.

[13]A repeated measures GLM analysis also indicated a highly significant effect for number of repeated plays: F(2.41) = 18.94, p<0.0001. There was no interaction between the effect of number of repeated plays and the lottery profile: F(6.37) = 1.24, p = 0.31.

4   Results: Estimation of Summary Statistics

The results shown in Table 2 provide strong support for hypotheses B1 and B2. Subjects correctly realized that the probability of a loss decreases as the number of repeated plays increased. Subjects also strongly overestimated the probability of loss, a result that is in line with Benartzi and Thaler (1999). This is highly significant for all lotteries using a binomial test. The only exception is Lottery 3 with 5 repeated plays where subjects' estimates of the probability of loss in the aggregate were relatively accurate.[14]

Table 2 shows the mean and median answers that respondents gave to the question of the lotteries' mean excess loss (MEL). Those responses showed a high standard deviation and a strong skew to the right. This suggests that MEL is a measure of risk that is difficult to evaluate. The skewed distribution of responses suggests that a minority of subjects had problems to understand the conditional nature of this risk measure. Some subjects gave absurdly high estimates, e.g., estimating an MEL of 12,000 for Lottery 3 with 50 repeated plays. As a result, mean and median MELs differ widely. Because medians are more resistant to outliers, we used the median responses in our statistical analyses.

We found no support for hypothesis B3, that is, MEL was underestimated instead of overestimated. Using a binomial test, median estimated values of MEL were significantly smaller than correct values for all lotteries and both sets of repeated plays, except for Lottery 3 at 5 repeated plays.

Table 2 shows support for hypothesis B4. Estimated median MELs are significantly higher for 50 than for 5 trials utilizing a Wilcoxon test, except for Lottery 3. The statistical insignificance in latter case comes from relatively high estimates for 5 iterations. Subjects might have anchored on the (high) value of the mean excess loss for the one-shot case (400) and may have adjusted only insufficiently for repeated plays.

From people's estimates of the median and 10th and 90th percentiles of the outcome distributions, we inferred their estimates of the outcome standard deviation using the following formula called the extended Swanson-Megill approximation (Keefer and Bodily 1983).

$$EstimatedVolatility = \sqrt{(0.3 \cdot Q_{0.1}^2 + 0.4 \cdot Q_{0.5}^2 + 0.3 \cdot Q_{0.9}^2) - mean^2}$$

---

[14]It should be noted, that the means and the medians of the probability of loss are always smaller for the direct question frame compared to the condition where it was inferred from the assessed probability of coming out ahead.

Probability of a loss

	Lottery 1		Lottery 2		Lottery 3		
Repeated Plays	5	50	5	50	5	50	
Mean	26.11%	21.86%	41.39%	27.45%	47.9%	29.79%	no losses
Median	25%	11%	37.5%	20%	40%	20%	for
Actual Value	18.75%	0.77%	23.75%	9.16%	36.72%	9.83%	Lottery 4
within Wilcoxon z	1.667		2.862		3.914		
p-value	(0.0955)		(0.0042)**		(0.0001)***		
Binomial p-value	0.0096**	0.0000***	0.0013**	0.0001***	0.6516	0.0000***	
Repeated Measures GLM Analysis				$F(1.42) = 4.09$, $p < 0.05$			
No interaction between the effect of No. of Rep. Plays and Lottery				$F(2.41) = 0.08$, $p = 0.91$			

Mean excess loss

	Lottery 1		Lottery 2		Lottery 3		
Repeated Plays	5	50	5	50	5	50	
Mean	148.75	786.7	211.89	910.05	524.61	1631.71	no losses
Median	110	100	100	250	300	275	for
Actual Value	250	406.54	500	687.33	396.28	860	Lottery 4
within Wilcoxon z	-2.105		-2.981		-0.718		
p-value	(0.0353)*		(0.0029)**		(0.4725)		
Binomial p-value	0.0001***	0.0488*	0.0000***	0.0488*	0.4514	0.0226*	
Repeated Measures GLM Analysis				$F(1.42) = 4.18$, $p < 0.05$			
No interaction between the effect of No. of Rep. Plays and Lottery				$F(2.41) = 0.94$, $p = 0.39$			

Standard deviation

	Lottery 1		Lottery 2		Lottery 3		Lottery 4	
Repeated Plays	5	50	5	50	5	50	5	50
Mean	264	1996	399	2672	406	2833	112	993
Median	310	1559	241	1629	347	1859	116	1084
Actual Value	335	1061	581	1837	581	1837	112	354
within Wilcoxon z	-5.473		-5.368		-5.153		-5.456	
p-value	(0.0000)***		(0.0000)***		(0.0000)***		(0.0000)***	
Binomial p-value	0.2912	0.2912	0.0096**	0.8804	0.0226*	1.0000	0.0488*	0.0013**
Repeated Measures GLM Analysis				$F(1.42) = 17.64$, $p < 0.0001$				
No interaction between the effect of No. of Rep. Plays and Lottery				$F(3.40) = 1.44$, $p = 0.24$				

Probability that the outcome of the lottery falls into a $\pm €100$ interval

	Lottery 1		Lottery 2		Lottery 3		Lottery 4	
Repeated Plays	5	50	5	50	5	50	5	50
Mean	47%	58%	45%	53%	46%	55%	56%	56%
Median	50%	60%	50%	60%	43%	60%	60%	60%
Approx. Value	21.27%	7.44%	11.98%	4.28%	11.98%	4.28%	58.28%	22.05%
within Wilcoxon z	-2.563		-2.461		-2.125		-0.358	
p-value	(0.0104)*		(0.0139)*		(0.0336)*		(0.7201)	
Binomial p-value	0.0001***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.6516	0.0000***
Repeated Measures GLM Analysis				$F(1.42) = 1.79$, $p = 0.13$				
Marginal interaction effect (with Lottery 4 showing no effect)				$F(3.40) = 1.979$, $p = 0.13$				

Perceived risk

	Lottery 1			Lottery 2		
Repeated Plays	1	5	50	1	5	50
Mean	49.27	38.43	30.16	65.20	48.69	41.57
Median	47.5	40	30	69	53	40
within Wilcoxon z p-value	2.873 (0.0041)**		2.721 (0.0065)**	4.192 (0.0000)***		2.950 (0.0032)**
		4.4378 (0.0000)***			5.153 (0.0000)***	

	Lottery 3			Lottery 4		
Repeated Plays	1	5	50	1	5	50
Mean	57.66	50.88	40.57	11.80	8.82	8.52
Median	60	51.5	40	0	0	0
within Wilcoxon z p-value	2.610 (0.0091)**		3.190 (0.0014)**	1.740 (0.0819)		0.539 (0.5901)
		3.969 (0.0001)***			1.822 (0.0684)	
Repeated Measures GLM Analysis				$F(2.41) = 2.53$, $p < 0.05$		
Marginal interaction effect (with Lottery 4 showing no effect)				$F(6.37) = 1.69$, $p = 0.14$		

Table 2: *Mean and Median Responses to All Elicited Possible Indices of Risk in Repeated Gambles.* The correct value of each statistic is also shown. Also shown are the z-value of a within Wilcoxon test for differences in the median and the corresponding p-values in parentheses. A positive z-value indicates that estimates are higher for 5 than for 50 trials. A negative z-value indicates the opposite. Also reported are the p-values of a two-sided binomial test of the hypothesis that the estimated risk measure was different from its actual value. For perceived risk, the upper left number belongs to a test between 1 and 5 plays and the upper right number to a test between 5 and 50 plays. The number in the middle gives the test statistic for a test between 1 and 50 plays. Also shown are the results of a repeated measure GLM analysis.

with

$$mean = 0.3 \cdot Q_{0.1} + 0.4 \cdot Q_{0.5} + 0.3 \cdot Q_{0.9}$$

Table 2 shows that subjects realized correctly that outcome standard deviation increases with the number of repeated plays, strongly supporting hypothesis B6. The p-value of a Wilcoxon test was p<0.0000*** for all four lottery profiles. Furthermore, we find no systematic overestimation of the standard deviation of the aggregated distributions. Subjects significantly underestimated this risk measure for lottery 2 and 3 with 5 iterations, and significantly overestimated the standard deviation for lottery 4 with both repeated plays. Beside this, no further comparison is significant. Note, that there is a large literature on overconfidence, which results in people reporting too narrow intervals on average (Lichtenstein et al. (1982), Griffin and Brenner (2004)). This general tendency may be at work here and it is possible that it cancelled out any systematic overestimation of outcome variability.

The next risk measure to be considered is people's estimate of the probability that the lottery outcome would fall into an interval of $\pm\text{\euro}100$ around the EV. As hypothesized by Samuelson (1963), subjects showed indeed evidence for the fallacy of large numbers, i.e., the estimated probability increased with the number of repeated plays. The result was significant for the first three lottery profiles and confirms hypothesis B5. Note that the overall effect of the repeated measure GLM analysis was only marginally significant, a result which was due to no effect for lottery 4. It should also be noted that people vastly overestimated the elicited probability and that the median estimates were remarkable stable across all four lottery profiles (i.e., 50% for 5 iterations and 60% for 50 iterations).

As discussed above, respondents' answers to the standard deviation and to the interval probability estimation question are inconsistent from a statistical perspective. That is, a correctly perceived increase in outcome standard deviation with a larger number of repeated plays should lead to lower estimates of the interval probability for more repeated plays. Our results show, however, that it is indeed possible that subjects overestimate the probability of a loss and believe in increases in outcome standard deviation, but nevertheless fallaciously believe in the law of large numbers. Given that respondents' estimates for these two possible indices of the risk of repeated-play lotteries point in different directions (i.e., one suggesting that the risk increases, the other that the risk decreases), it is interesting to examine what the self-reported perceived risk judgments of

16

our respondents tell us about the direction of perceived risk: Do subjects think that the risk increases with the number of trials because the standard deviation increases? Or do they believe that the risk decreases, possibly due to a fallacious application of the law of large numbers or because of the correct assessment that the probability of loss falls?

As shown in Table 2, repeated lotteries were perceived as less risky as the number of repeated plays increased.[15] The result is in line with hypothesis B7 and highly significant for the first three lottery profiles using a Page test (p<0.0000***), as is every pairwise comparison between perceived risk for the same lottery and different number of repeated plays. The insignificance of the result for lottery profile 4 (L=539; p=0.1205) comes mainly from the fact that some subjects perceive a lottery without losses as not risky at all (a result also observed for a subset of respondents by E. U. Weber et al. 1992), i.e. 24 out of 44 subjects assigned a value of 0 to the risk of Lottery 4 regardless of the number of repeated plays.

5   Results: Determinants of Risk Perception

As perceived risk is a psychological variable that might differ between individuals, the following analyses were performed at the individual subject level. Table 3 provides means and medians of Spearman and Pearson correlation coefficients between subjective judgments of perceived risk and each of six possible indices of risk. We consider two versions of the coefficient of variation. First, we define the CV as *true* standard deviation divided by the expected value. The second version is defined as the *estimated* standard deviation, i.e. the approximation of the standard deviation based on the elicited quantiles, divided by the expected value and is labelled "Estimated CV" in Table 3. For the probability of loss and the mean excess loss, we assumed that subjects correctly assessed the value from our graphical two-state distribution for single plays, i.e., that they inferred that the probability of loss was 0.5 and the mean excess loss €100 for a single play of Lottery 1. We also use correct values for the standard deviation and the fallacy questions for single

---

[15]Joag et al. (1990) report similar results. However, they used specifically constructed bets with the same expected values for single and multiple plays. This modification resulted in a much higher standard deviation in the single play condition because it makes it necessary to have a high loss with a non-negligible probability in the one-shot gamble. The results are therefore not directly comparable to our study.

plays, i.e., the probability that the outcome of an one-shot gamble of Lottery 3 lies in an $\pm \text{€}\,100$-interval around the expected value is 0.

Table 3 shows that the probability of a loss, mean excess loss, and both versions of the coefficient of variation are positively correlated with judgments of perceived risk. That is, the greater the probability of loss, the greater the mean excess loss, and the greater the coefficient of variantion, the greater are peoples judgments of perceived risk. The estimates of the probability that the overall outcome lies in a small interval around the long-run expected value are negatively correlated with judgments of perceived risk, i.e., the more people believe that the overall outcome lies in an $\pm \text{€}\,100$-interval around the long-run expected value, the less risky they consider the gamble to be. The coefficient of variation based on the true standard deviation and the probability of a loss showed the highest correlations with judgments of perceived risk. The only index of risk that does not show a significant correlation with judgments of perceived risk is the perceived standard deviation of outcomes.[16] This is disconcerting in light of the wide use of the standard deviation as a measure of risk in finance, but not at all surprising in light of previous behavioral evidence (Keller et al. 1986, E. U. Weber et al. 2004).[17]

Spearman correlation (perceived risk; risk measure) computed for every subject						
	Probability of Loss	Mean Excess Loss	Interval	standard deviation	CV	Estimated CV
Mean	0.64	0.55	-0.54	0.09	0.65	0.48
Median	0.67	0.63	-0.61	0.16	0.71	0.46
Pearson correlation (perceived risk; risk measure) computed for every subject						
	Probability of Loss	Mean Excess Loss	Interval	standard deviation	CV	Estimated CV
Mean	0.62	0.33	-0.52	-0.06	0.63	0.56
Median	0.68	0.37	-0.57	-0.08	0.64	0.57

Table 3: *Mean and Median of Spearman and Pearson Correlation Coefficients for the Correlation Between Perceived Risk and the Discussed Risk Measures, Computed Separately for Every Subject.* The analyses are based on 12 data points per subject and risk measure. Individual-subject correlations were significant at the (one-tailed) .05 level when the absolute value of the correlation exceeded 0.49. "Interval" refers to the probability that the overall outcome falls into a $\pm \text{€}\,100$-interval around the long-run expected value, "CV" stands for the coefficient of variation based on the true standard deviation and "Estimated CV" stands for the coefficient of variation based on the estimated standard deviation.

---

[16]Note that the true standard deviation is also only weakly correlated with perceived risk (Spearman: mean 0.16 and median 0.2; Pearson: mean 0.10 and median 0.16).

[17]There were sizeable individual differences in the ability of our risk indices to predict perceptions of risk, a fact that emphasizes the subjective nature of the concept of perceived risk. For example, for some subjects none of our risk measures correlated with the risk judgments; for several other subjects probability-of-a-loss and mean-excess-loss predict perceived risk nearly perfectly.

We also regressed perceived risk on all risk measures in a multiple regression at the aggregate level, i.e. across all subjects[18] and all lottery profiles. We used two regression equations, one for each definition of the coefficient of variation, i.e. "CV" and "Estimated CV", and investigated the incremental contribution of the predictor variables in explaining variability in perceived risk, given that all other variables are in the regression equation. Quadratic and interaction terms are also included. To be more precise, let X_i be one specific risk measure. We estimated the regressions (1) $PerceivedRisk = \alpha + \sum_{i=1}^{5} \beta_i \cdot X_i + \sum_{i=1}^{5} \gamma_i \cdot X_i^2 + \sum_{i=1}^{5} \sum_{j=i+1}^{5} \delta_{ij} \cdot X_i \cdot X_j$ and (2) $PerceivedRisk = \alpha + \sum_{i=1}^{4} \beta_i \cdot X_i + \sum_{i=1}^{4} \gamma_i \cdot X_i^2 + \sum_{i=1}^{4} \sum_{j=i+1}^{4} \delta_{ij} \cdot X_i \cdot X_j$, and tested then if the addition of risk measure X_5 in regression equation (1) explained significant more variation in the dependent variable than the regression equation (2).[19] This type of analysis does not allow us to conclude that one risk measure is more important as another. It simply states which risk measure is "unique" in the sense that this measure adds significantly to the explanation of variation in risk perception over and above the variables already in the regression equation. Note, that as we analyze the ability of the model to predict variation as a whole, and not coefficient estimation, multicollinearity is not an issue here. Table 4 reports on the results.[20] The results highlight the role of the probability of loss as a unique risk measure. Nevertheless, it also shows that every risk measure adds something and this further emphasizes the complexity of perceived risk.

6   Results: Does Perceived Risk Predict Preference?

Consistent with other recent perceived-risk and return analyses of risk behavior (e.g., E. U. Weber and Milliman (1997) and E. U. Weber and Hsee (1998)), the most important predictor of CEs in our study was the expected value of lotteries. We were interested in the additional ability of different measures of perceived risk to predict variability in CEs in a risk-return regression equation. To answer this question, we pooled the

---

[18]This was done because we did not have a sufficient number of data points for individual subject regressions. Problems with an aggregate analysis include the possibility of specification bias, i. e. the list of predictor variables may be too long or too short. In addition, we assume implicitly that all subjects are homogenous.

[19]See Gujarati (1995) or Chatterjee and Price (1991) for an intuitive description.

[20]We also did a regression analysis that included a dummy variable for every subject to account for the repeated measure design. This formulation of the regression model does not change the results qualitatively.

Regression A		
	F-value	p-value
Coefficient of variation (true standard deviation)	12.74	0.0000***
Probability of a loss	8.16	0.0000***
Prob. to end up in a interval around the long run expected value	3.58	0.0017**
Mean excess loss	2.60	0.0174*
Standard deviation	2.15	0.0466*
Regression B		
	F-value	p-value
Probability of a loss	24.42	0.0000***
Coefficient of variation (estimated standard deviation)	7.04	0.0000***
Mean excess loss	1.84	0.0899
Prob. to end up in a interval around the long run expected value	1.70	0.1187
Standard deviation	0.84	0.5371

Table 4: *Regression Analysis.* The table shows F- and p-value of a test that assessed the marginal contribution of every predictor variable to regression equation (1): $PerceivedRisk = \alpha + \sum_{i=1}^{5} \beta_i \cdot X_i + \sum_{i=1}^{5} \gamma_i \cdot X_i^2 + \sum_{i=1}^{5} \sum_{j=i+1}^{5} \delta_{ij} \cdot X_i \cdot X_j$ with the risk measures probability of a loss, mean excess loss, probability that the overall outcome falls into a small interval around the expected value, standard deviation, and cv. Regression B does exactly the same with the only difference that cv is substituted by the estimated CV.

data across subjects because of the lack of a sufficient number of observations on the individual level and estimated the linear regressions $CE = \alpha + \beta_1 \cdot EV + \beta_2 \cdot RiskMeasure$ for every risk measure in the spirit of a simple risk-value model (Sarin and M. Weber 1993). Table 5 gives F- and p-values for a test of the *marginal contribution* of every risk measure, analogous to the regression analysis in Section 5. We used this simple form of the regression model without interaction terms and an appropriate transformation of the predictor variables with respect to the simply linear form of risk-value models proposed in the literature.

	F-value	p-value
Perceived risk	8.05	0.0047**
Mean excess loss	6.81	0.0093**
Prob. to end up in a interval around the long run expected value	0.64	0.4239
Standard deviation	0.55	0.4579
Coefficient of variation (true standard deviation)	0.14	0.7061
Probability of a loss	0.12	0.7258
Coefficient of variation (estimated standard deviation)	0.04	0.8390

Table 5: *Regression Analysis.* The table shows F- and p-value of a test that assessed the marginal contribution of every risk measure to the regression equation $CE = \alpha + \beta_1 \cdot EV + \beta_2 \cdot RiskMeasure$.

As shown in Table 5, explicit judgments of perceived risk and mean excess loss are the only two measures of risk which significantly predict CE above and beyond the lotteries' expected value. Consistent with previous research (e.g., E. U. Weber et al. 2005), judgments of perceived risk (which are a complex amalgam of different aspects of risk, as shown in the previous section) are a better predictor of certainty equivalents than any single conventional index of riskiness.

7 Summary and Discussion

Our results have implications both for future research on risk taking and its determinants and for the practice of financial advisors. We summarize those lessons and implications by revisiting the research questions that we stated in the introduction.

(1) Are people able to correctly assess statistics of outcome distributions commonly used as indices of risk for repeated-play lotteries? We replicated the well-known fact that people overestimate the probability of a loss in repeated gambles. We also found that mean excess loss is underestimated, probably as the result of anchoring on the single-play loss and insufficient adjustment, and that this risk measure is difficult to evaluate. Our results thus confirm and add urgency to the recommendation by Benartzi and Thaler (2002) that investors ought to be presented with aggregated distributions that reflect the range of possible outcomes of their investment decisions, since they seem to lack the ability to infer many characteristics of final outcome distributions correctly. People who make decisions about their retirement savings may be simply not aware of the consequences their actions might have.

Aggregated distributions should help decision makers to assess the nature of risk and return of their decision alternatives and should help them to make "better" decisions. Computing, showing, and discussing aggregated distributions may have the potential to avoid utility losses in asset allocation decisions or other decisions involving repeated gambles. Further research needs to be done on how such computations may help, especially in financial decision making (see i. e. Klos and Weber 2005).

Furthermore, our data shows that people in fact fall prey to Samuelson's fallacy, while realizing correctly, and at the same time, that outcome standard deviation increases with the number of repeated plays. This paradoxical combination of beliefs is consistent with

a representation of multiple-play lottery outcome distribution that has fatter tails and more probability mass near the long-run expected value than it really has. Our results nevertheless show that the faulty application of the law of large numbers is a psychological reality and may well be responsible for people's greater willingness to choose between lotteries on the basis of expected value maximization, when the selected gamble could be played repeatedly, as in a study by Montgomery and Adelbratt (1982, p. 53).

Finally, our study is the first in the context of repeated gambles elicit more than a single risk measures and allows therefore a more complete picture of estimation errors utilizing a within subject design. Our results suggest that subjects show a general lack of coherence in judgments about probability distributions. Future research may profitably explore the degree of the lack of coherence by conducting an experiment where several clear-cut restrictions on the possible answers of the subjects are imposed.

(2) What constitutes perceived risk in repeated gambles? "Risk" rightfully is a core concept in financial decision making. Since the work of Markowitz (1952), risk has often been operationalized as standard deviation. Work by Keller et al. (1986), E. U. Weber (1997), E. U. Weber and Hsee (1988), and E. U. Weber and Milliman (1997) has questioned this operationalization from a descriptive perspective, showing that judgments of perceived risk that predict behavior do not coincide with standard deviation or standard deviation estimates. A natural follow-up that has already been discussed in different contexts by E. U. Weber (1988), Holtgrave and E. U. Weber (1993) and E. U. Weber et al. (2003, 2004) among others, is the question of what constitutes perceived risk. Our study provides an answer to this question in the context of repeated lotteries and thus, by extension, investment decisions with a longer time horizon. Interestingly and consistent with results in the above mentioned studies of single-play decisions, judgments of the perceived riskiness of repeated-play lotteries are largely unrelated to the estimated outcome standard deviation of these lotteries. Instead, our results are consistent with the Conjoint Expected Risk model by Luce and E. U. Weber (1986), which contains the probability of a loss and the mean excess loss as key ingredients in the mathematical specification of perceived risk and allows for differential weight to downside and upside variability. Furthermore, our data show that Weber et al's (2004) results that the CV (risk/standard deviation per unit of return/EV) better reflects people's intuitions about financial risk than the unstandardized standard deviation or variance also hold for the

domain of repeated gambles. Finally, our results show that perceived probability of ending up in a small interval around the long-run expected value also predicts perceived risk and that the fallacious application of the law of large numbers therefore is of relevance for how risky people think repeated gambles are.

The results in section 6 indicate that the effect of all indices of risk except for mean excess loss on preference or CEs is mediated by people's judgments of perceived risk. I.e., both probability of a loss and the CV strongly affect judgments of perceived risk (as described in section 5), but do not have any additional effect on CEs above and beyond their effect on perceived risk.

(3) Does perceived risk predict preference? Consistent with previous research in the context of single-play lotteries (e.g., E. U. Weber and Milliman 1997, E. U. Weber and Hsee 1998), respondents in our study provide lower certainty equivalents for gambles that they perceive to be riskier, validating the predictive power and practical relevance of perceptions of risk.[21] Further research needs to be done on risk-value models if risk is operationalized as perceived risk (E. U. Weber et al. 2005).

References

Albrecht, P., R. Maurer, U. Ruckpaul. 2001. Shortfall-risks of stocks in the long run. *Financial Markets and Portfolio Management* **15** 481 – 499.

Becker, G., M. DeGroot, J. Marschak. 1963. Measuring utility by a single-response sequential method. *Behavioral Sci.* **9** 226 – 232.

Benartzi, S., R. Thaler. 1999. Risk aversion or myopia? Choices in repeated gambles and retirement investments. *Management Sci.* **45** 364 – 381.

———— . 2002. How much is investor autonomy worth?. *J. Finance* **57** 1593 – 1616.

Brachinger, H.-W., M. Weber. 1997. Risk as a primitive: A survey of measures of perceived risk. *OR Spektrum* **19** 235 – 250.

Chatterjee, S., B. Price. 1991. *Regression Analysis by Example*. Hohn Wiley & Sons. New York.

---

[21]See e.g. E. U. Weber (1997) for a discussion on the practical relevance.

Goldstein, W., E. U. Weber. 1995. Content and discontent: Indications and implications of domain specificity in preferential decision making. in J. Busemeyer, R. Hastie, D. Medin. eds. *Decision Making from a Cognitive Perspective. The Psychology of Learning and Motivation.* Academic Press. New York. 83–136.

Gonzalez, R., G. Wu. 1999. On the shape of the probability weighting function. *Cognit. Psychol.* **38** 129 – 166.

Griffin, D., L. Brenner. 2004. Perspectives on probability judgment calibration. in D. Koehler, N. Harvey, eds. *Blackwell Handbook of Judgment and Decision Making.* Blackwell. 177 – 199.

Gujarati, D. 1995. *Basic Econometrics.* Mc Graw-Hill. Singapore.

Hertwig, R., G. Barron, I. Erev, E. U. Weber. 2004. Decisions from experience and the effect of rare events. *Psychological Sci.* **15** 534 – 539.

Holtgrave, D., E. U. Weber. 1993. Dimensions of risk perception for financial and health risks. *Risk Analysis* **13** 553 – 558.

Joag, S., J. Mowen, J. Gentry. 1990. Risk perception in a simulated industrial purchasing task: The effects of single versus multi-play decisions. *J. Behavioral Decision Making* **3** 91 – 108.

Klos, A. 2004. The investment horizon and dynamic asset allocation - some experimental evidence. *Econom. Letters* **85** 167 – 170.

——— , M. Weber. 2005. Portfolio choice in the presence of nontradeable income: An experimental analysis. Working Paper. University of Mannheim.

Keefer, D., S. Bodily. 1983. Three-point approximations for continuous random variables. *Management Sci.* **29** 595 – 609.

Keller, C. R. Sarin, R. K., M. Weber. 1986. Empirical investigation of some properties of the perceived riskiness of gambles. *Organ. Behavior and Human Decision Processes* **38** 114 – 130.

Kilka, M., M. Weber. 2001. What determines the shape of the probability weighting function under uncertainty? *Management Sci.* **47** 1712 – 1726.

Kritzman, M. 1994. What practitioners need to know about time diversification. *Financ. Anal. J.* **50** 14 – 18.

———— , D. Rich. 2002. The mismeasurement of risk. *Financ. Anal. J.* **58** 91 – 99.

Langer, T., M. Weber. 2001. Prospect theory, mental accounting, and differences in aggregated and segregated evaluation of lottery portfolios. *Management Sci.* **47** 716 – 733.

———— . 2005. Myopic prospect theory vs. myopic loss version – How general is the phenomenon?. *J. Economic Behavior and Organization* **56** 25 – 38.

Lichtenstein, S., B. Fischhoff, L. D. Phillips. 1982. Calibration of probabilities: The state of the art to 1980. in D. Kahneman, P. Slovic, A. Tversky. eds. *Judgment under uncertainty: Heuristics and Biases*. Cambridge University Press. 306 – 334.

Loewenstein, G., E. U. Weber, C. Hsee, E. Welch. 2001. Risk as feelings. *Psychological Bulletin* **127** 267 – 286.

Lopes, L., G. Oden. 1999. The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *J. Math. Psych.* **43** 286 – 313.

Luce, D., E. U. Weber. 1986. An axiomatic theory of conjoint, expected risk. *J. Math. Psych.* **30** 188 – 205.

Markowitz, H. 1952. Portfolio selection. *J. Finance* **7** 77 – 91.

Montgomery, H., T. Adelbratt. 1982. Gambling decisions and information about expected value. *Organ. Behavior and Human Performance* **29** 39 – 57.

Page, E. B. 1963. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *J. Amer. Statistical Assoc.* **58** 216 – 230.

Samuelson, P. 1963. Risk and uncertainty: A fallacy of large numbers. *Scientia* **98** 108 – 113.

———— . 1989. The $\sqrt{N}$ law and repeated risktaking. in T. W. Anderson, S. Karlin. *Probability, statistics, and mathematics: Papers in honor at Samuel Karlin*. Academic Press. 291 – 306.

Sarin, R., M. Weber. 1993. Risk-value models. *European J. Oper. Res.* **70** 135 – 149.

Shafir, S., A. Bechar, E. U. Weber. 2003. Cognition-mediated coevolution: Context-dependent evaluations and sensitivity of pollinators to variability in nectar rewards. *Plant Systematics and Evolution* **238** 195 – 209.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representations of uncertainty. *J. Risk and Uncertainty* **5** 297 – 323.

Weber, E. U. 1988. A descriptive measure of risk. *Acta Psychologica* **69** 185 – 203.

———. 1997. The utility of measuring and modelling perceived risk. in A. Marley, ed. *Choice, Decisions, and Measurements: Essays in Honor of R. Duncan Luce.* Kluwer Academic Press. Norwell. 45 – 57.

———. 2001a. Decision and choice: Risk, empirical studies. in N. Smelser, P. Baltes, eds. *International Encyclopedia of the Social and Behavioral Sciences.* Elsevier Science Limited. Oxford. 13347 – 13351.

———. 2001b. Personality and risk taking. in N. Smelser, P. Baltes, eds. *International Encyclopedia of the Social and Behavioral Sciences.* Elsevier Science Limited. Oxford. 11274 – 11276.

———. 2004. Perception Matters: Psychophysics for Economists. in J. Carrillo, I. Brocas, eds. *Psychology and Economics.* Oxford University Press. Oxford. 165 – 176.

———, C. Anderson, M. Birnbaum. 1992. A theory of perceived risk and attractiveness. *Organ. Behavior and Human Decision Processes* **52** 492 – 523.

Weber, E. U., A.-R. Blais, N. Betz. 2002. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *J. Behavioral Decision Making* **15** 263 – 290.

Weber, E. U., C. Hsee. 1998. Cross-cultural differences in risk perception, but cross-cultural similarities in attitudes towards perceived risk. *Management Sci.* **44** 1205 – 1217.

Weber, E. U., R. Milliman. 1997. Perceived risk attitude: Relating risk perception to risky choice. *Management Sci.* **43** 123 – 144.

Weber, E. U., S. Shafir, A.-R. Blais. 2004. Predicting risk-sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review* **111** 430 – 445.

Weber, E. U., N. Siebenmorgen, M. Weber. 2005. Communicating asset risk: How the format of historic volatility information affects risk perception and investment decisions. *Risk Analysis* forthcoming.

Weber. M. 2001c. Decision and choice: Risk, theories. in N. Smelser, P. Baltes, eds. *International Encyclopedia of the Social and Behavioral Sciences.* Elsevier Science Limited. Oxford. 13364 – 13368.

Wellek, S. 1989. Computing exact p-values in page's nonparametric test against trend. *Biometrie und Informatik in Medizin und Biologie* **20** 163 – 170.