How Do Beliefs about Skill Affect Risky Decisions? *

Adrian Bruhin

University of Lausanne

Luís Santos-Pinto

University of Lausanne

David Staubli

University of Lausanne

January 4, 2018

JEL Codes: C91; D81

Keywords: Individual risk taking behavior; Self-confidence; Laboratory experiment

* Luís Santos-Pinto (Corresponding author), University of Lausanne, Faculty of Business and Economics (HEC Lausanne), Internef 535, CH-1015, Lausanne, Switzerland, Ph: 41-216923658, Fax: 41-216923435, LuisPedro.SantosPinto@unil.ch; Adrian Bruhin, University of Lausanne, Faculty of Business and Economics (HEC Lausanne), Switzerland, Adrian.Bruhin@unil.ch; David Staubli, University of Lausanne, Faculty of Business and Economics (HEC Lausanne), Switzerland, David.Staubli@unil.ch

How do Beliefs about Skill Affect Risky Decisions?

Abstract

Beliefs about relative skill matter for risky decisions such as market entry, career choices, and financial investments. Yet in most laboratory experiments risk is exogenously given and beliefs about relative skill play no role. We use a laboratory experiment without strategy confounds to isolate the impact of beliefs about relative skill on risky choices. We find that low (high) skill individuals are more (less) willing to take risks on gambles where the probabilities depend on relative skill than on gambles with exogenously given probabilities. This happens because low (high) skill individuals overestimate (underestimate) their relative skill. Consequently, the wrong people may engage in risky activities where performance is based on relative skill while the right people may be crowded out.

1 Introduction

In most laboratory experiments risk is exogenously given and individuals' beliefs about relative skill play no role in their decisions. In the real world, however, beliefs about relative skill matter for many decisions. Examples include, entering a market where post-entry payoffs depend on relative skill (e.g., opening a restaurant) versus staying out and earning a certain amount (e.g., working as a waiter); following a career path where performance is highly dependent on relative skill (e.g., being a lawyer, or a musician) versus choosing a career path where performance does not depend much on it (e.g., taking a public sector job); or managing your own financial portfolio versus delegating this to an asset manager depends on one's perception of relative skill at picking financial assets.

This paper uses a laboratory experiment to investigate how beliefs about relative skill affect risky decisions. To answer this question, we elicit certainty equivalents (CEs) of luck and skill gambles. Both types of gambles are binary as they involve two possible prizes. In a luck gamble the probability of getting the higher prize is given while in a skill gamble it corresponds to the subject's relative skill, measured by relative performance in a cognitive ability test. Consequently, behavior in the luck gambles only depends on preferences towards risk whereas behavior in the skill gambles depends on preferences towards risk as well as beliefs about relative skill.

We begin with a model free analysis showing that low (high) skill subjects have higher (lower) CEs of skill gambles than of luck gambles that offer the same prizes and winning probabilities. This indicates that low (high) skill subjects are more (less) willing to take risks on gambles where the probabilities depend on relative skill than on gambles with exogenously given probabilities. In contrast, we find that intermediate skill subjects have similar CEs of skill gambles and luck gambles that offer the same prizes and winning probabilities.

Next, we investigate how these results relate to subjects' beliefs about relative skill. We do so by relying on two types of beliefs: stated and revealed. For eliciting the stated beliefs, we ask subjects to report the complete belief distribution about their relative performance in the cognitive ability test. This allows subjects to demonstrate their degree of confidence in their self-placement and sidesteps recent methodological concerns associated with previous research (Merkle and Weber (2011), Benoît et al. (2015)). We incentivize the stated beliefs with a quadratic scoring rule (QSR) which is valid under linear utility and the absence of probability weighting. The advantage of the stated beliefs is that they do not rely on a specific model. However, they may be biased if subjects are either risk averse or weight probabilities non-linearly.[1] In contrast, the revealed beliefs are directly estimated from the subjects' choices via a structural model and do not rely on the QSR.

Stated beliefs about relative skill display three main patterns. First, on average, there is a slight tendency towards overplacement, i.e. overestimation of relative skill. Second, however, the majority of subjects has biased beliefs: 70.8% of them state a belief which is more than one standard deviation away from their actual relative skill; 31.7% of them display a belief distribution which does not contain the actual relative skill. Third, the biases in stated beliefs correlate with relative skill: low skill subjects overplace themselves while high skill subjects underplace themselves. This suggests that subjects' beliefs about relative skill drive differences in behavior in the skill and luck gambles.

Next we estimate a structural model based on Cumulative Prospect Theory (Kahneman and Tversky (1979), Tversky and Kahneman (1992)).[2] The structural model has three advantages over the model free analysis. First, it allows us to disentangle the different components of risk preferences – utility curvature, likelihood sensitivity, and optimism/pessimism. Second, it allows us to rule out that potential correlations between the subjects' risk preferences and relative skill confounded our results. Finally, it provides us with estimates of the subjects' revealed beliefs about relative skill which do not rely on the QSR. Instead, the revealed beliefs are directly estimated from the subjects' choices. The identifying assumption is that subjects apply the same utility and probability weighting functions for evaluating the luck and skill gambles.

The structural model provides us with joint estimates of revealed beliefs and risk preference parameters. To model utility curvature we use the power utility function, and to model probability weighing we use the two-parameter probability weighting function in Goldstein and Einhorn (1987). The estimated parameters tell us that subjects display moderate degrees of concavity in the utility function, pronounced likelihood insensitivity, and a slight degree of pessimism. These estimates are plausible given the exist-

[1] Schlag et al. (2015) discuss the most prominent elicitation methods, their underlying assumptions, and provide theoretical comparisons.

[2] Since we only consider gambles in the gain domain, Cumulative Prospect Theory coincides with Rank Dependent Utility introduced by Quiggin (1982).

ing laboratory evidence on individual risk preferences (see Wakker (2010)). The revealed beliefs confirm the patterns obtained with the stated beliefs: overall, there is a slight tendency towards overplacement, the majority of subjects has biased beliefs about relative skill, low skill subjects overplace themselves while high skill subjects underplace themselves. However, the correlation between revealed and stated beliefs is just 0.46, and biases in revealed beliefs predict differences between CEs of luck and skill gambles better than biases in stated beliefs. This indicates that relying exclusively on stated beliefs may be misleading.

Our paper directly adds to the understanding of the observed low returns to entrepreneurship as documented by Hamilton (2000) and Moskowitz and Vissing-Jørgensen (2002).[3] The observed low returns to entrepreneurship stand in contrast to the predictions of the benchmark occupational choice model between paid employment and entrepreneurship by Lucas Jr (1978). In this model, subjects differ in their ability as entrepreneurs and choose between working for a wage or operating a firm. In equilibrium, those with low skills select into paid employment and those with high skills select into entrepreneurship. Our main result suggests that excess entry into entrepreneurship of low skill subjects due to overplacement and under entry of high skill subjects due to underplacement leads to a misallocation of talent which drives down the returns to entrepreneurship. For theoretical papers discussing similar mechanisms see De Meza and Southey (1996), Manove (1995), Fraser and Greene (2006), Rigotti et al. (2011)). In these models, excess entry into entrepreneurship of low skill subjects drives up input prices, which lowers the returns to entrepreneurship and, ultimately, crowds out high skill subjects.

The remainder of the paper is organized as follows. Section 2 discusses the paper's contribution to the literature. Section 3 describes our experimental design. Section 4 presents the model free results. Section 5 introduces the structural model and Section 6 presents its results. Finally, Section 7 concludes.

[3]For better readability, we use the term entrepreneurship to also refer to self-employment.

2 Literature

In this section, we outline the strands of literature to which our paper contributes. Regarding terminology, we follow Moore and Healy (2008) and distinguish between three different types of biases in subjects' beliefs about skill: (i) overestimation of absolute skill, (ii) overplacement (overestimation of relative skill or the "better-than-average effect"), and (iii) excessive confidence in the precision of ones beliefs (overprecision or miscalibration). Our paper focuses mostly on overplacement but also measures overprecision.

2.1 Market Entry

The first strand of literature the paper contributes to is about the decision whether or not to enter a market. Evidence gained from observational data indicates that overplacement might play an important role behind the low returns from entrepreneurship (Hamilton (2000), Moskowitz and Vissing-Jørgensen (2002)), over-entry into markets (Dunne et al. (1988), Mata and Portugal (1994)), as well as phenomena such as excess trading in financial assets (Odean (1999), Glaser and Weber (2007)).

This evidence has spurred several experimental studies that analyze the role of perceptions about skill on market entry. Most of these studies analyze market entry in strategic settings. Camerer and Lovallo (1999) study market entry games in which a small number of players chooses simultaneously whether to enter a market or to stay out. Post-entry payoffs depend either on a random device or on the subjects' relative skill in a trivia quiz. They find more entry in the condition where post-entry payoffs depend on relative skill and that subjects do not underestimate the number of entrants. However, they do not explicitly elicit subjects' beliefs about relative skill. Moore and Cain (2007) build on Camerer and Lovallo (1999) but manipulate and measure self-placement directly using simple and difficult trivia quizzes. Their evidence suggests that, relative to a condition with random ranks, overplacement in an easy trivia quiz leads to more entry, while underplacement in a difficult trivia quiz leads to less entry. Cain et al. (2015) investigate the entry decision into two markets, in which the chance of winning a single prize depends on the number and relative skill of entrants. In one market the relative skill depends on performance in an easy quiz, while in the other it depends on performance in a difficult quiz. They find overplacement in the easy quiz, underplacement in the difficult quiz, and a preference for entering the market where relative skill depends on the easy quiz.

Our paper contributes to this literature by studying the role that beliefs about relative skill play on risky decisions in a non strategic setting. This is relevant for understanding entry in competitive markets where post-entry payoffs depend on relative skill but strategic considerations play no role. Examples include occupational choices in competitive labor markets or the decision to become an entrepreneur (Lucas Jr (1978), Jovanovic (2004)). Moreover, as our setting is free of strategic considerations, we avoid confounds that may arise from ill-defined equilibrium concepts in entry games where post-entry payoffs depend on relative skill.

2.2 Measures of Beliefs about Skill

2.2.1 Economics

The paper also contributes to the literature in economics that aims at measuring beliefs about relative skill. Most of this literature relies on stated beliefs. Moore and Healy (2008) and Park and Santos-Pinto (2010) elicit beliefs about relative skill using a QSR with deterministic rewards for accurate guesses. Moore and Healy (2008) find overplacement on easy quizzes, underplacement on difficult quizzes, and no over- or underplacement on intermediate quizzes. Park and Santos-Pinto (2010) find overplacement by poker and chess players in real world tournaments. Clark and Friesen (2009) elicit subjective beliefs about relative skill using a QSR but replace the deterministic rewards with a probabilistic reward. This ensures truth-telling for any utility function under expected utility. Eil and Rao (2011) use the QSR with deterministic rewards and repeated feedback to obtain posterior beliefs about intelligence and beauty. They find that subjects' posterior beliefs are less sensitive to negative than to positive feedback. Mobius et al. (2011) use matching probabilities – a variation of the reservation-price elicitation mechanism of Becker et al. (1964) – and repeated feedback to obtain posterior beliefs about relative performance in an IQ test. The matching probabilities method induces truth telling for any utility function under expected utility. They find that subjects engage in biased Bayesian updating.

A study that does not exclusively rely on stated beliefs is Hoelzl and Rustichini (2005). They implement a choice based measure of beliefs about relative skill that relies on a voting game. Subjects vote on two payment procedures: a test and a lottery. If the majority votes for the test, subjects in the upper half of the scores in a skill test win a prize, otherwise each subject can win the same prize with 50% probability. The focal equilibrium is to vote for the test if one thinks to have an above-median skill, and to vote for the lottery otherwise. There is overplacement (underplacement) when more (less) than half of the subjects vote for the test. They find a substantial discrepancy between this choice based measure of beliefs and stated beliefs, which may be due to the strategic nature of the voting game.

We add to this literature by proposing a new choice based measure of beliefs about relative skill. Using a structural model based on revealed preference we jointly estimate risk preferences and revealed beliefs. This offers two advantages. First, the revealed beliefs avoid the potential problems of the QSR and matching probabilities caused by non-linear utility and probability weighting. Second, as the revealed beliefs are based on individual choices, they are free of strategic confounds.

2.2.2 Social Psychology

Social psychologists have also been interested in measuring subjects' beliefs about relative skills. In a typical experiment, subjects are either asked to judge whether they believe themselves to be above average in a certain domain (e.g., Svenson (1981)) or to provide a point estimate of the percentile of a distribution they believe themselves to belong to (e.g., Dunning et al. (1989)). The distribution of percentile point estimates is then compared to the uniform distribution. This yields a population measure of overplacement.

The main findings of the psychological studies can be summarized as follows.[4] First, subjects overplace themselves on the vast majority of positively connoted skills and traits (Myers (1987)). Second, the more ambiguous or vague is the definition of the skill or trait under evaluation, the greater is the tendency for overplacement (Felson (1981), Dunning et al. (1989)). Third, low skill subjects overplace themselves, high skill subjects underplace themselves, and low skill subjects are worse at evaluating their relative skills than high skill subjects (Kruger and Dunning (1999), Krueger and Mueller (2002)). Fourth, people display overplacement in easy tasks and underplacement in hard tasks (Kruger (1999)).

We contribute to this literature by (i) using monetary incentives to obtain stated beliefs about relative skill and (ii) eliciting the complete belief distribution which reveals the subjects' confidence in their self-placement. This sidesteps the methodological concerns arising from the lack of financial incentives for truth telling and the fact that with a single point estimate a lot of information gets lost (e.g., the degree of confidence in the stated belief). Our findings – a slight tendency towards overplacement; low skill subjects overplacing themselves, high skill subjects underplacing themselves; and low skill subjects displaying larger biases in beliefs than high skill subjects – confirm the "unskilled and unaware" hypothesis of Kruger and Dunning (1999) according to which the high skill individuals are better informed about their skills than the low skill individuals.

2.3 Errors in Self-Placement and Bayesian Rationality

Our paper is also related to the literature that investigates whether errors in self-placement are compatible with Bayesian updating. When information about relative skill is imperfect, errors in self-placement can arise from rational Bayesian updating. If people are better informed about their own skill than about the skill of others, the following pattern emerges: In difficult tasks, they overestimate their absolute skill and underestimate their relative skill, while in easy tasks, they underestimate their absolute skill and overestimate their relative skill. This is called the "regression" hypothesis (Dawes and Mulford (1996), Erev et al. (1994), Larrick et al. (2007), Moore and Small (2007), Moore and Healy (2008)).

[4] For a review of the psychological evidence on overplacement see Santos-Pinto and Sobel (2005). For a summary of the four main criticisms that have been made to this literature see Benoît et al. (2015).

Benoît and Dubra (2011) distinguish apparent from true overplacement. Apparent overplacement is consistent with rational Bayesian updating whereas true overplacement is not. They also show that rational Bayesian updating imposes testable implications on how the distribution of judgements about relative skill should be related to the distribution of true skill. Moore and Healy (2008) and Clark and Friesen (2009) find no evidence for true overplacement. In contrast, Merkle and Weber (2011) and Burks et al. (2013) find evidence for true overplacement.

Since we do neither manipulate the level of difficulty of the cognitive ability test nor collect estimates of absolute performance, our experiment has nothing to say about the "regression" hypothesis. However, by applying the allocation function of Burks et al. (2013) to the modes of the individual belief distributions, we find evidence of true overplacement of low skill subjects and true underplacement of high skill subjects.

2.4 Gender Effects

Finally, our experiment also contributes to the literature on gender differences in self-placement. Some papers focussing on strategic settings report pronounced gender differences in self-placement and in behavior. Niederle and Vesterlund (2007) study the effect of gender and overplacement on the willingness to compete. They find that the stronger preference of men for entering a competition can be partly explained by their stronger tendency to overplace themselves. Charness et al. (2017) study how gender interacts with statements about confidence in two-player tournaments. They find that only men inflate stated confidence when it is strategically optimal to deter the other player from entering. However, both men and women deflate stated confidence when it is strategically optimal to lure the other player into entering.

In contrast, in our non-strategic setting, gender differences are much weaker or even absent. There is only a a small but statistically significant gender difference in stated beliefs: on average, men overplace themselves whereas women do neither overplace nor underplace themselves. However, there is no gender difference in revealed beliefs: both men and women overplace themselves. This indicates that gender differences in self-placement may only arise under strategic interaction.

3 Experimental Design

The experiment consists of three stages which are explained in detail below. In the first stage, each subject completes a cognitive ability test and is assigned to a skill level according to her relative performance in the test. In the second stage, we elicit CEs of luck and skill gambles. Finally, in the last stage, we elicit subjects' beliefs about relative skill.

3.1 Cognitive Ability Test and Assignment of Subjects to Skill Levels

Subjects perform a Raven's matrices test with 20 questions (Raven (2000)). In each question, subjects are asked to identify the missing element that completes a pattern out of six possible choices. The difficulty of the questions varies in order to create heterogeneity in the test score results. Each subject had 20 minutes to complete the test. To incentivize the subjects to do their best, we paid CHF 0.50 per correct answer and told subjects that a good test result could help them earning more money later on during the experiment.[5]

A subject's relative performance in the cognitive ability test determines her relative skill among the 20 participants. The relative skill of each participant is assigned to one of 10 skill levels, θ, from the set $\{0.05, 0.15, \ldots, 0.85, 0.95\}$ as follows: the two best subjects (the tenth decile) in the room are allocated to the 0.95 skill level, the two next best subjects (the ninth decile) in the room to the 0.85 skill level, etc., and the two worst subjects (the first decile) in the room to the 0.05 skill level.[6] Later on, these skill levels will correspond to the winning probabilities in the skill gambles.

The assignment of subjects to skill levels was explained in detail. After reading the instructions subjects answered comprehension questions about the assignment. They could only go on after answering these questions correctly. While they were making their risky decisions, subjects could click on a button to see the list showing which winning probability was implied by which skill level. For further details please refer to the Experimental Material Appendix.

3.2 Luck and Skill Gambles

After the cognitive ability test and the assignment of subjects to skill levels, we elicited CEs of luck and skill gambles. A luck gamble pays CHF x_h^L with probability p and CHF x_l^L with probability $1-p$. We vary p from 5% to 95% in 10 percentage-points-increments across three prize combinations (x_h^L, x_l^L)

[5] At the time of the experiment, one CHF corresponded to roughly 1.05 USD.

[6] To break ties, subjects had to give their best guess about the sum of nine numbers that were displayed on matrix for 10 seconds. If this did not resolve the tie we broke the ties randomly. Note that since there were two subjects for each skill level, the tiebreaker was not relevant in every case. The performance in the estimation questions that breaks the tie between the 3^{rd} and the 4^{th} is inconsequential since rank 3 and rank 4 give rise to the same skill level.

from the set

$$\{(140, 0), (120, 20), (100, 40)\} . \tag{1}$$

The 10 winning probabilities p and the three prize combinations yield a total of 30 luck gambles.

In contrast, a skill gamble pays CHF x_h^S with probability θ_i and CHF x_l^S with probability $1 - \theta_i$, where θ_i is the skill level of subject i which is endogenously determined by her relative performance in the cognitive ability test. Since the winning probability θ_i corresponds to the subject's skill level we can only vary the prizes of the skill gambles. We consider nine different prize combinations (x_h^S, x_l^S) from the set

$$\{(180, 0), (150, 30), (120, 60),\ (140, 0), (120, 20), (100, 40),\ (100, 0), (80, 20), (60, 40)\} . \tag{2}$$

Note that for the prize combinations $(140, 0)$, $(120, 20)$, and $(100, 40)$ there are corresponding luck gambles. We will exploit this correspondence to directly compare CEs between luck and skill gambles.

To elicit CEs of both luck and skill gambles, we use the multiple price list format.[7] We have subjects choosing between a series of certain payoffs and either a luck or a skill gamble. The series of certain payoffs covers the payoff range of the gamble's prizes and decreases in 18 $(= 19 - 1)$ equally sized amounts. Subjects typically start by preferring the first certain payoff to the gamble and then switch to the gamble before the last certain payoff. The arithmetic mean of the last certain payoff preferred to the gamble and the first certain payoff not preferred to the gamble determines the CE of the gamble. In the example depicted in figure 1, the CE is approximated as $(84 + 91)/2 = 87.50$. If the subject always chooses the gamble the CE is approximated as $(140 + 133)/2 = 136.50$. Likewise, if she always chooses the certain payoff the CE is approximated as $(7 + 0)/2 = 3.50$.

We impose a unique switching point per multiple price list by automatically filling in all choices following the first switch (for details see Andersen et al. (2006) and Tanaka et al. (2010)). This has two advantages. First, it allows to determine the CE for every gamble. Second, it substantially reduces the number of clicks given the high number of gambles.[8]

The luck and skill gambles were presented in blocks. Across sessions the order of the blocks was balanced to cancel out potential order effects (see Charness et al. (2012)). Within a given block, the gambles were ordered according to two dimensions: attractiveness (first-order stochastic dominance) and prize spread $(x_h - x_l)$. To control for order effects across these two dimensions, we assign one

[7]Early applications of multiple price lists to elicit risk preferences include Binswanger (1980), Binswanger (1981), Murnighan et al. (1987), Murnighan et al. (1988), Beck (1994), Gonzalez and Wu (1999), and Holt and Laury (2002). See Andersen et al. (2006) for a more extensive discussion.

[8]In terms of additional assumptions, a unique switching point imposes weak monotonicity of the utility function, i.e. subjects weakly prefer a higher amount of money to a lower amount of money. Nevertheless, the unique switching points do not prevent choices across gambles that are inconsistent with first-order stochastic dominance.

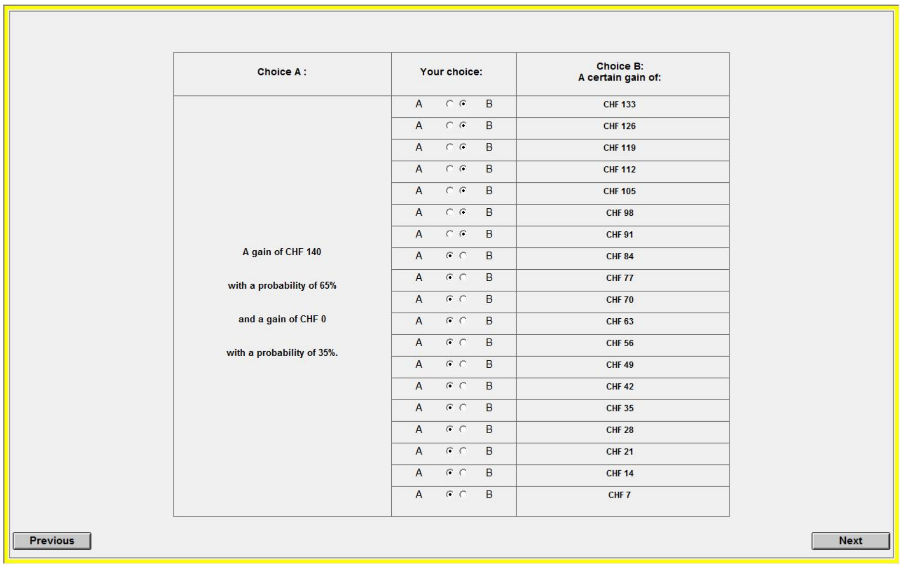


Figure 1: Interface of the luck gamble with x_h = 140 and x_l = 0 and a winning probability of 0.65. In this case, the elicited CE is CHF 87.50 (the arithmetic mean of CHF 84 and CHF 91).

fourth of the subjects to each of the four configurations: (i) increasing attractiveness and increasing spread, (ii) increasing attractiveness and decreasing spread, (iii) decreasing attractiveness and increasing spread, and (iv) decreasing attractiveness and decreasing spread.

As the main focus of the experiment is on risky choices we always elicit risky decisions before beliefs.

3.3 Belief Elicitation

After making risky decisions, we elicit each subject's entire belief distribution across the 10 skill levels. The subject can move sliders to shift probability mass across each skill level which are represented by a histogram in real-time. We incentivize the truthful disclosure of beliefs with a quadratic scoring rule (QSR) where the maximum payment is CHF 10 and the minimum is CHF 0 (Kadane and Winkler, 1988). Our main measure of subject i's stated belief about relative skill is the mean belief $\mu_i = \sum_{m=1}^{10} f_i(\theta_m)\theta_m$, where $f_i(\theta_m)$ stands for i's stated probability mass of having skill level θ_m, with $m = 1, ..., 10$. We also elicit each subject's mode of her belief distribution across skill levels. We ask each subject to provide a point estimate of her relative skill in the cognitive ability test and pay CHF 5 for a correct guess and CHF 0 for an incorrect guess. This allows us to distinguish rational from irrational placement using the allocation function of Burks et al. (2013). For further details see Appendix A.

Note, however, that the QSR assumes that subjects exhibit a linear utility function and weight probabilities linearly. If the subjects' utility function is concave, they will have an incentive to spread their beliefs too much compared to the true belief distribution and their mean beliefs will be less extreme than their true mean beliefs. If subjects engage in probability weighting their stated beliefs will reflect subjectively distorted probabilities instead of objective ones. In the discussion of the model free results we will first ignore these issues. However, we will come back to them in the context of the structural model.

3.4 Subjects and Payments

The experiment was performed at the University of Lausanne in April and May 2015 in the computer lab using the software z-Tree (Fischbacher, 2007). All subjects were students of the University of Lausanne and the École Polytechnique Fédérale de Lausanne (EPFL), recruited via ORSEE (Greiner, 2015). The experiment was conducted in 16 separate sessions with 20 subjects each. One subject left without finishing the experiment. Thus, our data comprises 319 subjects.[9]

In order to provide incentives for truthful revelation of preferences and beliefs, subjects were randomly paid for one of their risky decisions (Azrieli et al. (forthcoming) show that this is the only incentive-compatible way to use multiple price lists). We use the prior incentive scheme (Prince) proposed by Johnson et al. (2015). At the start of the experiment, subjects were told that they would be paid for only one of their decision situations and that decision situation was contained inside a closed envelope that they drew randomly from a box. Inside each envelope there is a decision sheet with the multiple price list corresponding to that decision. Subjects were told that they could not open the envelope before the payment stage. In the payment stage, each subject was paid according to the decision situation in the envelope, the subject's choice in that decision situation, the subject's performance on the cognitive ability test (if the decision situation involved a skill gamble and the subject preferred the skill gamble to the certain amount) and the realization of two ten-sided dice (if the subject preferred the luck or skill gamble to the certain amount).

Each subject received a show-up fee of CHF 10. Subjects were paid in private, one-by-one at the end of the experimental session. The information on a subject's earnings from each task and total earnings was printed on a sheet of paper that was shown to each subject. Total earnings per subject varied between CHF 17.50 and CHF 205.00 with a mean of CHF 105.85 and a median of CHF 113.50. Each session lasted for approximately two hours, comprising roughly 90 minutes for decisions and 30 minutes for payments. Total earnings paid to subjects across the 16 sessions were CHF $34,213$.[10] Compared to most other experimental work on individual decision making these are high stakes per hour which should ensure that, despite the large number of decisions, subjects take each of them properly into consideration. The instructions can be found in the Experimental Material Appendix.

[9] The subject left after finishing the cognitive ability test. The skill levels are thus determined including that subject.

[10] Since we needed exactly 20 subjects per session, we overbooked each session and sent home the supernumerous subjects after compensating them with the show-up fee of CHF 10.

4 Model Free Results

In this section, we discuss our model free results. We begin by comparing the CEs of skill and luck gambles that offer the same prizes and winning probabilities in order to investigate whether risk behavior differs if performance is determined by relative skill rather than by pure luck. Subsequently, we analyze whether the differences in risk behavior are related to the subjects' stated beliefs about relative skill.

4.1 Certainty Equivalents of Comparable Skill and Luck Gambles

Here, we compare the CEs of skill and luck gambles that offer the same prizes and winning probabilities. For each subject, there are always three comparable pairs of skill and luck gambles offering prizes from the set $\{(140, 0), (120, 20), (100, 40)\}$. For example, for a subject with skill level $\theta_i = 0.55$, we compare the CEs of the three skill gambles with the CEs of the three luck gambles that have a corresponding winning probability of $p = 0.55$. If subjects have correct beliefs about their skills, the CE of each comparable skill and luck gamble should be the same as the gambles would not differ in any dimension. This comparison of skill and luck gables with yields our first result.

Result 1 *On average, the CEs of skill gambles are similar to the CEs of comparable luck gambles. However, subjects in the lowest three skill quintiles (from the 5% to the 55% skill levels) have higher CEs of skill gambles than of luck gambles, and subjects in the two highest skill quintiles (from the 65% to the 95% skill levels) have lower CEs of skill gambles than of luck gambles.*

We find that the average CE of the skill gambles from the set $\{(140, 0), (120, 20), (100, 40)\}$ is CHF 57.105, while the average CE of the comparable luck gambles is CHF 57.807. The difference amounts to CHF -0.702 and is not significantly different from 0 (a t-test on the difference in means using subject cluster-robust standard errors yields a p-value = 0.576).

However, a more disaggregate look at the data reveals a correlation between relative skill and differences between CEs of skill and luck gambles. In figure 2 we divide subjects into quintiles according to their skill levels for skill and luck gambles offering prizes $(140, 0)$.[11] The figure reveals that low skill subjects have higher CEs of skill gambles than luck gambles, while for high skill subjects the reverse happens. The first regression shown in Table 1 confirms this correlation over all three comparable pairs of skill and luck gambles. The dependent variable, CE_{ij}^{diff}, represents the difference between subject i's CE of skill gamble j and her CE of luck gamble j. The independent variables are indicators for the skill quintiles. The constant is taken out to avoid perfect multicollinearity.

The interpretation of the coefficients is straightforward: subjects in the lowest skill quintile, corresponding to skill levels $\{0.05, 0.15\}$, value the skill gambles on average CHF 16.768 more than the

[11] The analogous figures for the skill and luck gambles with the prizes $(120, 20)$ and $(100, 40)$, respectively, are available in Appendix B and display a similar pattern.

CE_{ij}^{diff}	(1)	(2)	(3)
Skill {0.85, 0.95}	-25.380***		-25.268***
	(1.695)		(2.096)
Skill {0.65, 0.75}	-9.740***		-8.667***
	(1.592)		(1.955)
Skill {0.45, 0.55}	4.495***		5.058***
	(1.164)		(1.589)
Skill {0.25, 0.35}	10.709***		10.965***
	(1.433)		(1.523)
Skill {0.05, 0.15}	16.768***		17.892***
	(1.653)		(2.007)
Skill {0.85, 0.95}×Female			-0.311
			(3.577)
Skill {0.65, 0.75}×Female			-3.121
			(3.369)
Skill {0.45, 0.55}×Female			-1.503
			(2.274)
Skill {0.25, 0.35}×Female			-0.645
			(3.172)
Skill {0.05, 0.15}×Female			-3.559
			(3.535)
Female		-1.882	
		(1.755)	
Constant		-0.028	
		(1.040)	
Observations	955	955	955
R-squared	0.345	0.001	0.347

Subject cluster-robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1: Regressions where the dependent variable is the difference between the CE of skill gamble, i.e., CE_{ij}^{S}, and the CE of the corresponding luck gamble, i.e., CE_{ij}^{L}. The independent variables are skill quintile dummies, a female dummy and interaction terms of skill quintile dummies and the female dummy. Coefficients can be interpreted in CHF.

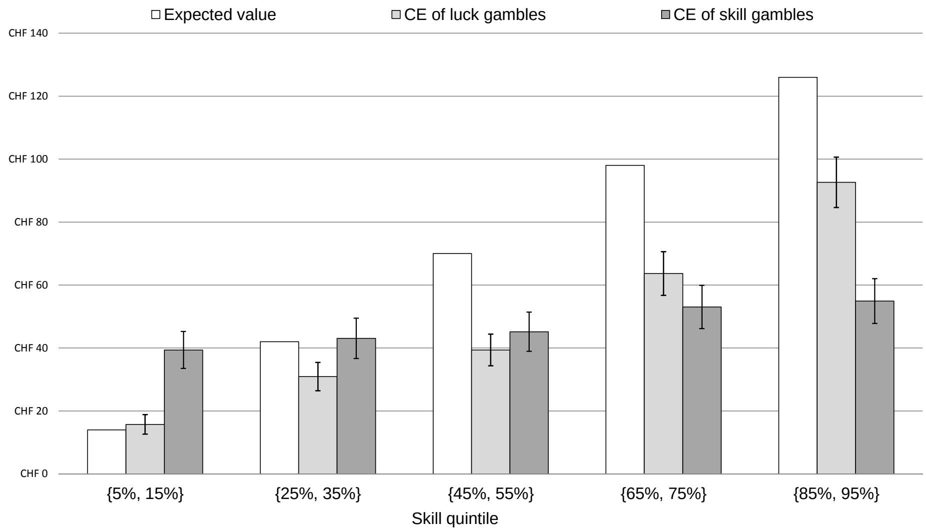


Figure 2: Average expected values of the luck gambles (white bars), average CEs of the luck gambles (light grey bars), and average CEs of the skill gambles (dark grey bars). Depicted is the gamble with winning prize $x_h = 140$ and losing prize $x_l = 0$. The 95%-confidence intervals are based on robust standard errors. Number of observations = 319, number of females = 114.

corresponding luck gambles. As skill quintiles increase, the difference between the CEs of skill and luck gambles decreases in absolute value down to CHF -25.380 for the top quintile. The numbers in parentheses display the subject cluster-robust standard errors. There are 955 observations, three observations per subject less two observations that were not recorded during the experiment ($955 = 3 \times 319 - 2$).

The second and third regressions in Table 1 show that gender has no effect on the differences in CEs between skill and luck gambles. At first glance this may seem at odds with previous literature showing gender differences in the tendency to compete (Niederle and Vesterlund (2007), Charness et al. (2017)). However, we conjecture that this discrepancy to the previous literature may be due to our non-strategic setting.

In sum, the first result indicates that low (high) skill subjects are more (less) willing to take risks on gambles where the winning probabilities depend on relative skill than in gambles where they depend on luck.

4.2 Relationship between Risk Behavior and Stated Beliefs

We now analyze if the differences in risk behavior across skill and luck gambles of low and high skill subjects are related to their stated beliefs. This yields our second result.

Result 2 *On average, subjects display a slight tendency towards overplacement. However, the majority of subjects state biased beliefs about their relative skill: Subjects in the two lowest skill quintiles (from* *the 5% to the 35% skill level) overplace themselves, and subjects in the two highest skill quintiles (from the 65% to the 95% skill level) underplace themselves. In addition, subjects are excessively certain about their relative skill.*

The average stated belief indicates a slight tendency towards overestimation of relative skill. On average, subjects state that their individual winning probability is 0.54. This is slightly but significantly different from the true average winning probability of 0.5 (a t-test on the mean difference between subjects' stated beliefs and subjects' true winning probabilities using robust standard errors yields a p-value = 0.005).

This slight tendency towards overplacement needs a comment. At first glance it seems to be at odds with a significant body of literature where overplacement is a well-established fact (Lichtenstein et al. (1977), Svenson (1981), Camerer and Lovallo (1999), Van den Steen (2004), Santos-Pinto and Sobel (2005), Burks et al. (2013)). However, overplacement is not an ubiquitous phenomenon. Experimental evidence shows that people overplace themselves on easy tasks but underplace themselves on difficult tasks (see Kruger (1999)). A likely reason why we only find a moderate tendency towards overplacement is that our cognitive ability test is a moderate difficulty task. On the one hand, in a trivially simple test, we would expect (close to) 20 correct answers for each subject. On the other hand, in a very difficult test we would expect, on average, $20/6 = 3.33$ correct answers as subjects would have to guess randomly. The observed average in our sample is 12.66 correct answers with a standard deviation of 2.80, i.e. only slightly above the middle of the two extremes.[12]

Although, on average there is only a slight tendency towards overplacement, a majority of 70.8% of subjects state biased beliefs about their relative skill. For these subjects, the true skill level falls outside the average stated belief plus/minus one standard deviation of the belief distribution. We also find that for 31.7% of subjects the true skill level falls outside the support of their stated belief distribution.

In addition, a disaggregate look at the stated beliefs reveals that low skill subjects overplace themselves and high skill subjects underplace themselves. Table 2 shows regressions estimating how over- and underplacement depend on relative skill. The dependent variable is the bias in stated beliefs of each subject, corresponding to the difference between stated beliefs and actual relative skill, i.e.

[12]The distribution of test scores is approximately normal (the p-value of a D'Agostino's Chi-squared test is 0.534). The minimum score was 2 and the maximum score was 19. That is, no subject got all 20 questions correct, so there is no bunching at the top of the score distribution. Moreover, there is no evidence of a gender difference in the average test score, wich is 12.63 for males and 12.71 for females.

Across the sixteen sessions, there were 87 ties between two or more subjects after answering the cognitive ability test. After the tiebreaker there were 18 ties left that had to be decided by means of the random number generator. Per session there are a total of $\binom{20}{2} = 190$ bilateral comparisons. Across the 16 sessions this amounts to 3040. The 87 ties account for 2.86% and the 18 random tie breaks account for 0.59% of all bilateral comparisons.

b_i^{stated}	(1)	(2)	(3)
Skill {0.85, 0.95}	-0.262***		-0.227***
	(0.017)		(0.019)
Skill {0.65, 0.75}	-0.141***		-0.133***
	(0.022)		(0.026)
Skill {0.45, 0.55}	0.031*		0.067***
	(0.018)		(0.023)
Skill {0.25, 0.35}	0.213***		0.260***
	(0.019)		(0.020)
Skill {0.05, 0.15}	0.381***		0.399***
	(0.017)		(0.019)
Skill {0.85, 0.95}×Female			-0.097***
			(0.034)
Skill {0.65, 0.75}×Female			-0.023
			(0.049)
Skill {0.45, 0.55}×Female			-0.095***
			(0.036)
Skill {0.25, 0.35}×Female			-0.118***
			(0.039)
Skill {0.05, 0.15}×Female			-0.058
			(0.038)
Female		-0.085***	
		(0.032)	
Constant		0.074***	
		(0.019)	
Observations	319	319	319
R-squared	0.714	0.022	0.736

Note: Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 2: Regressions where the dependent variable is the bias in stated beliefs, i.e., $b_i^{stated} \equiv \mu_i - \theta_i$. The independent variables are skill quintile dummies, a female dummy and interaction terms of skill quintile dummies and the female dummy. Coefficients can be interpreted in CHF.

$b_i^{stated} \equiv \mu_i - \theta_i$. The first regression reveals that subjects in the two lowest skill quintiles, $\{0.05, 0.15\}$ and $\{0.25, 0.35\}$, tend to overplace themselves. In contrast, subjects in the two highest skill quintiles, $\{0.65, 0.75\}$ and $\{0.85, 0.95\}$, tend to underplace themselves. Subjects in the intermediate skill quintile, $\{0.45, 0.55\}$, display slight but only marginally significant tendency towards overplacement. The second regression shows that, on average, women are less confident than men and do not overplace themselves. Thus, the slight tendency towards overplacement in the aggregate is driven by men. However, the interaction terms in the third regression show no clear pattern in this gender difference across skill levels.

The pattern of stated beliefs depicted in Table 2 could be broadly consistent with subjects being uniformed about their relative skill. To rule out this possibility, we apply a chi-square test (Snedeckor and Cochran (1989)) to compare each subject's stated belief distribution to the uniform distribution, which represents complete lack of information about relative skill. The corresponding chi-squared test has the following test statistic:

$$X = \sum_{m=1}^{10} \frac{(w_m - Nq_m)^2}{Nq_m}$$

where $N = 100$, as there are 100 percentage points to be allocated; $q_m = 0.1$, $\forall m$, as according to the null hypothesis of a uniform distribution there is a 10% probability weight on every skill level; and w_m is the number of percentage points that a subject puts on skill level m. The null hypothesis that stated beliefs are uniformly distributed can be rejected at both the 5%- and the 1%-significance level for all but 3 subjects. At 10%-significance level the null hypothesis can be rejected for all but one subject.

We also find support for the "unskilled-unaware" hypothesis by Kruger and Dunning (1999), according to which high skill subjects are better informed about their skill than low skill subjects. The absolute value of the bias in stated beliefs is on average 0.062 points smaller for high skill subjects, i.e. subjects with $\theta_i \geq 0.55$, than for low skill subjects, i.e. subjects with $\theta_i < 0.55$ (p-value < 0.001 for the corresponding t-test with robust standard errors).

In sum, subjects' stated beliefs indicate that on average subjects only slightly overplace themselves, however the majority is either over- or underplacing: low skill subjects overplace themselves whereas high skill subjects underplace themselves. In addition, subjects are excessively certain about their relative skill. Together with Result 1, this suggests that overplacement (underplacement) is the reason why low (high) skill subjects are more (less) willing to take risks on gambles where the winning probabilities depend on relative skill than in gambles where they depend on luck.

5 Structural Model

We now discuss the structural model which has three main advantages compared to the model free approach. First, it allows us to disentangle the different components of the subjects' risk preferences: (i) utility curvature, (ii) likelihood sensitivity, and (iii) optimism/pessimism. Second, it allows us to rule out that our results are confounded by potential correlation between subjects' risk preferences and their relative skill. Third, it allows us to elicit revealed beliefs which are immune to the potential issues affecting stated beliefs.

The structural model is based on Cumulative Prospect Theory (Kahneman and Tversky (1979), Tversky and Kahneman (1992)) since this model has proven to be descriptively superior to other models (see Starmer (2000)). According to Cumulative Prospect Theory (CPT), the value of a gamble X that pays x_h with probability p and x_l with probability $1-p$ is

$$V(X) = w(p)u(x_h) + [1 - w(p)]u(x_l),$$

where $u(\cdot)$ is the utility function and $w(\cdot)$ is the probability weighting function. We specify the utility function using the power function that implies constant relative risk aversion (CRRA), the benchmark approach on risk attitudes in most empirical work:

$$u(x;r) \begin{cases} \frac{x^{(1-r)}}{1-r}, & r \neq 1 \\ \ln(x), & r = 1 \end{cases},$$

where x is the monetary outcome of a gamble and the parameter r measures the curvature of the utility function: $r > 0$ corresponds to concavity, $r = 0$ to linearity, and $r < 0$ to convexity.

To specify the probability weighting function, we use the two-parameter function proposed by Goldstein and Einhorn (1987):

$$w(p; \delta, \gamma) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma},$$

where γ and δ reflect the degree of probability weighting. The parameter γ governs likelihood sensitivity. If $\gamma \in (0, 1)$ the function captures the inverse s-shape pattern where low winning probabilities are upweighted and high winning probabilities are downweighted. If $\gamma > 1$ we have an s-shape pattern where low winning probabilities are downweighted and high winning probabilities are upweighted. The parameter δ reflects the degree of optimism or pessimism. When $\delta \in (0, 1)$ the subject is a pessimist, since she downweights the probability of the high prize x_h and upweights the probability of the low prize x_l. When $\delta > 1$ the subject is an optimist, since she upweights the probability of the high prize x_h and downweights the probability of the low prize x_l. When $\gamma = \delta = 1$ there is no probability weighting and $w(p; 1, 1) = p$.

As we intend to estimate risk preferences at the individual level, we compute the CEs implied by our decision model for luck and skill gambles. Subject i's CE of a luck gamble $j = 1, ..., 30$ that pays

x^L_{hj} with probability p_j and x^L_{lj} with probability $1-p_j$, where p_j is the exogenously given winning probability, is implicitly defined by:

$$u(\hat{CE}^L_{ij}; r^L_i) = w(p_j; \gamma^L_i, \delta^L_i)u(x^L_{hj}; r^L_i) + [1 - w(p_j; \gamma^L_i, \delta^L_i)]u(x^L_{lj}; r^L_i), \tag{3}$$

where $\hat{CE}^L_{ij}$ is subject i's CE of luck gamble j predicted by the model. Similarly, subject i's CE of a skill gamble $k = 1, ..., 9$ that pays x^S_{hk} with probability θ_i and x^S_{lk} with probability $1 - \theta_i$, where θ_i is the probability of winning for subject i determined by her relative skill is implicitly defined by:

$$u(\hat{CE}^S_{ik}; r^S_i) = w(\theta_i; \gamma^S_i, \delta^S_i)u(x^S_{hk}; r^S_i) + [1 - w(\theta_i; \gamma^S_i, \delta^S_i)]u(x^S_{lk}; r^S_i),$$

where $\hat{CE}^S_{ik}$ is subject i's CE of skill gamble k predicted by the model. The prize combinations (x^L_{hj}, x^L_{lj}) are given by (1), (x^S_{hk}, x^S_{lk}) by (2), and $p_j = 0.05, 0.15, ..., 0.85, 0.95$.

In the luck gambles both prizes and winning probabilities are known to subjects. The estimation of the parameters is thus straightforward in this case and can be implemented in a similar way as in Bruhin et al. (2010). In skill gambles however, the subjects do not know their winning probability as they do not know their relative skill for sure. If we estimate the parameters of the utility and probability weighting functions using the objective winning probabilities, θ_i, we would obtain erroneous estimates since biases in beliefs would induce distortions in the estimates of these parameters. To illustrate, consider the case of a subject i who has an objective winning probability $\theta_i = 0.15$ but who believes her winning probability to be 0.45. If this subject evaluates the skill gambles based on her biased belief and we use the objective winning probability in our estimation, we would draw the wrong conclusion that subject i is very optimistic or insensitive to changes in probabilities. Thus, we replace the objective winning probability θ_i by a parameter ξ_i that reveals subject i's belief about her relative skill and needs to be estimated. We refer to the parameter ξ_i as subject i's revealed belief.

The main assumption for identifying the subject's revealed belief is that the utility and probability weighting functions remain stable across luck and skill gambles. Formally, this identifying assumption corresponds to:

$$r^S_i = r^L_i = r_i; \quad \gamma^S_i = \gamma^L_i = \gamma_i; \quad \delta^S_i = \delta^L_i = \delta_i; \forall i$$

Using the identifying assumption and rearranging equation (3) we obtain an expression for subject i's predicted CE of luck gamble j:

$$\hat{CE}^L_{ij} = u^{-1}\Big(w(p_j; \gamma_i, \delta_i)u(x^L_{hj}; r_i) + [1 - w(p_j; \gamma_i, \delta_i)]u(x^L_{lj}; r_i); r_i\Big).$$

Analogously, the predicted CE of skill gamble k corresponds to:

$$\hat{CE}^S_{ik} = u^{-1}\Big(w(\xi_i; \gamma_i, \delta_i)u(x^S_{hk}; r_i) + [1 - w(\xi_i; \gamma_i, \delta_i)]u(x^S_{lk}; r_i); r_i\Big).$$

To estimate the parameters of the structural model we decompose the CEs into the predicted CEs and a random error – $CE_{ij}^L = \hat{CE}_{ij}^L + \varepsilon_{ij}^L$ for luck gambles, and $CE_{ik}^S = \hat{CE}_{ik}^S + \varepsilon_{ik}^S$ for skill gambles – and apply the method of maximum likelihood (for further details see Appendix C).

The individual estimations allow us to directly compare the revealed beliefs ξ_i – estimated with the structural model – to the true skill level, θ_i, for each subject. We define the bias in revealed beliefs as $b_i^{revealed} = \xi_i - \theta_i$. If subject i's revealed belief is correct and she uses it for evaluating the skill gambles, then $b_i^{revealed} = 0$. However, if she evaluates the skill gambles based on a too high (low) revealed belief, then $b_i^{revealed} > 0$ ($b_i^{revealed} < 0$).

6 Structural Model Results

We now discuss the results from estimating the structural model. First, we decompose the subjects' risk preferences. Second, we rule out that correlations between risk preferences and relative skill are a potential confound of our results. Third, we examine the revealed beliefs and analyze how they correlate with the stated beliefs. Finally, we summarize some robustness checks.

6.1 Decomposition of Risk Preferences

Table 3 shows the summary statistics of the individual parameter estimates. These parameter estimates allow us to disentangle the different components of the subjects' risk preferences: (i) utility curvature, (ii) likelihood sensitivity, and (iii) optimism/pessimism.

As shown in the first columns, the average subject's risk preferences display the following components: a moderate degree of concavity in the utility function, pronounced likelihood insensitivity, and a slight degree of pessimism. This finding indicates that it is important to correct for risk preferences both through utility curvature and non-linear probability weighting.[13]

At the individual level, there is substantial heterogeneity in these components as indicated by the standard deviation in the third column and illustrated in figure 3. Regarding gender differences, the p-values in the last column of Table 3 indicate that on average women are pessimistic while men are neither optimistic nor pessimistic. There are no other significant gender differences which, as discussed earlier, may be due to our non-strategic setting.

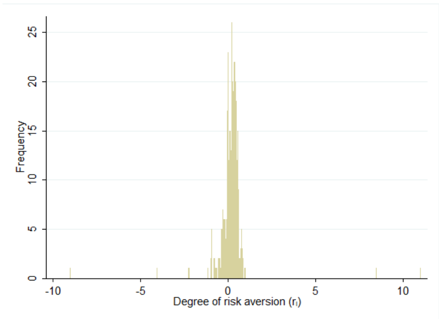
6.2 Potential Correlation between Risk Preferences and Relative Skill

The structural model also allows us to examine potential correlations between risk preferences and relative skill that could have confounded our main results. To do so, we estimate an OLS regression with the subjects' relative skill, θ_i, as the dependent variable and the risk preference parameters as independent variables.

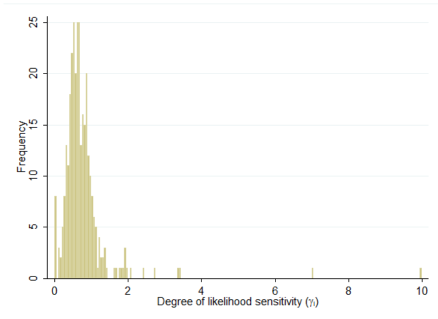
[13] Andersen et al. (2014) estimate subjective beliefs about ambiguous events concerning the 2008 U.S. presidential election using a similar approach. They jointly estimate risk preferences and subjective beliefs using a range of lottery choices and belief elicitation tasks and find, like we do, that it is important to correct for risk preferences.

	Mean	Median	Standard Deviation	Min	Max	p-value
Utility curvature (r_i)	0.20	0.24	1.03	-9.00	11.00	
Female	0.08	0.24	0.97	-9.00	1.00	0.123
Male	0.26	0.24	1.05	-4.04	11.00	
Likelihood sensitivity (γ_i)	0.77	0.65	0.76	0.00	10.00	
Female	0.79	0.66	0.73	0.00	7.02	0.676
Male	0.75	0.65	0.78	0.00	10.00	
Optimism-pessimism (δ_i)	0.90	0.71	1.03	0.00	10.00	
Female	0.72	0.61	0.58	0.00	3.39	0.006
Male	1.00	0.77	1.21	0.03	10.00	
Revealed beliefs (ξ_i)	0.54	0.55	0.21	0.05	0.95	
Female	0.53	0.53	0.23	0.05	0.95	0.590
Male	0.54	0.55	0.20	0.05	0.95	
Scaling of error term variance (luck) (κ_i^L)	0.08	0.07	0.05	0.00	0.32	
Female	0.09	0.07	0.05	0.03	0.32	0.034
Male	0.07	0.07	0.04	0.00	0.32	
Scaling of error term variance (skill) (κ_i^S)	0.11	0.10	0.06	0.00	0.47	
Female	0.12	0.11	0.06	0.00	0.31	0.175
Male	0.11	0.09	0.07	0.00	0.47	

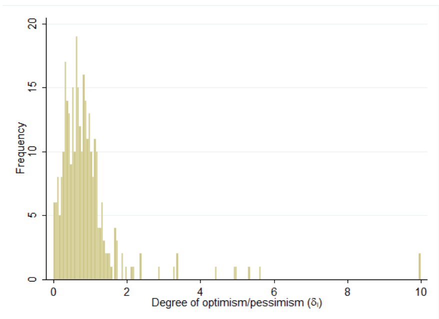
Table 3: Summary statistics of the individual parameter estimates across all subjects and for females and males separately. The p-value refers to a t-test on the difference in means between females and males using robust standard errors. Number of observations = 319, number of females = 114.



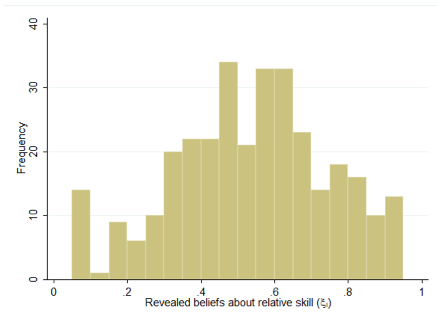
(a)



(b)



(c)



(d)

Figure 3: Histogram of estimated (a) degrees of utility curvature (r_i), (b) degrees of likelihood sensitivity (γ_i), (c) degrees of optimism-pessimism (δ_i), and (d) revealed beliefs (ξ_i). The bar width is 0.05. Number of observations = 319, number of females = 114.

Relative Skill (θ_i)	
Degree of risk aversion (r_i)	-0.003
	(0.014)
Degree of likelihood sensitivity (γ_i)	0.017
	(0.015)
Degree of optimism/pessimism (δ_i)	0.007
	(0.020)
Constant	0.482***
	(0.026)
Observations	319
R-squared	0.002
p-value of overall F-Test	0.721

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4: OLS regression of relative skill, θ_i on individually estimated risk preference parameters. Relative skill is not correlated with risk preferences as the overall F-Test has a p-value of 0.721, meaning that risk preference parameters are jointly insignificant.

Table 4 summarizes the result of the regression. The risk preference parameters – i.e., utility curvature, likelihood sensitivity, and optimism/pessimism – are insignificant both individually and jointly (the p-value of the corresponding overall F-Test is 0.721). Hence, there is no evidence for a correlation between risk preferences and relative skill, and we can rule out correlations between risk preferences and relative skill are a potential confound of our results.

6.3 Revealed Beliefs

We now discuss the subjects' revealed beliefs which are immune to the potential issues affecting the stated beliefs, as they are directly estimated by the structural model and do not rely on the QSR.

Table 5 exhibits regressions summarizing the bias in revealed beliefs per skill quintile. The first regression shows that subjects in the two lowest skill quintiles, $\{0.05, 0.15\}$ and $\{0.25, 0.35\}$, overplace themselves. In contrast, subjects in the two highest skill quintiles, $\{0.65, 0.75\}$ and $\{0.85, 0.95\}$, underplace themselves. Finally, subjects in the intermediate skill quintile, $\{0.45, 0.55\}$, display a slight but insignificant tendency towards overplacement. In sum, this yields Result 3.

Result 3 *The revealed beliefs confirm the pattern found in the stated beliefs (Result 2): On average, subjects display slight tendency towards overplacement; Subjects in the two lowest skill quintiles (from the 5% to the 35% skill level) overplace themselves, and subjects in the two highest skill quintiles (from*

$b_i^{revealed}$	(1)	(2)	(3)
Skill {0.85, 0.95}	-0.270***		-0.270***
	(0.025)		(0.031)
Skill {0.65, 0.75}	-0.175***		-0.144***
	(0.029)		(0.035)
Skill {0.45, 0.55}	0.063**		0.070**
	(0.025)		(0.031)
Skill {0.25, 0.35}	0.190***		0.201***
	(0.025)		(0.028)
Skill {0.05, 0.15}	0.374***		0.353***
	(0.026)		(0.029)
Skill {0.85, 0.95}×F			0.000
			(0.052)
Skill {0.65, 0.75}×F			-0.090
			(0.063)
Skill {0.45, 0.55}×F			-0.018
			(0.052)
Skill {0.25, 0.35}×F			-0.028
			(0.056)
Skill {0.05, 0.15}×F			0.068
			(0.062)
Female		-0.020	
		(0.038)	
Constant		0.043**	
		(0.021)	
Observations	319	319	319
R-squared	0.570	0.001	0.576

Note: Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Regression where the dependent variable is the bias in revealed beliefs, i.e., $b_i^{revealed} \equiv \xi_i - \theta_i$. The independent variables are skill quintile dummies, a female dummy and interaction terms of skill quintile dummies and the female dummy. Coefficients can be interpreted in CHF.

the 65% to the 95% skill level) underplace themselves.

Interestingly, in contrast to the regressions reported in Table 2 on the bias in stated beliefs, Regressions 2 and 3 show that there is no gender difference in the bias in revealed beliefs. This suggests that in non-strategic settings gender differences in self-placement may play only a minor or even no role. This suggestion is further supported by the lack of a significant gender effect in differences in CEs between luck and skill gambles (see Table 1).

Additionally, the results on revealed beliefs also support the "unskilled-unaware" hypothesis by Kruger and Dunning (1999). The absolute value of the bias in revealed beliefs is on average 0.037 points smaller for high skill subjects, i.e. subjects with $\theta_i \geq 0.55$, than for low skill subjects, i.e. subjects with $\theta_i < 0.55$ (p-value $= 0.077$ for the corresponding t-test with robust standard errors).

We now examine the correlation between stated and revealed beliefs, which yields Result 4.

Result 4 *Stated and revealed beliefs are positively correlated. However, this correlation is not perfect, and biases in revealed beliefs better predict differences in the subjects' CEs in the skill and luck gambles than do the biases in stated beliefs.*

Note that if subjects use their stated beliefs to evaluate the winning probabilities in skill gambles, we expect a positive correlation between stated, μ_i, and revealed beliefs, ξ_i. This is indeed the case, as shown in figure 4 which exhibits a scatterplot and a regression line illustrating this positive correlation. The estimated slope of the regression line (robust standard errors in parentheses),

$$\mu_i = \underset{(0.024)}{0.358} + \underset{(0.043)}{0.348}\xi_i,$$

confirms this positive correlation which amounts to 0.461.

However, the large dispersion of observations in the scatterplot indicates that the correlation between revealed and stated beliefs is not perfect. This indicates that many subjects make choices that are inconsistent with their stated beliefs. The estimates of the structural model point to two reasons behind this inconsistency which are both related to the shortcomings of the QSR. First, the average subject exhibits a concave utility function which gives her an incentive towards stating mean beliefs that are less extreme than her true mean beliefs. Second, the average subject also engages in non-linear probability weighting which invalidates the QSR. Even so, this does not affect our main results which are qualitatively robust across stated and revealed beliefs.

Moreover, the bias in revealed beliefs is a better predictor for the differences in the subjects' CEs in the skill and luck gambles than the bias in stated beliefs. Figure 5 shows the scatterplots and regression lines illustrating the relationship between the difference of the CEs in the skill and luck gambles and the biases in stated (left-hand side) and revealed (right-hand side) beliefs. The correlation of the differences of the CEs in the skill and luck gambles with the bias in revealed beliefs is 0.796, significantly higher

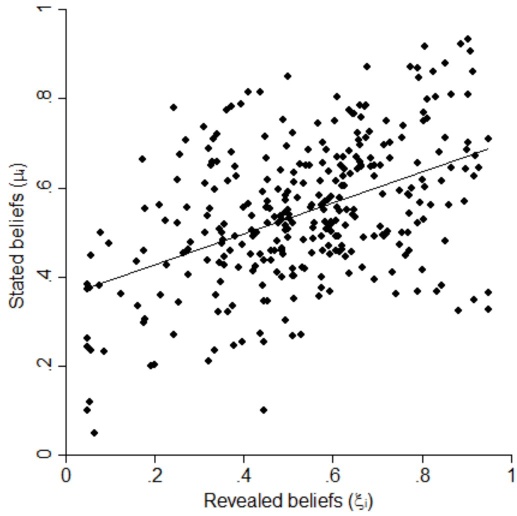
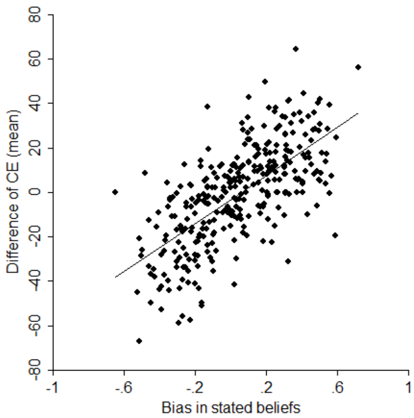
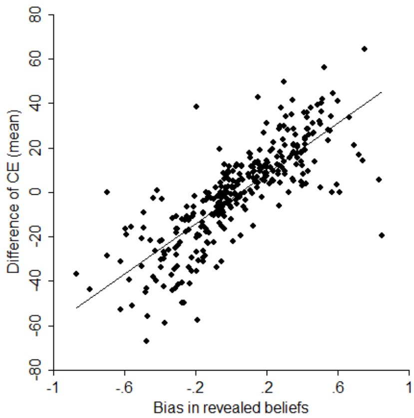


Figure 4: Relation between stated beliefs, μ_i, and revealed beliefs, ξ_i. Scatterplot and regression line. Number of observations = 319, number of females = 114.



(a)



(b)

Figure 5: (a) Relation between the mean of the differences of CEs between the three comparable pairs of skill and luck gambles and the bias in stated beliefs (b_i^{stated}). (b) Relation between the mean of the differences of CEs between the three comparable pairs of skill and luck gambles and the bias in revealed beliefs ($b_i^{revealed}$). Both panels show the scatterplot and the regression line. Number of observations = 319, number of females = 114.

than the corresponding correlation of 0.672 with the bias in stated beliefs. A Steiger's Z-test (Steiger (1980)) on the two correlation coefficients being equal yields a p-value < 0.001.

In sum, the comparison between stated and revealed beliefs provides two main insights: first, although stated and revealed beliefs are correlated, the former should be interpreted with caution as they are affected by the limitations of the QSR. Second, biases in revealed beliefs are a better predictor of the difference between the CEs of skill and luck gambles than biases in stated beliefs.

6.4 Robustness

We also tested several alternative specifications of the structural model. First, we specified the utility and probability weighting functions by using constant absolute risk aversion and the two-parameter-version of Prelec's (1998) probability weighting function, respectively. These alternative specifications all yield qualitatively identical results. In addition, we used an alternative measure of relative skill. Instead of ranking subjects within each session, we constructed a ranking within the full sample of 320 subjects. This did not affect our results. For further details, please refer to Appendix D.

7 Conclusion

We use a laboratory experiment to study the role of beliefs about relative skill on risky decisions. To answer this question and rule out strategic confounds, we compare subjects' behavior in luck and skill gambles offering the same prizes and winning probabilities. In addition, we introduce a structural model which provides us with a choice based measure of beliefs about relative skill and allows us to control for risk preferences. We find that low (high) skill subjects are more (less) willing to take risks on gambles where the winning probabilities depend on relative skill than in gambles where they depend on luck. This pattern is driven by systematic biases in the subjects' perceptions of their relative skill: low skill subjects overplace themselves, while high skill subjects underplace themselves.

This finding has important implications for career choices, market entry decisions, and financial choices. Even in the absence of strategic considerations, biased perceptions of relative skill can lead to misallocations of talent and, ultimately, low returns to entrepreneurship: the wrong people might enter professions where performance depends on relative skill and the right people might be crowded out.

References

Andersen, Steffen, Glenn W Harrison, Morten Igel Lau, and E Elisabet Rutström (2006) 'Elicitation using multiple price list formats.' *Experimental Economics* 9(4), 383–405

Andersen, Steffen, John Fountain, Glenn W Harrison, and E Elisabet Rutström (2014) 'Estimating subjective probabilities.' *Journal of Risk and Uncertainty* 48(3), 207–229

Azrieli, Yaron, Christopher P Chambers, and Paul J Healy (forthcoming) 'Incentives in experiments: A theoretical analysis.' *Journal of Political Economy*

Beck, John H (1994) 'An experimental test of preferences for the distribution of income and individual risk aversion.' *Eastern Economic Journal* 20(2), 131–145

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak (1964) 'Measuring utility by a single-response sequential method.' *Systems Research and Behavioral Science* 9(3), 226–232

Benoît, Jean-Pierre, and Juan Dubra (2011) 'Apparent overconfidence.' *Econometrica* 79(5), 1591–1625

Benoît, Jean-Pierre, Juan Dubra, and Don A Moore (2015) 'Does the better-than-average effect show that people are overconfident?: Two experiments.' *Journal of the European Economic Association* 13(2), 293–329

Binswanger, Hans P (1980) 'Attitudes toward risk: Experimental measurement in rural india.' *American journal of agricultural economics* 62(3), 395–407

_ (1981) 'Attitudes toward risk: Theoretical implications of an experiment in rural india.' *The Economic Journal* 91(364), 867–890

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper (2010) 'Risk and rationality: Uncovering heterogeneity in probability distortion.' *Econometrica* pp. 1375–1412

Burks, Stephen V, Jeffrey P Carpenter, Lorenz Goette, and Aldo Rustichini (2013) 'Overconfidence and social signalling.' *The Review of Economic Studies* 80(3), 949–983

Cain, Daylian M, Don A Moore, and Uriel Haran (2015) 'Making sense of overconfidence in market entry.' *Strategic Management Journal* 36(1), 1–18

Camerer, Colin, and Dan Lovallo (1999) 'Overconfidence and excess entry: An experimental approach.' *American Economic Review* pp. 306–318

Charness, Gary, Aldo Rustichini, and Jeroen Van de Ven (2017) 'Self-confidence and strategic behavior.' *Experimental Economics* pp. 1–27

Charness, Gary, Uri Gneezy, and Michael A Kuhn (2012) 'Experimental methods: Between-subject and within-subject design.' *Journal of Economic Behavior and Organization* 81(1), 1–8

Clark, Jeremy, and Lana Friesen (2009) 'Overconfidence in forecasts of own performance: An experimental study.' *The Economic Journal* 119(534), 229–251

Dawes, R.N., and M. Mulford (1996) 'The false consensus effect and overconfidence. flaws in judgement, or flaws in how we study judgement?' *Organizational Behavior and Human Decision Processes* 65, 201–211

De Meza, David, and Clive Southey (1996) 'The borrower's curse: optimism, finance and entrepreneurship.' *The Economic Journal* pp. 375–386

Dunne, Timothy, Mark J Roberts, and Larry Samuelson (1988) 'Patterns of firm entry and exit in us manufacturing industries.' *The RAND Journal of Economics* pp. 495–515

Dunning, David, Judith A Meyerowitz, and Amy D Holzberg (1989) 'Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability.' *Journal of personality and social psychology* 57(6), 1082

Eil, David, and Justin M Rao (2011) 'The good news-bad news effect: asymmetric processing of objective information about yourself.' *American Economic Journal: Microeconomics* 3(2), 114–138

Erev, Ido, T.S. Wallsten, and David V Budescu (1994) 'Simultaneous over- and underconfidence: The role of error in judgement processes.' *Psychological Review* 101, 519–527

Felson, Richard B (1981) 'Ambiguity and bias in the self-concept.' *Social Psychology Quarterly* pp. 64–69

Fischbacher, Urs (2007) 'z-tree: Zurich toolbox for ready-made economic experiments.' *Experimental economics* 10(2), 171–178

Fraser, Stuart, and Francis J Greene (2006) 'The effects of experience on entrepreneurial optimism and uncertainty.' *Economica* 73(290), 169–192

Glaser, Markus, and Martin Weber (2007) 'Overconfidence and trading volume.' *The Geneva Risk and Insurance Review* 32(1), 1–36

Goldstein, William M, and Hillel J Einhorn (1987) 'Expression theory and the preference reversal phenomena.' *Psychological Review* 94(2), 236–254

Gonzalez, Richard, and George Wu (1999) 'On the shape of the probability weighting function.' *Cognitive psychology* 38(1), 129–166

Greiner, Ben (2015) 'Subject pool recruitment procedures: organizing experiments with orsee.' *Journal of the Economic Science Association* 1(1), 114–125

Hamilton, Barton H (2000) 'Does entrepreneurship pay? an empirical analysis of the returns to self-employment.' *Journal of Political Economy* 108(3), 604–631

Hey, John D, and Chris Orme (1994) 'Investigating generalizations of expected utility theory using experimental data.' *Econometrica* pp. 1291–1326

Hoelzl, Erik, and Aldo Rustichini (2005) 'Overconfident: Do you put your money on it?' *The Economic Journal* 115(503), 305–318

Holt, Charles A, and Susan K Laury (2002) 'Risk aversion and incentive effects.' *American Economic Review* 92(5), 1644–1655

Johnson, Cathleen A, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie Van Dolder, and Peter P Wakker (2015) 'Prince: An improved method for measuring incentivized preferences.' *Available at SSRN 2504745*

Jovanovic, Boyan (2004) 'The pre-producers.' Technical Report, National Bureau of Economic Research

Kadane, Joseph B, and Robert L Winkler (1988) 'Separating probability elicitation from utilities.' *Journal of the American Statistical Association* 83(402), 357–363

Kahneman, Daniel, and Amos Tversky (1979) 'Prospect theory: An analysis of decision under risk.' *Econometrica* pp. 263–291

Krueger, Joachim, and Ross A Mueller (2002) 'Unskilled, unaware, or both? the better-than-average heuristic and statistical regression predict errors in estimates of own performance.' *Journal of personality and social psychology* 82(2), 180

Kruger, Justin (1999) 'Lake wobegon be gone! the "below-average effect" and the egocentric nature of comparative ability judgments.' *Journal of personality and social psychology* 77(2), 221

Kruger, Justin, and David Dunning (1999) 'Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments.' *Journal of personality and social psychology* 77(6), 1121

Larrick, R.P., K.A. Burson, and J.B. Soll (2007) 'Social comparison and confidence. when thinking you're than average predicts overconfidence (and when it does not).' *Organizational behavior and Human Decision Processes* 102, 76–94

Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence D Phillips (1977) 'Calibration of probabilities: The state of the art.' In 'Decision making and change in human affairs' (Springer) pp. 275–324

Lucas Jr, Robert E (1978) 'On the size distribution of business firms.' *The Bell Journal of Economics* pp. 508–523

Manove, Michael (1995) *Entrepreneurs, Optimism, and the Competitive Edge* (Universitat Autònoma de Barcelona, Departament d'Economia i d'Història Econòmica)

Mata, Jose, and Pedro Portugal (1994) 'Life duration of new firms.' *The Journal of Industrial Economics* pp. 227–245

Merkle, Christoph, and Martin Weber (2011) 'True overconfidence: The inability of rational information processing to account for apparent overconfidence.' *Organizational Behavior and Human Decision Processes* 116(2), 262–271

Mobius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat (2011) 'Managing self-confidence: Theory and experimental evidence.' Technical Report, National Bureau of Economic Research

Moore, D.A., and D.A. Small (2007) 'Error and bias in comparative judgement: On being better and worse than we think we are.' *Journal of personality and social psychology* 92(6), 972–989

Moore, Don A, and Daylian M Cain (2007) 'Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition.' *Organizational Behavior and Human Decision Processes* 103(2), 197–213

Moore, Don A, and Paul J Healy (2008) 'The trouble with overconfidence.' *Psychological Review* 115(2), 502

Moskowitz, Tobias J, and Annette Vissing-Jørgensen (2002) 'The returns to entrepreneurial investment: A private equity premium puzzle?' *The American Economic Review* 92(4), 745–778

Murnighan, J Keith, Alvin E Roth, and Françoise Schoumaker (1987) 'Risk aversion and bargaining: Some preliminary results.' *European Economic Review* 31(1-2), 265–271

Murnighan, J Keith, Alvin E Roth, and Francoise Schoumaker (1988) 'Risk aversion in bargaining: An experimental study.' *Journal of Risk and Uncertainty* 1(1), 101–124

Myers, David G (1987) *Social psychology* (McGraw-Hill New York)

Niederle, Muriel, and Lise Vesterlund (2007) 'Do women shy away from competition? do men compete too much?' *The Quarterly Journal of Economics* 122(3), 1067–1101

Odean, Terrance (1999) 'Do investors trade too much?' *The American Economic Review* 89(5), 1279–1298

Park, Young Joon, and Luís Santos-Pinto (2010) 'Overconfidence in tournaments: evidence from the field.' *Theory and Decision* 69(1), 143–166

Prelec, Drazen (1998) 'The probability weighting function.' *Econometrica* pp. 497–527

Quiggin, John (1982) 'A theory of anticipated utility.' *Journal of Economic Behavior & Organization* 3(4), 323–343

Raven, J.C. (2000) *Ravens advanced progressive matrices (APM)* (Upper Saddle River: Pearson)

Rigotti, Luca, Matthew Ryan, and Rhema Vaithianathan (2011) 'Optimism and firm formation.' *Economic Theory* 46(1), 1–38

Santos-Pinto, Luis, and Joel Sobel (2005) 'A model of positive self-image in subjective assessments.' *American Economic Review* pp. 1386–1402

Schlag, Karl H, James Tremewan, and Joel J Van der Weele (2015) 'A penny for your thoughts: A survey of methods for eliciting beliefs.' *Experimental Economics* 18(3), 457–490

Snedeckor, GW, and WG Cochran (1989) *Statistical Methods,* 8 ed. (Iowa State University Press)

Starmer, Chris (2000) 'Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk.' *Journal of economic literature* pp. 332–382

Steiger, James H (1980) 'Tests for comparing elements of a correlation matrix.' *Psychological bulletin* 87(2), 245–251

Svenson, Ola (1981) 'Are we all less risky and more skillful than our fellow drivers?' *Acta Psychologica* 47(2), 143–148

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen (2010) 'Risk and time preferences: linking experimental and household survey data from vietnam.' *The American Economic Review* 100(1), 557–571

Tversky, Amos, and Daniel Kahneman (1992) 'Advances in prospect theory: Cumulative representation of uncertainty.' *Journal of Risk and Uncertainty* 5(4), 297–323

Van den Steen, Eric (2004) ‘Rational overoptimism (and other biases).’ *American Economic Review* pp. 1141–1151

Wakker, Peter P (2010) *Prospect theory: For risk and ambiguity* (Cambridge University Press)

Appendices

A Elicitation of Beliefs

This Appendix provides details on the elicitation of stated beliefs about the subjects' relative performance in the cognitive ability test. We elicited the stated beliefs in two ways: first, by asking the subjects to state their belief as a probability distribution over all possible skill levels and, second, by asking them to provide a point estimate about their most likely, i.e. modal, skill level. Both ways were incentivized.

To state their belief as a probability distribution over all possible skill levels subjects had to adjust a set of sliders as shown in figure 6. The elicited belief distribution corresponds to a vector $q = (q_1, \ldots, q_m)$, where q_m is the probability by which the subjects believes to belong to skill level m.

We applied the following QSR to incentivize the subjects to reveal their true beliefs. If a subject belongs to skill level m, the QSR offers her a payment equal to

$$QSR_m(q) = 5 + 10q_m - 5\sum_{v=1}^{10} q_v^2 \, .$$

The subject obtains the maximum payoff of CHF 10 only if she puts the entire probability mass on the correct skill level, i.e. $q_m = 1$. By contrast, her payoff is CHF 0 if she puts the entire probability mass on a wrong skill level. In all other cases, the payment is somewhere between CHF 10 and CHF 0.

figure 7 shows the distribution of the realized payments. Two subjects received the lowest possible payment of CHF 0, and two subjects received the highest possible payment of CHF 10. The average realized payment for the QSR was CHF 4.85, the median of realized payment was CHF 4.5, and the standard deviation of realized payments was CHF 1.71.

We also asked the subjects directly to provide a point estimate about their most likely skill level. If their answer corresponded to their actual skill level, they earned an additional CHF 5 and CHF 0 otherwise. This gives the subjects an incentive to reveal their modal belief, even if they are not risk neutral.

Table 6 shows the joint distribution of the stated modal beliefs and the true skill levels. Benoît and Dubra (2011) and Burks et al. (2013) show that the Bayesian model imposes testable implications on how the distribution of relative skill judgements should be related to true skill. Because individuals pick the most likely skill level, the largest (modal) group of individuals thinking they are in a given skill level must belong to that skill level. This is called the allocation function (see Burks et al. (2013) for a more detailed description of the properties of Bayesian updating). As Table 6 shows, of the 73 subjects who believe to be in skill level $\{0.25, 0.35\}$ only 17 truly belong to that skill level. This is less than

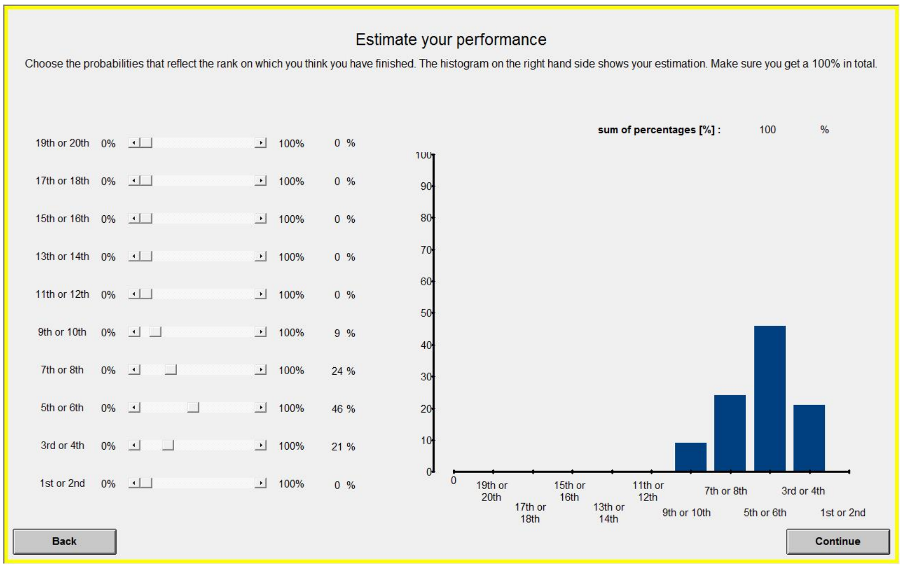


Figure 6: Interface of the elicitation of stated beliefs (full probability distribution)

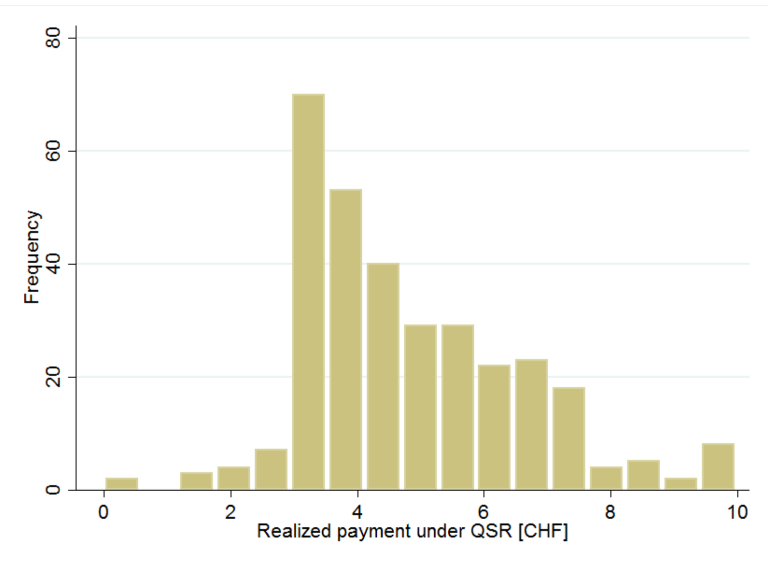


Figure 7: Histogram of realized payments under QSR in Swiss francs (CHF). The bar width is 0.50 (50 Swiss cents). Number of observations = 319, number of females = 114.

True skill level \ Stated beliefs (mode)	$\{0.05, 0.15\}$	$\{0.25, 0.35\}$	$\{0.45, 0.55\}$	$\{0.65, 0.75\}$	$\{0.85, 0.95\}$	sum
$\{0.85, 0.95\}$	0	7	8	38	**11**	64
$\{0.65, 0.75\}$	0	14	16	**31**	3	64
$\{0.45, 0.55\}$	0	15	**27**	15	7	64
$\{0.25, 0.35\}$	0	**17**	23	22	1	63
$\{0.05, 0.15\}$	**5**	20	25	13	1	64
sum	5	73	99	119	23	319

Table 6: Absolute frequencies of stated beliefs about relative skill level (mode)

20 individuals who belong to skill level $\{0.05, 0.15\}$. Similarly, of the 119 subjects who believe to be in skill level $\{0.65, 0.75\}$ only 31 truly belong to that skill level. This is less than 38 individuals who belong to skill level $\{0.85, 0.95\}$. In contrast, the allocation function is not violated for subjects in the skill levels $\{0.05, 0.15\}$, $\{0.45, 0.55\}$, and $\{0.85, 0.95\}$.

B figures for Prizes (120, 20) and (100, 40)

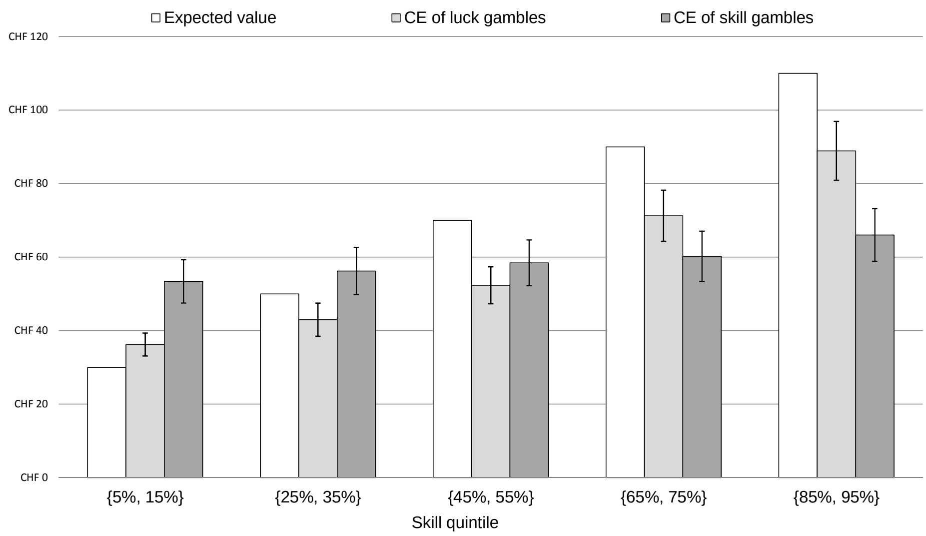


Figure 8: Average expected values of the luck gambles (white bars), average CEs of luck gambles (light grey bars), and average CEs of skill gambles (dark grey bars). Depicted is the gamble with winnining prize $x_h = 120$ and losing prize $x_l = 20$. The 95%-confidence intervals are based on robust standard errors. Number of observations = 319, number of females = 114.

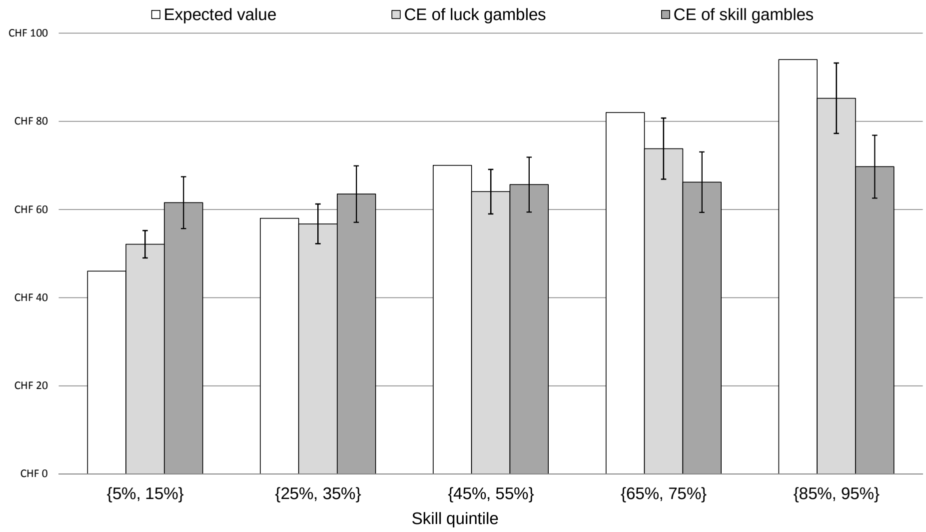


Figure 9: Average expected values of the luck gambles (white bars), average CEs of luck gambles (light grey bars), and average CEs of skill gambles (dark grey bars). Depicted is the gamble with winning prize $x_h = 100$ and losing prize $x_l = 40$. The 95%-confidence intervals are based on robust standard errors. Number of observations = 319, number of females = 114.

C Maximum Likelihood Estimation

This appendix describes how we estimate individual risk preferences using the method of maximum likelihood. The observed CEs can be decomposed into the predicted CEs and a random error: $CE^L_{ij} = \hat{CE}^L_{ij} + \varepsilon^L_{ij}$ for luck gambles, and $CE^S_{ik} = \hat{CE}^S_{ik} + \varepsilon^S_{ik}$ for skill gambles.

There are several reasons why the observed CEs may differ from the predicted ones. The most obvious one is that the multiple price list approach only allows us to approximate the observed CEs by taking an average of the certain amounts just above and below the switching point. Furthermore, Hey and Orme (1994) point out that carelessness, hurry, or inattentiveness can lead to accidentally wrong answers.

Following Hey and Orme (1994) we assume that subject i's random error for luck gamble j follows the distribution $\varepsilon^L_{ij} \sim N(0, \sigma^L_{ij})$, where the standard deviation σ^L_{ij} equals $\kappa^L_i(x^L_{hj} - x^L_{lj})$. Analogously, for skill gamble k: $\varepsilon^S_{ik} \sim N(0, \sigma^S_{ik})$, where $\sigma^S_{ik} = \kappa^S_i(x^S_{hk} - x^S_{lk})$.[14] The variance of the random error is proportional to the corresponding gamble's payoff range, as the subject's CE is elicited with respect to 19 certain amounts that are equally spread out within the gamble's payoff range.[15]

[14]A likelihood-ratio test shows that the data doesn't support the assumption of equal variance of errors in luck and skill gambles. The null hypothesis that $\kappa^S_i = \kappa^L_i$ is rejected at the 5%-level for 49% of the subjects.

[15]For instance, in the gamble with $x_h = 140$ and $x_l = 0$ a shift of the switching point by one line translates into a change in the CE of 7. On the other hand, if $x_h = 100$ and $x_l = 40$ a shift by one line translates into a change in the CE of 3.

The likelihood function is given by:

$$L(\psi_i; CE^L_{ij}, CE^S_{ik}, x^L_{hj}, x^L_{lj}, x^S_{hk}, x^S_{lk}) = \prod_{j=1}^{30} \frac{1}{\sigma^L_{ij}} \phi\Big(\frac{CE^L_{ij} - \hat{CE}^L_{ij}}{\sigma^L_{ij}}\Big) \prod_{k=1}^{9} \frac{1}{\sigma^S_{ik}} \phi\Big(\frac{CE^S_{ik} - \hat{CE}^S_{ik}}{\sigma^S_{ik}}\Big),$$

$\psi_i = (r_i, \delta_i, \gamma_i, \xi_i, \kappa^S_i, \kappa^L_i)'$ is the vector of parameters that are estimated by numerically maximizing L. To increase numerical precision and ensure convergence of the individual estimations, we apply the following restrictions on the parameters: $-10 \leq 1 - r_i \leq 10$; $0 \leq \gamma_i \leq 10$; $0 \leq \delta_i \leq 10$; $0.05 \leq \xi_i \leq 0.95$.[16] The reason for the restriction on ξ_i is that individual winning probabilities are 0.95 in the best case and 0.05 in the worst case. In order to account for the possibility that a subject's choices are serially correlated, we estimate cluster-robust standard errors at the subject-level.

D Robustness Checks

D.A Specifications of Utility and Probability Weighting Functions

We tested the robustness of the results using alternative functional forms for the utility and the probability weighting function. We estimated alternatively a utility function with constant absolute risk aversion (CARA):

$$u(x; r) = -e^{-xr}$$

where x is the monetary outcome of a gamble and the parameter r measures utility curvature: $r > 0$ corresponds to concave utility, $r = 0$ to linear utility, and $r < 0$ to convex utility.

In terms of the probability weighting function we estimated alternatively the two-parameter function proposed by Prelec (1998) (PLC):

$$w(p; \alpha, \beta) = e^{-\beta(-lnp)^{\alpha}}$$

where α and β reflect the degree of probability weighting. The parameter α governs likelihood sensitivity. If $\alpha \in (0, 1)$, the function captures the inverse s-shape pattern where low probabilities are upweighted and high probabilities are downweighted. If $\alpha > 1$ an s-shape pattern where low probabilities are downweighted and high probabilities are upweighted. The parameter β reflects the degree of optimism or pessimism. When $\beta > 1$, the subject is a pessimist, since she downweights the probability of the high prize x_h and upweights the probability of the low prize x_l. When $\beta \in (0, 1)$, the subject is an optimist, since she upweights the probability of the high prize x_h and downweights the probability of the low prize x_l. When $\alpha = \beta = 1$ there is no probability weighting and $w(p; 1, 1) = p$

[16]The bounds for the utility curvature and the probability weighting parameters are hit 11 times (nine subjects hit one bound and one subject hits two bounds). The lower bound for γ is hit six times, the upper bound for δ is hit twice, and the remaining bounds are hit once.

	Mean	Median	Stand. Dev.	Min	Max
CRRA & PLC					
Utility curvature (r_i)	0.27	0.25	1.00	-2.00	11.00
Likelihood sensitivity (α_i)	0.68	0.62	0.47	0.00	4.69
Optimism-pessimism (β_i)	1.34	1.10	1.04	0.00	10.00
Revealed beliefs (ξ_i)	0.54	0.56	0.21	0.05	0.95
Scaling of error term variance (luck) (κ_i^L)	0.08	0.07	0.04	0.00	0.32
Scaling of error term variance (skill) (κ_i^S)	0.11	0.10	0.06	0.00	0.46
CARA & GE					
Utility curvature (r_i)	0.02	0.01	0.04	-0.04	0.51
Likelihood sensitivity (γ_i)	0.86	0.69	0.95	0.00	10.00
Optimism-pessimism (δ_i)	1.59	0.98	2.1	0.00	10.00
Revealed beliefs (ξ_i)	0.52	0.54	0.21	0.05	0.95
Scaling of error term variance (luck) (κ_i^L)	0.08	0.07	0.04	0.00	0.32
Scaling of error term variance (skill) (κ_i^S)	0.11	0.09	0.06	0.00	0.50
CARA & PLC					
Utility curvature (r_i)	0.04	0.01	0.18	-0.04	2.11
Likelihood sensitivity (α_i)	1.00	0.70	1.40	0.00	10.00
Optimism-pessimism (β_i)	1.18	0.93	1.20	0.00	10.00
Revealed beliefs (ξ_i)	0.52	0.54	0.21	0.05	0.95
Scaling of error term variance (luck) (κ_i^L)	0.08	0.06	0.04	0.00	0.32
Scaling of error term variance (skill) (κ_i^S)	0.11	0.09	0.07	0.00	0.50

Table 7: Summary statistics of the individual parameter estimates for the alternative specifications of the structural model. Number of observations = 319, number of females = 114.

The estimation procedure remains the same as described in Section 5 and Appendix C, including the parameter restrictions. The alternative functional forms give rise to three additional specifications of the structural model: CRRA & PLC, CARA & Goldstein and Einhorn (1987) (GE), and CARA & PLC.

D.A.1 Summary Statistics on Structural Estimations

Table 7 reports the summary statistics of the individual parameter estimates of the alternative specifications of the structural model.

D.A.2 Correlations between Estimated Revealed Beliefs

Table 8 reports the correlation coefficients between the revealed beliefs, ξ_i obtained under the four combinations of utility function and probability weighting function specifications. The high correlations provide evidence for robustness of our baseline specification.

	CRRA & GE	CRRA & PLC	CARA & GE	CARA & PLC
CRRA & GE	1			
CRRA & PLC	0.976	1		
CARA & GE	0.942	0.938	1	
CARA & PLC	0.935	0.963	0.957	1

Table 8: Correlation coefficients between the revealed beliefs, ξ_i, obtained under the four different combinations of utility and probability weighting function specifications. Number of observations = 319, number of females = 114.

D.A.3 Revealed Beliefs per Skill Quintile

Table 9 reports the bias in revealed beliefs per quintile level. Depicted are the three additional specifications of the utility and the probability weighting function. The estimates of under- and overestimation of relative skill are robust to alternative specifications of the utility and the probability weighting function. The coefficients change only little compared to the baseline estimates (exhibited in Table 5).

	CRRA & PLC	CARA & GE	CARA & PLC
Skill {0.85, 0.95}	-0.268***	-0.296***	-0.290***
	(0.024)	(0.026)	(0.025)
Skill {0.65, 0.75}	-0.171***	-0.193***	-0.197***
	(0.029)	(0.029)	(0.030)
Skill {0.45, 0.55}	0.065***	0.048*	0.042
	(0.024)	(0.025)	(0.026)
Skill {0.25, 0.35}	0.208***	0.180***	0.194***
	(0.025)	(0.023)	(0.025)
Skill {0.05, 0.15}	0.383***	0.342***	0.359***
	(0.026)	(0.025)	(0.026)
Observations	319	319	319
R-squared	0.585	0.567	0.574

Note: Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 9: Regression where the dependent variable is the bias in revealed beliefs, i.e., $b_i^{revealed} \equiv \xi_i - \theta_i$, and the independent variables are skill quintile dummies. Depicted are the alternative specifications of the structural model.

	CRRA & GE	CRRA & PLC	CARA & GE	CARA & PLC
Skill {0.85, 0.95}	-0.242***	-0.240***	-0.267***	-0.262***
	(0.024)	(0.024)	(0.026)	(0.025)
Skill {0.65, 0.75}	-0.164***	-0.160***	-0.182***	-0.186***
	(0.031)	(0.031)	(0.031)	(0.032)
Skill {0.45, 0.55}	0.080***	0.082***	0.065**	0.059*
	(0.030)	(0.030)	(0.031)	(0.032)
Skill {0.25, 0.35}	0.177***	0.195***	0.168***	0.182***
	(0.026)	(0.027)	(0.025)	(0.028)
Skill {0.05, 0.15}	0.335***	0.344***	0.303***	0.320***
	(0.029)	(0.029)	(0.028)	(0.029)
Observations	319	319	319	319
R-squared	0.487	0.498	0.475	0.478

Note: Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

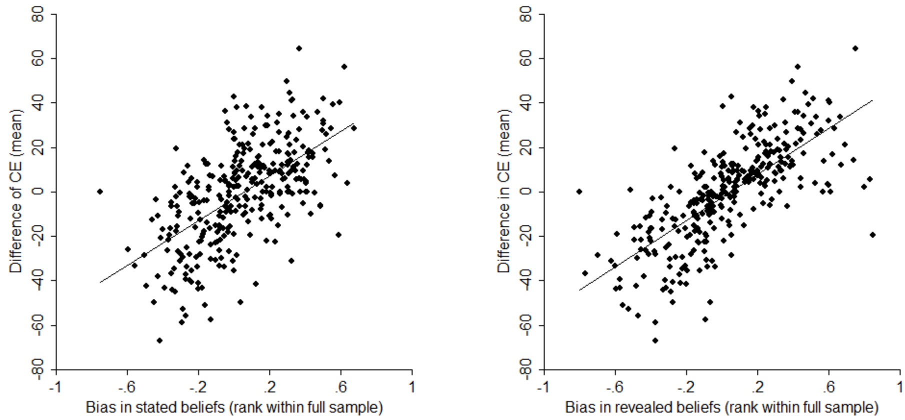
Table 10: Regression where the dependent variable is the bias in revealed beliefs, i.e., $\tilde{b}_i^{revealed} \equiv \xi_i - \tilde{\theta}_i$, and the independent variables are skill quintile dummies. Depicted are the four specifications of the structural model. True skill levels determined by ranking within the full sample.

D.B Ranking within the Full Sample

In this section we report the results obtained by determining true skill levels based on ranking within the full sample of 320 subject rather than within each session. We denote this measure of the true skill level of subject i by $\tilde{\theta}_i$. The corresponding bias in stated and revealed beliefs are denoted by $\tilde{b}_i^{stated} \equiv \mu_i - \tilde{\theta}_i$ and by $\tilde{b}_i^{revealed} \equiv \xi_i - \tilde{\theta}_i$, respectively. The correlation between $\tilde{\theta}_i$ and the test score $(= 0.963)$ is significantly higher (p-value < 0.001) than the correlation between θ_i and the test score $(= 0.903)$ because $\tilde{\theta}_i$ does not depend on the random sampling of subjects in each session. However, $\tilde{\theta}_i$ is still not a strictly monotonic transformation of the test score because of the tiebreaker and the random number generator that apply to break ties. It is thus possible that two subjects have the same test score but not the same $\tilde{\theta}$.

Table 10 reports the bias in revealed beliefs per quintile level. Depicted are the four combinations of utility and the probability weighting function specifications. The coefficients change only little compared to the baseline estimates (see Tables 5 and 9). In contrast to the baseline estimates, however, the weak overestimation of relative skill in the intermediate quintile is statistically significant at at least the 10%-level in all four specifications.

finally, we exhibit in figure 10 the scatterplots and regression lines illustrating the correlation of the difference in CEs with both the bias in stated beliefs, $\tilde{b}_i^{stated}$, and with the bias in revealed beliefs, $\tilde{b}_i^{revealed}$. Depicted are the estimates obtained by our baseline specification with CRRA utility functions and GE probability weighting functions.



(a) (b)

Figure 10: (a) Relation between the mean of the differences of CEs between the three comparable pairs of skill and luck gambles and the bias in stated beliefs, $\tilde{b}_i^{stated}$. (b) Relation between the mean of the differences of CEs between the three comparable pairs of skill and luck gambles and the bias in revealed beliefs, $\tilde{b}_i^{revealed}$. True skill levels determined by ranking within the full sample. Both panels show the scatterplot and the regression line. Number of observations = 319, number of females = 114.

As in the baseline estimate where subjects are ranked within each session, the correlation of the differences of the CEs with the bias in revealed beliefs ($= 0.722$) is significantly higher (p-value = 0.006) than the correlation of the differences in CEs with the bias in stated beliefs ($= 0.601$).