CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Discussion Paper No. 2014-05

Zahra Murad, Martin Sefton and Chris Starmer
April 2014

How do risk attitudes affect measured confidence?

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
Fax: +44 (0) 115 95 14159
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

How do risk attitudes affect measured confidence?

Zahra Murad[a b], Martin Sefton[a] and Chris Starmer[a]

April 2014

Abstract

We examine confidence in own absolute performance using two elicitation procedures: self-reported (non-incentivised) confidence and an incentivised procedure that elicits the certainty equivalent of a bet based on performance. The former procedure reproduces the "hard-easy effect" (overconfidence in easy tasks and underconfidence in hard tasks) found in a large number of studies using non-incentivised self-reports. The latter procedure produces general underconfidence, which is somewhat reduced when we filter out the effects of risk attitudes. However, even after controlling for risk attitudes our incentivised procedure leads to significant underconfidence, and does not lead to better calibration between confidence and performance than non-incentivised self-reports. Finally, we find that self-reported confidence correlates significantly with features of individual risk attitudes including parameters of individual probability weighting.

Keywords: Overconfidence, Underconfidence, Experiment, Risk Preferences

JEL: C91, D81

Acknowledgements: We are grateful to Thomas Epper for providing us the data from their experiments and the R-code for the econometric estimation. We also thank Steve Burks, Robin Cubitt, Simone Quercia, Francesco Fallucchi, Antonio A. Arechar and seminar participants at the CCC 2012 meeting in Norwich, TIBER XI in Tilburg, and Workshop on Human Decision Making in Birmingham for helpful comments. Martin Sefton and Chris Starmer acknowledge the support of the ESRC funded Network for Integrated Behavioural Science (ES/K002201/1).

[a] Centre for Decision Research and Experimental Economics (CeDEx), School of Economics, University of Nottingham
[b] Department of Health Care Management and Policy, University of Surrey

Corresponding author: chris.starmer@nottingham.ac.uk

1. Introduction

A large literature dating back to the 1970s documents systematic biases in individuals' confidence assessments of their own performance. In this paper we report an experiment investigating possible relationships between miscalibration and risk attitudes.

Our study has two primary motivations. The first flows from an apparent clash between the established results from the psychological literature and more recent evidence emerging from experimental economics. A large volume of research in psychology suggests that individuals have predictable tendencies towards either overconfidence or a hard-easy effect (over-estimating own performance in hard tasks and under-estimating it in easy tasks). By contrast, more recent research in experimental economics has found either much less confidence miscalibration or, when it occurs, strikingly different patterns of miscalibration (we discuss the evidence in more detail in the next section). What might account for this difference? One distinctive feature of much of the newer literature is that it employs various (financial) incentive mechanisms to motivate revelation of confidence, whereas the psychology studies rely on non-incentivised self-reports of confidence. So, one possible diagnosis is that the newer evidence provides more accurate confidence measurement as a consequence of incentivised revelation techniques. In this paper, however, we investigate another possibility: that some of the differences between findings of economists and psychologists may be a consequence of biases in measured confidence induced by incentive mechanisms which fail to control for the influence of individual risk attitudes.

A second motivation for our study is to explore the possibility that confidence judgements may be intrinsically related to risk attitudes. It seems intuitively plausible that there could be a positive association between individuals being more confident and being more willing to take risks. For example, overconfidence about own abilities and a willingness to take risks might be common consequences of particular personality traits (e.g. egotism) or emotional states or dispositions (e.g. optimism). While these considerations suggest a possible linkage between individual confidence assessment and risk attitudes, as far as we know, our study is the first to directly test for it.

In pursuit of these objectives, we elicit confidence via two distinct methods and we also independently measure individual risk attitudes. One of our confidence measurement tools is a non-incentivised task designed to be analogous to standard procedures that have been used extensively in psychological research; the other is a simple incentivised choice based procedure. We designed the latter to be incentive compatible for revelation of confidence for risk neutral subjects, but, in common with other incentive mechanisms that have been used in the recent literature, our procedure will result in biased confidence measurements for non-risk neutral subjects. Thus, we complement this procedure with a method that uses elicited risk attitudes to correct incentivised confidence measures for departures from risk-neutrality. Section 3 describes our experimental methodology in more detail.

In Section 4 we present our results. There are three primary findings. First, our two tools produce markedly different patterns of confidence miscalibration, mimicking the stylised facts of existing research (the non-incentivized tool reproduces the familiar hard-easy effect, while our incentivised tool reveals general underconfidence). Second, when we filter out the effects of risk attitudes on incentivised measurements of confidence, we find that measured miscalibration is much reduced. This shows that incentivised mechanisms for confidence elicitation can be significantly biased in the absence of suitable controls for individual risk preferences. Finally, we find that confidence as measured by the standard psychological technique correlates significantly with features of individual risk attitudes including parameters of individual probability weighting functions. Moreover the directions of association are intuitively plausible: for example, reported confidence is positively associated with 'optimism' in probability weights. Section 5 discusses these results and concludes.

2. Literature Review

There is a large literature in psychology on biases in individual assessments of their own abilities, both relative to others' and in absolute terms. Findings of overconfidence in own performance relative to that of others (e.g. Svenson 1980) has motivated many studies by experimental economists on the relationship between relative confidence, relative ability, and willingness to take risks in strategic environments (e.g. Camerer & Lovallo 1999; Hoelzl & Rustichini 2005; Moore & Cain 2007; Niederle & Vesterlund 2007). In our paper we focus on the calibration of own *absolute* performance. This is the more suitable measure given our purpose of studying the relationship between confidence judgements and risk attitudes since both are in the domain of individual choice. By contrast, miscalibration of relative performance may reflect miscalibration of own performance, or of the performance of others, and its measurement may be complicated by strategic and/or social comparison concerns.

Early studies by psychologists on individuals' self-assessment of own performance document systematic miscalibration, usually towards overconfidence or a hard-easy effect (Fischoff, Slovic & Lichtenstein 1977; Lichtenstein & Fischoff 1981; Lichtenstein, Fischoff

& Phillips, 1982; see Keren 1991 or Alba & Hutchinson 2000 for a review). In a typical study (e.g. Fischoff et al. 1977), individuals are given quiz questions and asked to give an answer and an assessment of the chances of their answer being correct. A common finding is that on questions where, say, 90% of individuals get the correct answer, average confidence is substantially lower, whereas in questions where, say, 60% get the correct answer, average confidence is substantially higher. A variety of explanations for these findings have been given including, response scale effects, stochastic errors in decision making or regression towards the mean (Erev, Wallstein & Budescu 1994; Suantak, Bolger & Ferrel 1996; Juslin, Winman & Olsson 2000; Brenner 2000).

Regardless of the source of confidence miscalibration, it has important implications from an economics perspective: confidence about own abilities affect many important economic decisions such as trading behaviour (Biais, Hilton, Mazuier & Pouget 2005), job search (Dubra 2004), investment in education (Dunning, Heath & Suls 2004) and bargaining behaviour in binding arbitration (Dickinson 2006). Thus it is not surprising that economists have begun to incorporate overconfidence into economic models (Compte & Postlewaite 2004; Gervais, Heaton & Odean 2011, Herz, Schunk & Zehnder 2014). But, recent research by experimental economists on miscalibration of (absolute) own confidence has revealed rather different patterns to the earlier psychology literature.

One of the first papers in the experimental economics literature using incentivized elicitation tools to study absolute confidence calibration is Blavatskyy (2009). He has subjects answer a set of 10 multiple choice quiz questions after which they choose from two payment schemes. Either one question is selected at random and the subject receives a payoff if he or she has answered this question correctly, or the subject receives the same payoff with a stated probability set by the experimenter to be equal to the percentage of correctly answered questions (although the subject does not know this is how the probability is set). Subjects could also indicate indifference. He finds that the majority chose the second payment scheme, which he interprets as underconfidence. Blavatskyy also elicited risk attitudes in a separate part of the experiment and found no significant relationship between elicited risk attitudes and choices of payment scheme. In a related contribution, Urbig, Stauf & Weitzel (2009) elicit confidence about own performance over a set of 10 multiple choice quiz questions. They find the majority of subjects are well-calibrated. Both of these studies note the difference between their findings and those from the earlier psychology literature, and speculate that the difference may be due to the introduction of financial incentives.

However, both studies lack a benchmark treatment for comparing the elicited confidence with an unincentivized tool. Our study includes such a comparison.

Clark & Friesen (2009) study subjects' confidence in relation to two types of real effort task involving verbal and numerical skills. They study calibration over a set of tasks elicited through unincentivized self-reports or quadratic scoring rule (QSR) incentives. They find underconfidence more prevalent than overconfidence and find better calibration with incentives. Moreover, they find that underconfidence was greatest among those using greater effort. One potential limitation of their analysis is that, unless subjects are risk neutral, QSR may result in biased measurements of confidence (we return to this point below in more detail).

A potentially significant feature of all three of the experiments discussed in the last two paragraphs is that they elicit confidence in relation to performance across *sets* of tasks. By contrast, much of the earlier psychological literature investigating confidence calibration assessed it with reference to performance in *single tasks*. This may be a significant distinction because there is evidence that miscalibration varies between measurements based on single versus sets of tasks. For example, Gigerenzer, Hoffrage & Keinbolting (1991), Liberman (2004) and Griffin & Brenner (2008) report that when beliefs are elicited about aggregate performance in sets of tasks most subjects are either well-calibrated or underconfident whereas overconfidence is evident when elicitation is at the single task level. We study confidence on a single task level. Hence our evidence is more directly comparable with the original confidence calibration studies.

The two studies most closely related to ours are Offerman, Sonnemans, van de Kuilen & Wakker (2009) and Hollard, Massoni & Vergnaud (2010). Hollard et al. (2010) elicit absolute confidence on a disaggregate task-level and compare confidence in visual perception and quiz tasks comparing three elicitation tools: unincentivized self-reports; the QSR; and the Becker-deGroot-Marschak (BDM) mechanism. They find highest overconfidence in the unincentivized self-reports followed by BDM and then QSR. That BDM-elicited confidence is higher than QSR-elicited confidence is consistent with the effects of risk aversion, but since they do not elicit risk attitudes it is not possible to say whether the difference between these elicitation tools is caused by risk attitudes or something else, such as differences in understanding of the elicitation procedures. Offerman et al. (2009) study biases in additivity of elicited beliefs relative to two mutually exclusive events whose occurrence is determined by nature. They hypothesize that the additivity bias in elicited beliefs arises because of the effect of risk attitudes on (QSR) elicited beliefs. In a two-step process, they elicit subjects' beliefs about uncertain events using QSR, and then use estimates of risk attitudes to filter out the effect of risk attitudes on measured beliefs. They find that the frequency of biases slightly decreases.

Our research strategy shares some features in common with Offerman et al., in particular that we explicitly estimate risk attitude parameters to filter out risk attitudes from beliefs. The key difference is that we are concerned with biases in subjective estimates of confidence in own performance (not biases in assessments of naturally determined chance events).[1] We use an elicitation tool for inferring confidence from incentivised choice behaviour that *will* be affected by risk attitudes if subjects are not risk neutral. By explicitly measuring risk attitudes we are able to observe the effect of risk attitudes on elicited beliefs and, more importantly, filter out risk attitudes and obtain risk-attitude-adjusted measures of confidence. By comparing risk adjusted to unadjusted confidence, we will be able to track an effect of risk attitudes on elicited confidence.

By explicitly measuring risk attitudes we are also able to investigate how those attitudes correlate with self-reported confidence. Previous studies investigating the link between individual characteristics and confidence have mostly focused on gender differences and find that women are less confident than men in relative terms but not in absolute terms (Clark & Friesen 2009; Biais et al. 2005; Lundeberg, Fox, Brown & Elbedour 2000). Campbell, Goodie & Foster (2004) find that narcissism predicts higher self-reported confidence and more willingness to bet on one's own performances. More recently, economists have become interested in how personality traits and economic preferences interact. It has been found that personality traits such as openness and extraversion predict confidence and overconfidence respectively (Schaefer, Williams, Goodie & Campbell 2004), neuroticism and cognitive ability predict risk taking (Rustichini, DeYoung, Anderson & Burks 2012), and personality traits complemented by risk preferences are successful in predicting many life outcomes such as health, earnings and education (Becker, Deckers, Dohmen, Falk & Kosse 2012). None of these studies, however, report how risk attitudes are correlated with elicited confidence at the individual level. Our methodology allows us to study the connection between risk attitudes and confidence directly and, conditional on there being some correlation, we will be able to probe how different components of risk attitude (i.e. curvature of utility or probability weighting) contribute towards it.

[1] Another difference from Offerman et al. is that we use the method developed by Fehr-Duda, Gennaro & Schubert (2006) to estimate individual risk parameters under the two leading models of decision under risk – expected utility (EU) and rank-dependant utility (RDU) theories.

3. Methods

We measure confidence about own performance in the context of a standard quiz framework. A subject responds to a series of two-item multiple-choice questions and, for each one, we elicit her subjective probability that her answer is correct. As a benchmark treatment we elicit confidence using self-reported non-incentivised confidence assessments. In another treatment we infer confidence from responses to a new incentivised procedure that employs pairwise choices between bets on own performance and certain amounts of money.

In both treatments we also estimate individual risk attitudes from a sequence of binary lottery choices. We use these estimates to filter out the effects of risk attitudes on elicited confidence in our incentivized procedure and to study the relationship between individual confidence and risk attitude.

3.1. Inferring Confidence and Eliciting Risk Preferences

We measure confidence about one's own performance using a multiple price list format.[2] Across a series of tasks, subjects have to say which of two cities has the higher population and then complete a table as in Figure 1.

[Insert Figure 1 around here]

Given the construction of the table, subjects are expected to choose Option B in the first row and Option A in the last row. At some point they will likely switch from option B to A, and this switchpoint is used to measure their confidence in their answer. For example, suppose a subject thinks she has a 67% chance of being correct. Her expected earnings from option A are £6.70 and so if she wants to maximise her expected earnings she should switch from B to A at row 8. We will refer to these switchpoints as *certainty equivalents (CE)* and under expected value maximisation (EV) the CE can be interpreted as revealing an individual's subjective probability of success (± 2.5%).

More generally, the CE picks up some mix of assessment of their chances of success with (possibly several) aspects of risk attitudes including non-linear attitudes to consequences and probabilities. For example, if the subject is a risk averse expected utility maximiser she will switch at a later row. If we were to incorrectly assume that this subject makes choices

[2] Andersen, Harrison, Lau & Rutstrom (2006) and Isoni, Loomes & Sugden (2011) extensively discuss the advantages and disadvantages of using multiple price list (MPL) elicitation tools. We choose to use MPL mainly because of the clear interpretable framework of the decision environment (the value of betting on own answer) and the relative ease for subjects to see that truthful revelation is in their best interest.

according to the EV model, we would interpret this later switchpoint as indicating a low subjective probability of success. In this case our estimate of subject confidence would be biased and, even if the individual is perfectly calibrated in that her subjective probability accurately reflects her underlying performance, we would find systematic underconfidence. Similarly, if choices are made based on non-linear attitudes to probabilities, we would obtain biased measures of confidence if we were to infer confidence through the lens of a model that fails to incorporate these attitudes, and as a result we would attribute systematic miscalibration to well-calibrated subjects.

To allow for non-linear attitudes to consequences and/or probabilities we infer confidence from CE's using the two most common specifications for risk preferences: Expected Utility (EU) and Rank Dependent Utility (RDU) theories. For both theories, there should be a unique switchpoint at which the utility of the certainty equivalent will be (approximately) equal to the utility of the lottery.[3] Hence, under the RDU model (which contains EU and EV as special cases) we may write:

$$U(CE_i) = U\ (£10)w(Conf_i) + U(£0)\big(1 - w(Conf_i)\big) \tag{1}$$

where CE_i is an individual's certainty equivalent for question i, $U(.)$ is a value function defined on money payoffs and $w(.)$ is an RDU probability weighting function. In expression (1) we treat confidence as a subjective probability judgement that underlies choices, but may be prone to misperceptions. In our analysis here, these misperceptions are equivalent to confidence miscalibration. The function $w(.)$ is then interpreted as capturing attitudes to chance distinct from misperceptions.[4] Rearranging equation (1) we obtain the probability that a subject assigns to being correct in question i, $Conf_i$, as:

$$Conf_i = w^{-1}\left(\frac{U(CE_i) - U(£0)}{U(£10) - U(£0)}\right) \tag{2}$$

Under the EV model both the value function and the probability weighting function are linear so confidence can be inferred directly from an observed CE as $Conf_i = CE_i/10$. Estimation of confidence under the EU model requires knowledge of the value (utility) function while estimation under the RDU model requires knowledge of both the value function and the probability weighting function.

[3] For compactness, the discussion now proceeds as if CE is revealed accurately by our procedure but the reader should keep in mind that there is, of course, an element of approximation.

[4] In the literature on prospect theory, probability weights are sometimes interpreted as reflecting misperception of underlying probabilities, sometimes reflecting subjective attitudes to chance, and sometimes a mixture of the two. For discussion and a formalisation following the latter mixed approach, see Abdellaoui, L'Haridon & Paraschiv (2011). For a thorough discussion of prospect theoretic models see Wakker (2010).

For the purpose of estimating $U(.)$ and $w(.)$, we use a simple and easy to understand procedure introduced in Fehr-Duda, Gennaro & Schubert (2006) and successfully employed to estimate value function and probability weighting function parameters in several subsequent studies (including: Bruhin, Fehr-Duda & Epper 2010; Fehr-Duda, Bruhin, Epper & Schubert 2010; and Epper, Fehr-Duda & Bruhin 2011). Because it uses a multiple price list elicitation task which is very similar in structure to our confidence elicitation task, it is particularly well suited to our study as its use minimises the cognitive load involved in subjects learning how to respond to the two types of task.

[Insert Figure 2 around here]

The procedure requires each subject to complete 25 tables of the form given in Figure 2. Each table consists of 20 rows, where each row is a choice between a two-outcome lottery and a guaranteed amount of money, with the guaranteed amount of money decreasing from the high outcome to the low outcome of the lottery in equal increments moving down the rows. The subject's certainty equivalence, CE_L, of lottery L can be written as in (3), where the high prize of the lottery x_{1L}occurs with probability p_{1L} and the low prize of the lottery x_{2L} occurs otherwise:

$$U(CE_L) = U(x_{1L})w(p_{1L}) + U(x_{2L})\big(1 - w(p_{1L})\big). \tag{3}$$

We use the switching point from choosing the guaranteed amount (Option B) to the lottery L (Option A) as our estimate of the subject's certainty equivalent of the lottery.

To estimate $U(.)$ and $w(.)$ we first specify functional forms for value and probability weighting functions. We follow Bruhin, Fehr-Duda, & Epper (2010) in their choice of flexible and interpretable functions which have been widely used elsewhere in the empirical literature. On this basis we use the power function for the value function:

$$U(x) = x^{\alpha}. \tag{4}$$

This specification is parsimonious in modelling risk attitudes via a single curvature parameter, α, and has been shown to provide a good fit to a wide range of choice data. To allow for non-linear probability weighting in the estimation of RDU parameters, we use the linear-in-log-odds function of Goldstein & Einhorn (1987):

$$w(p) = \frac{\beta p^{\gamma}}{\beta p^{\gamma} + (1-p)^{\gamma}}. \tag{5}$$

This specification is credited with providing a good account of individual heterogeneity (Wu, Zhang & Gonzalez 2004) and its two parameters have the advantage of having clear intuitive interpretations (Lattimore 1992; Bruhin, Fehr-Duda & Epper 2010): the parameter β captures 'elevation' of the probability weighting function (with greater β reflecting more 'optimism'); the parameter γ controls curvature (the smaller is γ, the stronger is deviation from linearity).

Finally, to operationalize the model requires specification of the stochastic decision process. Following Epper, Fehr-Duda & Bruhin (2011) we assume that the observed switching point, $\widehat{CE}_L$, is given by:

$$\widehat{CE}_L = CE_L + \epsilon_L, \tag{6}$$

where the error terms are independent draws from a normal distribution with zero mean. Heteroskedasticity in the error variances across tables is accounted for by assuming the standard deviation of the error is proportional to the difference between the guaranteed amounts in option B as one moves down the rows of the table. The normalized standard deviation and the parameters of *U(.)* and *w(.)* are then obtained by maximum likelihood estimation.

3.2. Experimental Procedures

The experiment consisted of two parts where Part 1 was the same for all the subjects and Part 2 varied according to the treatment. We use Part 1 for eliciting subjects' utility and probability weighting functions. The 25 lotteries of Part 1 are summarized in Table 1 and were adapted from Fehr-Duda, Gennaro & Schubert (2006). The order of the lotteries was randomized to avoid order effects.

After completing Part 1 of the experiment, subjects were asked to answer quiz questions where they had to choose the city with the highest population out of two options provided. They could earn £0.50 for each correct answer. In the Reported Confidence treatment, subjects were asked to provide a confidence judgement for each question by filling in the blank "I am ___% confident that my answer is correct". In the Inferred Confidence treatment, we introduced our new elicitation tool where subjects were asked to complete a table as in Figure 1. They had to complete one table for each quiz question.

After answering all quiz questions and providing their confidence levels (either by reporting or filling in the table), subjects were asked to complete a short post-study questionnaire while we checked their answers. We used a random lottery incentive system to pay subjects.[5] Subjects were paid based on one randomly drawn row in one randomly drawn table in one randomly drawn part of the experiment. We used physical objects (dice,

[5] The random lottery incentive system is widely used because, despite evidence showing failure of the independence axiom, empirical tests broadly support its use. (see Starmer & Sugden 1991, Cubitt, Starmer & Sugden 1998).

numbered balls and poker chips) to make the independence of the randomization devices very salient, and we explained the randomization procedures with simple examples and diagrams. The full experimental instructions are available on request.

[Insert Table 1 around here]

The experiment was conducted at the University of Nottingham, CeDEx lab in 2011. Subjects were recruited using Orsee (Greiner 2004). In total 86 subjects participated; 40 in the inferred confidence treatment (25 male), and 46 in the reported confidence treatment (23 male). The experiment was conducted in pen and paper format with subjects seated in cubicles. The experiment lasted approximately 1 hour and the average payment to a participant was £9.

4. Results

We structure the results under three subheadings. In Section 4.1, we compare and contrast the data on average confidence elicited in the two treatments. In Section 4.2, we present our findings on individual risk attitudes and filtered inferred confidence levels. And finally in Section 4.3, we present results looking at the relationship between risk attitudes and reported confidence.[6]

4.1. Reproducing Standard Results

Figure 3 provides a quick eye-balling tool for comparing confidence measured using the standard psychological tool with confidence elicited using our incentivised mechanism (on the assumption that individuals are risk neutral). Consider first the top left panel. This plots, for each quiz question, the mean of reported confidence against the average success rate. The 45-degree line provides a natural benchmark in the sense that a general tendency to overconfidence would result in points located above the line whereas a general tendency towards underconfidence would result in points below it.

[Insert Figure 3 around here]

[6] Before proceeding with the analysis, we dropped the data for four quiz questions that were potentially misleading because the success rate on each of these questions was less than 40% (whereas reported confidence judgements were constrained to the interval 50-100%). We also excluded data from tables where subjects switched on one row and then switched back again at a later row. Reassuringly, however, less than 2% of the tables included such non-monotonic responses.

The reported confidence data have a pattern consistent with the familiar 'hard-easy effect'. To highlight this, we have drawn a vertical (dashed) line through the question which is the median in terms of its success rate (at around 68%). If we define 'hard' ('easy') questions as those with lower (higher) than median success rates it is then apparent that, on average, there is overconfidence for all but one of the easy questions and underconfidence for all of the hard ones. For each question we measure miscalibration bias as average confidence minus the proportion of correct answers. We then test whether the mean of the distribution of biases is equal to zero using a simple t-test. For easy questions there is significant underconfidence (average bias = -0.115, p = 0.002) while for hard questions there is significant overconfidence (average bias = 0.070, p = 0.001). Pooling hard and easy questions we cannot reject the null of zero expected bias (average bias = -0.027, p = 0.312), evidently because the positive bias on easy questions offsets the negative bias on hard questions.

The top right panel of Figure 3 provides corresponding analysis for confidence inferred from our incentivised elicitation tool, but on the assumption that individuals are *risk neutral*. We refer to this measure as $Conf_{EVi}$ for short and, from expression 2 above, it is easy to see that this can be calculated directly from an individual's switch point in any given table because $Conf_{EVi} = CE_i/10$. Here, all of the observations sit below the 45 degree line indicating a systematic and highly significant tendency towards underconfidence (average bias = -0.212, p = 0.000).

The bottom two panels provide corresponding analysis, but in this case, each dot represents an individual, plotting individual average reported confidence across tasks against actual success rate in them. For individuals with less than median success rate there is marginal overconfidence (p = 0.085) and for individuals with more than median success rate there is significant underconfidence (p = 0.041) in reported confidence. Across all individuals in the inferred confidence treatment there is general underconfidence (p = 0.000).

Taken together, the results presented in Figure 3 reproduce a standard pattern of findings that has motivated our study. Using a procedure based on non-incentivised self-reports of confidence, similar to those used in a range of psychological studies, we reproduce a hard-easy effect; in contrast, by using an incentivised procedure to elicit confidence we find a marked tendency towards underconfidence.

4.2. Risk Preferences and Risk-Filtered Confidence

As we explained above, if individuals are not risk neutral, then confidence measures elicited via our incentivised mechanism may be biased because they may capture a mixture of confidence assessments and risk attitudes (and similarly so for other incentive mechanisms that have so far been used for this purpose in the literature). This section takes account of this possibility by implementing analysis to filter out the effects of risk attitudes in our incentivised confidence measures.

To this end, we exploit the data that we obtained from Part 1 of the experiment which allows us to fit risk preference models separately for each individual. We do this using two leading models of risk preference: expected utility theory (EU) and rank-dependent utility theory (RDU). Considering the EU and RDU models together, we have 6 parameters to estimate per experimental subject: the value function parameter under EU (α_{EU}); the value function and probability weighting parameters under RDU ($\alpha_{RDU},\ \beta, \gamma$); and the normalized standard deviations of the decision errors (σ_{EU} and σ_{RDU}). We will omit the discussion of error parameters from the results since they are not central to our analysis.

Figure 4 summarises the results of fitting these models to individuals in our ('Nottingham') study and, as a benchmark for our estimates, we also report parameters obtained by applying the same econometric method to the data reported in Bruhin, Fehr-Duda & Epper (2010) and Epper, Fehr-Duda & Bruhin (2011) (these are labelled the 'Zurich' estimates). The mean estimate of α_{EU} for Nottingham is substantially less than one, and for 85% of our sample we reject the null hypothesis of α_{EU} = 1, indicating concave utility function (i.e. risk aversion).[7] This is in line with standard findings (Zurich results are perhaps slightly unusual in finding risk neutrality in the EU specification). For the RDU model, the results for Nottingham and Zurich are qualitatively very similar. The mean of the value function parameter distribution is close to one in both cases and for 75% of Nottingham subjects we cannot reject the null hypothesis that $\alpha_{RDU} = 1$. The means of the parameter estimates for the probability weighting function are also qualitatively similar across Nottingham and Zurich. The graph presented in Figure 4 plots the probability weighting function based on the median estimates of β and γ of the sample and for 45% of subjects we reject the null hypothesis of $\beta = \gamma = 1$. The two plots are clearly qualitatively similar in displaying the inverse-s shape which overweights (underweights) small (large) probabilities; this is quite typical of the broader empirical literature estimating probability weighting functions, at least for data gathered from tasks with stated (as opposed to learned) probabilities (for a review see Starmer 2000; Fehr-Duda, Gennaro & Schubert 2006). This correspondence between our estimates and those obtained in Zurich (and the broader

[7] The estimate of the α parameter of the EU model did not converge to plausible values for 19 subjects in our data set (e.g. negative estimates) which we drop from the analysis when we assume EU model.

literature) provides some reassurance that our procedures for estimating the risk preference measures are reliable (or at least comparably reliable to those based on similar procedures elsewhere in the literature).

[Insert Figure 4 about here]

The significant non-linearity in utility and probability weighting functions for the majority of our subjects strongly suggests that $Conf_{EVi}$ is a *biased* measure of confidence. Also notice that from the bottom right panel of Figure 3 it is apparent that $Conf_{EVi} < 0.5$ for a significant proportion of individuals (47.5%). Given that each task involved a choice between two options, one of which was right, confidence below 50% is implausibly low. In our incentivised task, however, risk aversion (say as measured by concavity of the utility function) would tend to depress $Conf_{EVi}$. In other words, the data obtained from our incentivised mechanism might seem *more* plausible were we to filter out potential biases attributable to departures from risk neutrality.

Since we have independent measures of individuals' risk parameters (based on responses to Part 1 of the experiment) we can estimate 'decontaminated' or risk-filtered measures of inferred confidence. To be more specific, based on expression (2) above, we calculate inferred confidence, filtered for either EU or RDU as follows:

$$\mathrm{C}onf_{EU_i} = \left(\frac{CE_i}{10}\right)^{\alpha_{EU}} \tag{7}$$

$$Conf_{RDU_i} = w^{-1}\left(\left(\frac{CE_i}{10}\right)^{\alpha_{RDU}}\right) = \frac{1}{\left(\beta * \left(\frac{CE}{10}\right)^{-\alpha} - \beta\right)^{\frac{1}{\gamma}} + 1} \tag{8}$$

Here, $Conf_{EU_i}$is the confidence measure for question *i*, estimated on the assumption that the subject is an expected utility maximiser (and similarly, for $Conf_{RDU_i}$).

The results of filtering out risk in this way are shown in Figure 5. This plots inferred confidence against actual success rates for each question, with separate panels for the EV, EU and RDU models. For comparison, we also reproduce the reported confidence in the bottom right panel. We observe that (i) the extent of underconfidence falls as we move from EV to RDU ($p = 0.025$), (ii) the difference between mean biases of reported and inferred confidence decreases as we filter out risk attitudes ($p = 0.023$), and (iii) inferred confidence is significantly more noisy than reported confidence (Levene (1960) variance equality test: $p = 0.009$). These results suggest that, in the absence of filters for risk attitude, the extent of underconfidence is exaggerated. By filtering out components of these measures attributable to risk attitudes, the overall mean bias falls from -0.212 (inferred confidence under EV) to -0.086 (inferred confidence under RDU).

[Insert Figure 5 about here]

We should emphasise, however, that while confidence miscalibration is reduced as a consequence of allowing for risk attitudes, it is not eliminated and the mean (underconfidence) bias remains significant for all three measures of inferred confidence. Averaging across questions, subjects' success rates are 8.6 percentage points higher than their inferred confidences under our most general (RDU) specification. For comparison, success rates are 2.7 percentage points higher than reported confidence. However, as previously noted, the bias in reported confidence varies with difficulty of the question. Thus a better overall measure of miscalibration is the average absolute bias (i.e. the sum of vertical deviations from the 45 degree line). This is not significantly different for inferred RDU confidence (10.2%) compared to reported confidence (11.6%) (p = 0.666).

4.3. Relationship between Reported Confidence and Risk Attitudes

So far we have focussed on the possibility that risk attitudes may bias confidence measured in an incentivised mechanism. As our second research objective we explore a possible connection between risk attitudes and confidence: that one's (correctly measured) confidence in a given task is related to one's risk attitude. On the face of it, it seems plausible that confidence might be related to risk attitude. For example, some popular contemporary theories of risk preference can be interpreted as allowing some departures from risk neutrality to arise as consequences of the way that people assess and/or respond to probabilities. For example, prospect theory (Kahneman & Tversky 1979, Tversky & Kahneman 1992) can be interpreted as allowing for both misperception of objective probabilities and subjective attitudes to whatever probabilities are perceived. To the extent that such processes reflect generic properties of the way that humans perceive and respond to risks, that provides reason to expect that similar processes might operate in relation to confidence judgements because those judgments *are* assessments of probabilities. In our data set, the cleanest way to investigate this is by looking for an association between individual level risk parameters and reported confidence; the latter is the best confidence measure for our purposes here because it is the only one of our four measures which is independent of risk attitudes (we have already concluded that $Conf_{EV}$ is biased by risk attitudes, while $Conf_{EU}$ and $Conf_{RDU}$ use individual risk parameters as inputs to their estimation).

[Insert Table 2 about here]

Table 2 presents the results of OLS regression where the dependent variable is average reported confidence (subject level). The table reports three specifications which differ according to which of the risk parameters from Part 1 (α_{EU}, α_{RDU}, β , γ) are included. The first specification excludes them all, the second includes just the EU parameter (i.e. α_{EU}), while the third model includes all of the parameters of the RDU model. The latter two allow us to assess whether, and if so by how much, risk attitudes (as captured by EU or RDU models) affect reported confidence judgements. In addition, we also include controls for gender, age, and success rate.

Across all three models, there is no significant association between average reported confidence levels and average success rates across subjects.[8] Females are slightly less confident than males, although the effect is only marginally significant in the specification that includes EU risk parameters; we further discuss the gender results below. There is a small and negative effect of age on reported confidence levels. Turning to our central interest in these estimates, the risk preference parameters are all highly significant predictors of confidence. Moreover, the signs of the coefficients all have quite natural interpretations. For both EU and RDU models, greater risk aversion in the form of curvature of utility (as captured by α_{EU} and α_{RDU}) is associated with lower confidence. From the third specification, incorporating RDU parameters, we find significant effects of the probability weighting parameters. The β parameter controls the elevation of probability weighting function and so has a natural interpretation as "probabilistic optimism" (Bruhin, Fehr-Duda & Epper 2010). The positive (and significant) β coefficient thus suggests a positive association between probabilistic optimism (revealed, in our experiment in choices among lotteries) and confidence (as revealed in judgements about one's own success in quiz tasks). The positive effect of γ also has a natural interpretation. Recall that γ controls curvature of the weighting function, then notice that, for our tasks, success rates are such that we are typically operating

[8] We also checked the relation between confidence and success in a more disaggregate analysis using responses to each question (rather than averages) as the dependent variable. In this analysis, there is a positive and significant association between success and expressed confidence levels; confidence is about 8.5% higher when a subject's answer to a question is correct. This relationship fades away in average subject-level analysis which is consistent with the findings by e.g. Kruger & Dunning (1999) and Massoni & Roux (2012).

in a region where the median subject's weighting function underweights probabilities. In this region, increases in γ reduce underweighting. Hence, the positive sign here is consistent with a positive association between underweighting and underconfidence.[9]

As a coda to this analysis, it may be interesting to note that while there is some evidence of a gender difference in confidence (with females having a tendency towards lower confidence in the EU specification), that difference disappears when we introduce individual-specific parameters of the probability weighting function as controls. This suggests that differences in reported confidence in our data set may be explained by gender-specific differences in attitudes to chance as captured by features of probability weighting functions. Consistent with this, and in line with Fehr-Duda, Gennaro & Schubert (2006), we find that females (compared to men) are less 'optimistic' in the sense of having significantly lower elevation parameters (mean β = 0.609 for females compared with 0.823 for males, Wilcoxon ranksum p = 0.027). As we see it, the primary significance of this coda lies not in identifying a gender effect per se, but rather in underscoring that confidence appears to co-vary with features of individual's risk preferences including both their attitudes to consequences (as captured by curvature of utility) and their attitudes to chance (as captured by the shape of their probability weighting functions). We believe this is a novel, and scientifically interesting, finding suggesting the possibility of common psychological mechanisms underpinning risk attitudes and confidence judgements.

5. Discussion

There is a very large empirical literature investigating confidence judgements and much of this point to the presence of overconfidence in a range of judgements or the existence of a hard-easy effect. The bulk of this literature, however, rests on data generated from non-incentivised self-reports of confidence and, more recently, the robustness of conclusions from this line of research has been challenged by the emergence of a small number of studies by experimental economists which use incentivised tasks to elicit confidence judgements and

[9] We also studied whether confidence and risk attitude parameters co-varied with another standard psychological measure of optimism. This was the Life Orientation Scale (LOT) adopted from Scheier & Carver (1985), and included in our post experimental questionnaire, which classifies individuals according to their optimism (people with positive scores up to a maximum of 16) or pessimism (people with negative scores down to a minimum of -16). We find the LOT score is positively correlated with all four individual risk preference parameters and significantly so in the cases of α_{RDU} and γ (p-values for the respective Pearson correlation coefficients are 0.082 and 0.008). We also find the LOT score is positively correlated with reported confidence (p = 0.013). These results support the interpretation that elicited confidence and risk attitudes reflect common psychological traits.

find that overconfidence bias is considerably reduced. Indeed, in these recent studies, underconfidence is the typical finding.

Our study contributes to this literature, and its central novelty lies in combining two key design features. Like the recent contributions to the economics literature on this topic, we compare confidence miscalibration across incentivised and non-incentivised confidence elicitation tasks. We build into our design procedures for measuring the risk attitudes of our participants coupled with techniques that allow us to track how filtering out risk attitudes affects the measurement of confidence. We are also able to investigate a possible link between reported confidence and risk attitudes at the individual level.

Using a non-incentivised procedure, designed to be very similar to those used in much of the background psychology literature, we reproduce the standard finding of a hard-easy effect. With our new incentivised confidence measurement, regardless of whether or not we filter for risk attitudes, and in line with the recent experimental economics literature, we observe a general tendency towards underconfidence and the hard-easy effect disappears.

Our primary novel findings then relate to the impacts of risk aversion on measured confidence. In the context of incentivised confidence elicitation, we find that filtering out risk attitudes from inferred confidence reduces the degree of underconfidence. We also observe a striking association between risk attitudes inferred from incentivised decisions about lotteries and confidence measured using the standard psychologist's tool. Specifically, individuals who are more risk averse (based on curvature of a best fitting EU function) or more pessimistic (based on best fitting estimates of their RDU probability weighting function) tend to express lower confidence. We also find evidence that gender differences in reported confidence (women tend to be less confident) may be explained by gender differences in specific components of risk attitudes (women tend to be less optimistic, that is they tend to show lower elevation of the probability weighting function).

As far as we know, we are the first to identify that probability weighting may play a significant role in determining confidence judgements. Should we be surprised by this finding? We suspect that priors will differ considerably across economists. To those who tend to think of measured probability weighting as a consequence of more general underlying principles of cognition, the manifestation of those principles in another domain will be reassuring, but not, perhaps especially surprising. We suspect, however, that many other economists aware of evidence for probability weighting may, quite reasonably, think of it as an essentially empirical regularity derived, mainly, from observing choices among simple gambles, with stated probabilities. To those who do interpret it in this, more limited, way our results are arguably much more surprising by establishing a clear empirical connection between responses to probabilities in two very different domains: one involving attachment of certainty equivalents to gambles with stated probabilities (Part 1 of our experiment); the other involving self-reported probability judgements about one's own success rate in a given question (Part 2 of our experiment). We suggest that the ability of measured probability weighting to predict behaviour in these very different tasks and domains should lead to positive reconsideration of the explanatory scope and significance of the concept of probability weighting within economics.

Given that probability weighting does appear to influence confidence judgements, it is natural to ask whether other 'non-standard' aspects of preference in relation to risk or uncertainty might affect confidence judgements. In this respect, an obvious candidate to consider is ambiguity aversion, particularly since confidence judgments appear to be intrinsically ambiguous (as opposed to risky). Although this raises issues beyond the boundaries of the present study, our debriefing questionnaire did include two tasks intended to provide a preliminary assessment of whether, and if so how much, ambiguity attitudes impact confidence judgments. These preliminary investigations failed to reveal any significant relationship between ambiguity attitudes and confidence as measured by our new tool. Nor indeed did we find any relationship between ambiguity attitudes and self-reported confidence. This is, of course, far from conclusive evidence that there is no relationship to discover, and we would certainly support calls for further research into this issue and the broader question - previously highlighted by Hoelzl & Rustichini (2005), Offerman et al. (2009) and Kothiyal et al. (2011) - of how to assess and control the potential impact of ambiguity attitudes in the context of incentivised belief elicitation.

We conclude the present paper with a brief cautionary remark. Economists have, understandably, shown an interest in the large volume of evidence supporting overconfidence. While it seems entirely appropriate to analyse the consequences of confidence miscalibration, it now looks naïve to proceed, as some have done in the past, by simply assuming overconfidence as a reasonable empirical assumption (Odean 1999; Compte & Postlewaite 2004; Malmendier & Tate 2005; Galasso & Simcoe 2011; Gervais, Heaton & Odean 2011). In contrast, our results, alongside other recent work (e.g., Hoelzl & Rustichini 2005; Moore & Healy 2008; Blavatskyy 2009; Clark & Friesen 2009; Merkle & Weber 2011), support the following conclusion: while miscalibration of confidence judgements is a real phenomenon which persists in controlled incentivised decisions, there is currently – and perhaps ironically – apparent overconfidence regarding the empirical significance of overconfidence. We hope that our work provides a helpful input to recalibration.

References

Abdellaoui, M., L'haridon, O., & Paraschiv, C. (2011). "Experienced vs. Described Uncertainty: Do We Need Two Prospect Theory Specifications?" *Management Science*, Vol. 57, pp.1879-1895.

Alba J. W. & Hutchinson J. W. (2000). "Knowledge Calibration: What Consumers Know and What They Think They Know", *Journal of Consumer Research*, Vol. 27, pp. 123-156.

Andersen, S., Harrison, G. W., Lau, M. I. & Rutstrom, E. E. (2006). "Elicitation Using Multiple Price List Formats", *Experimental Economics*, Vol. 9, pp. 383-405.

Becker, A., Deckers, T., Dohmen, T., Falk, A., & Kosse, F. (2012). "The Relationship Between Economic Preferences and Psychological Personality Measures", IZA Discussion Paper Series No. 6470.

Biais, B., Hilton, D., Mazurier, K. & Pouget, S. (2005). "Judgemental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market", *Review of Economic Studies*, Vol. 72, pp. 287-312.

Blavatskyy, P. (2009). "Betting on Own Knowledge: Experimental Test of Overconfidence", *Journal of Risk and Uncertainty*, Vol. 38, pp. 39-49.

Brenner, L. (2000). "Should Observed Overconfidence Be Dismissed as a Statistical Artifact? Critique of Erev, Wallstein, and Budescu (1994)", *Psychological Review*, Vol. 107, pp. 943-946.

Bruhin, A., Fehr-Duda, H. & Epper, T. (2010). "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion", *Econometrica*, Vol. 78, No.4, pp.1375-1412.

Campbell, W. K., Goodie, A. S. & Foster, J. D. (2004). "Narcissism, Confidence, and Risk Attitude", *Journal of Behavioural Decision Making*, Vol. 17, pp. 297-311.

Camerer, C. & Lovallo, D. (1999). "Overconfidence and Excess Entry: An Experimental Approach", *The American Economic Review*, Vol. 89, pp. 306-318.

Clark, J. & Friesen, L. (2009). "Overconfidence in Forecasts of own Performance: an Experimental Study", *The Economic Journal*, Vol.119, pp. 229 - 251.

Compte, O. & Postlewaite, A. (2004). "Confidence-Enhanced Performance", *The American Economic Review*, Vol. 94, pp. 1536-1557.

Cubitt, R., Starmer, C. & Sugden, R. (1998). "On the Validity of the Random Lottery Incentive System", *Experimental Economics*, Vol. 1, pp. 115-131.

Dickinson, D.L. (2006). “The Chilling Effect of Optimism: the Case of Final-offer Arbitration”, *Journal of Socio-Economics*, Vol. 35, pp. 17–30.

Dubra, J. (2004). “Optimism and Overconfidence in Search”, *Review of Economic Dynamics*, Vol. 7, pp. 198-218.

Dunning, D., Heath, C. & Suls, J. M. (2004). “Flawed Self-Assessment: Implications for Health, Education, and the Workplace”, *Psychological Science in the Public Interest*, Vol. 5, pp. 69–106.

Epper, T., Fehr-Duda, H. & Bruhin, A. (2011). “Viewing the Future through a Warped Lens: Why Uncertainty Generates Hyperbolic Discounting”, *Journal of Risk and Uncertainty*, Vol. 43, pp. 169-203.

Erev, I., Wallstein, T.S., & Budescu, D.V. (1994). “Simultaneous Over- and Underconfidence: The Role of Error in Judgement Processes”, *Psychological Review*, Vol. 101, pp. 519-527.

Fehr-Duda, H., Bruhin, A., Epper, T. & Schubert, R. (2010). “Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size”, *Journal of Risk and Uncertainty*, Vol. 40, pp. 147-180.

Fehr-Duda, H., de Gennaro, M. & Schubert, R. (2006). “Gender, Financial Risk and Probability Weights”, *Theory and Decision*, Vol. 60, pp. 283-313.

Fischoff, B., Slovic, P. & Lichtenstein, S. (1977). “Knowing with Certainty: The Appropriateness of Extreme Confidence”, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 3, pp. 552-564.

Galasso, A. & Simcoe, T. S. (2011). “CEO Overconfidence and Innovation”, *Management Science*, Vol. 57, pp. 1469-1484.

Gervais, S., Heaton, J.B. & Odean, T. (2011). “Overconfidence, Compensation Contracts and Capital Budgeting”, *The Journal of Finance*, Vol. 66, pp. 1735-1777.

Gigerenzer, G., Hoffrage, U. & Kleinbolting, H. (1991). “Probabilistic Mental Models: A Brunswikian Theory of Confidence”, *Psychological Review*, Vol. 98, pp. 506-528.

Goldstein, W.M. & Einhorn, H.J. (1987). “Expression Theory and the Preference Reversal Phenomena”, *Psychological Review*, Vol. 94, pp. 236-254.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In: Kurt Kremer, Volker Macho (Eds.): Göttingen :Ges. fürsWiss. Datenverarbeitung, pp. 79-93.

Griffin, D. & Brenner, L. (2008). "Perspectives on Probability Judgment Calibration", in Blackwell Handbook of Judgment and Decision Making (eds D. J. Koehler and N. Harvey), Blackwell Publishing Ltd, Malden, MA, USA, pp.177-199.

Herz, H., Schunk, D. & Zehnder, C. (2014). "How Do Judgmental Overconfidence and Overoptimism Shape Innovative Activity?" *Games and Economic Behavior*, Vol. 83, pp. 1-23.

Hoelzl, E. & Rustichini, A. (2005). "Overconfident: Do You Put Your Money on It?" *The Economic Journal*, Vol. 115, pp. 305-318.

Hollard G., Massoni S. & Vergnaud J.C. (2010). "Subjective Beliefs Formation and Elicitation Rules: Experimental Evidence", Working Paper, CES Paris.

Isoni, A., Loomes, G. & Sugden, R. (2011). "The Willingness to Pay—Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment", *The American Economics Review*, Vol. 101, pp. 991-1011.

Juslin, P., Winman A. & Olsson, H. (2000). "Naive Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard-Easy Effect", *Psychological Review*, Vol. 107, pp. 384-396.

Kahneman, D. & Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, Vol. 47, pp. 263–91.

Keren, G. (1991). "Calibration and Probability Judgements: Conceptual and Methodological Issues", *ActaPsychologica*, Vol. 77, pp. 217-273.

Kothiyal, A., Spinu, V., & Wakker P.P. (2011). "Comonotonic Proper Scoring Rules to Measure Ambiguity and Subjective Beliefs", *Journal of Multi-Criteria Decision Analysis*, Vol. 17, pp. 101-113.

Kruger, J. & Dunning, D. (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*, Vol. 77, pp. 1121-1134.

Lattimore, P. K. (1992). "The Influence of Probability on Risky Choice: A Parametric Examination", *Journal of Economic Behavior & Organization*, Vol. 17, pp. 377-400.

Levene, H. (1960). "Robust tests for equality of variances", in Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, ed. I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann, 278-292. Menlo Park, CA: Stanford University Press.

Liberman, V. (2004). “Commentary: Local and Global Judgements of Confidence”, *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 30, pp. 729-732.

Lichtenstein, S. & B. Fischoff (1981). “The Effects of Gender and Instructions on Calibration”, Decision Research Technical Report 81-5, Decision Research, Eugene, OR.

Lichtenstein, S., B. Fischoff & L. D. Phillips (1982). “Calibration of Probabilities: The State of the Art to 1980”, in D. Kahneman, P. Slovic and A. Tversky (eds), Judgment Under Uncertainty): Heuristics and Biases, Cambridge Press. Cambridge. MA. pp. 306-334.

Lundeberg, M. A., Fox, P. W., Brown, A. C. & Elbedour, S. (2000). “Cultural Influences in Confidence: Country and Gender”, *Journal of Educational Psychology*, Vol. 92, pp. 152-159.

Malmendier, U. & Tate, G. (2005). “CEO Overconfidence and Corporate Investment”, *Journal of Finance*, Vol. 60, pp. 2661-2700.

Massoni, S. & Roux, N. (2012). “Optimal Group Decision: A Matter of Confidence Calibration”, Working Paper, CES Paris.

Merkle, C. & Weber, M. (2011). “True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence”, *Organizational Behaviour and Human Decision Processes*, Vol. 116, pp. 262-271.

Moore, D. A. & Cain, D. M. (2007). “Overconfidence and Underconfidence: When and Why People Underestimate (and Overestimate) the Competition”, *Organizational Behaviour and Human Decision Processes*, Vol. 103, pp. 197-213.

Moore, D. & Healy P. J. (2008). “The Trouble with Overconfidence”, *Psychological Review*, Vol. 115, pp. 502-517.

Niederle, M. & Vesterlund, L. (2007). “Do Women Shy Away from Competition? Do Men Compete too Much?” *Quarterly Journal of Economics*, Vol. 122, pp. 1067-1101.

Odean, T. (1999). “Do Investors Trade Too Much?” *The American Economic Review*, Vol. 89, pp. 1279-1298.

Offerman, T., Sonnemans, J., van de Kuilen, G. & Wakker, P. (2009). “A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes”, *Review of Economic Studies*, Vol. 76, pp. 1461-1489.

Rustichini, A., De Young, C. G., Anderson, J. & Burks, S. V. (2012). “Towards the Integration of Personality Theory in the Explanation of Economic and Health Behavior”, IZA Discussion Paper Series No. 6750.

Schaefer, P. S., Williams, C. C., Goodie, A. S. & Campbell, W. K. (2004). “Overconfidence and the Big Five”, *Journal of Research in Personality*, Vol. 38, pp. 473-480.

Scheier, M. F., & Carver, C. S. (1985). “Optimism, Coping, and Health: Assessment and Implications of Generalized Outcome Expectancies”, *Health Psychology*, Vol. 4, pp. 219-247

Starmer, C. (2000). “Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk”, *Journal of Economic Literature*, Vol. 38, pp. 332-382.

Starmer, C. & Sugden, R. (1991). “Does the Random-Lottery Incentive Systems Elicit True Preferences? An Experimental Investigation”, *The American Economic Review*, Vol. 81, pp. 971-978.

Suantak, L., Bolger, F. & Ferrel, W.R. (1996). “The Hard-Easy Effect in Subjective Probability Calibration”, *Organizational Behaviour and Human Decision Processes*, Vol. 67, pp. 201-221.

Svenson, O. (1981). “Are we all less risky and more skilful than our fellow drivers?” *Acta Psychologica*, Vol. 47, pp. 143-148.

Tversky, A. & Kahneman, D. (1992). “Advances in Prospect Theory: Cumulative Representation of Uncertainty”, *Journal of Risk Uncertainty*, Vol. 5, pp. 297–323.

Urbig, D., Stauf, J. & Weitzel, U. (2009). “What is your level of overconfidence? A strictly incentive compatible measurement of absolute and relative overconfidence”, Discussion Paper Series 09-20, Utrecht School of Economics.

Wakker, P. P. (2010). “Prospect Theory: For Risk and Ambiguity”, Cambridge University Press, Cambridge, UK.

Wu, G., Zhang, J., & Gonzalez, R. (2004). “Decision under Risk”, in The Blackwell Handbook of Judgment and Decision Making, ed. by D. Koehler and N. Harvey. Oxford: Oxford University Press, 399–423.

Figures

Which of the following cities has the larger population?
□ City X □ City Y
Tick one of the boxes to indicate your answer.

In each row of the table choose either Option A or B.

Row	Option A: Lottery	Your Choice A	Your Choice B	Option B: Guaranteed Amount
1		□	□	£10.00
2		□	□	£9.50
3		□	□	£9.00
4		□	□	£8.50
5		□	□	£8.00
6		□	□	£7.50
7		□	□	£7.00
8		□	□	£6.50
9	You get £10.00 if your city choice is correct and £0.00 if not	□	□	£6.00
10		□	□	£5.50
11		□	□	£5.00
12		□	□	£4.50
13		□	□	£4.00
14		□	□	£3.50
15		□	□	£3.00
16		□	□	£2.50
17		□	□	£2.00
18		□	□	£1.50
19		□	□	£1.00
20		□	□	£0.50

Figure 1: Our Confidence Elicitation Tool

For each row of the table please choose either Option A or B				
Row	Option A: Lottery	Your Choice		Option B: Guaranteed amount of
		A	B	
1	50% chance of £10.00 and 50% chance of £0.00	□	□	£10.00
2		□	□	£9.50
3		□	□	£9.00
4		□	□	£8.50
5		□	□	£8.00
6		□	□	£7.50
7		□	□	£7.00
8		□	□	£6.50
9		□	□	£6.00
10		□	□	£5.50
11		□	□	£5.00
12		□	□	£4.50
13		□	□	£4.00
14		□	□	£3.50
15		□	□	£3.00
16		□	□	£2.50
17		□	□	£2.00
18		□	□	£1.50
19		□	□	£1.00
20		□	□	£0.50

Figure 2: Sample Risk Elicitation Tool

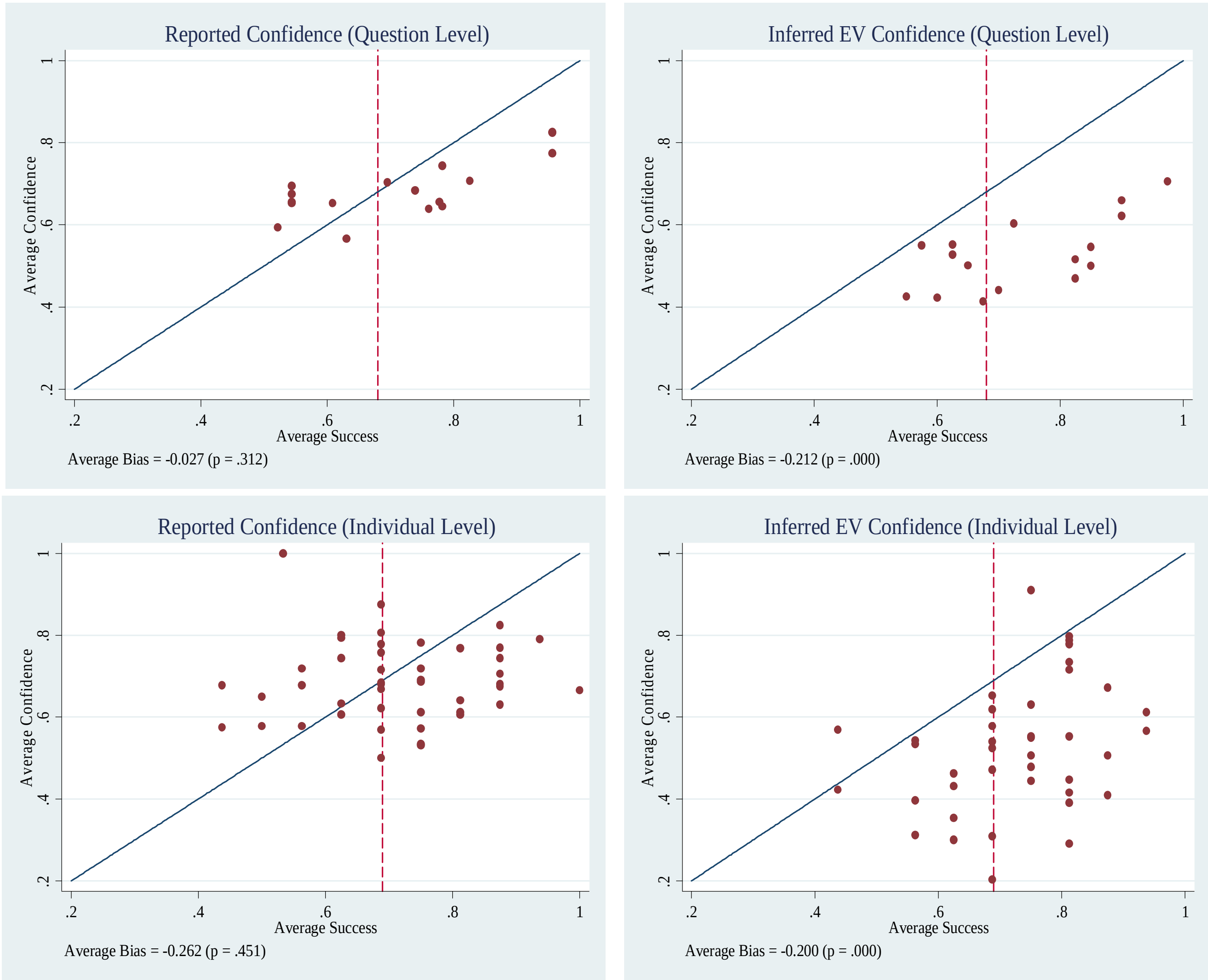
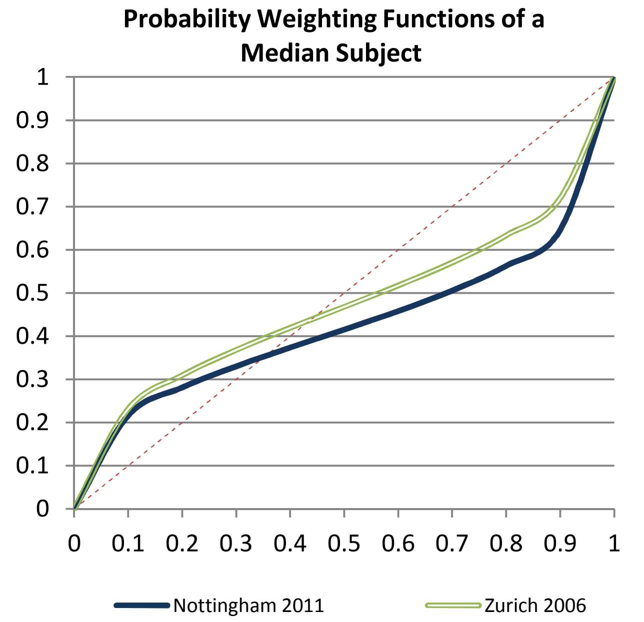


Figure 3: Confidence and Success

Top panels: Each dot represents a question. For a given question Bias = (average confidence) – (average success) across subjects. Average bias is the average bias across questions and the reported p-value is for a two-tailed t-test that mean of distribution of biases is zero. Bottom panels: Each dot represents a subject. For a given subject bias = (average confidence across questions) – (success rate across questions). Average bias is the average across subjects and the reported p-value is for a two-tailed t-test that the mean of distribution of biases is zero.



The mean of the Estimated Parameters for Nottingham and Zurich Samples

	Nottingham (n = 86)	**Zurich (n = 134)**
α_{EU}*	0.868 (0.24)	1.055 (0.26)
α_{RDU}	1.047 (0.22)	1.104 (0.29)
β	0.729 (0.16)	0.884 (0.17)
γ	0.477 (0.07)	0.533 (0.10)

Mean standard errors are in parentheses
* 19 (Nottingham) and 5 (Zurich) subjects excluded due to non-convergence

Figure 4: Estimates of risk preference parameters: the plot is the weighting function based on the median estimates of β and $\gamma$of the sample

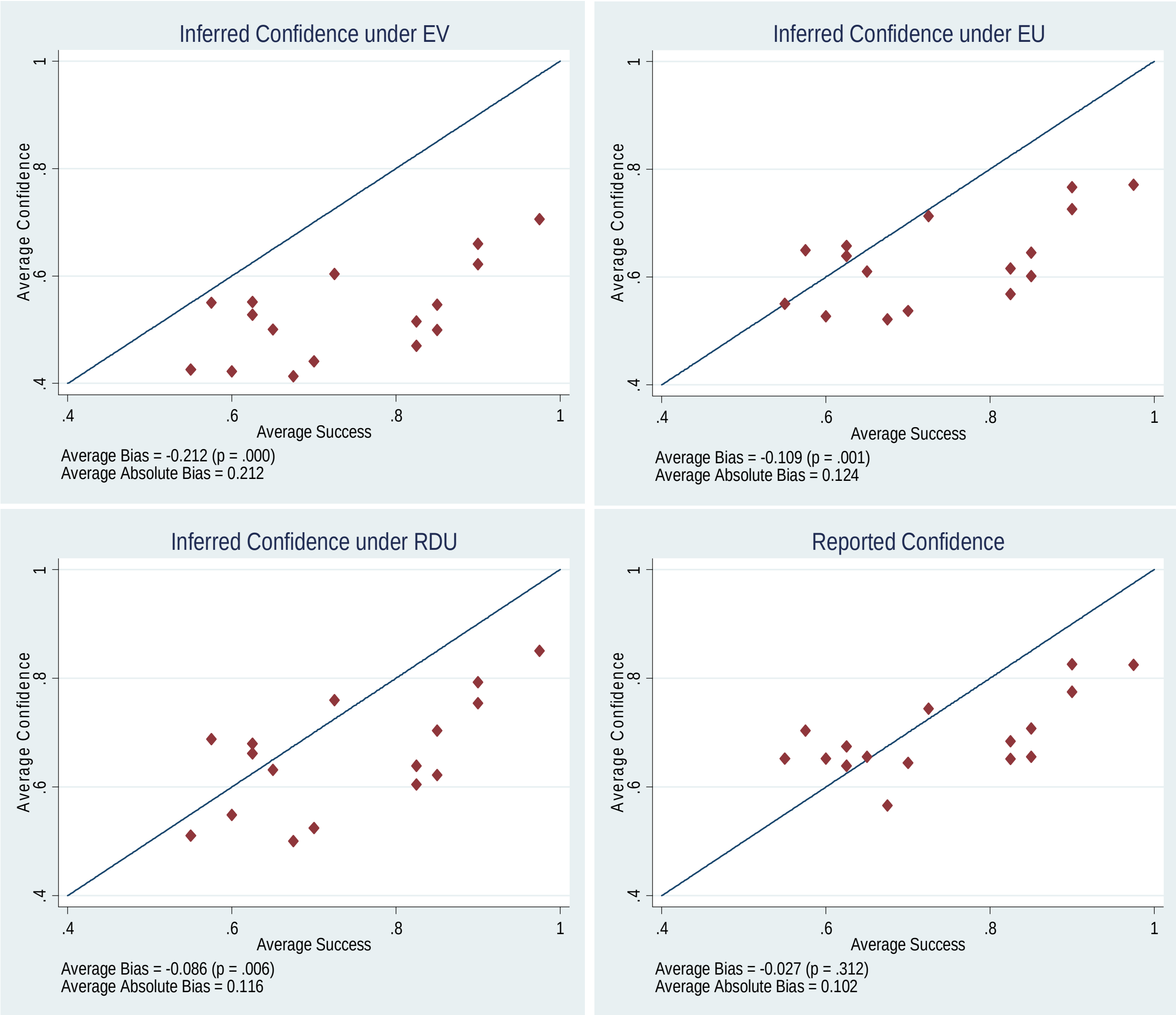


Figure 5: Risk adjusted Confidence and Success

Each dot represents a question. For a given question Bias = (average confidence – average success) across subjects. Average bias is the average of biases across all questions and the reported p-value is for a two-tailed t-test that mean of distribution of biases is zero. Absolute Bias = Absolute (average confidence– average success) across subjects.

Tables

Table 1: Risky Prospects of Part 1 of the experiment

Lottery	*p*	x_1	x_2	Lottery	*p*	x_1	x_2
1	0.05	£4	£0	**14**	0.5	£10	£0
2	0.05	£8	£2	**15**	0.5	£10	£4
3	0.05	£10	£4	**16**	0.5	£30	£0
4	0.05	£30	£10	**17**	0.75	£4	£0
5	0.1	£2	£0	**18**	0.75	£8	£2
6	0.1	£4	£2	**19**	0.75	£10	£4
7	0.1	£10	£0	**20**	0.9	£2	£0
8	0.25	£4	£0	**21**	0.9	£4	£2
9	0.25	£8	£2	**22**	0.9	£10	0
10	0.25	£10	£4	**23**	0.95	£4	£0
11	0.5	£2	£0	**24**	0.95	£8	£2
12	0.5	£4	£2	**25**	0.95	£10	£4
13	0.5	£8	£2				

p denotes the probability of the first outcome, x_1

Table 2: Determinants of Average Reported Confidence

Explanatory Variables	*No risk controls*	*EU risk controls*	*RDU risk controls*
$\boldsymbol{\alpha_{EU}}$		.084***	
		(0.02)	
$\boldsymbol{\alpha_{RDU}}$			.119***
			(0.05)
$\boldsymbol{\beta}$			.090**
			(0.43)
$\boldsymbol{\gamma}$			.111**
			(0.04)
Average Success	.006	.-0.003	.034
	(0.11)	(0.14)	(0.10)
Female	-.047	-.053*	-.038
	(0.03)	(0.03)	(0.03)
Age	-.011**	-.015***	-.013*
	(0.01)	(0.01)	(0.01)
Constant	.872***	1.14***	.69***
$\boldsymbol{\bar{R}^2}$	.034	.209	.267
$\boldsymbol{n}$	43	33	43

** 10%, ** 5%, *** 1% significance levels*
Standard errors are in parentheses
43 subjects in Model EV and RDU, and 33 subjects in Model EU (because of missing α_{EU} parameter for some) with pooled OLS regression