# EXPERIMENTS ON CONFIDENCE CALIBRATION AND DECISION MAKING

Zahra Murad, M.A.

THESIS SUBMITTED TO THE UNIVERSITY OF NOTTINGHAM
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

February 2014

# Experiments on Confidence Calibration and Decision Making

## *Abstract*

This thesis reports on three experiments studying subjects' confidence about performance on a task and how it relates to decision-making under uncertainty. Chapter 1 introduces the thesis providing an overview of the common themes and methods underlying this research.

Chapter 2 reports the first experiment, investigating the relationship between risk attitudes and confidence judgements. We measure confidence in two different ways, with an incentivized elicitation tool and with unincentivized self-reports. Using our incentivised tool we find that, in the absence of controls for risk attitudes, subjects tend to be underconfident about their own performance. When we filter out the effects of risk attitudes we find that underconfidence is reduced, but not eliminated. We also identify an interesting link between self-reported confidence and risk attitudes in that experimental subjects with less concave utility functions and more elevated probability weighting functions tend to report higher confidence levels.

Chapter 3 reports the second experiment, investigating the role of information in experimental market entry games. We look at whether individual over-entry to simple and under-entry to difficult markets disappears when subjects make entry decisions in groups or are given statistical information about performance of previous subjects. We find that individuals and groups are both susceptible to the same type of biases in entry and both fail to learn from repetition and feedback. We find that individuals learn to de-bias their entry

decisions in the second half of the experiment when given explicit information about the performance of others.

Chapter 4 reports an experiment investigating "snowballing of confidence" in hierarchical tournaments. We analyse how high/low scorers of a group in one stage of the tournament change their confidence levels in the next stage when they are re-grouped with other high/low scorers. We find that all subjects start the tournament assigning an equal chance to being high or low scorers in their groups. As they proceed through the stages, low scorers become more underconfident whereas high scorers become more overconfident about their relative performances. We also identify an interesting difference in the perceptions of the task between high and low scorers that is linked to self-serving causal attribution biases previously found in the psychology literature.

Chapter 5 summarizes the findings of this dissertation and concludes.

# ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisors, Professor Chris Starmer and Professor Martin Sefton. Both Professor Starmer and Professor Sefton taught me valuable skills in experimental methodology and guided me through every step on the road of becoming an academic. I have thoroughly enjoyed taking their modules as well which have significantly contributed to the development of this dissertation.

A number of people other than my supervisors were also prepared to discuss any questions I had or advice I needed. In particular, I would like to thank Professor Robin Cubitt, Dr Daniele Nosenzo, Professor Fergus Bolger, and Simone Quercia.

My gratitude also extends to the other members of the Centre for Decision Research and Experimental Economics (CeDEx) for their many useful suggestions at various stages of this research, their financial support during the period of my research, and providing the resources and facilities for running experiments.

I also would like to thank the School of Economics at the University of Nottingham in general for offering me an inspiring and encouraging environment to complete this dissertation. I would like to thank my colleagues Francesco, Simone, Jianan, Debasmita, Emilija, Jonathan, Samantha, Steven and Jeroen for making the PhD years an extremely pleasurable experience.

I owe special thanks to my family for their unconditional love and unlimited patience from the very beginning. And, last but not least, I would like to thank Osman, who has always supported me through thick and thin.

**CONTENTS**

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1: INTRODUCTION

This thesis is a collection of three papers (Chapters 2, 3 and 4) reporting studies that contribute to different research areas in the field of behavioural and experimental economics. While each paper is self-contained and can be read independently from the others, there is a common theme of *confidence judgements* underlying the research questions examined in these studies, as well as commonalities in the methodology and research strategies used to address these questions.

The research questions in the three papers of this thesis are inspired by considerations about the importance of confidence judgements for understanding the outcomes of economic decisions and interactions among individuals. While standard economic theory relies on the simplifying assumptions that all economic agents are perfectly aware of their skills, abilities and relative standings in the distribution of other agents, there is a large body of evidence gathered through survey and experimental studies that shows that people make systematic mistakes in judging their absolute and relative skill levels. The literature in confidence judgements is divided into two main strands that study either absolute confidence or relative confidence. *Absolute confidence* is the judgement of one's own performance independent of others' performances whereas *relative confidence* is the judgement of one's performance relative to that of others.

The early studies on absolute confidence judgements (e.g. Fischoff, Slovic & Lichtenstein 1977; Lichtenstein & Fischoff 1981; Lichtenstein, Fischoff & Phillips, 1980) mostly find overconfidence or a "hard-easy" effect (overestimation of performance in hard tasks, and underestimation of performance in easy tasks). The robustness of the findings has been subsequently challenged by, for example Suantak, Bolger, & Ferrel (1996) for response scale effect, by Gigerenzer, Hoffrage & Keinbolting (1991) for frequentist versus probabilistic belief elicitation and by Erev, Wallstein, & Budescu (1994) for possible asymmetries in random error in judgement. Irrespective of the source of a systematic miscalibration of confidence from actual performance, it is still the case that the miscalibration can have important effects on economic decisions such as job search, bargaining behaviour, investment decisions and trading behaviour (Dubra 2004; Biais, Hilton, Mazuier & Pouget 2005; Dickinson 2006). Most of the studies measuring absolute confidence calibration have used a number of elicitation devices to elicit confidence levels and measure calibration between confidence and performance. They report that with incentives confidence levels are lower than without and overcconfidence may be a product of incentivized elicitation procedures (Blavatskyy 2009; Clark & Friesen 2009; Urbig, Stauf & Weitzel 2009; Hollard, Massoni & Vergnaud 2010).

The second chapter, entitled "How Do Risk Attitudes Affect Measured Confidence?", uses a laboratory experiment to study how absolute confidence is affected by individual risk attitudes. Specifically, we study how risk attitudes (attitudes to consequences and to probabilities) contaminate elicited confidence with incentivized elicitation tools and how we can filter out risk attitudes and infer de-contaminated confidence. We also study whether confidence elicited through non-incentivized self-reports correlates with individual risk attitudes, to explore whether there is an intrinsic relationship between confidence and risk preferences measurable at the individual level. We design our experiment with two between-subject treatments to address both of our research questions. In our unincentivized self-reported confidence treatment, we replicate the standard finding in the psychology literature of "hard-easy" effect – overestimation of success rate in hard and underestimation of success rate in easy tasks. We find that risk attitudes bias confidence downwards in the incentivized confidence treatment and filtering out risk attitudes decreases observed underconfidence and brings average confidence closer to well calibration. We find a significant relationship between individual risk attitudes and self-reported confidence judgements, such that people with more risk seeking utility functions and more optimistic probability weighting functions tend to report higher confidence levels.

The third and fourth chapters of the thesis study relative confidence judgements. The previous research in relative confidence judgements has found that people tend to judge themselves as performing better than average in common and/or easy tasks such as driving or completing an easy quiz (Svenson 1981; Camerer & Lovallo 1999; Moore 2007) and rate themselves as performing worse than average in rare and/or difficult tasks such as graduating in the top of their grade, computer programming, and unicycle riding (Windschitl, Kruger & Simms 2003; Kruger & Burrus 2004; Kruger, Windschitl, Burrus, Fessel & Chambers 2008).

In the third chapter, entitled "The Role of Implicit and Explicit Information in Entry to Competitions", we replicate and extend the research of Moore & Cain (2007) investigating whether, and if so how, individuals learn to avoid coordination failures in entry decisions to experimental markets where their performance in a task decides their final outcome. Their study shows that giving subjects full feedback in a repeated market entry game does not help to de-bias entry decisions: people over/underenter to markets where a participants's ranking in a prior easy/difficult task is a determinant of their success in a market compared to a case where there is no performance task. This has been shown to be caused by subjects' beliefs that they are better than average in easy and worse than average in difficult tasks which is not mitigated by repetition and feedback. We study whether providing subjects with historical average performance information (explicit information) or having subjects make decisions in groups (implicit information) will de-bias their relative confidence judgements and eliminate coordination failures in entry behaviour to experimental markets.

We use three treatments to answer our research question of how more information can help to overcome coordination failures in entering competitive experimental markets. In the control treatment, we use the design of Moore & Cain (2007) to study the individual entry decisions across 12 rounds of market entry game with varying difficulty level tasks. We replicate their finding that there is significant overentry to simple and underentry to difficult markets and subjects do not learn to coordinate their entry decisions across difficulty levels with feedback and repetition. In our explicit information treatment, we give subjects average historical performance information at each skill round, so that subjects know how previous participants have scored in a given task. In the implicit information treatment, we seat subjects in groups of three and ask them to make their decisions as a single group unit. Our hypothesis is that groups can serve as implicit informational channels about performance of others and help to decrease the coordination failures in entry. We test this hypothesis and study differences between groups and individuals in a competitive decision making context. This chapter contributes to the literature studying group versus individual decision making and beliefs as important determinants of competitive decisions. Our results show significant treatment differences in both entry decisions and confidence levels of subjects. We find that only explicit information is successful in de-biasing subjects' entry decisions but only combined with feedback and repetition. Groups make significantly better judgements about others' performance (and so do serve as implicit informational channels) but fail to incorporate those judgements in their behaviour to make better entry decisions.

In the fourth chapter, "Snowballing Confidence in Hierarchical Tournaments", we study snowballing of relative confidence in a context of hierarchical tournaments through multiple stages. We introduce a novel confidence elicitation device in the context of a skill-based performance task, and we use if to track how confidence changes from one stage to another as subjects learn about their relative performance in the previous stage and are grouped with those similar to them in the new stage. Ours is one of the first studies to measure and track confidence in such a multiple stage setting. Our elicitation device is incentive compatible under non-EU risk preferences. This study is related to the different strands of literature such as tournaments and contests (e.g. Dargnies 2012; Ludwig, Wichardt & Wickhorst 2011; Park & Santos-Pinto 2010), asymmetric belief updating (e.g. Mobius, Niederle, Niehaus & Rosenblat 2011; Eil & Rao 2011; Eberlein, Ludwig & Nafziger 2011; Grossman & Owens 2012) and differences in relative confidence across simple and difficult tasks (e.g. Moore & Cain 2007; Hoelzl & Rustichini 2005) however, is unique as it is the first study explicitly study confidence snowballing.

The results of the fourth chapter show that subjects on average do not display any systematic relative confidence bias in the first stage of the tournament where on average 50% confidence is assigned to being in the top scoring half. In the next stages, we observe significant snowballing of confidence, where top scorers are matched with other top scorers increase their confidence levels and bottom scorers matched with other bottom scorers decrease their confidence levels from one stage to another. Furthermore we find significant differences between top and bottom scorers' perceptions of the task, where top scorers rate the task as a more skill task and bottom scorers rate the task as a more luck task which exhibits a bias in attribution of causality to successes and failures.

All of the studies use novel experimental strategies to elicit confidence levels. The advantage of using experimental methodology is the control we have over the data generating process and the ability to manipulate the contexts we want to measure confidence in. There are a number of field studies on confidence literature that are very ingenious in their designs and research questions. Some early studies for example measure confidence calibration of meteorological weather forecasters using the experience of the forecaster and the past inaccuracy as proxies for the amount of feedback and repetition (Murphy & Daan 1984; Murphy & Winkler 1984). The field studies on relative confidence usually lack objective performance measures to study confidence as an objective social comparison measure and hence pose a difficulty in interpreting elicited confidence (Svenson 1981; Hoorens & Buunk 1993). A recent study by Park & Santos-Pinto (2010) measured relative confidence of poker and chess players in national championships by asking them their rank forecasts before the start of the tournaments. However due to limitations in the structure of the tournaments, they are not able to elicit and track the changes in confidence of the tournament winners and losers. Experimental methods are especially useful for studying confidence in a tournament setting, since in field settings players (winners and especially losers) may be hard to track, self-selection may play an important role and controlling for the causality and in the environment may be more difficult where more than one variable may change from one stage to another. In this dissertation, we are able to systematically elicit confidence, manipulating both the treatment conditions to check for causality and specific features of the decision environment that we are specifically interested in with the help of experimental tools.

The abundance of studies in the psychology literature studying confidence both in absolute terms (calibration studies) and relative terms (social comparison studies) has definitely attracted attention of economists interested in how beliefs shape economic decisions. Little is known about how individual risk attitudes are related to confidence judgements or how people update their confidence when they move from competing within one reference group to the other. The focus of this thesis is to contribute to extending the application of confidence research in the discipline of economics and in the domain of economically relevant settings where confidence impacts decisions.

Chapter 5 concludes by summarizing the results of chapters 2 to 5, pointing out their limitations and suggesting directions for further research.

## *References*


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

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

Camerer, C., & Lovallo, D. (1999). "Overconfidence and Excess Entry: An Experimental Approach". *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.

Dargnies, M. P. (2012). "Men Too Sometimes Shy Away from Competition: The Case of Team Competition". *Management Science*, Vol. 58, pp. 1982-2000.

Dickinson, D. L. (2006). "The Chilling Effect of Optimism: The Case of Final-offer Arbitration". *The 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.

Eberlein, M., Ludwig, S., & Nafziger, J. (2011). "The Effects Of Feedback On Self-Assessment". *Bulletin of Economic Research*, Vol. 63, pp. 177-199.

Eil, D., & Rao, J. M. (2011). "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself". *American Economic Journal: Microeconomics*, Vol. 3, pp. 114-138.

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

Fischhoff, 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.

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

Grossman, Z., & Owens, D. (2012). “An Unlucky Feeling: Overconfidence and Noisy Feedback”. *Journal of Economic Behavior & Organization*, Vol. 84, pp. 510-524.

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.

Hoorens, V., & Buunk, B. P. (1993). “Social Comparison of Health Risks: Locus of Control, the Person-Positivity Bias, and Unrealistic Optimism”. *Journal of Applied Social Psychology*, Vol. 23, pp. 291-302.

Kruger, J., & Burrus, J. (2004). “Egocentrism and Focalism in Unrealistic Optimism (and Pessimism)”. *Journal of Experimental Social Psychology*, Vol. 40, pp. 332-340.

Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2008). “The Rational Side of Egocentrism in Social Comparisons”. *Journal of Experimental Social Psychology*, Vol. 44, pp. 220-232.

Lichtenstein, S., & Fischhoff, B. (1981). “The Effects of Gender and Instructions on Calibration” (No. PTR-1092-81-7). Decision Research Eugene OR.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. “Calibration of Probabilities: The State of the Art to 1980”, Kahneman D., Slovic P., Tversky A., Judgment under Uncertainty: Heuristics and Biases, 1982, pp. 306-334.

Ludwig, S., Wichardt, P. C., & Wickhorst, H. (2011). “Overconfidence can Improve an Agent's Relative and Absolute Performance in Contests”. *Economics Letters*, Vol. 110, pp. 193-196.

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). “Managing Self-confidence: Theory and Experimental Evidence”. National Bureau of Economic Research working paper No. w17014.

Moore, D. A. (2007). “Not so Above Average After All: When People Believe They are Worse than Average and its Implications for Theories of Bias in Social Comparison”. *Organizational Behavior and Human Decision Processes*, Vol. 102, pp. 42-58.

Murphy, A. H., & Daan, H. (1984). “Impacts of Feedback and Experience on the Quality of Subjective Probability Forecasts". *Comparison of Results from the First and Second Years of the Zierikzee Experiment. Monthly Weather Review*, Vol. 112, pp. 413-423.

Murphy, A. H., & Winkler, R. L. (1984). “Probability Forecasting in Meteorology”. *Journal of the American Statistical Association*, Vol. 79, pp. 489-500

Park, Y. J., & Santos-Pinto, L. (2010). “Overconfidence in Tournaments: Evidence from the Field”. *Theory and Decision*, Vol. 69, pp. 143-166.

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

Svenson, O. (1981). “Are We All Less Risky and More Skillful than Our Fellow Drivers?”. *Acta Psychologica*, Vol. 47, pp. 143-148.

Urbig, D., Stauf, J., & Weitzel, U. (2009). “What is Your Level of Overconfidence? A Strictly Incentive Compatible Measurement of Absolute and Relative Overconfidence”. Tjalling C. Koopmans Institute Discussion Paper Series, Vol. 9, pp. 1-40.

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). “The Influence of Egocentrism and Focalism on People's Optimism in Competitions: When What Affects us Equally Affects Me More”. *Journal of Personality and Social Psychology*, Vol. 85, pp. 389-408.

# CHAPTER 2: HOW DO RISK ATTITUDES AFFECT MEASURED CONFIDENCE?

## *2.1. 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 confidence 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 2.3 describes our experimental methodology in more detail.

In Section 2.4 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 2.5 discusses these results and concludes.

## *2.2. 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 study, 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

[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.

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.

## *2.3. 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.

### *2.3.1. 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.

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

[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.

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%).

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**

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 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:

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

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.

[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.

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.

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.

[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 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**

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:

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

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, $\alpha$, 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 $\beta$ captures 'elevation' of the probability weighting function (with greater $\beta$ reflecting more 'optimism'); the parameter $\gamma$ controls curvature (the smaller is $\gamma$, 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.

### *2.3.2. 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, 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.

**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,* $x_1$

The experiment was conducted at the University of Nottingham, CeDEx lab in 2011. Subjects were recruited using Orsee (Greiner 2004). In total 86

[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).

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.

## *2.4. 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]

### *2.4.1. Reproducing Standard Results*

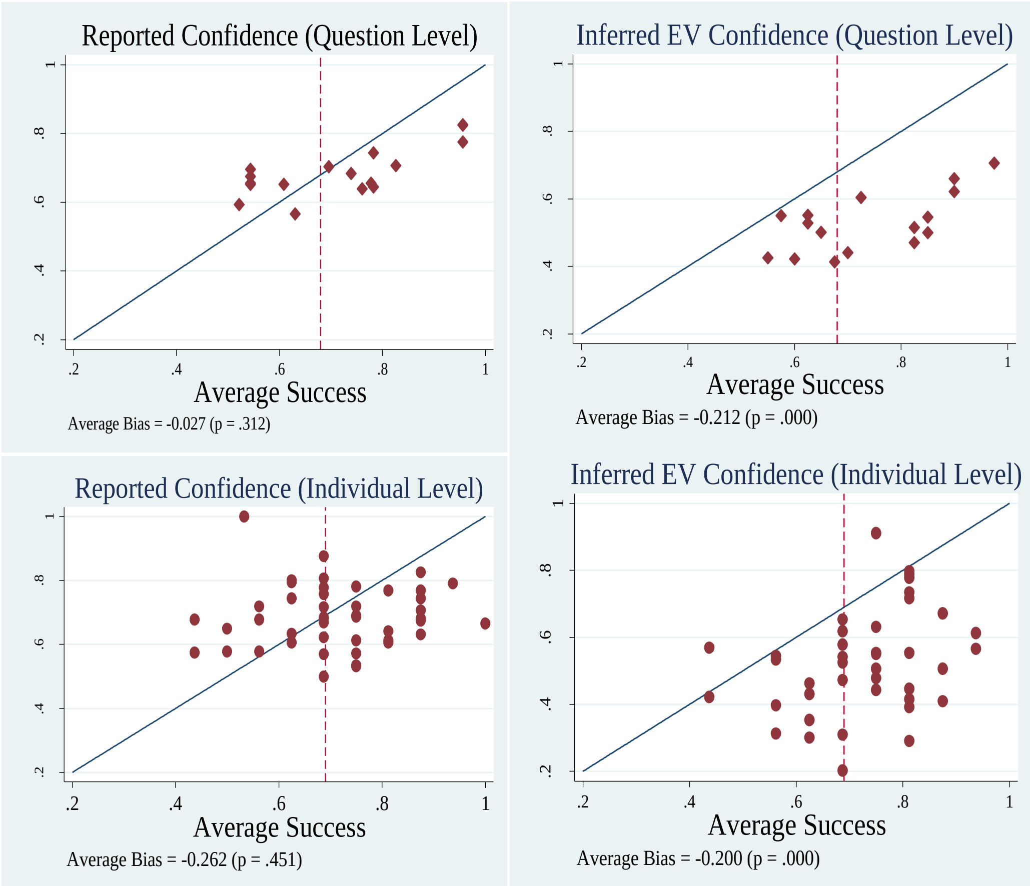
Figure 3 provides a quick eye-balling tool for comparing confidence measured using the standard psychological tool with confidence elicited using our incentivised mechanism (on the assumption that individuals are risk neutral). Consider first the top left panel. This plots, for each quiz question, the 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.

The reported confidence data have a pattern consistent with the familiar 'hard-easy effect'. To highlight this, we have drawn a vertical (dashed) line

[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.

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).




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.

**Figure 3: Confidence and Success**

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.

### *2.4.2. Risk Preferences and Risk-Filtered Confidence*

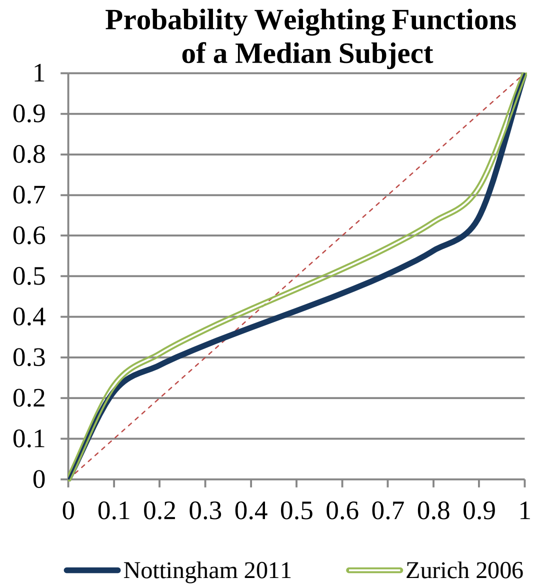
As we explained above, if individuals are not risk neutral, then confidence measures elicited via our incentivised mechanism may be biased because they may capture a mixture of confidence assessments and risk attitudes (and similarly so for other incentive mechanisms that have so far been used for this 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 ($\alpha_{EU}$); the value function and probability weighting parameters under RDU ($\alpha_{RDU},\ \beta,\ \gamma$); and the normalized standard deviations of the decision errors ($\sigma_{EU}$ and $\sigma_{RDU}$). 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 $\alpha_{EU}$ for Nottingham is substantially less than one, and for 85% of our sample we reject the null hypothesis of $\alpha_{EU}$ = 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 $\beta$ and $\gamma$ 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 literature) provides

[7] The estimate of the $\alpha$ 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.

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).




**The mean of the Estimated Parameters for Nottingham and Zurich Samples**

| | Nottingham (n = 86) | Zurich (n = 134 ) |
|---|---|---|
| $\alpha_{EU}$* | 0.868 (0.24) | 1.055 (0.26) |
| $\alpha_{RDU}$ | 1.047 (0.22) | 1.104 (0.29) |
| $\beta$ | 0.729 (0.16) | 0.884 (0.17) |
| $\gamma$ | 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

The plot is the weighting function based on the median estimates of β and γ of the sample.

**Figure 4: Estimates of Risk Preference Parameters**

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:

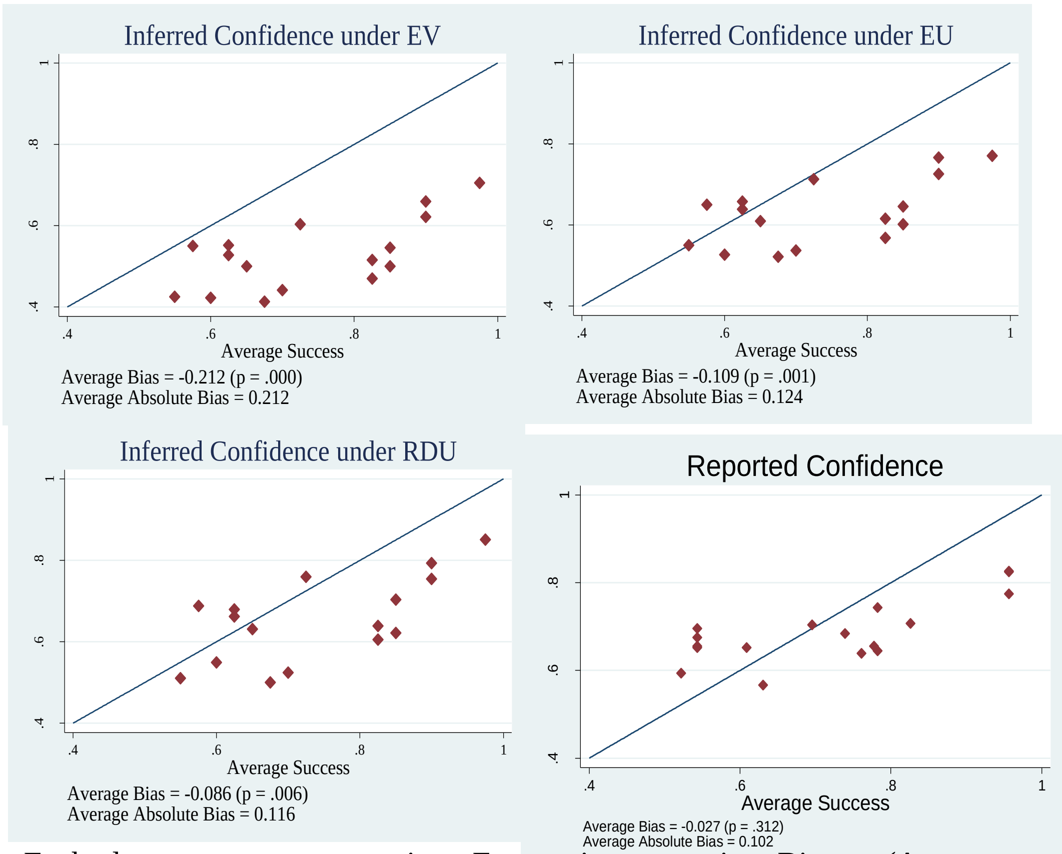
$$\mathrm{C}onf_{EU_i} = \left(\frac{CE_i}{10}\right)^{\alpha_{EU}} \tag{7}$$

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

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).

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




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.

**Figure 5: Risk Adjusted Confidence and Success**

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).

### *2.4.3. 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).

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 ($\alpha_{EU}$, $\alpha_{RDU}$, $\beta$ , $\gamma$) are included. The first specification excludes them all, the second includes just the EU parameter (i.e. $\alpha_{EU}$), 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

[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).

interpretations. For both EU and RDU models, greater risk aversion in the form of curvature of utility (as captured by $\alpha_{EU}$ and $\alpha_{RDU}$) is associated with lower confidence. From the third specification, incorporating RDU parameters, we find significant effects of the probability weighting parameters. The $\beta$ 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) $\beta$ 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 $\gamma$ also has a natural interpretation. Recall that $\gamma$ controls curvature of the weighting function, then notice that, for our tasks, success rates are such that we are typically operating in a region where the median subject's weighting function underweights probabilities. In this region, increases in $\gamma$ reduce underweighting. Hence, the positive sign here is consistent with a positive association between underweighting and underconfidence.[9]

[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 $\alpha_{RDU}$ and $\gamma$ (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.

**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 $\alpha_{EU}$ parameter for some) with pooled OLS regression*

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 $\beta$ = 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.

## *2.5. 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 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 Economic 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 and 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.

# CHAPTER 3: THE ROLE OF IMPLICIT AND EXPLICIT INFORMATION IN MARKET ENTRY GAMES

## *3.1 Introduction*

Previous research has shown that the industries with high entry rates also tend to have high rates of business failure (Dunne, Roberts & Samuelson 1988; Mata & Portugal 1994). The often cited phenomenon in the literature labelled as the *better than average* (BTA) effect (also known as *relative overconfidence*) has been used as one of the explanations for high entry and failure rates of businesses (Camerer & Lovallo 1999). More recently though, researchers have found evidence of a worse than average (WTA) effect in situations involving difficult tasks where people are usually underconfident relative to others (for a review see Moore 2007). Observing the two effects, Cain, Moore & Haran (2013) study whether the perceived ease of operating in the market can determine entry rates to the market. They use real market entry data in various industries and show that there is excessive entry to industries classified as "simple" (e.g. food stores, hobby shops and restaurants) and insufficient entry to industries classified as "difficult" (e.g. forestry, agriculture and fabricated metal production) controlling for costs of entry.

In this paper we study the BTA/WTA effect in the context of experimental market entry games where decision makers play a role of an entrepreneur deciding whether to enter a market or stay out. More specifically, we look at the effect of implicit and explicit information on the BTA/WTA effect in the context of repeated decisions in experimental markets where we manipulate difficulty levels. In two between-subject treatments, we either manipulate subjects being informed about others' performances *implicitly* via being a part of a group or *explicitly* via receiving statistical information about historical performance in a task. We test the extent of BTA/WTA effect across information treatments and how it interacts with repetition and feedback over multiple rounds of market entry game.

In almost all previous BTA/WTA studies, individuals have some information about themselves (their general ability, memories and experiences) and relatively little information about others (Svenson 1981; Weinstein 1980; Camerer & Lovallo 1999). This is also the case in one of our treatments where individuals make decisions about whether or not to enter a market only knowing about their own performances in a task. In another treatment, entry decisions are made by groups. In this case the interaction within a group gives individuals additional information about others' abilities in a task and so information about others is implicitly embedded in the decision environment via group decision making. By comparing entry decisions and beliefs of groups to that of individuals we test whether groups can serve as implicit informational channels about competitors' abilities and whether this will reduce BTA/WTA effects.

A broader motivation is to study the behaviour of groups in competitive environments. Many important real world competitions are between *collections* of individuals (teams or groups) rather than *single* individuals, such as sport tournaments, inter-organizational grant competitions and inter-firm competitions to capture higher market share. Observing different patterns of behaviour and beliefs between individuals and groups may suggest whether groups are prone to certain biases more than individuals (or vice versa) and hence justify the use of either individual or group decision making in certain environments. There are relatively few studies in the literature looking at the behaviour of groups compared to individuals in competitive settings (exceptions include Sutter & Strassmair 2009; Sheremeta & Zhang 2010; Healy & Pate 2011; Dargnies 2012). Our paper adds to this small literature by studying competitive behaviour and relative performance assessment of groups compared to individuals. The related literature is discussed in detail in Section 2.

To study the effect of explicit information on experimental entry behaviour and beliefs, we directly manipulate the knowledge of individuals about population performance by giving individuals statistical information about average historical performance in a task. We choose to study historical information instead of real time information in the lab for two reasons. First, outside of the lab, historical information can be retrieved much more easily and is more accessible than real time information, and thus has more relevance from an external validity standpoint. Second, the higher accessibility of historical information, in turn, makes it a more relevant tool for both agents and policy makers. For example, entrepreneurs revising their business strategies can base their strategies on historical performance information across sectors and competitors, while competition organizers can give historical information to potential competitors to deter or encourage entry to sports tournaments. Comparing entry decisions and beliefs of individuals' with and without historical performance information, we will test whether explicit information mitigates or eliminates BTA/WTA effect. The full details of our experimental design are presented in Section 3.

Section 4 presents our results. The results show that both groups and individuals choose to enter excessively to simple markets and too often stay out of the difficult markets. Groups are better at predicting the entry rate and average performance of others than individuals but nevertheless demonstrate a similar biased pattern in their entry decisions to individuals. When individuals are given explicit historical information about performance, the BTA/WTA effect in entry behaviour is much less in the first half of the experiment and eventually dies out entirely in the second half of the experiment. We find that beliefs about own relative standings can explain entry behaviour better than other elicited beliefs. We also identify an interesting gender effect in our study which shows that females are less confident than men but only when they are deciding individually rather than deciding in a group. Section 5 discusses the implications of our findings and concludes.

### *3.2 Related Literature*

In an experimental market entry game, subjects decide to enter a market or stay out and each of those who entered receives a ranking. Depending on a pre-determined market capacity, a certain number of high ranked players receive a positive payoff and those ranked below that number receive a negative payoff which is less than a safe payoff of staying out. It has been shown that the Nash equilibrium solution predicts very well the aggregate entry behaviour of subjects in these types of games where rankings are randomly assigned to each entrant (Sundali, Rapoport, & Seale 1995; Rapoport, Seale, Erev, & Sundali 1998). Camerer & Lovallo (1999) exploit this property of experimental market entry games to measure the relative confidence of individuals in a competitive setting. In contrast to previous findings, subjects fail to coordinate in their entry decisions when their rankings are determined according to their scores in a quiz task. The authors conclude that people are overconfident in their relative abilities and hence enter excessively.

Camerer & Lovallo's finding is consistent with the relative overconfidence bias often cited in the psychology literature. Recent developments, however, have suggested a more complex pattern in relative confidence judgements by identifying situations where people usually believe they are worse than others. These situations involve difficult tasks such as computer programming, unicycle riding, coping with a death of a loved one and graduating in the top 1% of a class (Windschitl et al. 2003; Kruger & Burrus 2004; Kruger, Windschitl, Burrus & Fessel 2008). Don Moore and his colleagues in a series of recent papers show that Camerer and Lovallo's finding is valid for tasks that are perceived as simple. For tasks perceived as difficult, underconfidence is more prevalent and causes insufficient entry rate below market capacity and hence, foregone potential payoffs (Moore & Cain 2007; Moore, Oesch & Zietsma 2007; Moore & Small 2007; Radzevick & Moore 2008; Cain, Moore & Haran 2013).

Moore & Cain (2007) investigate whether repetition and feedback eliminate overentry/underentry to markets with simple/difficult tasks. They show that neither overentry nor underentry is reduced through repetition and feedback. Looking at the beliefs of participants, they find that people are accurate in predicting the number of entrants (the same is true in Camerer and Lovallo's paper) but demonstrate significant BTA/WTA effects in beliefs about own percentile rankings relative to competitors. Hence, they conclude that it is the relative confidence that primarily determines entry decisions. In this paper, we adopt the same protocol of entry decisions and belief elicitations as Moore & Cain (2007) to test if over/underentry to easy/difficult markets is reduced or eliminated with the informational treatment manipulations.

Whether group decision making provides implicit informational channels about others' performances and enhance assessments of own relative performances is, as far as we are aware, a novel research question. The previous literature finds that groups are found to make more self-interested, outcome oriented, strategic decisions and are cognitively more sophisticated than individuals, hence make less errors in their decisions and learn faster (for a review see Charness & Sutter 2012). Groups of three have been shown to perform better when competing against individuals in beauty contest games because they are one step ahead of individuals in hierarchical reasoning (Kocher & Sutter 2005; Sutter 2005; Kocher, Strauss, & Sutter 2006). Market entry games are similar to beauty contest games in the way that one should engage in hierarchical reasoning to predict others' beliefs about entry behaviour by responding to these beliefs accordingly in their entry decisions. Also teams of two players have been shown to behave more strategically in signalling game experiments demonstrating more learning transfer between games compared to individuals (Cooper & Kagel 2005). Given these findings, we expect that groups will be better informed about the performances of the competitors and more successful in strategizing the decision environment in market entry games.

The literature on competition between groups versus individuals is scarce and mostly concentrates on differences in the strategies chosen in games.

For example, McCallum, Harring, Gilmore, Drenan, Chase, Insko & Thibaut (1985) look at the differences between dyads and individuals in prisoner's dilemma and "mutual fate control" games and find that groups are more competitive and less cooperative than individuals mostly caused by groups' stronger desire to win or avoid losing. Sutter & Strassmair (2009) look at the effect of communication in tournaments between and within teams and find that communication increases chosen effort levels within teams and decreases between teams. Healy & Pate (2011) and Dargnies (2012) study how competition in teams can help to close the gender gap in competitive preferences. They compare tournament entry behaviour of individuals to dyads and show that women tend to enter competitions more often while men tend to avoid competitions when in teams. Sheremeta & Zhang (2010) show that dyads make less risky decisions, which explains why they make lower bids in contests. Our study will be one of the first studies to directly compare individuals' competition entry behaviour and relative confidence to that of groups. As in many real world group interaction settings and differently from the other existing studies we do not restrict communication within the group.

One of the causes for the emergence of BTA/WTA effects in competitive environments is argued to be "differential information" - the asymmetry of information one possesses about self versus others (Moore 2007). In our third treatment, we look at the effect of explicit information about the average historical performance of others on market entry behaviour. This treatment is closely related to the second study of Moore & Cain (2007) and a more recent study by Ewers (2013). Moore and Cain check how historical information of the distribution of performance affects relative percentile rankings of individuals. Subjects were asked the question "What percentage of the group will have scores below yours?" before taking the quiz, after taking the quiz but before receiving historical distributional information, and after receiving information. They show that after subjects were given information, the BTA/WTA effect in their percentile rankings disappears, but slight overconfidence persists: subjects on average place themselves on 55$^{th}$ percentile in both difficult and simple tasks. We will be looking at the effect of historical information on percentile beliefs of our subjects and *also* whether it translates into their market entry decisions. Ewers (2013) match subjects in pairs who complete a task and can choose to enter the competition with the other or stay out. The author shows that giving information about opponents' average or distribution of performances eliminates competition failures: potential losers successfully stay out of the competition. Differently from Ewers' study, we choose to use historical performance information instead of real time information about the performance of the current opponents. Our motivation in using historical information was that historical performance information is more available in real world settings than real time performance information and hence is a more available tool from a perspective of policy makers.

### *3.3 Experimental Design*

We extended Moore & Cain's (2007) experimental design which in turn was built on the design of Camerer & Lovallo (1999). Our experiment consisted of 12 rounds of a market entry game. In each round, 7 players decided simultaneously and without communication whether to enter a market or stay out. The market capacity, which was equal to 3, determined how many of the entrants would earn money. Entrants were ranked either randomly or by their scores in a trivia quiz. Those ranked below 3$^{rd}$ lost money, while non-entrants lost nothing. The payoff depended on the rank of an entrant, such that higher ranked entrants earned more than lower ranked ones; payoffs were determined according to Table 1 which was presented to subjects during instructions. The three types of rounds were Simple – ranking according to a simple quiz, Difficult – ranking according to a difficult quiz and Random – ranking randomly. Since subjects participated in both random- and quiz conditions, their decisions in the random rounds act as a within-subject control for risk attitudes. The difference in the number of entrants in the random and quiz rounds is the primary measure of interest of over and under entry.

To study the effect of group decision making as a possible implicit information channel, we compare the entry decisions and beliefs of individual decision maker (treatment *Individual*) to those of a group of three subjects (treatment *Group*). We also test how explicit information affects entry decisions and beliefs of our subjects by giving them historical average performance information (treatment *IndividualInfo*) and comparing to those of subjects in the *Individual* treatment.

In each round we gave our subjects a decision sheet according to which they knew whether it was a random round or a quiz round. We did not explicitly separate quiz rounds as simple and difficult to avoid framing effects. In a random round, a decision maker had to choose whether to enter a market or stay out and received a random score which determined his ranking in case he entered. In a quiz round, in addition to deciding about entry, a decision maker answered a 5 question quiz. He received a score according to the number of his correct answers to a quiz and was ranked according to this score if he chose to enter the market. Each quiz had a tiebreaker question which eliminated any possibility of a tie depending on the answer's distance from the correct numerical answer.

**Table 1: Experimental Payoff from Entering a Market**

| **Rank** | **Payoff** |
|---|---|
| 1st | £14 |
| 2nd | £10 |
| 3rd | £5 |
| 4th | -£10 |
| 5th | -£10 |
| 6th | -£10 |
| 7th | -£10 |

We wanted to observe the effect of our treatment variables on entry decisions and beliefs interacting with feedback and repetition to see whether there is any learning throughout the experimental session. Each session had four blocks of three rounds. Each block consisted of one Random, one Simple and one Difficult round. The order was counterbalanced either in *SimpleRandomDifficult* or *DifficultRandomSimple* order across experimental sessions to avoid order effects and stayed the same for all four blocks within a session. Simple round quizzes had an average of 4.4 correct answers whereas difficult round quizzes had an average of 0.7 correct answers out of 5. The quiz questions were designed and pretested so that the variance of the number of correct answers was small across subjects and at least 5 of 7 subjects would get the same number of correct answers. The four simple and four difficult quizzes appear in Appendix A. The order of the quizzes across blocks was also randomized.

Each round of a market entry game consisted of an entry decision, belief elicitation and feedback stages. The entry decision stage was timed as 3 minutes if it was a quiz and as 1 minute if it was a random round. After taking their entry decisions and before getting feedback, subjects answered the following five questions about their beliefs regarding their own and others' performance.

1. How many entrepreneurs[10] in total do you think entered the market this round? Include yourself in this figure if you chose to enter.
2. How many of the other six entrepreneurs in this round do you think scored higher than you did (regardless of whether anyone entered)?
3. How many quiz questions (out of questions 1-5) do you think you got correct in this round?
4. How many quiz questions (out of questions 1-5) do you think the average entrepreneur got correct this round?
5. If you chose to enter the market this round, what rank do you think you will get?

These questions measured beliefs of individuals as well as groups. By contrasting the beliefs of groups and individuals, we will be able to test whether groups appear more informed than individuals and if so we will say that groups are acting as implicit informational channels. Belief elicitation is not incentivized as we wanted to replicate exactly the Camerer & Lovallo (1999) and Moore & Cain (2007) design so as to be able to compare our results to theirs. After answering these questions, every individual/group received full feedback on their own and others' scores, entry decisions and rankings.[11] The feedback for all previous rounds stayed on the computer screens throughout the experimental session. The individual/group numbers were anonymous and could not be linked to participants' identities. Each individual/group knew their

[10] Moore and Cain's (2007) instructions were framed such that subjects played a role of an entrepreneur who could open a restaurant and we followed their procedure. Instructions can be found in Appendix B.

[11] The sample feedback screen can be seen in Appendix C.

own number but did not know other individuals'/groups' numbers and in which cubicle they were seated.

The Individual and IndividualInfo treatments were identical with the only difference between treatments being one sentence in their entry decision sheets following the quiz. The sentence read as "The average number of correct answers in previous sessions was … out of 5". A sample decision sheet can be seen in Appendix D. In the Group treatment, 21 participants were randomly allocated 7 group numbers and seated in a cubicle in groups of three.[12] Group members could discuss their answers and decisions face-to-face among themselves and submit one decision sheet per group. Subjects were given enough time to submit their decisions and were free to do anything they want during this time as long as they did not try to communicate to other groups or use any electronic devices.

There were 18 experimental sessions in total, each with 7 decision makers: there were 6 Individual, 6 IndividualInfo and 6 Group treatment sessions with a total of 210 participants.[13] The experiment was conducted in the CeDEx laboratory where participants were seated in cubicles and lasted on average 60 minutes. The experiment was in pen and paper format, where the decision sheets were immediately analysed and feedback was given through Z-tree (Fischbacher, 2007) via computer monitors on participants' desks. The experiment was not fully computerized because of the open-ended nature of the quiz questions.

---

[12] Groups were seated so that there was ample space between each group to prevent contamination and attempt of interaction between groups. They were also visually segregated and could not see anyone except their own group members.

[13] All participants were British students at the University of Nottingham recruited through Orsee (Greiner 2004).

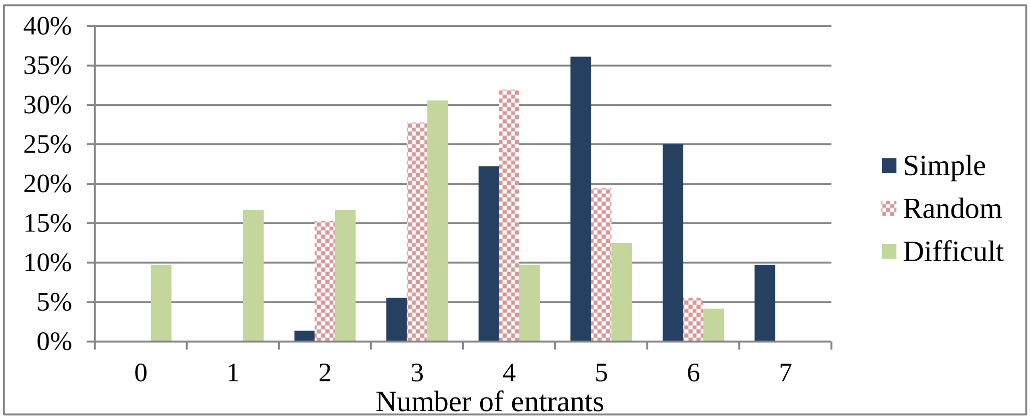
At the end of the 12 rounds, three rounds were randomly chosen for payoff. The earnings (or losses) from these three rounds were averaged and this amount was added to (or subtracted from) the participant's £10 endowment. In the Group condition, each group member earned the same amount of money which was calculated identically to the individual treatments. The maximum possible payoff to a participant was £24, if a player entered and ranked first in all three payoff rounds (£10 endowment plus £14 earning). It was also possible for a subject to leave the experiment empty-handed if in all three payoff rounds he entered and ranked below third (-£10 average loss plus £10 endowment). The actual average payment to a participant was £11.7 with a range of £0 to £24.

### *3.4 Results*

In this section, we present our results looking at the market entry behaviour and beliefs of our subjects. We check whether we replicate the standard finding in the literature of overentry to simple and underentry to difficult markets across our treatment conditions and if there are significant differences between treatments in entry behaviour. By looking at entry behaviour and beliefs of individuals with and without historical information, we will examine whether additional information mitigates or eliminates BTA/WTA effect and how it interacts with repetition and feedback over rounds. We check whether groups are more informed than individuals in their beliefs and whether this translates into better entry decisions by comparing groups' behaviour and beliefs to that of individuals. We then examine whether elicited beliefs, or other individual characteristics, predict entry behaviour across our treatment conditions.

### *3.4.1 Entry Rates*

First we look at whether we replicate the previous finding that people enter excessively to simple markets and stay out too often from difficult markets. Figure 1 presents the entry frequencies to different markets for all three treatment conditions pooled together. There is a significant difference in entry rates between simple and difficult rounds. The distribution of entry rates is skewed to the left for difficult rounds and to the right for simple rounds. In random rounds, subjects successfully coordinate their entry decisions on the mixed Nash equilibrium such that in most of the random rounds (80%), between 3 and 5 people enter the market.[14] In simple rounds, on the other hand, we observe coordination failure where in most of the simple rounds (72%), 5 or more subjects entered. In difficult rounds, on the other hand, subjects forego potential profits by entering too little: in 70% of difficult rounds 3 or fewer players entered.




Pooled across Individual, IndividualInfo and Group treatments. There are 72 rounds per simple, random and difficult conditions.

**Figure 1: Histogram of Entry Rates in Three Different Ranking Conditions**

[14] Coordinating on market capacity is Pareto Efficient equilibrium of market entry games but there is a positive expected payoff of entering if less than five players enter. Mixed Nash equilibria and theoretical predictions of market entry games are discussed in Moore & Cain (2007).

**Table 2: The Mean Difference Between Entry to Easy and Difficult Rounds**

| | Block 1 | Block 2 | Block 3 | Block 4 |
|---|---|---|---|---|
| **Individual** | 4 | 4 | 1 | 2.34 |
| | (*p=.000*) | (*p=.002*) | (*p=.111*) | (*p=.008*) |
| **Individual_Info** | 2 | 2.5 | 0.5 | -0.33 |
| | (*p=.010*) | (*p=.042*) | (*p=.363*) | (*p=.611*) |
| **Group** | 5.16 | 4 | 0.66 | 2.17 |
| | (*p=.000*) | (*p=.009*) | (*p=.394*) | (*p=.041*) |

*The p-values are from a two tailed t-test to test for the hypothesis of the difference being equal to zero.*

To assess the effect of repetition and feedback on entry behaviour, we look at the dynamics of entry rates across rounds for each of the three experimental treatments (Figure 2). Block 1 reports the first three rounds of the simple, random and difficult conditions, Block 2 reports the second three rounds and so forth. The entry rate to random rounds serves as a benchmark entry rate for a non-risk neutral decision maker. Hence the significant difference in entry between random and quiz rounds is an evidence of over/under entry.

In the first block of the experiment, we observe a significant difference in entry rates to easy and difficult rounds in all of the treatments. Table 2 reports the mean difference between entry rates to easy and difficult rounds across blocks and treatment conditions. In Blocks 1 and 2, there are significant differences between simple and difficult rounds in all of the treatments. In Block 3, the difference disappears and this is mainly driven by the entry rate to difficult markets which increases across blocks. This pattern in Block 3 was also evident in Moore & Cain's (2007) data. In the last block, we observe a different pattern for each treatment. Whereas in the Individual and Group treatments there is significant positive difference in entry rates to easy and difficult markets, we observe higher entry to difficult than to simple markets in IndividualInfo treatment (although it is not statistically significant) in Block 4. Given these results, we conclude that repetition and feedback work only when combined with the provision of additional historical performance information that subjects receive in the IndividualInfo treatment which eliminates the bias completely in the last two blocks of the experiment.

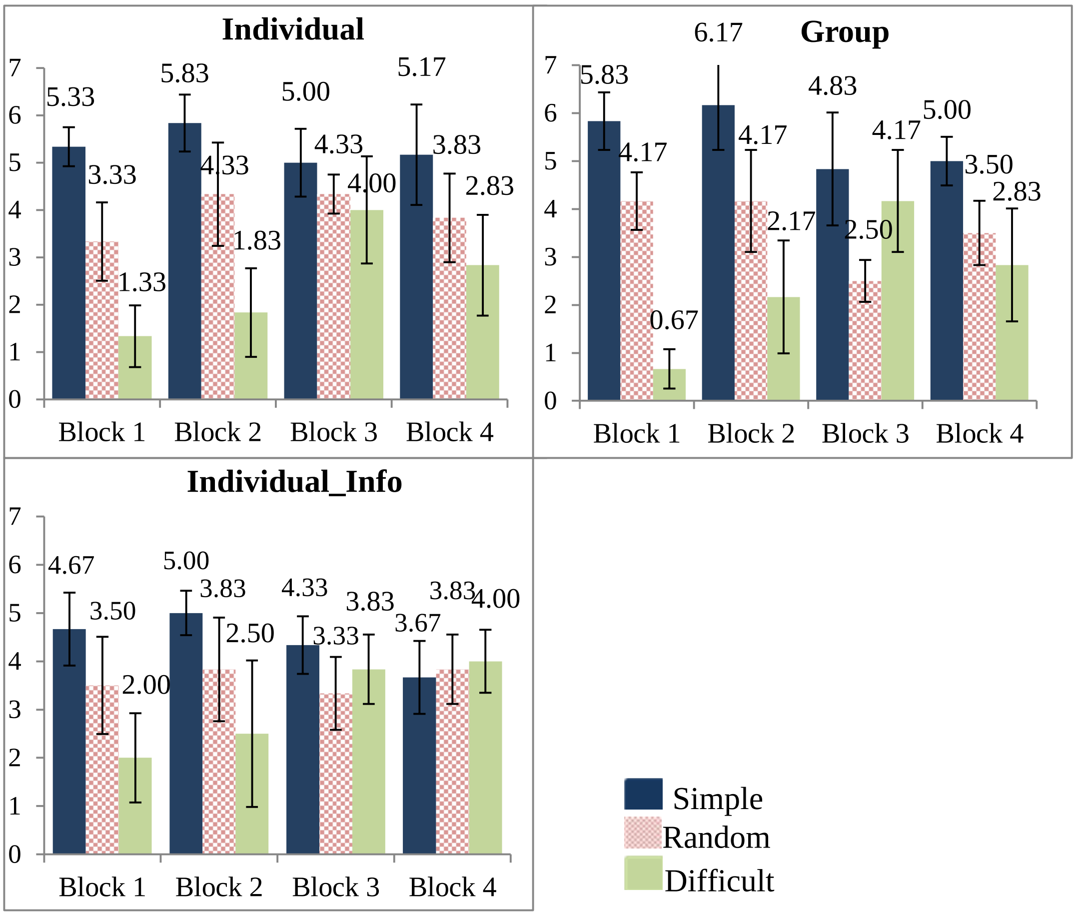



**Figure 2: Entry Rates in Three Different Ranking Conditions Across Four Blocks**.
Error bars show 95% confidence intervals

### *3.4.2 Beliefs*

We elicited subjects' beliefs with a questionnaire at the end of each round after entry decisions were taken. We analyse the accuracy of subjects to predict the entry rate in a given round and check whether beliefs differ across treatments and blocks. To study the BTA/WTA effect, we elicit subjects' beliefs about their relative performance rankings which serve as a measure of relative confidence with respect to competitors. We test whether groups serve as implicit information channels for decision makers so that groups are more informed about the skills and entry decisions of competitors compared to individuals. We also test whether giving individuals historical information about past performances makes them more accurate in their beliefs.

#### *3.4.2.1 Beliefs About Entry Rates*

After entry decisions were taken, we asked subjects to predict how many entrants there were in a given round. We measure the accuracy of subjects' beliefs by the absolute difference between predicted and actual entry rates, where the smaller the difference the more accurate one is. Over all rounds, the average absolute difference between predicted and actual entry rate was 1.5 in the Individual treatment, 1.32 in the Group treatment and 1.17 in the IndividualInfo treatment indicating that subjects in the IndividualInfo treatment were the most accurate. If we test the equality of the predicted entry rate to the actual entry rate for each round, we reject the null hypotheses of equality in 9 out of 12 rounds in Individual, 7 out of 12 rounds in Group and 6 out of 12 rounds in IndividualInfo treatments (Table E1 of Appendix E). This is in contrast to the findings of Camerer & Lovallo (1999) and Moore & Cain (2007) who find that individuals are able to correctly predict entry rates of others in almost all rounds. In Table 3, we regress predictive accuracy of the entry rate on treatment dummies, dummy variables for the first or the second part of the experiment, the difficulty level and their interaction terms. We observe that groups and individuals with information are better at predicting entry rates compared to individuals without information. The positive coefficient of the *Block* dummy indicates a significant improvement of predictions of entry rates in the second half of the experiment compared to the first half. Moreover, subjects are less accurate in predicting entry rates in difficult rounds than in simple rounds as seen by a significant coefficient of the *Difficulty* dummy.

**Table 3: Dependant Variable Absolute Difference Between Predicted and Actual Entry Rates**

| | Model 1 | Model 2 |
|---|---|---|
| **IndividualInfo** | -0.334*** | -0.452*** |
| **Group** | -0.181* | -0.384*** |
| **Block** | -0.129** | -0.345*** |
| **Block*IndividualInfo** | | 0.23* |
| **Block*Group** | | 0.408*** |
| **Difficult** | | 0.331*** |
| **Female** | | 0.059 |
| **Constant** | 1.56*** | 1.67*** |
| **N** | 1505 | 1505 |
| **Adj-Rsq** | 0.018 | 0.022 |

*OLS regression clustered by individual and group id. * 10%, ** 5%, *** 1% significance level. Block is a dummy variable for the first and second half of the experiment. Difficult is a dummy variable for the difficult and simple tasks. Female variable in the Group treatment is measured as the proportion of females in a group.*

### *3.4.2.2 Relative Confidence Beliefs*

To assess BTA/WTA effect in beliefs across difficulty levels, we elicit subjects' relative confidence beliefs by asking them "How many of the other six entrepreneurs in this round do you think scored higher than you did?". The lower the number they report, the more confident they are in their assessment of scoring relative to others. Answering 3 to the question means that a subject places himself on the median of ranking $4^{th}$ out of 7 participants and hence we will test the null hypotheses of average expected ranks being equal to 4. The change of expected relative confidence across blocks is graphed in Figure 4 showing the histogram of mean beliefs about their ranks (1 added to their answers to the above question to test rank equals 4).

On average, subjects in the Individual treatment believed to rank 4.1 in simple rounds (Wilcoxon sign rank test p=0.443) and on average 4.5 in difficult rounds (Wilcoxon sign rank test p=0.001). This indicates that individuals, on average, were well calibrated in simple tasks but underconfident in difficult tasks and this result is valid in all four blocks of the Individual treatment. However, we have to note that the significant underconfidence in difficult tasks is mainly driven by underconfidence in Block 2 and is not observed in the other blocks. Removing Block 2, on average we observe well calibration both in simple and difficult rounds of the Individual treatment. We observe a similar pattern in the IndividualInfo treatment. We cannot reject the null hypothesis of subjects being well calibrated in simple rounds ranking themselves on 3.96 (Wilcoxon sign rank test p=0.735) and on average being significantly underconfident in difficult rounds ranking themselves on 4.51 (Wilcoxon sign rank test p=0.003). Differently from the Individual treatment, there is significant difference in confidence between simple and difficult rounds in the first half of the IndividualInfo treatment (3.76 versus 4.77, Wilcoxon sign rank test p=0.002) and no significant difference between simple and difficult rounds in the second half of the treatment (4.15 versus 4.25, Wilcoxon sign rank test p=0.619) This result is caused by a significant decrease in relative confidence in simple rounds (from 3.76 to 4.15, Wilcoxon sign rank test p=0.051) and a significant increase in relative confidence in difficult rounds (from 3.77 to 4.25, Wilcoxon sign rank test p=0.063) from first to the second half of the experiment. We conclude that giving subjects historical information helps subjects to *learn* to calibrate their beliefs about their relative performance from first to the second half of the experiment, whereas we do not observe the same learning for the subjects in the Individual treatment.

Groups on average are relatively overconfident in simple rounds reporting ranking on 3.05 (Wilcoxon sign rank test p=0.000) and we cannot reject the null hypothesis of them being relatively well calibrated in difficult rounds reporting ranking on 4.10 (Wilcoxon sign rank test p=0.401). There is a more pronounced difference in confidence between simple and difficult rounds in the first two blocks (2.74 versus 3.62, Wilcoxon sign rank test p=0.000) which disappears in the third block (3.55 versus 3.38, Wilcoxon sign rank test p=0.193) but remains significant in the last block of the experiment (3.14 versus 3.79, Wilcoxon sign rank test p=0.001).

We identify an interesting gender effect in beliefs about relative performance. We find that females, on average, are relatively less confident and place themselves in lower ranks (and more specifically below median rank) than males both in the Individual (4.83 versus 3.72, Wilcoxon ranksum p=0.000) and the IndividualInfo (4.33 versus 4.00, Wilcoxon ranksum p=0.022) treatments. In the Group treatment, we compare the relative confidence of the groups containing only females to the groups containing only males and find no significant difference between them (3.91 versus 3.71, Wilcoxon ranksum p=0.628). It has been previously found that when in groups females are as competitive as males and do not shy away from competition (Healy & Pate 2011; Dargnies 2012). Our finding that women in groups have similar beliefs about their relative performances as men in groups offers an explanation why gender gap closes when competition is between groups than between individuals.

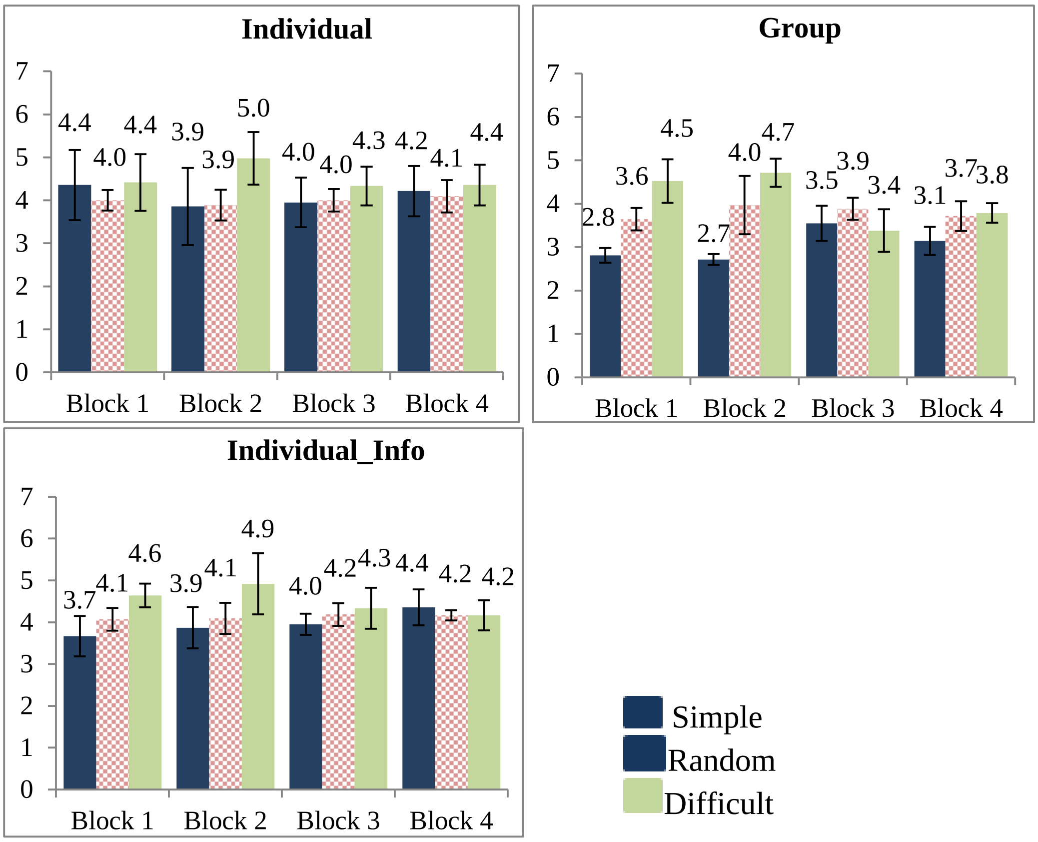



**Figure 3: Relative Confidence Rankings in Three Different Ranking Conditions across the Four Blocks**.
Error bars show 95% confidence intervals

### *3.4.2.3 Beliefs About Absolute Performances*

Are groups better than individuals at predicting others' performance and hence serve as implicit informational channels? We answer this question by testing the equality of predicted and actual scores in quizzes across the three treatments (Table 4).[15] Groups significantly overestimate the number of answers they got right whereas individuals are quite accurate in predicting their own absolute performance in both Individual and IndividualInfo treatments. The inaccuracy of the groups to predict their own score is quite surprising in the context of our

[15] We also investigated whether the accuracy of subjects to predict scores in a quiz changes from first to the second half of the experiment. We find the consistent results to the reported results in Table 4.

study: since the questions were both open-ended and either very simple or very difficult, we would expect that all subjects would have a high degree of certainty about whether they knew the answer to the question or not. Predicting the performance of the average subject on the other hand could be more difficult as one should correctly predict the abilities of the subject pool in a given quiz. Groups perform much better at predicting performance of the average subject than individuals for whom we reject the hypothesis that mean difference between predicted and actual average score is zero (the third column of Table 2). This result is in support for the hypothesis that groups perform as an implicit informational channels about others' performances, but also that they are less accurate in predicting their own scores.

**Table 4: Testing for the Equality of Predicted Scores to Actual Scores in a Quiz**

| | **Mean Difference Own Score** | **Mean Difference Average Score** |
|---|---|---|
| **Individual** | 0.036 | 0.627*** |
| **IndividualInfo** | -0.045 | 0.372*** |
| **Group** | 0.122*** | 0.086 |

*Mean Difference Own Score = Mean (Predicted Own score – Actual Score). Significances are according to two-tailed t-test. * 10%, ** 5%, *** 1% significance levels.*

### *3.4.3 Explaining Entry with Beliefs*

In a strategic situation like a market entry game, we expect beliefs to be important determinants of behaviour. We investigate whether and if so which beliefs are significant predictors of entry across our treatment manipulations. In Table 5, we regress the binary variable of entry decision on the range of beliefs elicited through the round questionnaire. These are measured expectations of entry rates, own percentile rankings, own score, average subject's score in a quiz, own rank and the difference between own actual score: for the IndividualInfo treatment only we also have the historical average score.

In a causal sense, we would predict that higher predicted entry will discourage entry and there will be a negative relationship between predicted entry and actual entry. In the Group and IndividualInfo treatments, the predicted entry rate significantly and positively correlates with the decision to enter a market. This suggests that in these two treatments, where subjects are more accurate in predicting actual entry rates (compared to Individual treatment; Table 3), the expectation about the entry rate of others was not a factor to strategically discourage or encourage entry decision in a round.[16] A similar result was previously found by Camerer & Lovallo (1999) and Moore & Cain (2007) as well, who also show that entry decisions are affected by the beliefs about relative confidence and not by the expectations about entry decisions of others. We find that the more confident subjects were about their performance in a quiz, the more likely they were to enter a market in IndividualInfo and Group treatments (significant coefficient of Relative Confidence). We do not find a similar effect in the Individual treatment, although we find that subjects believe they obtained a lower rank when they decided to enter a market than when they decided to stay out.[17] Looking in more detail at the pattern of relative confidence and entry behaviour in Figure 2 and Figure 3 across rounds and treatments, we see that in the Group and IndividualInfo treatments the relative confidence mirrors the actual entry behaviour which is another manifestation of

[16] We also have done additional analysis separately for each difficulty level and random rounds. In all, we find either no significant effect or positive correlation between expected entry rates and entry decisions.
[17] Note that a lower rank means a better rank, such that ranking 1st is better than ranking 2nd.

the significant correlations between relative confidence and entry behaviour in these two treatments but not in the Individual treatment.

In the second column of the Table 5, we check whether experimental manipulation of giving subjects historical information significantly affects the entry decisions of individuals. We find that the greater is the difference between one's own score and average historical score in a quiz, the more likely is an individual to enter a market. For this treatment, we also observe a significant negative effect of predicted average score on entry. However, Table 4 suggests that individuals with information are just as accurate in predicting average scores in a quiz round as individuals without information. This may indicate that historical information manipulation in IndividualInfo treatment has primed individuals to take into account expectations about average performances of others when deciding to enter whereas in the other treatments this variable was not a significant predictor of entry.

To conclude this section with summarizing the results, we observe the following. Different beliefs affect entry behaviour differently across treatments. In the Individual treatment, no other belief except expected rank significantly predicts entry. In the Group and IndividualInfo treatments, subjects' expected entry rates are significantly correlated with the entry behaviour. Also relative confidence is a significant predictor of entry behaviour, and its pattern mirrors the pattern of entry rates across rounds. In the IndividualInfo treatment, we additionally observe that subjects' entry behaviour is significantly predicted by expected average score. Since there is no evident under/overconfidence in the last two blocks of the IndividualInfo treatment (Figure 3), the predictive power of expected average score on entry behaviour may be the key to understanding why in this treatment the BTA/WTA effect disappears both in entry and in relative confidence judgements in the last two blocks.

**Table 5: Regression Analysis of Entry**

| | Individual | IndividualInfo | Group |
|---|---|---|---|
| **Expected Entry Rate** | 0.025 | 0.131*** | 0.049*** |
| **Relative Confidence** | 0.007 | 0.072*** | 0.045*** |
| **Expected Own Score** | -0.034 | 0.197 | 0.025 |
| **Expected Average Score** | 0.029 | -0.092** | -0.021 |
| **Expected Own Rank** | -0.174*** | -0.001 | -0.095*** |
| **Difference_Info_Score** | | 0.097*** | |
| | | | |
| **Other variables** | | | |
| **N** | 274 | 251 | 253 |
| **Pseudo_$R^2$** | 0.21 | 0.54 | 0.66 |

*Probit regression clustered by session. Reported coefficients are average marginal effects. Quiz rounds only * 10%, ** 5%, *** 1% significance levels. Difference_Info_Score is the difference between one's own score and the information received about the average historical score in a given quiz. Other variables contain controls for age, gender, self-reported risk attitudes, feelings of confidence and competitiveness measured by post-study questionnaire.*

### *3.5 Discussion*

More than 50% of businesses fail after three years of operating and the rate of failure is significantly different across industries. The highest rate of failure occurs in fields such as Information, Retail and Transportation/Communication Services and the lowest rate of failure is in the fields of Education/Health, Agriculture and Finance/Insurance (Statistic Brain 2012). A recent research agenda by Moore and his colleagues has tried to explain why these differences appear in markets by experimental investigation of why individuals decide to enter certain competitions and avoid others. They show that individuals underestimate their relative performance in difficult tasks and overestimate it in easy tasks which in turn translate into varying entry decisions to difficult and simple competitions.

Our findings contribute to this line of literature by investigating factors that may help to de-bias entry decisions. We have shown that the bias in entry behaviour is not inherent to individual decisions but that groups are prone to the same type of bias as well. We explored whether information helps to alleviate overentry and underentry to simple and difficult markets. Our answer is “Yes, it does, but under specific conditions”. We show that groups as implicit information channels can, on average, predict competitors’ entry behaviour and performances better than individuals. However, being more informed about competitors’s performance does not guarantee the mitigation of the coordination failures in entry behaviour: groups similar to individuals over-enter to simple rounds and under-enter to difficult rounds and fail to learn. Groups’ entry decisions are explained by relative confidence in their beliefs about percentile rankings which are not de-biased through repetition and feedback over the course of the experiment. We show that only explicit information about average historical performance in a given task eliminates the difference in entry behaviour and relative confidence rankings between simple and difficult rounds. The difference between simple and difficult rounds in this treatment is initially much smaller than in the other two treatments but is still significant in the first half of the experiment. The interaction of historical performance information with repetition and feedback de-biases individuals in both their entry decisions and relative confidence rankings in the last two blocks.

An important policy recommendation following our results is that explicit statistical information about past performance is a useful tool to eliminate biased overentry and underentry to competitions. Excessive entry to industries perceived as simple and insufficient entry to industries perceived as difficult can be overcome if the information about revenues, profits and failure rates (as an indicator of firm performance) are publicly available and transmitted to potential entrepreneurs in the initial stages of their start-up ventures.

Although in our study we do not find any gender differences in entry behaviour, many studies in the literature show that women are less willing to compete than men regardless of their actual performances (e.g. Barber & Odean 2001; Niederle & Vesterlund 2010; Buser, Niederle & Oosterbeek 2012). We find that women are less confident than men in their relative performance rankings in both of our individual treatments, but this difference disappears when they are in groups. This may explain the increase in competitive preferences of women when in groups that encourages women to compete especially in situations where they are as skilled as men (Healy & Pate 2011; Dargnies 2012). As another application, for example, schools can use historical performance information of grades to encourage more entry of females to competitive science degrees held back by females' misjudged relative performance beliefs (Buser, Niederle & Oosterbeek 2012; Dargnies 2012). Further research could usefully explore what mechanisms contribute to the emergence of this difference in relative confidence between female individuals and groups to provide insight into how relative confidence beliefs are formed.

## *References*


Barber, B. M., & Odean, T. (2001). "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment". *Quarterly Journal of Economics*, Vol. 116, pp. 261-292.

Buser, T., Niederle, M., & Oosterbeek, H. (2012). "Gender, Competitiveness and Career Choices". National Bureau of Economic Research (No. w18576).

Cain, D. M., Moore, D. A. and Haran, U. (2013). "Making Sense of Overconfidence in Market Entry", *Strategic Management Journal*, forthcoming.

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

Charness, G. and Sutter, M. (2012). "Groups Make Better Self-Interested Decisions". *Journal of Economics Perspectives*, Vol. 26, pp. 157-176.

Cooper, D. J. and Kagel, J. H. (2005). "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games." *The American Economic Review*, Vol. 95, pp. 477-509.

Dargnies, M.P. (2012). "Men too Sometimes Shy Away from Competition: the Case of Team Competition". *Management Science,* Vol. 58, pp. 1982-2000.

Dunne, T., Roberts, M. J., & Samuelson, L. (1988). "Patterns of Firm Entry and Exit U.S. Manufacturing Industries". *Rand Journal of Economics,* Vol. 19, pp. 495-515.

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

Ewers, M. (2013). "Information and Competition Entry". IZA Discussion Paper. No 6411.

Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Ready-made Economic Experiments". *Experimental Economics,* Vol. 10, pp. 171-178.

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

Greiner, B. (2004): “An Online Recruitment System for Economic Experiments”. Published in: Forschung und wissenschaftlichesRechnen 2003. GWDG Bericht , Vol. 63, pp. 79-93.

Hayward, M.L.A., Shepherd, D.A. and Griffin, D. (2006). “A Hubris Theory of Entrepreneurship”. *Management Science*, Vol. 52, pp. 160-172.

Healy, A. and Pate, J. (2011). “Can Teams Help to Close the Gender Competition Gap?” *The Economic Journal*, Vol. 121, pp.1192-1204.

Kocher, M., Strauß, S., & Sutter, M. (2006). “Individual or Team Decision-Making—Causes and Consequences of Celf-selection”. *Games and Economic Behavior*, Vol. 56, pp. 259-270.

Kocher, M. G., & Sutter, M. (2005). “The Decision Maker Matters: Individual Versus Group Behaviour in Experimental Beauty-Contest Games”. *The Economic Journal*, Vol. 115, pp. 200-223.

Kruger, J., & Burrus, J. (2004). “Egocentrism and Focalism in Unrealistic Optimism (and Pessimism)”. *Journal of Experimental Social Psychology,* Vol. 40, pp. 332-340.

Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2008). “The Rational Side of Egocentrism in Social Comparisons”. *Journal of Experimental Social Psychology*, Vol. 44, pp. 220-232.

Malmendier, U. and Tate, G. (2008). “Who Makes Acquisitions? CEO Overconfidence and the Market’s Reaction”. *Journal of Financial Economics*, Vol. 89, pp. 20-42.

Mata, J., & Portugal, P. (1994). “Life Duration of New Firms”. *The Journal of Industrial Economics,* Vol. 42, pp. 227-245.

McCallum, D. M., Harring, K., Gilmore, R., Drenan, S., Chase, J. P., Insko, C. A. and Thibaut, J. (1985). “Competition and Cooperation Between Groups and Between Individuals”. *Journal of Experimental Social Psychology,* Vol. 21, pp. 301-320.

Moore, D. A. (2007), “Not So Above Average After All: When People Believe They are Worse than Average and its Implications for Theories of Bias in Social comparison”. *Organizational Behaviour and Human Decision Processes*, Vol. 102, pp. 42-58.

Moore, D. A. and 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. A., & Small, D. A. (2007). "Error and Bias in Comparative Judgment: on Being Both Better and Worse Than We Think We Are". *Journal of Personality and Social Psychology*, Vol. 92, pp. 972-989.

Moore, D. A., Oesch, J. M., & Zietsma, C. (2007). "What Competition? Myopic Self-focus in Market-entry Decisions". *Organization Science*, Vol. 18, pp. 440-454.

Niederle, M. and Vesterlund, L. (2010). "Explaining the Gender Gap in Math Test Scores: The Role of Competition", *Journal of Economic Perspectives*, Vol. 24, pp. 129 – 144.

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

Rapoport, A., Seale, D. A., Erev, I., & Sundali, J. A. (1998). "Equilibrium Play in Large Group Market Entry Games". *Management Science*, Vol. 44, pp. 119-141.

Radzevick, J. R., & Moore, D. A. (2008). "Myopic Biases in Competitions". *Organizational Behavior and Human Decision Processes*, Vol. 107, pp. 206-218.

Sheremeta, R. M., & Zhang, J. (2010). "Can Groups Solve the Problem of Over-bidding in Contests?". *Social Choice and Welfare*, Vol. 35, pp. 175-197.

Statistic Brain 2012. "Startup Business Failure Rate By Industry". Statistic Brain Research Institute, publishing as Statistic Brain. 01.01.2014, http://www.statisticbrain.com/startup-failure-by-industry/

Sundali, J. A., Rapoport, A., & Seale, D. A. (1995). "Coordination in Market Entry Games with Symmetric Players". *Organizational Behavior and Human Decision* Processes, Vol. 64, pp.203-218.

Sutter, M. (2005). "Are four heads better than two? An Experimental Beauty-contest Game with Teams of Different Size". *Economics Letters*, Vol. 88, pp. 41-46.

Sutter, M., & Strassmair, C. (2009). "Communication, Cooperation and Collusion in Team Tournaments—an Experimental Study". *Games and Economic Behavior*, Vol. 66, pp. 506-525.

Svenson, O. (1981), “Are We All Less Risky and More Skillful than Our Fellow Drivers?” *Acta Psychologica*, Vol. 47, pp. 143–148.

Weinstein, N. D. (1980), “Unrealistic Optimism about Future Life Events,” *Journal of Personality and Social Psychology*, Vol. 39, pp. 806–820.

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). “The Influence of Egocentrism and Focalism on People's Optimism in Competitions: When What Affects Us Equally Affects Me More”. *Journal of Personality and Social Psychology*, Vol. 85, pp. 389-408.

## *Appendix A: Quizzes*

**Simple A**

What is the capital city of Greece?
Who is the author of the “Harry Potter” books?
What is the name of the animated star of the computer game and movie Tomb Raider?
What was the first name of Bonaparte, a French military and political leader of late 18$^{th}$ early 19$^{th}$centuries France?
Which country, located in the northern Eurasia, is the largest country in the world?
**Tiebreaker**: What is the height of Eiffel Tower in metres?

**Simple B**

What is the capital city of Argentina?
Who wrote the play “Hamlet”?
Which actress played the title role in the 1990 film “Pretty Woman”?
What was the name of the admiral famously known for his part in the victory of Britain in the Battle of Trafalgar?
Which river, located in Northern Africa, is the longest river in the world?
**Tiebreaker**: How many member states are there in United Nations?

**Simple C**

What is the capital city of France?
Who wrote the novel “The Life and Adventures of Nicholas Nickleby”?
Which cartoon show included characters called Thelma and Shaggy?
What was the name of the famous ship that left the British port of Southampton with the aim of reaching New York City in the United States in 1912 but tragically sank on its Maiden Voyage?
Which mountain peak, located in the range Himalayas, is the highest peak in the world?
**Tiebreaker**: How many films did Alfred Hitchcock direct (including short films and documentaries)?

**Simple D**

What is the capital city of China?
Please complete the title of the Tolstoy’s novel: “War and ….
What type of animal was Stuart, in the 1999 film “Stuart Little”?
Who served as a Prime Minister of the United Kingdom from 2007 to 2010?
What mountain peak, located in Scotland, is the highest peak in Britain?
**Tiebreaker**: How many men signed the American Declaration of Independence?

**Difficult A**

What is the capital city of Bahrain?
Who wrote the novel “Master and Margarita”?
Who is the voice of adult Simba in the 1994 film “The Lion King”?
What is the name of the Battle also known as Battle of the Nations fought in 1813 by Russia Prussia, Austria and Sweden against Napoleon?
Which mountain peak, located in the range of Guiana Highlands, is the highest peak in Brazil?

**Tiebreaker**: How many days did British-American astronaut Michael Foalespend in space?

**Difficult B**
What is the capital city of Togo?
What Chilean author wrote Sub Terra?
What is the real name of the actor playing the character Cramden in the 2008 film "Toe Tactic"?
Who was the Prime Minister of United Kingdom from 1937 to 1940?
Which river, with its source in the Great Slave Lake, is the longest river in Canada?
**Tiebreaker**: How many thousands of squared kilometres is the area of Madagascar?

**Difficult C**
What is the capital city of Suriname?
What is the surname of the German brothers known for their collections of fairy tales? Brothers …..
What is the real name of the actress playing Aunt Tina Little in the 1999 film "Stuart Little"?
Who was the Prime Minister of the United Kingdom from 1905 to 1908?
What country in Africa borders only with Senegal?
**Tiebreaker**: How many thousand kilometres is the coastline of Canada?

**Difficult D**
What is the capital city of Russia?
What was the surname of the literary character Nana in Emil Zola's novel titled "Nana"?
What is the real name of the actor playing The Tin Man in the 1939 film "Wizard of Oz"?
Who was the king of England between 1327 and 1377?
What country surrounds two other countries?
**Tiebreaker**: What year was the ex-Prime minister of United Kingdom Arthur Balfour born?

## *Appendix B: Instructions for the Group Treatment*

**Welcome!**

You are about to participate in an experiment. There are 21 people participating in the experiment and participants have been divided into 7 groups of 3. This group composition will remain fixed for the whole experiment. You will be allowed to communicate with people in your group but you must not communicate with anyone outside your group. If you have a question at any time during the experiment, please raise your hand and someone will come to your desk to answer it.

The use of electronic devices such as mobile phones, music players, and tablet computers is strictly prohibited. Please make sure that all such devices are turned off and put away out of sight.

If you break these rules, you will be excluded from the experiment without receiving any payment and be disqualified from future experiments with CeDEx.

**Market Entry Experiment**

In this experiment, your group will be playing the role of an entrepreneur who has to choose whether to enter into new markets. Entering a new market can be highly profitable if your group does well. On the other hand, if competition is too tough, your group may lose money. The experiment will consist of a number of rounds. In each round, your group will make one entry decision. The decisions made by groups will affect your final payoff.

Each round, market rankings will be determined in one of two ways. In some rounds, the rankings of all entrants will be determined by scores on a trivia quiz. In other rounds, all entrants will be ranked randomly (by being assigned a random score). In each round, the groups that have entered will be ranked according to their scores and their group members will receive payoffs according to this table:

| Your Group's Rank | Point Payoff for Each Member of |
|---|---|
| 1st | 14 |
| 2nd | 10 |
| 3rd | 5 |
| 4th | -10 |
| 5th | -10 |
| 6th | -10 |
| 7th | -10 |

The top 3 groups who decided to enter the market will each earn points. Higher-ranked entrants will earn more, according to the table above. If more than 3 groups enter the market, the members of groups ranked below 3rd will lose 10 points each. Your group may always choose to stay out of the market; staying out means you neither gain nor lose anything.

For participating in this experiment, each of you will receive a £10 base payment. In addition, you can earn points in each round as described above. At the end of the experiment, three of the rounds will be randomly selected and your points from those three rounds will be averaged and then converted into cash at a rate of £1 per point. For each of you, this amount will be added to (or, if you lost points, subtracted from) your £10 base payment.

## *Appendix C: Sample Feedback Screen as seen after Round 8*

| | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 | Round 6 | Round 7 | Round 8 |
|---|---|---|---|---|---|---|---|---|
| **Entrepreneur 1** | | | | | | | | |
| Score | 56 | 7 | 7 | 55 | 2 | 27 | 56 | 6 |
| Entry | Enter | Enter | Stay Out | Enter | Enter | Enter | Enter | Enter |
| Rank | 1 | 1 | -- | 3 | 3 | 1 | 2 | 1 |
| **Entrepreneur 2** | | | | | | | | |
| Score | 41 | 4 | 3 | 56 | 7 | 15 | 57 | 3 |
| Entry | Enter | Stay Out | Enter | Enter | Stay Out | Stay Out | Enter | Enter |
| Rank | 5 | -- | 1 | 2 | -- | -- | 1 | 3 |
| **Entrepreneur 3** | | | | | | | | |
| Score | 52 | 1 | 1 | 57 | 6 | 6 | 55 | 4 |
| Entry | Stay Out | Enter | Enter | Enter | Enter | Enter | Enter | Stay Out |
| Rank | -- | 4 | 2 | 1 | 1 | 3 | 3 | -- |
| **Entrepreneur 4** | | | | | | | | |
| Score | 54 | 2 | 15 | 53 | 5 | 2 | 54 | 1 |
| Entry | Enter | Stay Out | Stay Out | Stay Out | Stay Out | Stay Out | Enter | Stay Out |
| Rank | 3 | -- | -- | -- | -- | -- | 4 | -- |
| **Entrepreneur 5** | | | | | | | | |
| Score | 57 | 6 | 2 | 41 | 4 | 13 | 51 | 5 |
| Entry | Stay Out | Enter | Stay Out | Enter | Stay Out | Stay Out | Stay Out | Enter |
| Rank | -- | 2 | -- | 5 | -- | -- | -- | 2 |
| **Entrepreneur 6** | | | | | | | | |
| Score | 53 | 3 | 6 | 52 | 1 | 11 | 43 | 7 |
| Entry | Enter | Enter | Stay Out | Enter | Stay Out | Enter | Enter | Stay Out |
| Rank | 4 | 3 | -- | 4 | -- | 2 | 5 | -- |
| **Entrepreneur 7** | | | | | | | | |
| Score | 55 | 5 | 14 | 44 | 3 | 4 | 42 | 2 |
| Entry | Enter | Stay Out | Stay Out | Stay Out | Enter | Stay Out | Enter | Stay Out |
| Rank | 2 | -- | -- | -- | 2 | -- | 6 | -- |

## *Appendix D: Manipulation of Information Treatment*

| | | Answers |
|---|---|---|
| 1. | What is the capital city of France? | |
| 2. | Who wrote the novel "The Life and Adventures of Nicholas Nickleby"? | |
| 3. | Which cartoon show included characters called Thelma and Shaggy? | |
| 4. | What was the name of the famous ship that left the British port of Southampton with the aim of reaching New York City in the United States in 1912 but tragically sank on its Maiden Voyage? | |
| 5. | Which mountain peak, located in the range Himalayas, is the highest peak in the world? | |
| **Tiebreaker**: What is the height of Eiffel Tower in metres? | | |

For questions 1-5, for each correct answer you will get 10 marks.

For the tie-breaker question, the closest answer to the correct answer will get 7 marks, the second closest will get 6 marks, the third closest 5 marks and so on.

Your score in this round is the total of these marks in this round.

The average number of correct answers for this quiz in previous sessions was 4.2 out of 5 questions.

Do you choose to enter the market and compete or do you choose to stay out?

☐ Enter the market ☐ Stay out

## *Appendix E: Other Results*

**Table E1: Testing for the equality of expected to actual entry rates**

| | Rounds | Average Actual Entry | Average Expected Entry | *p-value* |
|---|---|---|---|---|
| **Individual** | **Round 1 (Simple)** | 5.333 | 3.809 | *.000* |
| | **Round 2 (Random)** | 3.333 | 4.285 | *.000* |
| | **Round 3 (Difficult)** | 1.333 | 3.5 | *.000* |
| | **Round 4 (Simple)** | 5.833 | 4.428 | *.000* |
| | **Round 5 (Random)** | 4.333 | 4.261 | *.767* |
| | **Round 6 (Difficult)** | 1.833 | 3.666 | *.000* |
| | **Round 7 (Simple)** | 5 | 4.476 | *.023* |
| | **Round 8 (Random)** | 4.333 | 4.428 | *.599* |
| | **Round 9 (Difficult)** | 4 | 3.976 | *.929* |
| | **Round 10 (Simple)** | 5.116 | 4.285 | *.001* |
| | **Round 11 (Random)** | 3.833 | 4.476 | *.012* |
| | **Round 12 (Difficult)** | 2.833 | 3.928 | *.000* |
| **Group** | **Round 1 (Simple)** | 5.833 | 5.642 | *.438* |
| | **Round 2 (Random)** | 4.116 | 4.285 | *.522* |
| | **Round 3 (Difficult)** | .666 | 2.785 | *.006* |
| | **Round 4 (Simple)** | 6.166 | 5.595 | *.008* |
| | **Round 5 (Random)** | 4.166 | 4.333 | *.493* |
| | **Round 6 (Difficult)** | 2.166 | 3.024 | *.002* |
| | **Round 7 (Simple)** | 4.833 | 5.21 | *.114* |
| | **Round 8 (Random)** | 2.5 | 4.19 | *.000* |
| | **Round 9 (Difficult)** | 4.166 | 3.286 | *.003* |
| | **Round 10 (Simple)** | 5 | 4.95 | *.843* |
| | **Round 11 (Random)** | 3.5 | 4.047 | *.014* |
| | **Round 12 (Difficult)** | 2.833 | 3.357 | *.031* |
| **Individual Info** | **Round 1 (Simple)** | 4.666 | 4.833 | *.492* |
| | **Round 2 (Random)** | 3.5 | 3.904 | *.133* |
| | **Round 3 (Difficult)** | 2 | 2.57 | *.003* |
| | **Round 4 (Simple)** | 5 | 4.761 | *.229* |
| | **Round 5 (Random)** | 3.833 | 3.857 | *.921* |
| | **Round 6 (Difficult)** | 2.5 | 2.76 | *.433* |
| | **Round 7 (Simple)** | 4.333 | 5.047 | *.000* |
| | **Round 8 (Random)** | 3.4 | 4.457 | *.000* |
| | **Round 9 (Difficult )** | 3.833 | 3.071 | *.000* |
| | **Round 10 (Simple)** | 3.66 | 4.619 | *.000* |
| | **Round 11 (Random)** | 3.8333 | 4.119 | *.116* |
| | **Round 12 (Difficult)** | 4 | 3.191 | *.000* |

*For each round there are 42 observations. p-values are from two tailed t-test*

# CHAPTER 4: SNOWBALLING CONFIDENCE IN HIERARCHICAL TOURNAMENTS

## *4.1 Introduction*

Confidence about own ability relative to others has important effects on many economic decisions such as job search, entrepreneurial activity and effort choices in competitive settings. In this paper, we study confidence in the context of a three-stage hierarchical tournament, in which top (bottom) scorers of a group are matched with the top (bottom) scorers of another group as the tournament progresses from one stage to the next. The construct of a hierarchical tournament allows us to study whether subjects adjust their confidence levels according to the reference group they are competing against. Learning that one is a top scorer of his group is a positive signal of his relative ability with respect to the bottom scorers of his group. However, in the new stage top scorers matched together have all received the same positive signal. We explore the conjecture that people may underweight this observation, and as a consequence their confidence may increase from one stage of the tournament to the next, snowballing their confidence. The opposite may be true for bottom scorers who move down the tournament ladder and are matched with the other bottom scorers. To our knowledge, ours is the first study on confidence snowballing in a tournament setting.

There are many real life settings where confidence snowballing may be evident. Educational attainment is a good example where confidence judgements are decisive and may play an important role on how much effort students spend: a first year student at an elite university will be overconfident if he ignores the change in the difficulty of the new competition from his high school to university, and this may even strengthen further if he proceeds to highly selective graduate schools. Another example is managerial overconfidence: the promotion tracks and corporate governance policies in themselves may give origins to managerial overconfidence and explain the widely cited phenomenon in the literature of *CEO overconfidence* (e.g. Paredes 2005; Malmendier & Tate 2005; Malmendier & Tate 2008; Brown & Sarma 2007; Galasso & Simcoe 2011). The opposite can be true for the losers of tournaments that enter consolation tournaments against other losers. Recently fired workers may be underconfident competing for newly posted positions if they neglect how easy the new competition is, and hence may exert suboptimal effort to job search (Dubra 2004; Koszegi 2006).

The previous literature, which we discuss in Section 2, has studied confidence mostly within the context of single stage tournaments and contests (Niederle & Vesterlund 2007; Moore & Cain 2007; Ludwig, Wickhardt & Wickhorst 2011) and has not looked how confidence evolves through stages. Only a small strand of the literature investigates the effect of feedback on confidence (Eberlein, Ludwig and Nafziger 2010; Mobius, Niederle, Niehaus & Rosenblat 2011; Grossman & Owens 2012; Eil & Rao 2011), but even here the environment is very different from ours since the reference group one compares himself to does not change from one stage to the next.

To study confidence snowballing we develop a three stage hierarchical tournament where in each stage subjects perform a skill task. The skill task involved seeing a pair of black circles with white dots for 1 second and judging which circle had more dots in it. This task has been previously used to study absolute confidence calibration individuals by Hollard, Massoni & Vergnaud

(2010) and Massoni & Roux (2012). By manipulating the difference in the number of dots, we run easy and difficult tasks as a between subject treatment. On each stage subjects are grouped with others who have received similar feedback about their relative performances in previous stages, and before performing the task, we elicit their relative confidence levels. We elicit their confidence levels with an incentive compatible device that has a clear behavioural interpretation of confidence – a subjective probability a person assigns to being in the top half of his group. Section 3 describes our experimental design and procedures in more detail.

In Section 4 we present our findings. In the first stage of the tournament, we observe neither overconfidence nor underconfidence: on average subjects assess their likelihood of being in the top half as fifty percent. As the tournament progresses to later stages however, we observe asymmetric change in confidence between top and bottom scorers. By the last stage of the tournament, top scorers matched in a group with the other top scorers significantly increase their confidence both in easy and difficult tasks. Bottom scorers matched in a group with the other bottom scorers significantly decrease their confidence in the difficult but not in the easy task. We do not find any gender differences in terms of either performance or confidence, but we do find interesting gender differences in how top and bottom scorers attribute their success to skill or luck. Men top scorers perceive the task as significantly more a skill than a luck task compared to men bottom scorers, whereas we cannot tell the same about the women top and bottom scorers. Section 5 discusses the implications of our findings, possible further research directions and concludes.

## *4.2 Literature Review*

In this paper, we study relative confidence and more specifically the subjective probability a person assigns to being in the top half of his group. The previous experimental economics literature has studied relative confidence within the settings of tournaments and contests where beliefs about one's chances of success can determine whether to enter a competition, how much effort to exert when competing, what strategies to use while competing and if and when to leave a competition. It has been shown that people tend to be overconfident relative to others and bet on themselves more in easy and familiar tasks and are underconfident relative to others in difficult and non-familiar tasks (Hoelzl & Rustichini 2005; Moore 2007; Moore and Cain 2007; chapter 3); females are less confident than males and enter competitions less often (Niederle & Vesterlund 2007; Dargnies 2012; Kamas & Preston 2012); being overconfident increases exerted effort in contests and hence the chance of winning (Ludwig, Wichardt & Wickhorst 2011) and people tend to neglect competition when making tournament entry decisions (Camerer & Lovallo 1999; Radzevick & Moore 2008). In our study, by manipulating the difficulty level of a task as a between subject treatment variable, we test for the presence of over/underconfidence with a novel incentivized elicitation device and more importantly how confidence changes from one stage of a hierarchical tournament to another.

Two leading explanations have been given in the literature for why BTA/WTA effect occurs. One is asserting that, in competitive situations, people neglect the competition and focus on themselves. Camerer & Lovallo (1999) propose a *reference group neglect* hypothesis - "the tendency to underadjust to changes in the reference group one competes against" to explain overentry to tournaments. The main support for their hypothesis was the observation that trivia experts recruited to compete against each other were more overconfident than the general student sample- indicating that subjects took into account their own knowledge and competence and ignored that of the competition. Moore and his colleagues in a series of papers introduce an alternative but related - *differential information* hypothesis to explain their findings of overentry to easy and underentry to difficult competitive tasks (Moore & Cain 2007; Radzevick & Moore 2008; Moore & Small 2007). They argue that because there is greater accessibility and quality of information about the self than about others people make more regressive estimates about others which leads them to further underweight those regressive estimates.[18] Differential information hypothesis aims to explain reference group neglect hypothesis by asserting that people tend to neglect competition because they don't have enough information to base their decision on. With our experimental design we will be able to partly disentangle these two hypotheses because the differential information hypothesis predicts no change in confidence since in the new stage there is symmetric information about self and reference group and reference group neglect hypothesis predicts that subjects will disregard the information about others and hence confidence will snowball.

Our study is also related to the literature investigating how confidence is affected by feedback. When competing against the same opponent repeatedly, subjects learn to calibrate their confidence levels with the help of feedback but

[18] Both hypotheses are based on a broader finding that self is evaluated more egocentrically compared to others. This could be due to a number of reasons such as self being a natural focal point (Kruger & Burrus 2004), superior memory for the self-related events (Symons & Johnson 1997), motivational reasons (Benabou & Tirole 2002; Koszegi 2006) and insufficient anchoring and adjustment to shared circumstances between self and others (Windschitl, Kruger & Simms 2003).

only given explicit information about others' performances (Rose & Windschitl 2008; Moore & Cain 2007; Chapter 3 of this dissertation). When feedback is noisy, people overweight positive feedback, update both their absolute and relative confidences insufficiently for negative feedback (Mobius, Niederle, Niehaus & Rosenblat 2011; Grossman & Owens 2012) and exhibit dislike or indifference to new information when expecting negative feedback (Eil & Rao 2011, Burks, Carpenter, Goette & Rustichini 2010). There is also heterogeneity among subjects on how they react to feedback, some refusing to update their relative confidence and some overreacting to the feedback (Eberlein, Ludwig & Nafziger 2010). In our study, subjects are given feedback both about their own and their group members' relative performance at each stage of the tournament to investigate how people update their relative confidence from one stage to the next. Positive and negative feedback may differently affect confidence because of the self-serving attribution bias which has been extensively studied in the psychology literature. It has been shown that people tend to attribute positive feedback to internal factors such as their skill and abilities and negative feedback to external factors such as luck and destiny (Arkin, Appelman & Burger 1980; Pyszczynski, Greenberg, Solomon, Arndt & Schimel 2004) and tend to rate a task they succeeded at as more important than a task they failed at (von Hippel, Lakin & Shakarchi 2005). In our study, we ask subjects about their perceptions of the task in a post study questionnaire in order to ascertain whether self-serving bias might explain observed changes of confidence.

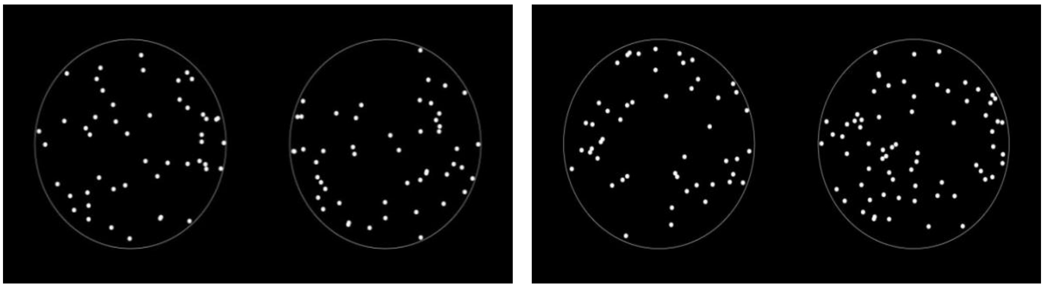
Finally, our study is related to previous studies of behavioural anomalies such as the "hot hand" fallacy and gambler's fallacy. These have been studied in real world sports tournaments investigating betting behaviour and performance in basketball, golf and bowling showing that players believe that after winning they have more chance of winning because they are having a "hot hand" or that people tend to bet on the outcomes that have not occurred before in a series of random choices (Camerer 1989; Livingston 2012; Abrevaya 2002). In our study we will check whether winners will increase their confidence after winning even if they know that the task is of a random nature. One of the studies to explicitly study confidence in a real world tournament setting is by Park & Santos-Pinto (2010) where the authors elicit confidence of participants in the form of forecasts of ranks in chess and poker tournaments differentiating tournaments as more skill and more luck based, respectively. Our study complements the existing field studies by controlling for self-selection into different tournament types, and also by systematically measuring and tracking confidence at each stage of the tournament and eliminating any other confounds of effort.

Based on these previous literatures, we can posit a number of hypotheses. We hypothesize that subjects will be overconfident in easy and underconfident in difficult tasks in the first stage of the tournament, consistent with the existing empirical evidence of BTA/WTA effect. We hypothesize that as top and bottom scorers get feedback about their relative performances and are matched with the other top and bottom scorers, confidence of top scorers will increase and that of bottom scorers will decrease. This hypothesis of snowballing is based on the evidence of egocentric valuations of the self found in other settings. Finally we hypothesize that top and bottom half scorers will update their confidence asymmetrically due to self-serving attribution bias.

### *4.3 Experimental Design*

The experiment consisted of three stages. At each stage, 16 participants were divided into four groups. Before the 1st stage began, subjects were read instructions and informed that they will be completing a set of Circle tasks where their performance will affect their earnings. The Circle task (adapted from Hollard, Massoni & Vergnaud 2010; Massoni & Roux 2012) involved seeing a pair of black circles with white dots in them for 1 second and judging which circle had more dots. One of the circles had 50 dots and the other had either 51 (difficult task) or 70 (easy task) dots (see Figure 1). The easy and difficult tasks were a between subject treatment manipulation to check for the BTA/WTA effect and how differently confidence snowballs when performance in a task is low versus high in absolute terms. To control for potential cross contamination between confidence, effort and performance, we aimed to have a skill rather than an effort task. Hence, we pre-tested the task to calibrate for difficulty levels and evaluate the role of effort. The results of the pre-test showed that (i) performance varied across individuals, (ii) did not show any evidence of learning across stages and (iii) varying incentive levels (£0.05 versus £0.50 per correct answer) did not affect performance.[19] The latter finding in particular convinced us that effort plays a minimal role in performance.

[19] We discuss the task and further results of the pre-test in more detail in the Appendix A.



50/51 Dots: Difficult task 50/70 Dots: Easy task

**Figure 1: The Circle Task**

After receiving the instructions describing the Circle task, subjects had a chance to practice the task on their computer terminals. One set of the Circle task involved 20 pairs of circles with a known dot difference (either 50/51 or 50/70, varied across sessions as a between subject treatment variable). The purpose of the practice set was to introduce the task to the subjects. We did not give any feedback to subjects about the performance in the practice set. After the practice set, the rest of the instructions were read to subjects and Stage 1 began.[20]

Each stage consisted of explaining to subjects what group they were in, filling out the confidence elicitation table, completing the Circle task and receiving relative performance feedback. In Stage 1, four groups of four subjects were randomly formed. Subjects were told which group they were in (Group A, B, C or D) and that they would all be completing the same Circle task similar to the practice set. Before proceeding to the Circle task, they were asked to complete a table as in Figure 2 which served as our confidence elicitation tool. The tool infers confidence from observed choices subjects make and has a clear behavioural interpretation of what subjective probability a person assigns to

[20] The instructions can be found in the Appendix B.

being in the top half of his group. For every row of the table subjects had to make a choice between Option A and Option B. Option A paid off £10 if a person was one of the top two scorers of his group in that stage and £3 if he was one of the bottom two scorers of his group in that stage. Option B was a lottery which paid either £10 or £3 with the chance of £10 progressively decreasing down the rows. The row where a subject switched from choosing Option B to Option A is our measure of confidence that he will be in the top two of his group.[21] This is incentive compatible under both EU and prospect theoretic models of choice under risk and so is independent of individual risk attitudes.[22] A person who thinks that he has an equal chance of being in the top or bottom two of his group will switch in the middle of the table and we say he has 50% confidence. Any deviation from switching in the middle of the table indicates whether a person assigns higher or lower probability to being in the top half of his group. [23]

After subjects filled in the confidence elicitation table at the beginning of a stage, they completed the Circle task that determined their performance in that stage. Each correct answer earned them £0.50 so that they could earn up to £10. When all subjects had completed the Circle task, they received feedback about whether they were in the top or bottom of their groups. They were then regrouped according to their performances for the next stage.[24] The grouping at

[21] We eliminated the possibility of double switching by not letting subjects proceed to the next screen and asking them to modify their choices if they switched more than once.

[22] A similar procedure has been previously used by Urbig, Stauf & Weitzel (2009) to elicit absolute and relative confidence levels.

[23] It can be argued that ambiguity attitudes can affect the choice in the elicitation procedure. However this is tangential to our analysis assuming that ambiguity attitudes are inherently part of any confidence measure and are constant within an individual and we are specifically interested in the change of confidence within individual.

[24] The feedback read as "You were in the TOP/BOTTOM two of your group" and the next screen informed them of their new groups and reminded them of how the regrouping was done.

each stage is displayed in Figure 3 and was common knowledge for all subjects before the experiment began. They were reminded about the grouping at the beginning of each stage to make sure they understand who they were in a group with.

| Row | Option A: Bet | Your Choice | Option B: Lottery |
|---|---|---|---|
| 1 | You get £10 if you are one of the top two scorers of you group and £3 if you are one of the bottom two scorers of your group | A o o B | £10 with 100% chance |
| 2 | | A o o B | £10 with 95% chance and £3 with 5% chance |
| 3 | | A o o B | £10 with 90% chance and £3 with 10% chance |
| 4 | | A o o B | £10 with 85% chance and £3 with 15% chance |
| 5 | | A o o B | £10 with 80% chance and £3 with 20% chance |
| 6 | | A o o B | £10 with 75% chance and £3 with 25% chance |
| 7 | | A o o B | £10 with 70% chance and £3 with 30% chance |
| 8 | | A o o B | £10 with 65% chance and £3 with 35% chance |
| 9 | | A o o B | £10 with 60% chance and £3 with 40% chance |
| 10 | | A o o B | £10 with 55% chance and £3 with 45% chance |
| 11 | | A o o B | £10 with 50% chance and £3 with 50% chance |
| 12 | | A o o B | £10 with 45% chance and £3 with 55% chance |
| 13 | | A o o B | £10 with 40% chance and £3 with 60% chance |
| 14 | | A o o B | £10 with 35% chance and £3 with 65% chance |
| 15 | | A o o B | £10 with 30% chance and £3 with 70% chance |
| 16 | | A o o B | £10 with 25% chance and £3 with 75% chance |
| 17 | | A o o B | £10 with 20% chance and £3 with 80% chance |
| 18 | | A o o B | £10 with 15% chance and £3 with 85% chance |
| 19 | | A o o B | £10 with 10% chance and £3 with 90% chance |
| 20 | | A o o B | £10 with 5% chance and £3 with 95% chance |

**Figure 2: Confidence Elicitation Table**

At the end of the experiment subjects filled out a post study questionnaire about demographics, personality measures and beliefs, after which the experimenter approached each subject with randomization devices to determine according to which stage/part/row a subject would be paid for.[25] The random incentive procedure either paid for the performance in the Circle task or confidence elicitation table to control for the hedging opportunities between

[25] We used physical randomization devices such as a 6-sided and a 20-sided dice and numbered balls to make the independence of the randomization procedure to subjects as clear as possible.

these two parts.[26] Subjects were undergraduate and postgraduate students recruited via Orsee (Greiner 2004) from an online database of CeDEx lab at the University of Nottingham. There were 192 subjects divided into two treatments with 6 sessions per treatment. The subjects were 51% female from various disciplines with 22% from Economics and Business majors. The experiment was fully computerized using the software Ztree (Fischbacher 2007) and lasted around 30 minutes. An average payment to a subject was £7.60 ranging from £3 to £10.

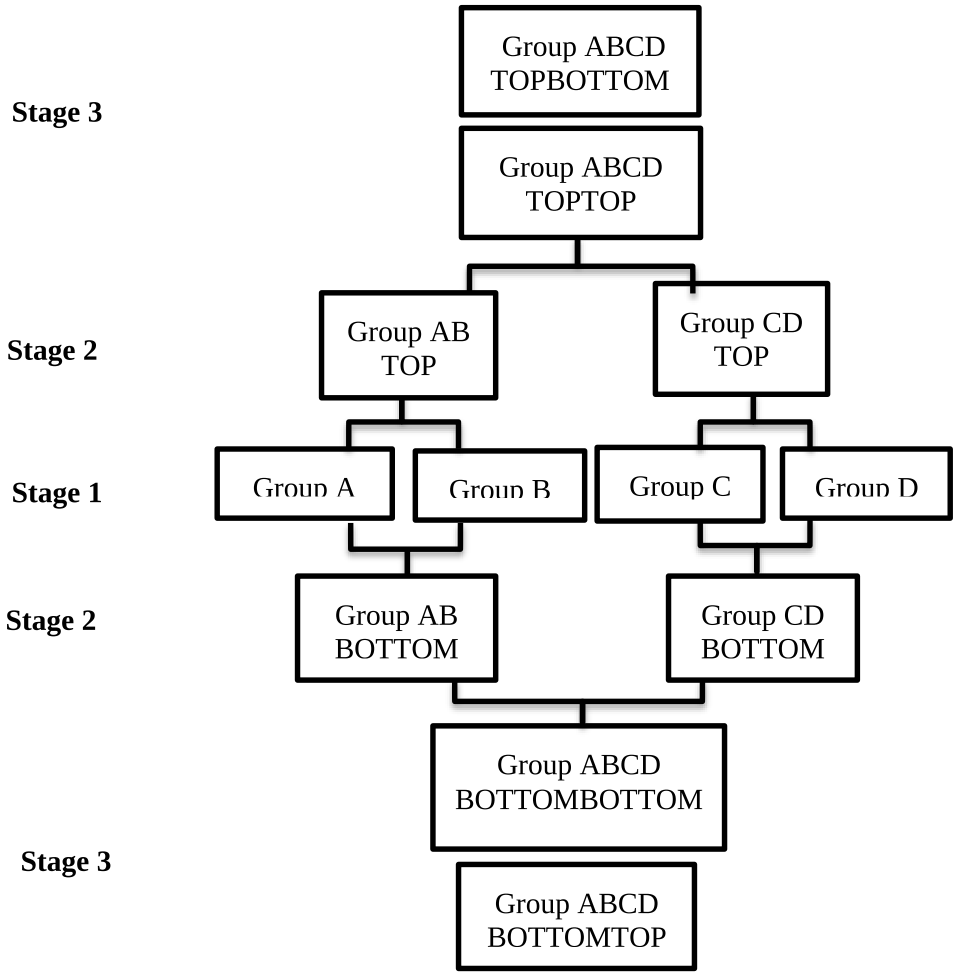



**Figure 3: The Stages and Groupings in the Experimental Session**

[26] The control question in our post study questionnaire also confirms the non-existence of intentional underperformance or hedging between performance and belief elicitation.

## *4.4 Results*

We first present our findings about how confident our subjects were in the first stage of the tournament across difficulty levels and hence checking for overconfidence in easy and underconfidence in difficult tasks (difficulty effect). In Section 4.4.2., we present evidence for if confidence snowballs from one stage to another and whether there are differences between difficulty levels. Finally in Section 4.4.3, we will check whether top and bottom scorers have differing perceptions of the task, more specifically whether they perceive the task more skill or a luck task.

### *4.4.1. Stage 1 Confidence Levels and the Difficulty Effect*

Previous relative confidence elicitation mechanisms in the literature usually ask subjects for their relative percentile rankings, point estimates of performance for self and others, or infer confidence from observed entry behaviour into competitions. The usual finding is that people are overconfident in easy and underconfident in difficult tasks (Moore 2007; Chapter 3 of this dissertation). We elicit relative confidence with a choice list and interpret the switch point as a subjective probability subject assigns to being in the top half of his group.

The histograms of Stage 1 confidence presented in Figure 4 shows the following. (i) Although the modal confidence is 50% in both treatments, the distribution in the Easy treatment appears slightly skewed to the right. Comparing distributions across two treatments, we observe slightly higher confidence for the Easy treatment compared to the Difficult treatment (Wilcoxon rank-sum p-value = 0.061). (ii) Examining averages, we find mild overconfidence in the Easy treatment: the average subjective probability subjects assign to being in the top half is 53.1% (Wilcoxon rank-sum p-value = 0.098). In the difficult treatment, we do not find any evidence of underconfidence: the average subjective probability that subjects assign to being in the top half is 48.9%, which is not significantly different from 50% (Wilcoxon sign rank p-value=0.741).

This result is interesting as it contributes to the debate in the overconfidence literature of how elicitation devices, incentives and the nature of tasks can affect the degree of observed over/underconfidence. Eliciting relative confidence with our novel elicitation device, we observe a bias only in average level data in the easy treatment. We conjecture that this is due to features of elicitation device or the type of the task used which we further discuss in the discussion section of the paper.

To further test for the difficulty effect, we compare average confidence levels between Easy and Difficult treatments for each type of subject depending on whether they were top or bottom in one stage or another (Table 1). Overall, we find lower levels of confidence for the difficult than for the easy task and, in 5 out of 7 cases the difference is significant at a 10% level.

**Table 1: Testing for Difficulty Effect for Each Group Type**

| **Average Confidence** | **n** | **Difficult** | **Easy** | ***p-value*** |
|---|---|---|---|---|
| **Stage 1** | 96 | 49.0 | 53.1 | *0.061* |
| **Stage 2 – Top** | 48 | 54.1 | 60.5 | *0.015* |
| **Stage 2 – Bottom** | 48 | 44.5 | 43.2 | *0.095* |
| **Stage 3 – TopBottom** | 24 | 54.6 | 45 | *0.135* |
| **Stage 3 – BottomTop** | 24 | 55 | 56.3 | *0.519* |
| **Stage 3 – TopTop** | 24 | 57.1 | 64.6 | *0.063* |
| **Stage 3 -- BottomBottom** | 24 | 32.5 | 45.8 | *0.058* |

*p-values are from Wilcoxon rank sum test*

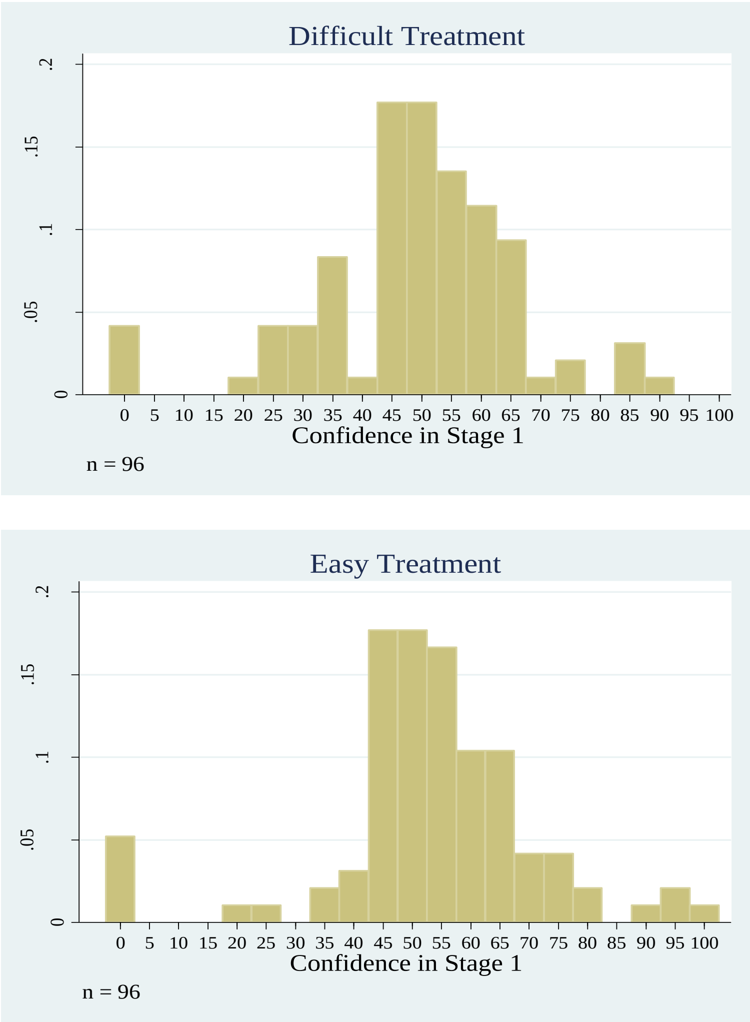



**Figure 4: The Distribution of the Elicited Confidence in Stage 1 of the Tournament**

#### *4.4.2. Snowballing of Confidence*

Our main purpose is to investigate whether and, if so how, confidence snowballs as one moves up or down a tournament ladder. Table 2 reports how subjects' confidence changes from one stage to the next after a positive feedback (being in a *top half***)** or after a negative feedback (being in a *bottom half* of his group). The reported numbers are average elicited confidence levels at each stage for the subjects who were in the relevant group: for example, $Top_1$ are the subjects who were in the top half of their group in Stage 1 and $TopBottom_{1-2}$ are the subjects who were in a Top group in Stage 1 and in a Bottom group in Stage 2. We observe a significant effect of positive feedback from first stage to the second stage: average confidence of top scorers increased significantly from 50.3% to 54.1% in the difficult treatment and from 52.6% to 60.5% in the easy treatment. The change of confidence of bottom scorers from Stage 1 to Stage 2 is significantly negative: average confidence decreases from 47.6% to 44.5% in the difficult treatment and from 53.6% to 43.2% in the Easy treatment. Looking at the change of confidence through all three stages we note the following: subjects react to the feedback received in the previous stage significantly decreasing confidence after receiving negative feedback and increasing confidence after receiving positive feedback (except the TopBottom subjects in the difficult and the BottomBottom subjects in the easy treatment). Hence, we conclude that it is the last feedback that subjects receive which matters in the formation of the next stage confidence levels. Snowballing of confidence is particularly evident when a subject receives only positive or negative feedback in all past stages.

**Table 2: Confidence Levels Across Stages**

| **Difficult** | ***n*** | **Stage 1** | **Stage 2** | **Stage 3** | ***Stage1-2 p-value*** | ***Stage2-3 p-value*** |
|---|---|---|---|---|---|---|
| **Top$_1$** | *48* | 50.3% | 54.1% | | *0.005* | |
| **Bottom$_1$** | *48* | 47.6% | 44.5% | | *0.009* | |
| **TopBottom$_{1-2}$** | *24* | | 54.6% | 54.6% | | *0.747* |
| **BottomTop$_{1-2}$** | *24* | | 45% | 55% | | *0.003* |
| **TopTop$_{1-2}$** | *24* | | 53.5% | 57.1% | | *0.017* |
| **BottomBottom$_{1-2}$** | *24* | | 44.0% | 32.5% | | *0.003* |
| **Easy** | ***n*** | **Stage 1** | **Stage 2** | **Stage 3** | ***Stage1-2 p-value*** | ***Stage2-3 p-value*** |
| **Top$_1$** | *48* | 52.6% | 60.5% | | *0.001* | |
| **Bottom$_1$** | *48* | 53.6% | 43.2% | | *0.000* | |
| **TopBottom$_{1-2}$** | *24* | | 61.5% | 45% | | *0.002* |
| **BottomTop$_{1-2}$** | *24* | | 39.4% | 56.3% | | *0.024* |
| **TopTop$_{1-2}$** | *24* | | 59.6% | 64.6% | | *0.005* |
| **BottomBottom$_{1-2}$** | *24* | | 47.1% | 45.8% | | *0.371* |

*The p-values are from Wilcoxon matched-pairs signed-ranks test.*

Figure 5 shows the plot of confidence across stages of *absolute top* and absolute *bottom* scorers i.e. those subjects who ended up in the groups TopTop and BottomBottom in Stage 3. We test for a significant trend from one stage to the other with a non-parametric test developed by Cuzick (1985). We find that the trend in panel (a) of Figure 5 is significantly positive for the top scorers (p=0.015) and significantly negative for the bottom scorers (p=0.003). In panel (b) of Figure 5, we find a significant positive trend in the confidence of top scorers (p=0.007) but no significant trend in the confidence of bottom scorers (p=0.156). This result is consistent with the results reported in Table 2 of upward confidence snowballing for the top scorers and downward confidence snowballing for the bottom scorers only in the difficult task.

To test the robustness of our results, we test the equality of average confidence of absolute top and bottom scorers for each stage of the tournament (Table 3). We find that at Stage 1, top and bottom scorers do not have significantly different confidence levels either in the easy or in the difficult treatments. The confidence between top and bottom scorers diverges starting from Stage 2 such that top scorers always have significantly higher confidence than bottom scorers.

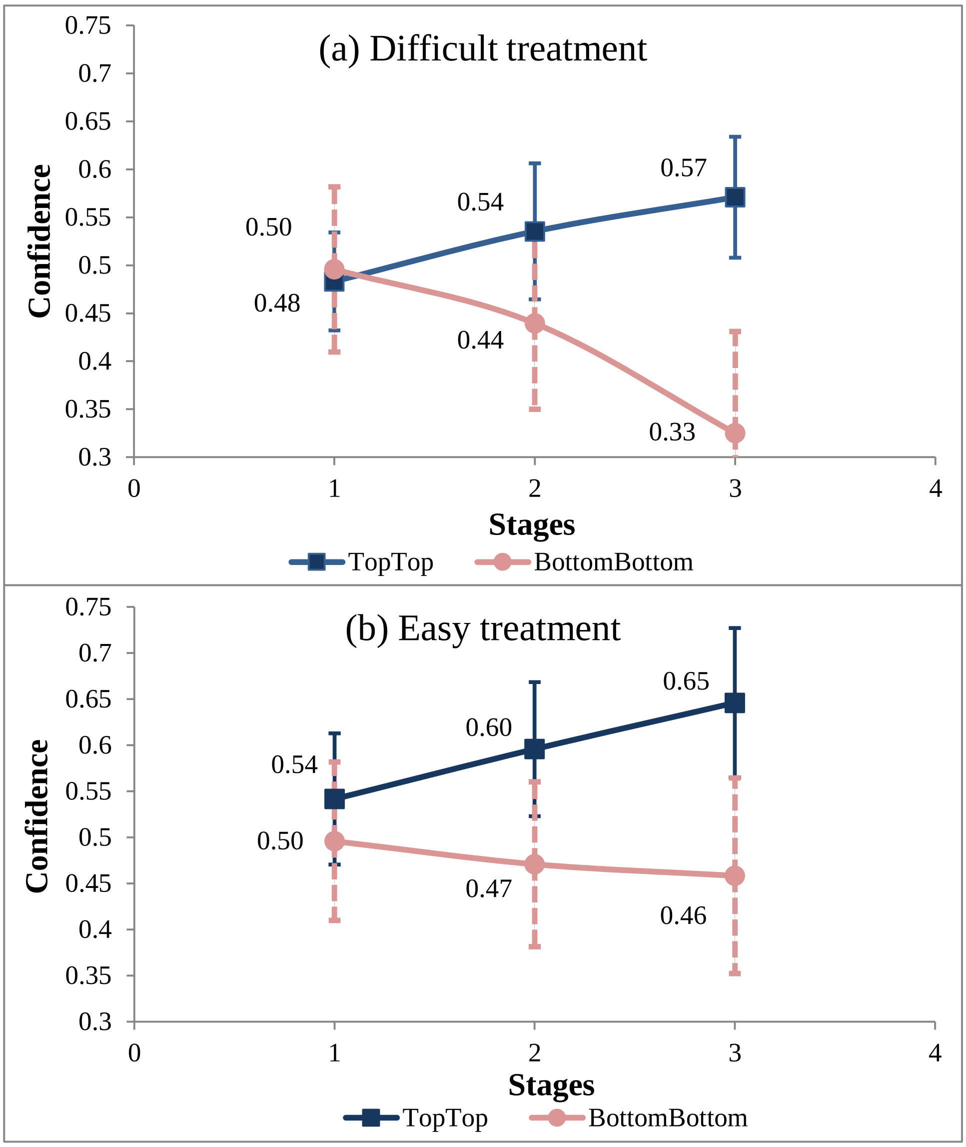



**Figure 5: Snowballing of Confidence of Absolute Top and Bottom Scorers** (n = 24 per each data point)

**Table 3: Confidence Across Stages of Top and Bottom Scorers**

| | **Difficult** | | | **Easy** | | |
|---|---|---|---|---|---|---|
| | TopTop | Bottom Bottom | *p-value* | TopTop | Bottom Bottom | *p-value* |
| **Stage 1 Confidence** | 48.3% (11.9) | 49.6% (18.7) | *0.715* | 54.2% (54.2) | 49.6% (20.4) | *0.786* |
| **Stage 2 Confidence** | 53.5% (13.7) | 44% (17.8) | *0.019* | 59.6% (17.2) | 47.1% (21.2) | *0.005* |
| **Stage 3 Confidence** | 57.1% (14.8) | 32.5% (23.7) | *0.000* | 64.6% (19.3) | 45.8% (25.1) | *0.001* |

*Standard deviations are in parenthesis, p-values are from Wilcoxon ranksum test, n=24*

#### *4.4.3. Self-Serving Attribution Biases and Gender*

Do absolute top scorers perceive the task differently from absolute bottom scorers? In a post-study questionnaire, we asked our subjects "On a scale of 1 to 7, did luck or skill determine your performance in the Circle task? (1 completely luck, 7 completely skill)". In Table 4, we test the equality of reported perceptions about the task between top and bottom scorers to test for self-serving biases in the attribution of causality of successes and failures. Top scorers reported that it was skill rather than luck that determined their results in a task (score of 3.25 in Difficult and 5.79 in Easy treatment) compared to the bottom scorers' reports (score of 1.875 in Difficult and 4.04 in Easy). We hence find a significant effect supporting the hypothesis that subjects attribute success to own dispositions (such as skill) and failures to external forces (such as luck).

One might argue that the observed effect was due to bottom scorers getting discouraged by Stage 3 and hence decreasing their effort by performing randomly (not trying) in the Circle task. However, we do not observe this; performance does not decrease across stages for bottom scorers and even slightly increases in the difficult treatment (from 9.4 to 10.6 in difficult, Wilcoxon sign rank $p=0.345$ and from 18.5 to 18.6 in easy, $p=0.784$). Moreover, since subjects were incentivized by a piece rate in the circle task, it is highly unlikely that they would be willing to sacrifice their performance by not trying hard enough in the Circle task.

A further investigation of the result that top scorers perceive the task as more skill than a luck task compared to bottom scorers revealed that this difference is strongly significant for males only ($p = 0.009$ in the Difficult and $p = 0.000$ in the Easy treatment). For females the result is less clear: there is no difference between top and bottom scorers at 5% significance level ($p = 0.221$

in the Difficult and p=0.072 in the Easy treatment). This result is consistent with the existing literature, which has previously found that women tend to attribute successes to external and failures to internal causes compared to men (e.g. Feather 1969 and Frieze, Whitley Jr, Hanusa, & McHugh 1982; Mezulis, Abramson, Hyde, & Hankin 2004). This observation was partly scrutinized by Niederle & Yestrumskas (2008) speculating that the differences between genders' choice of tasks, could be driven by the fact that women may attribute success to luck, and failure to ability, and the other way around for men. We also test whether performance and confidence levels are different between males and females across stages and within top and bottom scorers: we do not find that any significant difference either in performance or in confidence. The only gender difference was in causal attributions of successes and failures which could be one of the underlying psychological causes for the gender difference in entry to competitions independent of confidence and performance.

**Table 4: Biases in Perception of the Task between Top and Bottom Scorers**

| | **Difficult** | **Easy** |
|---|---|---|
| **TopTop** | 3.25 | 5.79 |
| | (1.85) | (1.49) |
| **BottomBottom** | 1.88 | 4.04 |
| | (1.29) | (1.32) |
| ***p*** | 0.004 | 0.000 |

*Mean answers of top and bottom scorers to the question "On a scale of 1 to 7, did luck or skill determine your performance in the Circle task? (1-entirely luck, 7-entirely skill)". Standard deviations are in parenthesis, reported p-values are from Wilcoxon rank sum test*

### *4.5 Discussion*

In their seminal paper, Camerer & Lovallo (1999) introduced the concept of overconfidence into the economics literature to explain entrepreneurial overentry to markets and high failure rates. They proposed a hypothesis of “reference group neglect” conjecturing that when deciding to enter a competition people only focus on their own performance and neglect that of the competition. A related and complementary hypothesis of “differential information” was proposed in several papers by Don Moore and his colleagues. They argue that subjects’ relative confidence judgements regress towards their own absolute performances because they have more information about themselves than about others. Our study enables us to partially distinguish between these two hypotheses. In our setting, subjects have symmetric information about their own past relative performance and the performance of those who are in their reference group. We find support for the reference group neglect hypothesis: confidence snowballs from one stage to the other as people receive positive or negative performance feedback about themselves neglecting the information about others in his current group. More specifically, we find that confidence of top scorers increased from one stage to the next in both easy and difficult task treatments and that of bottom scorers decreased in the difficult but not in the easy treatment. The differential information hypothesis cannot account for this observation of snowballing confidence as it would predict no change in confidence because of the symmetric information one receives about self and others.

We check for overconfidence and underconfidence in easy and difficult task treatments. The usual finding in the literature is overconfidence in easy and familiar tasks and underconfidence in difficult and unfamiliar tasks. In the first stage of the tournament, we find that there is mild overconfidence in the easy treatment and no underconfidence in the difficult treatment where on average subjects assign 50% confidence to being in top of their groups. We conjecture that these results are partly due to our confidence elicitation tool which is a colder and more cognitive incentivized elicitation device than non-incentivized self-reports. We also conjecture that these results are partly due to the nature of the task which was initially unfamiliar to the majority of our subjects and may have caused more uncertain confidence judgements which resulted in a "fifty-fifty" confidence of being in the top half of the group.[27] Further investigation is needed to answer whether it is the task type or elicitation device that results in well calibration of relative confidence in Stage 1 than what has been found in previous psychology and experimental economics literature. Interestingly, in the later stages, after subjects receive initial feedback about their relative performances we do find results in line with previous literature that sbjects are more confident in the easy than in the difficult task.

Another interesting finding of our study is that subjects perceive the task differently, depending on whether they ended up being in the top or bottom scoring groups. Top scorers rate the task more a skill task than a luck task and bottom scorers do the opposite. This is another manifestation of causal attribution bias which suggests that people tend to attribute their success to internal factors (such as skill) and failures to external factors (such as luck). This may explain why confidence snowballing happens in the first place and why upwards snowballing is more robust than the downward snowballing. Thus our finding provides an example where feedback worsens calibration rather than helps it and makes subjects over/underconfident. We find an interesting gender difference in the attribution bias. While males exhibit the bias, females do not

[27] Clark & Friesen (2009) also show evidence of accurate calibration in the two initially unfamiliar computerized tasks.

differ in their perceptions of the task as a skill or luck task depending on them being in the top or bottom scoring groups. Observing gender difference in causal attribution bias in our setting has an important implication for real world settings where corporate governance policies can cause overconfidence of male managers more than of female ones which can further be aggravated by self-selection to managerial positions by more confidence males as well as by reference group neglect.

### *References*


Arkin, R. M., Appelman, A. J., & Burger, J. M. (1980). "Social Anxiety, Self-presentation, and the Self-serving Bias in Causal Attribution". *Journal of Personality and Social Psychology*, Vol. 38, pp. 23-35.

Benabou, R., & Tirole, J. (2003). "Intrinsic and Extrinsic Motivation". *The Review of Economic Studies*, Vol. 70, pp. 489-520.

Berlin, N., & Dargnies, M. P. (2012). "Linking Beliefs to Willingness to Compete". CES Working papers 75.

Brown, R., & Sarma, N. (2007). "CEO Overconfidence, CEO Dominance and Corporate Acquisitions". *Journal of Economics and Business*, Vol. 59, pp. 358-379.

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2013). "Overconfidence and Social Signalling". *The Review of Economic Studies*, Vol. 80, pp. 949-983.

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

Cárdenas, J. C., Dreber, A., Von Essen, E., & Ranehill, E. (2012). "Gender Differences in Competitiveness and Risk Taking: Comparing Children in Colombia and Sweden". *Journal of Economic Behavior & Organization*, Vol. 83, pp. 11-23.

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

Cuzick, J. (1985). "A Wilcoxon-Type Test for Trend". *Statistics in Medicine*, Vol. 4, pp. 543-547.

Dargnies, M. P. (2012). "Men too Sometimes Shy away from Competition: The Case of Team Competition". *Management Science*, Vol. 58, pp. 1982-2000.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences". *Journal of the European Economic Association*, Vol. 9, pp. 522-550.

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

Eil, D., & Rao, J. M. (2011). “The Good News-Bad News effect: Asymmetric Processing of Objective Information about Yourself”. *American Economic Journal: Microeconomics*, Vol. 3, pp. 114-138.

Ertac, S., & Szentes, B. (2011). “The Effect of Information on Gender Differences in Competitiveness: Experimental Evidence”. TÜSİAD-Koç University Economic Research Forum working paper series.

Feather, N. T. (1969). “Attribution of Responsibility and Valence of Success and Failure in Relation to Initial Confidence and Task Performance”. *Journal of Personality and Social Psychology*, Vol. 13, pp. 129-144.

Fischbacher, U. (2007). z-Tree: “Zurich Toolbox for Ready-made Economic Experiments”. *Experimental Economics*, Vol. 10, pp. 171-178.

Frieze, I. H., Whitley Jr, B. E., Hanusa, B. H., & McHugh, M. C. (1982). “Assessing the Theoretical Models for Sex Differences in Causal Attributions for Success and Failure”. *Sex Roles*, Vol. 8, pp. 333-343.

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.

Goel, A. M., & Thakor, A. V. (2008). “Overconfidence, CEO Selection, and Corporate Governance”. *The Journal of Finance*, Vol. 63, Vol. 2737-2784.

Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. *University of Cologne, Working paper series in economics*, Vol. 10, pp. 63-104.

Grossman, Z., & Owens, D. (2012). “An Unlucky Feeling: Overconfidence and Noisy Feedback”. *Journal of Economic Behavior &* Organization, Vol. 84, pp. 510-524.

Große, N. D., & Riener, G. (2010). “Explaining Gender Differences in Competitiveness: Gender-task Stereotypes”. Jena economic research papers, (No. 2010, 017).

Günther, C., Ekinci, N. A., Schwieren, C., & Strobel, M. (2010). “Women Can’t Jump?—An Experiment on Competitive Attitudes and Stereotype Threat.” *Journal of Economic Behavior & Organization*, Vol. 75, pp. 395-401.

von Hippel, W., Lakin, J. L., & Shakarchi, R. J. (2005). "Individual Differences in Motivated Social Cognition: The Case of Self-serving Information Processing". *Personality and Social Psychology Bulletin*, Vol. 31, pp. 1347-1357.

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". Paris School of Economics working paper.

Kamas, L., & Preston, A. (2012). "The Importance of Being Confident; Gender, Career Choice, and Willingness to Compete". *Journal of Economic Behavior & Organization*, Vol. 83, pp. 82-97.

Kisfalvi, V. (2000). "The Threat of Failure, the Perils of Success and CEO Character: Sources of Strategic Persistence". *Organization Studies*, Vol. 21, pp. 611-639.

Kruger, J., & Burrus, J. (2004). "Egocentrism and Focalism in Unrealistic Optimism (and Pessimism)". *Journal of Experimental Social Psychology*, Vol. 40, pp. 332-340.

Köszegi, B. (2006). "Ego Utility, Overconfidence, and Task Choice". *Journal of the European Economic Association*, Vol. 4, pp. 673-707.

Levinson, H. (1994). "Why the Behemoths Fell: Psychological Roots of Corporate Failure". *American Psychologist*, Vol. 49, pp. 428-436.

Ludwig, S., Wichardt, P. C., & Wickhorst, H. (2011). "Overconfidence Can Improve an Agent's Relative and Absolute Performance in Contests". *Economics Letters*, Vol. 110, pp. 193-196.

Malmendier, U., & Tate, G. (2005). "Does Overconfidence Affect Corporate Investment? CEO Overconfidence Measures Revisited". *European Financial Management*, Vol. 11, pp. 649-659.

Malmendier, U., & Tate, G. (2008). "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction". *Journal of Financial Economics*, Vol. 89, pp. 20-43.

Massoni, S., & Roux, N. (2012). "Optimal Group Decision: A Matter of Confidence Calibration". Paris School of Economics working paper.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). "Is There a Universal Positivity Bias in Attributions? A Meta-analytic Review of

Individual, Developmental, and Cultural Differences in the Self-serving Attributional Bias”. *Psychological Bulletin*, Vol. 130, pp. 711-747.

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). “Managing Self-confidence: Theory and Experimental Evidence”. (No. w17014). National Bureau of Economic Research.

Moore, D. A. (2007). “Not so Above Average After All: When People Believe They are Worse than Average and its Implications for Theories of Bias in Social Comparison”. *Organizational Behavior and Human Decision Processes*, Vol. 102, pp. 42-58.

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

Moore, D. A., & Kim, T. G. (2003). “Myopic Social Prediction and the Solo Comparison Effect”. *Journal of Personality and Social Psychology*, Vol. 85, pp. 1121-1135.

Moore, D. A., & Small, D. A. (2007). “Error and Bias in Comparative Judgment: on Being Both Better and Worse than We Think We are”. *Journal of Personality and Social Psychology*, Vol. 92, pp. 972-989.

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

Niederle, M., & Yestrumskas, A. H. (2008). “Gender Differences in Seeking Challenges: The Role of Institutions”. National Bureau of Economic Research, (No. w13922).

Paredes, T.A. (2005). “Too Much Pay, Too Much Deference: Behavioural Corporate Finance, CEOs, and Corporate Governance”, *American Law & Economics Association Annual Meetings*, bepress, p.40

Park, Y. J., & Santos-Pinto, L. (2010). “Overconfidence in Tournaments: Evidence from the Field”. *Theory and Decision*, Vol. 69, pp. 143-166.

Pyszczynski, T., Greenberg, J., Solomon, S., Arndt, J., & Schimel, J. (2004). “Why Do People Need Self-esteem? A Theoretical And Empirical Review”. *Psychological Bulletin*, Vol. 130, pp. 435-468.

Radzevick, J. R., & Moore, D. A. (2008). “Myopic Biases in Competitions”. *Organizational Behavior and Human Decision Processes*, Vol. 107, pp. 206-218.

Rose, J. P., & Windschitl, P. D. (2008). “How Egocentrism and Optimism Change in Response to Feedback in Repeated Competitions”. *Organizational Behavior and Human Decision Processes*, Vol. 105, pp. 201-220.

Roth, J. (1975). “Heads I Win, Tails It's Chance: The Illusion of Control as a Function of the Sequence of Outcomes in a Purely Chance Task”. *Journal of Personality and Social Psychology*, Vol. 32, pp. 951-955.

Symons, C. S., & Johnson, B. T. (1997). “The Self-Reference Effect in Memory: a Meta-Analysis”. *Psychological Bulletin*, Vol. 121, pp. 371-394.

Urbig, D., Stauf, J., & Weitzel, U. (2009). “What is Your Level of Overconfidence? A Strictly Incentive Compatible Measurement of Absolute and Relative Overconfidence”. *Tjalling C. Koopmans Institute Discussion Paper Series*, No: 9, pp. 1-40.

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). “The Influence of Egocentrism and Focalism on People's Optimism in Competitions: When What Affects us Equally Affects Me More”. *Journal of Personality and Social Psychology*, Vol. 85, pp. 389-408.

## *Appendix A: The Circle Task*

The Circle task has been used previously in Hollard, Massoni & Vergnaud (2010) and Massoni & Roux (2012) to study calibration between absolute confidence and success rates. The task involved seeing a pair of black circles with white dots in them for one second and judging which circle had more dots: a set of tasks comprised of 20 pairs of circles. We were specifically attracted to this task because we wanted to measure confidence in a "skill" rather than "effort" task. As a skill task it would require a minimum effort level to achieve the maximum performance after which additional effort would not improve performance. We also aimed for a task that was gender neutral both in performance and in perceptions about the task. We pre-tested the task for these properties in a standard experimental session format with subjects recruited via Orsee (Greiner 2004) to the lab and completing the task on computer terminals programmed with Ztree (Fischbacher 2007). The pre-test was conducted one month prior to the main study.

Four difficulty levels of a task were pretested: 50/51, 50/55, 50/60, 50/65, 50/70 dot circles. We randomized the order in which they were presented to subjects. We also presented two of the difficulty levels (50/51 and 50/60) to subjects three times, in order to check whether there were learning effects. We had two between subject treatments, low and high incentives, to check whether additional effort improves performance. Subjects were paid for one randomly selected set. In low incentive scheme were paid £0.05 per correct answer plus a fixed payment of £5. The high incentive scheme paid £0.50 per correct answer.

In none of the difficulty levels, do we find difference in performance between the two incentive levels.[28] The maximum performance was in 50/70 dot circles (19.81 in low and 19.75 in high incentive, Wilcoxon ranksum p=0.729) and the minimum performance was in 50/51 dot circles (11.3 in low and 10.9 in high incentive, Wilcoxon ranksum p=0.437) and hence we chose these two difficulty levels as our easy and difficult tasks for the main experiment. We also test for learning effects in our pre-test data and find that performance within a difficulty level does not improve from earlier to later sets (Cuzick (1985) test for trend p=0.432).

**Table A1: Performance at Each Difficulty and Incentive Level**

| **Difficulty Level** | **High Pay** | **Low Pay** | ***p-value*** |
|---|---|---|---|
| 50/51 | 10.9 | 11.3 | *0.437* |
| 50/55 | 15.0 | 15.0 | *0.941* |
| 50/60 | 18.1 | 18.2 | *0.767* |
| 50/65 | 19.0 | 19.7 | *0.119* |
| 50/70 | 19.8 | 19.8 | *0.729* |

*p-values are from Wilcoxon ranksum test*

Previous studies have presented mixed evidence of whether task type being perceived to be "female" or "male" in nature can affect confidence levels between genders (Grosse & Riener 2010; Cardenas, Dreber, Essen & Ranehill 2011; Gunther, Ekinci, Schwieren & Strobel 2010). To avoid a possible contamination of gender effects in our study we checked whether there were differences in performance, response times, and perceptions about the difficulty level of the task across genders. To check for the perceptions about the tasks we asked a number of questions after each set of a task and at the end of the pre-test session (e.g. on a scale of 1 to 7, "how difficult did you find the task", "how

[28] We do not find any difference in the answers of subjects to the questions "how hard did you try", and "how focused were you on a task" across incentive levels either, indicating subjects did not consciously exert effort level depending on the rewards they would get.

much did luck or skill determine your performance in the set"). We do not find any gender differences neither in performance (Wilcoxon ranksum p=0.125), nor in response times (p=0.587) nor in perceptions (p=0.214).

Results of the pre-test convinced us that this task could be usefully employed to study confidence. In particular, we concluded that

i) incentives play no role in performance,

ii) learning effects are absent,

iii) there are no gender differences.

These conditions attested that the task was more skill than effort task and can be further used in studies interested in measuring behaviour relating to individual abilities.

## *Appendix B: Instructions*

Welcome! You are about to participate in an experiment. There are 16 people participating in the experiment. You must not communicate with anyone. If you have a question at any time during the experiment, please raise your hand and someone will come to your desk to answer it.

The use of electronic devices such as mobile phones, music players, and tablet computers is strictly prohibited. Please make sure that all such devices are turned off and put away out of sight.

If you break these rules, you will be excluded from the experiment without receiving any payment and be disqualified from future experiments with CeDEx.

If you agree to these terms please press the "Next" button and proceed to the Instructions.

| Next |
|---|

In this experiment you will complete the Circle task. In this task you will see a pair of circles on your screen for 1 second. One of the circles contains 50 dots and the other contains 70 dots. After the circles disappear from your screen, you will be asked to judge whether the right or the left circle contained more dots. You have to indicate your judgement by pressing on the "Left" or "Right" button. When you press the button of your choice, you will move to the next pair of circles.

To acquaint you with the task, you can practice the task by pressing START THE PRACTICE button. You will practice one set of 20 tasks. The practice set will not affect your final outcome. You can start the practice when you are ready. We will give you more information about the experiment, when you finish the practice set.

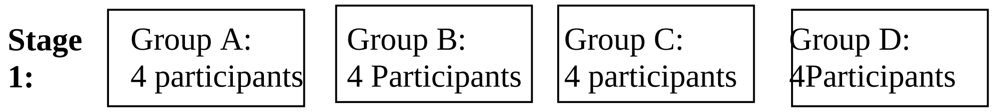
| START THE PRACTICE |
|---|

**Instructions:**

There are 3 stages in this experiment. At each stage, the participants will be divided into groups of 4. At each stage, you will complete the Circle task and receive a score based on your performance. You will complete one set of 20 pairs of circles and score 1 point for each correct answer, so you can score up to 20 points. Similarly, each other group member will complete the Circle task, seeing an identical set of circles, scoring 1 point for each correct answer, and so scoring up to 20 points.

**Stage 1:**

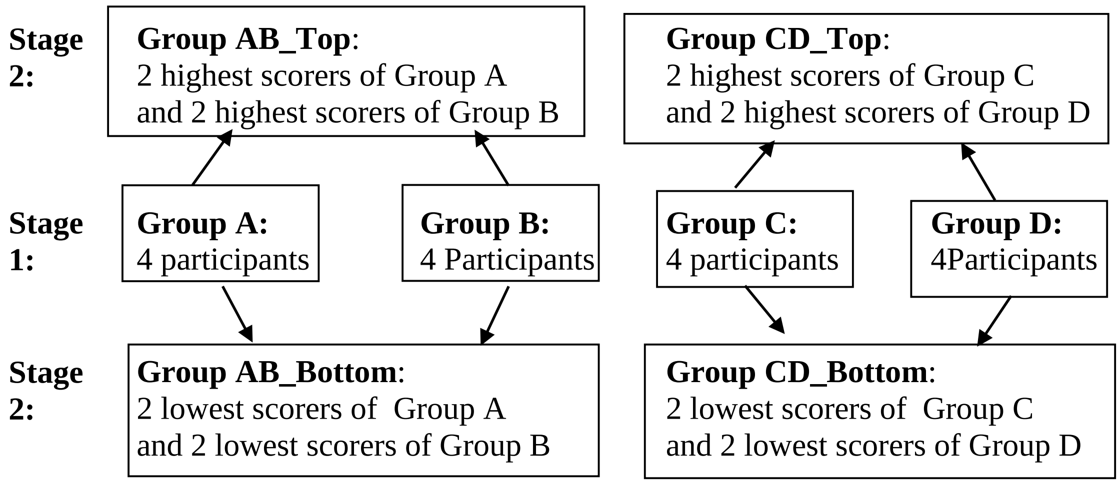
In Stage 1, you will be randomly matched with 3 other participants to form a group. There will be four groups: Group A, Group B, Group C and Group D. You will see which group you are in on your screens.




In Stage 2, you will be placed in Back ord Stage 2 Instructions ge 1.

- The top two scorers of Group A will be matched with the top two scorers of Group B to form the **Group AB_Top**.
- The top two scorers of Group C will be matched with the top two scorers of Group D to form the **Group CD_Top**.
- The bottom two scorers of Group A will be matched with the bottom two scorers of Group B to form the **Group AB_Bottom**.
- The bottom two scorers of Group C will be matched with the bottom two scorers of Group D to form the **Group CD_Bottom**.

Ties will be randomly broken.



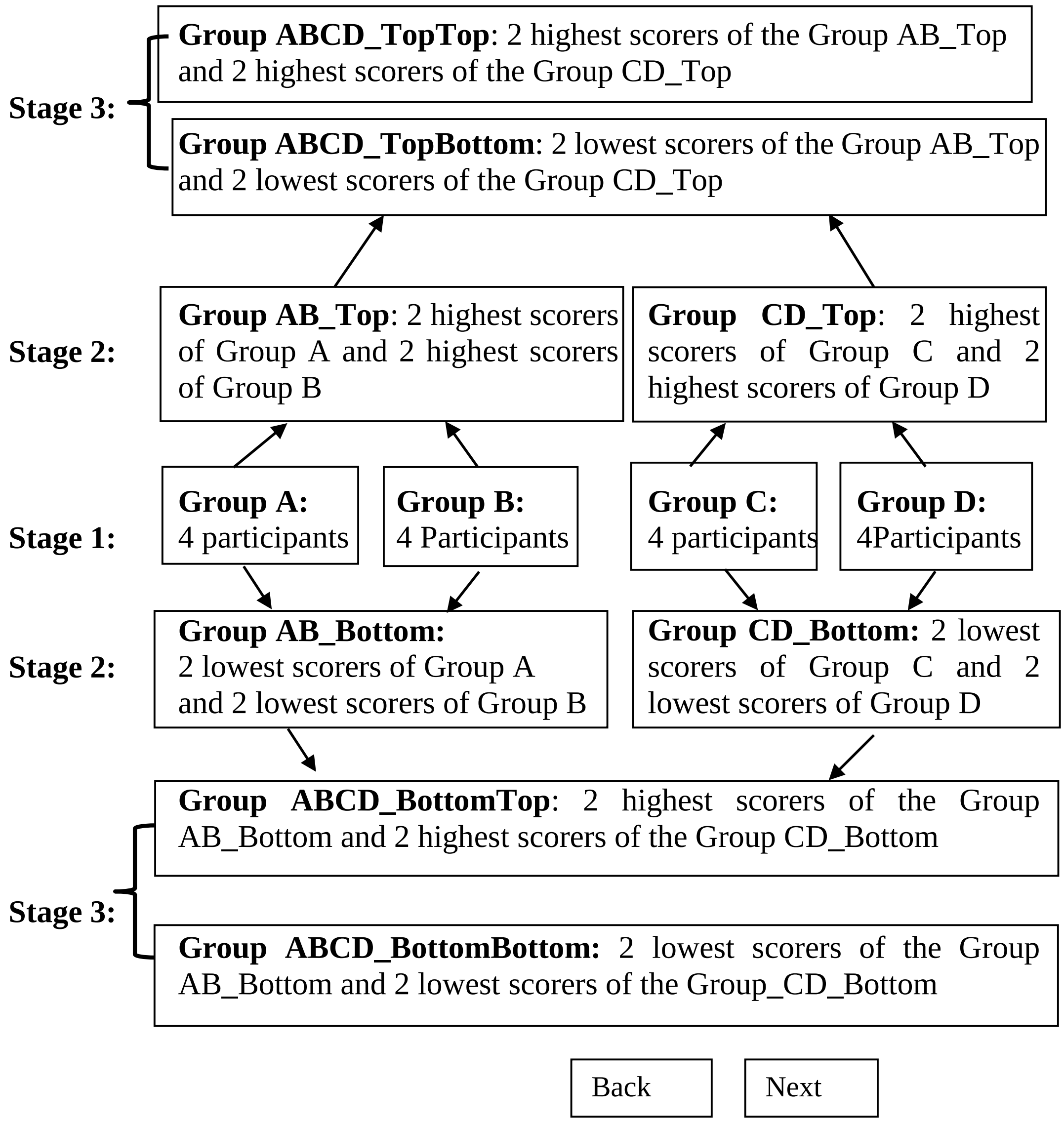

Back    Stage 3 Instructions

**Stage 3:**

In Stage 3, you will be placed in a new group according to your scores in Stage 2.

- The top two scorers of Group AB_Top will be matched with the top two scorers of Group CD_Top to form the **Group ABCD_TopTop**.
- The bottom two scorers of Group AB_Top will be matched with the bottom two scorers of Group CD_Top to form the **Group ABCD_TopBottom**.

- The top two scorers of AB_Bottom will be matched with the top two scorers of Group CD_Bottom to form the **Group ABCD_BottomTop**.
- The bottom two scorers of the Group AB_Bottom will be matched with the bottom two scorers of Group CD_Bottom to form the **Group ABCD_BottomBottom**.




At each stage you will be informed of which group you are in. You will have to do two things: fill out a table as below and then complete the Circle task

**Table:**

For every row of the table you must make a choice between Option A and Option B. You should consider which of these two options you would prefer to have for each row then mark your choice by ticking the circle corresponding to your preferred option. In each table, we ask that you start with the top row and work your way down the rows. Option A will pay off £10 if you are one of the top two scorers of your group in this stage and £3 if you are one of the bottom two scorers of your group in this stage. (You will complete the Circle task and find out if you are in the top two or bottom two after you have filled the table.)Option B is a lottery which will pay off £10 or £3 with the chance of £10 progressively decreasing as you move down the rows. Since the chance of winning £10 for Option B in the first row is 100%, we think that you will want to choose Option B in the first row. But, since Option B gets progressively worse as you move down the rows (while Option A stays exactly the same), there may come a row where the chance of winning £10 of Option B is sufficiently small, that you prefer Option A. If you find such a row, you should then choose Option A for that row and the rows below it (since Option B continues to get worse all the way down the table).

| Row | Option A:Bet | Your Choice: | Option B: Lottery |
|---|---|---|---|
| 1 | | A ○ ● B | £10 with 100% chance |
| 2 | | A ○ ○ B | £10 with 95% chance and £3 with 5% chance |
| 3 | | A ○ ○ B | £10 with 90% chance and £3 with 10% chance |
| 4 | | A ○ ○ B | £10 with 85% chance and £3 with 15% chance |
| 5 | | A ○ ○ B | £10 with 80% chance and £3 with 20% chance |
| 6 | | A ○ ○ B | £10 with 75% chance and £3 with 25% chance |
| 7 | | A ○ ○ B | £10 with 70% chance and £3 with 30% chance |
| 8 | | A ○ ○ B | £10 with 65% chance and £3 with 35% chance |
| 9 | You get £10 if | A ○ ○ B | £10 with 60% chance and £3 with 40% chance |
| 10 | you are one of the | A ○ ○ B | £10 with 55% chance and £3 with 45% chance |
| 11 | top two scorers of your group | A ○ ○ B | £10 with 50% chance and £3 with 50% chance |
| 12 | and you get £3 if | A ○ ○ B | £10 with 45% chance and £3 with 55% chance |
| 13 | you are one of the | A ○ ○ B | £10 with 40% chance and £3 with 60% chance |
| 14 | bottom two scorers of your group | A ○ ○ B | £10 with 35% chance and £3 with 65% chance |
| 15 | | A ○ ○ B | £10 with 30% chance and £3 with 70% chance |
| 16 | | A ○ ○ B | £10 with 25% chance and £3 with 75% chance |
| 17 | | A ○ ○ B | £10 with 20% chance and £3 with 80% chance |
| 18 | | A ○ ○ B | £10 with 15% chance and £3 with 85% chance |
| 19 | | A ○ ○ B | £10 with 10% chance and £3 with 90% chance |
| 20 | | A ○ ○ B | £10 with 5% chance and £3 with 95% chance |

**Circle Task**

After you have completed the table you will complete the Circle task. You will have 20 tasks to complete similar to the practice set. After everyone has completed the task, your scores will be calculated. For each correct answer, 1 point will be added to your score so that you can score up to 20 points. Your score will be compared to the scores of the other members of your group and at the end of the stage you will be informed of whether you were in the TOP or BOTTOM two of your group for that stage. Ties will be randomly broken. You will then be assigned to a new group according to your performance, as explained above, and the next stage will begin.

**Payment:**

At the end of the session, a 6-sided die will be rolled. Depending on the outcome, you will be paid based on one of the stages and either the table or your performance in the Circle task.

| Roll | Paid according to: |
|---|---|
| 1 | Stage 1 – Table |
| 2 | Stage 1 – Circle |
| 3 | Stage 2 – Table |
| 4 | Stage 2 – Circle |
| 5 | Stage 3 –Table |
| 6 | Stage 3 – Circle |

If a table is selected, a 20-sided die will be rolled to select a row of that table and you will be paid according to your choices on the selected row. If you have chosen Option B on the selected row, you will play out the lottery of the selected row. If you have chosen Option A on the selected row, you will be paid £10 if you were in the top two of your group and £3 if you were in the bottom two of your group according to your scores in the selected stage. If Circle task is selected, you will get £0.50 for each correct answer so that you can earn up to £10.

This payment structure has been designed so that you have an incentive to do your best in the Circle task and to think carefully about each and every choice you have to make in the Table.

If you have read and understood the instructions you may start the experiment. You may press the Back button to go back and read the instructions once more. If you have any questions, please raise your hand and someone will come to your desk to answer it.

| Back | Start the Experiment |
|---|---|

# CHAPTER 5: CONCLUSIONS

In this thesis, we have reported three studies that use laboratory experiments to investigate different topics in the field of behavioural and experimental economics. In one of the studies (reported in chapter 2) we have explored the relationship between absolute confidence levels and individual risk attitudes. The two other studies included in this thesis (reported in chapters 3 and 4) studied relative confidence in competitive environments.

In Chapter 2, we looked at two potential ways in which risk attitudes may affect measured confidence levels. In the first, we investigated whether individual risk attitudes, more specifically whether attitudes to consequences or to probabilities, were related to self-reported confidence levels. We replicated previous findings of a hard-easy effect where subjects underestimate their success in easy and overestimate their success in difficult tasks (Fischoff, Slovic & Lichtenstein 1977; Lichtenstein & Fischoff 1981) using a self-reported elicitation procedure. We found strong association between individual risk attitudes and self-reported confidence. 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.

In the second, we investigated how risk attitudes affect confidence elicited through incentivized device where confidence was inferred from choices subjects made. With our new incentivized elicitation procedure, the inferred confidence exhibited general level of underconfidence, which has also been found in more recent experimental economics literature that has used incentives to elicit confidence (e.g. Blavatskyy 2009; Clark & Friesen 2009).

Only when we corrected for individual risk attitudes, observed underconfidence decreased and confidence became better calibrated.

In Chapter 3, we studied how information affected entry decisions in experimental market entry games. In these games, market entrants receive a payoff that depends on a ranking that is determined either randomly or by their scores in a trivia quiz. Camerer & Lovallo (1999) and Moore & Cain (2007) have shown that individuals overenter markets with a simple quiz and underenter markets with a difficult quiz and explain this pattern by relative confidence. We asked whether implicit information through individuals deciding in a group or explicit information through provision of historical average performance information affects entry patterns to simple and difficult markets. We replicated Moore & Cain's experimental design as one of the treatments and added two between-subject information manipulation treatments to study this research question. We hypothesized that groups could serve as implicit information channels, in that interaction between group members would result in groups possessing more information about others' performances than individuals. We found that when information was implicit via group decision making, groups predicted others' entry behaviour and performances more accurately than individuals. However, in spite of this, groups exhibited a similar bias in entry to individuals and the bias was not mitigated through repetition and feedback. Giving explicit average historical performance information, we observed a lower degree of difference in entry between simple and difficult markets, which was completely eliminated in the second half of the experiment. In this treatment, we also observed the bias in confidence disappeared in the second half of the experiment.

In Chapter 4, we investigated whether relative confidence snowballs as subjects move up or down a hierarchical tournament ladder. The previous literature has studied confidence mostly within the context of single stage tournaments and contests (Niederle & Vesterlund 2007; Moore & Cain 2007; Ludwig, Wickhardt & Wickhorst 2011) but has not looked how confidence evolves through stages of a tournament. We considered two between-subject treatment conditions, where a task was either very easy or very difficult and elicited confidence using a novel incentivized elicitation tool. We found that confidence was well calibrated in the first stage of the tournament; on average subjects assigned 50% to being in the top half of their groups in the difficult treatment and were mildly overconfident in the easy treatment. As subjects progressed to the next stages of the tournament, their confidence snowballed. Top scorers, who received positive feedback in the previous stage, assigned higher probability to being in the top half and the bottom scorers, who receive negative feedback in the previous stage, assigned lower probability to being in the top half of their new groups. We also found an asymmetric pattern of snowballing: top scorers changed their confidence upwards more consistently than bottom scorers did downwards. We also identified an interesting bias in causal attribution of successes and failures between top and bottom scorers, such that top scorers rated the task more as a skill than a luck task. Further investigation showed that this bias was mainly due to males' ratings of the task rather than females'.

While we present novel and stimulating evidence of causal and correlational relationships that affect confidence (risk attitudes, information, feedback), there are many other potential relationships that could be considered to affect confidence. In our first paper presented in the Chapter 2, we find a positive and significant association between risk attitudes and elicited confidence, and identify parameters of utility and probability weighting functions that affect confidence. 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). Further research could seek to investigate whether ambiguity attitudes about uncertain events are related to confidence judgements subjects make about their own performance.

In the last paper of the dissertation, we present compelling evidence of confidence snowballing in the hierarchical tournament setting. Further research on confidence snowballing might look in more detail at the interaction between group formation and snowballing. For example, one could explore whether subjects' confidence snowballs when subjects receive feedback about their own relative performance in the previous stage but are randomly grouped with other subjects in the subsequent stage. In this case, subjects have information about themselves but no information about the reference group they are competing against. Comparing the snowballing of confidence from one stage to the next in this condition with that observed in chapter 4 would enable one to identify the extent of reference group neglect where either subjects completely neglect competition (the snowballing of confidence is identical across two studies) or they only underweight competition (the snowballing is stronger in the randomly grouped study). This will further test the robustness of the reference group neglect hypothesis against the differential information hypothesis as possible psychological mechanisms explaining biases in relative confidence judgements and competitive decision making.

## *References*


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

Camerer, C., & Lovallo, D. (1999). "Overconfidence and Excess entry: An Experimental Approach". *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.

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.

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

Ludwig, S., Wichardt, P. C., & Wickhorst, H. (2011). "Overconfidence Can Improve an Agent's Relative and Absolute Performance in Contests". *Economics Letters,* Vol. 110, pp. 193-196.

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

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.