Do positive illusions predict overconfidence in judgment? A test using interval production and probability evaluation measures of miscalibration

Denis Hilton
University of Toulouse

Isabelle Régner
University of Aix-Marseille I and University of Toulouse

Laure Cabantous
Nottingham University Business School

Laetitia Charalambides
University of Toulouse

Stéphane Vautier
University of Toulouse

This paper has been published in the *Journal of Behavioral Decision-Making*

To quote this paper

Hilton D., I. Régner, L. Cabantous, L. Charalambides, & S. Vautier (2011). Do positive illusion predict overconfidence in judgment? A test using interval production and probability evaluation measures of miscalibration. *Journal of Behavioral Decision Making*, 24(2), p.117-139

http://onlinelibrary.wiley.com/doi/10.1002/bdm.678/abstract

Correspondence should be addressed to Denis Hilton, Cognition, Langage, Langues et Ergonomie (CLLE), Université de Toulouse Le Mirail, Maison de la Recherche, 5 allées Antonio-Machado, 31058 Toulouse Cedex 9, FRANCE, Tel: +33 561 50 49 81, email: hilton@univ-tlse2.fr

Do positive illusions predict overconfidence in judgment?
A test using interval production and probability evaluation measures of miscalibration

<u>Abstract</u>

We address the question as to whether judgmental overconfidence, as assessed by probability miscalibration, is related to positive illusions about the self. We first demonstrate that judgmental overconfidence measured with interval production procedures can be considered a trait, due to correlations observed in miscalibration scores in two sets of general-knowledge questions of varying difficulty administered at different times. In addition, the hard-easy effect operated in different ways on overprecision and self-placement of one's performance relative to others: the more difficult the calibration task, the greater the *over*precision but the greater the *under*placement of one's performance. Finally, there was no evidence that miscalibration was related to dispositional optimism and self-efficacy. A second study extended these results by including further measures of disposition to experience positive illusions such as unrealistic optimism, a general tendency to consider oneself "better-than-average", and two indexes of dispositional perception of control. The positive illusion measures showed considerable inter-correlations, but did not correlate with miscalibration on the interval production task, and correlated negatively with optimism concerning societal risks. A final study replicated this pattern of findings, but showed that disposition to positive illusions did predict miscalibration on the same questions measured with a probability evaluation technique. Our research demonstrates that "overconfidence" is not a unitary construct, but a series of overlapping ones.

Keywords: Overconfidence, Miscalibration, Positive Illusions, Worse-than-average Effect

Do positive illusions predict overconfidence in judgment?

A test using interval production and probability evaluation measures of miscalibration

Overconfidence has become one of the mainstays of the heuristics and biases programme in behavioral judgment and decision-making (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). There are many manifestations of this overconfidence. For example, people are miscalibrated (i.e. overestimate the probability that their judgments are correct; Alpert & Raiffa, 1982) and suffer from "positive illusions" (Taylor & Brown, 1988) such as illusion of control (i.e. they overestimate their control over events; Langer & Roth, 1975), unrealistic optimism (i.e. they consider that bad events are more likely to happen to others than to themselves; Weinstein, 1982) and the "better-than-average" effect (i.e. they overestimate their achievements and abilities relative to others; Svenson, 1981). The above literature in psychology has been cited by behavioral finance researchers to support the claim that people are overconfident, and that this is likely to lead them to make errors in financial markets (Barber & Odean, 2000; 2001; Odean, 1998; for a review, see Glaser, Nöth, & Weber, 2004).

Nevertheless one may ask whether these different forms of overconfidence are actually related to each other. Moore and Healy (2008) have distinguished between three kinds of overconfidence: *Overprecision* (overestimating the precision of one's knowledge), *overplacement* (overestimating one's ranking in a group) and *overestimation* (overestimating the quality of one's performance), and present experimental evidence to show that these judgments can be dissociated. At a more specific level, Griffin and Brenner (2004) distinguish five theoretical perspectives on calibration of probability judgment, of which one is motivational in character (optimistic overconfidence), while the others describe various possible sources of cognitive bias (confirmatory bias, case-based judgment, ecological probability, and error in expression of judgments).

In this paper, we ask whether the kind of self-enhancement biases that presumably motivate positive illusions also predict overconfidence on measures of probability calibration. We address this question through an individual difference approach that measures disposition to experience positive illusions (illusion of control, unrealistic optimism, the better-than-average effect, etc.) and correlates this with measures of judgmental overconfidence (as assessed by miscalibration scores on interval production and probability evaluation tasks). We review evidence that judgmental overconfidence measured by an interval production task can be considered to be a stable trait with cognitive origins that predicts real-world outcomes. Our first study demonstrates the temporal stability of miscalibration scores on an interval production task, and shows that these are not related to positive illusions such as self-placement on the calibration task. In our second study, we use a wider range of measures of positive illusions, and find that while the positive illusions are intercorrelated, they do not predict miscalibration on the interval production task. In our final study, we further replicate these patterns, but also find that positive illusions predict miscalibration on a probability evaluation task. We conclude that judgmental overconfidence should be distinguished from positive illusions, and end by reviewing relevant work in psychology and behavioural finance that illustrates the importance of this distinction.

Is Judgmental Overconfidence a Stable Trait?

The core finding that supports the existence of judgmental overconfidence is that people are miscalibrated: that is, they tend to overestimate the probability that their judgments are correct (Alpert & Raiffa, 1982; for a review, see Lichtenstein, Fischhoff, & Phillips, 1982). However, miscalibration depends at least in part on the format used to measure it. In particular, higher overconfidence is observed in the *interval production* task wherein participants are asked to state an interval such that they are XX% (e.g. 90%) sure that the correct response to the question falls in that interval. The repeated and robust finding is that people use confidence intervals that are too narrow (Soll & Klayman, 2004), leading to a high error rate due to correct answers falling outside the confidence interval. Overconfidence is still observed, but at a lower level, in the *probability estimation* format where participants are required to evaluate the probability that the chosen answer from a set of forced-choice options is correct (Klayman, Soll, Gonzales-Vallejo, & Barlas, 1999; Winman, Hanson, & Juslin, 2004). While the evidence for overconfidence in judgment using the above methodologies is very robust, it is important to note that research that uses other methods reliably suggests general underconfidence rather than overconfidence in judgment (e.g. Erev, Wallsten, & Budescu, 1994).

Miscalibration on Interval Production Tasks

Interval production techniques have been found to have cross-domain consistency: Glaser, Langer and Weber (2005) report considerable consistency in miscalibration scores across general and financial knowledge questions in both students and finance professionals. In addition, interval production tasks have been found to have predictive validity. Thus Klayman and Burt (1998) find that superior calibration in American MBA students is negatively associated with self-report measures of membership of constrained social networks (which show “groupthink” characteristics, such as being small, having many strong interconnections, with someone in a central co-ordinating position, and weak connections to outsiders). Biais and Weber (2005) find that higher miscalibration on an interval production task predicts lower earnings in traders but not in other specializations in a major bank.

Using an interval production questionnaire (Russo & Schoemaker, 1992), Biais et al. (2005) found that miscalibrated participants earn less profits on an experimental trading game, a finding replicated by Deaves, Lüder and Luo (2005). Using the same interval production measure, Bonnefon, Hilton and Molian (2005) found that highly overconfident entrepreneurs run businesses that are less profitable than those run by less overconfident entrepreneurs. These results encourage belief that individual differences revealed by the interval production task have external validity, as they successfully predict economically significant outcomes in three different samples. Moreover, these effects seem to be independent of intelligence, as Biais et al. (2005) found no correlation between an index of intelligence and correct calibration using this measure.

Juslin, Winman and Hansson (2006) provide a cognitive explanation for miscalibration on the interval production task. They argue that in the interval production task, people act as naïve statisticians and generate a "naïve sample" of relevant exemplars in response to the probe question. For example, if the task is to define a 90% confidence interval for the age at death of Martin Luther King, participants may consider that the relevant reference class is "leaders of a people". Participants may then generate a few relevant exemplars of this class: e.g. Nicolas Sarkozy, George Bush, and Angela Merkel. They may then naïvely conclude that this sample is representative of the relevant population, and by lopping a few years off the youngest exemplar and adding a few years to the oldest, they could conclude that 90%, of leaders are (say) aged between 45 and 65. This would lead to error, as the correct answer in this case is 39. However, participants who generate larger naïve samples would be less prone to this error, as they will have more chances of generating both older and younger exemplars, thus leading them to set wider confidence intervals. Thus we can suppose that in addition to thinking of Bush, Merkel and Sarkozy, they think of J.F. Kennedy, José Zapatero, Vladimir Putin and Nelson Mandela. This might lead them to define an interval between 35 and 85 years of age, in which case the correct answer (39) would indeed fall in the confidence interval. In line with this reasoning, Juslin et al. find that participants with higher working memory capacity, who may be expected to generate more exemplars, are indeed less prone to miscalibration using the interval production technique.

In sum, findings encourage the belief that judgmental overconfidence as measured on the interval production task may be a stable characteristic, as it is predicted by short-term memory capacity, correlates with social network measures, and is capable of predicting economic performance.

Miscalibration on Probability Evaluation Tasks

Miscalibration has also been measured using probability evaluation tasks, whereby respondents are asked to choose from two or more possible responses to a question, and then state a probability between 50% (for two-option questions) and a 100% that they think this response is correct. For example, they might be asked “What is the capital of Pakistan?”, and given two responses: a) Hyderabad, or b) Islamabad. They would then assign a probability (e.g., “80%”) to express their confidence that they have chosen the right answer. Research shows that the probability evaluation technique systematically produces overconfidence in judgment, though to a lesser extent than with the interval production technique (Klayman et al., 1999), a phenomenon referred to as “format dependence” (Winman et al., 2004).

There is some evidence that miscalibration revealed through probability evaluation techniques has the characteristics of a stable trait. First, miscalibration using probability evaluation shows both test-retest consistency and consistency across tasks (Bruine de Bruin, Parker, & Fischhoff, 2007; Jonsson & Allwood, 2003). Second, miscalibration using probability evaluation measures correlates with other judgmental characteristics. Thus Jonsson and Allwood (2003) report that higher need for cognition is correlated with higher confidence and accuracy, but not with better calibration. Superior calibration using a probability evaluation technique was correlated with measures of decision-making competence such as resistance to preference reversals induced by reframing choices, recognition of social norms, the tendency to make choices consistent with Luce's choice axioms, and the tendency to apply decision rules in choice situations, but not with indices of risk perception or resistance to sunk costs in a representative adult population (Bruine de Bruin et al., 2007). Similar results were obtained in a teenage population at risk for drug use (Parker & Fischhoff, 2002), with the exception of no correlation with the tendency to recognize social norms. Finally, several studies find that superior calibration of judgments using probability evaluation techniques is modestly correlated with higher cognitive ability (Bruine de Bruin et al., 2007; Parker & Fischhoff, 2005; Stanovich & West, 1998).

Are Positive Illusions Related to Judgmental Overconfidence?

There is some reason to expect that positive illusions about the self such as the "better-than-average" effects, optimism bias and the illusion of control (see Taylor & Brown 1988), should be correlated with each other. In their review, Alicke and Govorlin (2005) conclude that the better-than-average effect and optimism biases principally have motivational sources. Indeed, all the positive illusions appear to imply biases of self-enhancement (for a review, see Sedikides & Strube, 1997): belief that the self is better than the average person on valued characteristics; overestimation of the self's chances of experiencing positive outcomes and underestimation of the self's chances of experiencing negative outcomes; and overestimation of the self's chances of influencing the course of events. Some previous research has shown that unrealistic optimism and LOC beliefs are correlated (Hoorens & Buunk, 1993; see also McKenna, 1993), and that self-efficacy is correlated with measures of dispositional optimism (Macko & Tyska, 2006). We thus address the question of whether all the positive illusions identified by Taylor and Brown (1988) are themselves inter-correlated and can be seen as having a common origin, as to our knowledge there has been no research that addresses this question.

Should positive illusions be correlated with overconfidence in probability calibration? If we follow writers such as Barber and Odean (2002), who appear to treat judgmental overconfidence and positive illusions as part of the same phenomenon, then we should expect correlations between the two. However, if positive illusions have chiefly motivational origins, whereas judgmental overconfidence has cognitive origins, then there is no reason to expect significant correlations between the two. In addition, other research on the "hard-easy" effect has drawn attention to dissociations between probability miscalibration and one of the positive illusions, namely the better-than-average effect. Thus while increasing difficulty of a task leads to greater probability miscalibration in the form of overprecision (e.g. Lichtenstein et al., 1982), it leads people to judge themselves to have done worse than others on the task in question, thus producing a "worse-than-average" (WTA) effect (Larrick Burson & Soll, 2007; Moore & Healy, 2008). This appears to be because people tend to make ability-related social comparison judgments on the basis of how easy or difficult they themselves find the task, without taking into account whether others also find the task easy or difficult (Kruger, 1999). For this reason people consider themselves to be better-than-average at driving (because they find it easy) but worse-than-average at computer programming (because they find it difficult).

Summary and Plan of Research

Although test-retest validity has been demonstrated for calibration on the probability evaluation method (Jonsson & Allwood, 2003), this has not been demonstrated for the interval production method. Accordingly, a first question is whether probability miscalibration on the interval production task can be measured reliably, by assessing the same individual using interval production tasks on different occasions. In Study 1, we therefore presented the same participants with two interval production tasks at a two-week interval. In order to avoid recall problems, our two interval production tasks used different sets of questions.

We designed three studies whose aim is to advance our understanding of a simple quantitative index of calibration using the interval production task on general knowledge questions used by Biais et al. (2005), derived from Russo and Schoemaker (1992). In the first study, we examine some psychometric properties of our indices of individual differences in calibration, and their relation to metacognitive social comparison beliefs about one's performance on the calibration task, dispositional optimism and self-efficacy. In the second study, we extend our study of the relation of judgmental calibration to positive illusions, by adding measures of the better-than-average effect, unrealistic optimism and locus of control, as well as societal risk perception. In the third study, we compare miscalibration scores obtained using the interval technique and the probability evaluation technique on the same set of general knowledge questions, counterbalanced across the same group of participants.

Study 1: Internal and Discriminant Validity of the Interval Production Task: Is Judgmental Overconfidence Stable and Distinct over Time?

In the present research, we aimed to examine the psychometric characteristics of a measure of overconfidence in judgment that we knew had predictive validity. In particular, we administered the Russo and Schoemaker (1992) miscalibration questionnaire. We also developed another similar 10-item interval production questionnaire, which we administered after an interval of two weeks. This allowed us to assess the internal validity of the interval production task by examining test-retest reliabilities of miscalibration scores between the two questionnaires.

In addition we assessed the relationship between calibration and participants' self-placement on the calibration task. This can be evaluated through examining whether there is a relationship between actual performance on the calibration task (as quantified by the average proportion of actual values falling outside the range, that is by the number of surprises) and self-placement on the calibration task (as quantified by the degree to which a person considers him/herself to be better/worse than average at this calibration task).

The difference in difficulty of the two interval production tasks allowed us take into account the possibility that the "hard-easy" effect may lead miscalibration to be dissociated from self-evaluation. Immediately after each calibration task, we included a question requiring participants to evaluate their performance and that of others who did the task. This enabled us to calculate a calibration self-placement (cSP) score, by subtracting the perceived difficulty for others from the perceived difficulty for the self. Our hypothesis was that the more difficult the task (as indexed by higher overprecision on the calibration task), the more that participants should consider themselves to be worse-than-average (Kruger, 1999), leading to lower self-placement scores. Because task difficulty is of course related to ability, we controlled for differences in participants' ability. Specifically, as the calibration tasks used here tapped general knowledge, we retained French students majoring in different domains and schools and who could be expected to vary in general knowledge or general academic ability.

Finally, we included two personality measures that should be related to positive illusions: dispositional optimism (as indexed by the LOT-R, Life Orientation Test-Revised, Scheier, Carver, & Bridges, 1994) and self-efficacy (Schwarzer & Jerusalem, 1995). We included these measures because both unrealistic optimism (Weinstein, 1980) and illusion of control (Langer, 1975) have been linked to overconfidence in judgment (Barber & Odean, 2002).

Method

Participants

A total of 291 French undergraduate students completed two questionnaires on two separate occasions, both during classes. These students were enrolled in 3 different academic courses, in ascending level of selectivity (and thus *a priori* of academic ability): 133 third year Psychology students from the University of Toulouse-II, 133 second or third year Managerial economics students from The University of Toulouse-I, and 25 Management students from a *Grande Ecole,* the ESC Toulouse. Our assumption was that the ESC students would be more likely to have the kind of prior ability and/or mathematical training that should facilitate judgmental calibration.

Procedure

The first questionnaire consisted of a calibration scale in a confidence range format (as in Klayman et al., 1999; Lichtenstein et al., 1982). The scale was followed by self-placement questions concerning participants' performance on this scale, and a dispositional optimism scale (i.e., the LOT-R, Scheier et al., 1994). In the second questionnaire proposed two weeks later, participants completed another similar calibration scale, self-placement questions concerning performance on this calibration scale, and a general self-efficacy scale (Schwarzer & Jerusalem, 1995).

Instruments

Calibration tasks. The first calibration task was Russo and Schoemaker's (1992) 10-item scale using general knowledge questions (See the first ten questions in Appendix 1). For each item, participants were asked to give an upper and lower limit such that they were 90% sure that the actual value would fall within the range specified. For instance, the first item of this scale asked participants to provide range predictions for Martin Luther King's (MLK) age at death. We will refer to this scale as the "MLK scale". For the purpose of the present study, a second scale was created with 10 different but comparable items (See the first ten questions in Appendix 2). We will refer to this scale as the "Einstein scale" as the first item asked participants to provide range estimates for Einstein's expected age at death. Internal consistency, which was estimated by the Kuder-Richardson coefficient (K-R 20), was low for the MLK scale (.37) and modest for the Einstein scale (.50). For both scales, the average proportion of actual values falling outside the range (i.e., surprises) was computed. Participants are called overconfident when they give ranges such that actual values fall outside the range more than 10% of the time. The usual proportion of surprises obtained with this scale in previous studies is quite high, between 40 and 60% (see Biais et al., 2005; Klayman et al., 1999).

Self-placement on calibration tasks. Each calibration scale was followed by two questions designed to capture self-placement on the calibration task (cSP). The first question asked participants to indicate to what extent they thought they had succeeded at the previous calibration task (self-evaluation of calibration). The second one asked them to indicate to what extent they thought most other students had succeeded at the same task (evaluation of others' calibration). For both questions ratings were made on an 11-point scale (0 = *entirely failed*, 10 = *entirely succeeded*). The calibration self-placement (cSP) index was computed by subtracting the others-evaluation score from the self-evaluation score. A positive score indicates that participants evaluate themselves as being better than average (BTA effect), whereas a negative score means that they evaluate themselves as being worse than average (WTA effect).

Dispositional Optimism. The Life Orientation Test-Revised (LOT-R, Scheier et al., 1994) is a 10-item scale designed to assess dispositional optimism. It consists of 6 items used to derive an optimism score and 4 filler items. Among the 6 target items, 3 are keyed in a positive direction (e.g., "In uncertain times, I usually expect the best") and 3 are keyed in a negative direction (e.g., "If something can go wrong for me, it will"). The 3 negatively worded items were reverse coded before scoring. Participants were asked to indicate to what extent they agreed with each statement [1], with higher scores indicating higher dispositional optimism (Cronbach α =.83).

Self-Efficacy. The Generalized Self-Efficacy Scale is a 10-item scale designed to assess global confidence in one's ability to cope with a variety of difficult events in life (e.g., "I can always manage to solve difficult problems if I try hard enough"). Participants were asked to indicate to what extent they agreed with each statement, with higher scores indicating higher self-efficacy (Cronbach α = .89).

Results

Miscalibration

Table 1 summarizes overconfidence rates for the calibration tasks used in all three studies. The average proportion of surprises [2] was higher for the MLK scale (M = 78.14%, SD = 15.13) than for the Einstein scale (M = 62.41%, SD = 19.82), t (290) = 13.46, $p < .0001$, Cohen's d = .89). This indicates that both scales were difficult, with the MLK scale being significantly more difficult than the Einstein scale. This finding held whatever the student samples examined [3] (ps < .001, Cohen's ds > .80). As the sub-samples did not show different patterns of correlation between miscalibration scores and positive illusions, we will systematically report findings for the whole sample.

Self-Placement Scores on the Calibration Tasks

The perceived difficulty of these scales was supported by the findings on the calibration self-placement scores. Participants reported for both scales that they had performed *worse* than most other students (WTA effect), as indicated by the negative value of the cSP index (M = -1.11, SD = 1.64, t(290) = -11.52, $p < .001$ for the MLK scale and M = -.77, SD = 1.85, t(290) = -7.08, $p < .001$ for the Einstein scale). This shows the expected dissociation (Larrick et al., 2007): task difficulty increased overconfidence on the calibration task but increased the WTA effect on the self-placement task. Comparing across tasks, paired-*t* tests indicated that the WTA effect was stronger on the more difficult MLK scale than on the Einstein scale ($t(290) = -3.66$, $p < .001$). [4]

Correlation Analyses: Positive illusions and miscalibration [5]

Table 2 shows that the Pearson correlations between variables are small, which indicates good discriminant validity for these indices. First, the two calibration scales were positively correlated, although this relationship was modest, $r = .37$, $p < .01$. No other correlation involving the calibration scales reached significance. Second, all the other indices were positively correlated with each other, except the calibration self-placement specific to the Einstein scale, whose correlation with self-efficacy did not reach significance.

Discussion

First, our results suggest that the interval production task has internal validity in that the data exhibits some test-retest consistency (at an interval of two weeks) in miscalibration. Second, the miscalibration score obtained from the interval production task is not correlated to the other measures, although the measures of positive illusions are substantially correlated amongst themselves.

Importantly, our results suggest that judgmental overconfidence in the form of miscalibration measured on the interval production task is not a "positive illusion" in the sense of Taylor and Brown (1988). The first reason is that when asked to evaluate their performance on the calibration task, participants generally think they have done *worse* than average (WTA effect). Second, task difficulty affects judgmental overconfidence (overprecision) and self-placement differently. Thus we find the most judgmental overconfidence (overprecision), the greatest WTA (underplacement) on difficult tasks. The third reason is that the miscalibration scores are independent of dispositional optimism, self-efficacy and WTA effects in self-evaluated performance on the calibration task, whereas these latter characteristics appear to be intercorrelated with each other. Moreover, the fact that the pattern of correlations held whatever the samples of students suggests that the lack of relationship between overconfidence and positive illusions does not depend on the level of participants' academic abilities.

These results are consistent with those of Oberlechner and Osler (2008), who found that miscalibration in professional traders' currency exchange rate predictions (measured using an interval production task) is not correlated with the tendency to consider oneself better than others at trading. However, it is of course possible that our failure to obtain correlations between our measure of judgmental overconfidence and our other measures may simply have been due to our failure to include the appropriate measures of positive illusions. It is possible however, that some other appropriate measures of positive illusions might correlate with judgmental overconfidence. For this reason, we conducted a replication and extension of Study 1 using a modified procedure and additional measures.

Study 2: Replication and Extension: General BTA Measures, Unrealistic Optimism, Locus of Control and Perception of Societal Risks

The second study extends the first one in several ways. First, we varied the order of presentation of the calibration tasks to control for practice effects due to participants simply finding the first task more difficult. Second, we assessed social comparison beliefs about one's general abilities compared to others. We used tasks that are likely to induce classic better-than-average effects, unlike our calibration self-placement task which induced a worse-than-average effect in Study 1. Third, we included an index of unrealistic (comparative) optimism based on Weinstein's (1980) items, which assess an individual's perception of his or her chances *relative to others* of obtaining positive outcomes and avoiding negative ones. Fourth, as an alternative index of illusion of control to self-efficacy, we included a locus of control scale to assess control beliefs. This scale has been shown by Aspinwall and Taylor (1992) to predict social adjustment and self-reported health symptoms.

Finally, we included an index of perceptions of societal risks (Satterfield, Mertz, & Slovic, 2004) to assess optimism/pessimism about risks independently of social comparisons of one's chances relative to others of experiencing positive or negative events. This scale has the advantage of making a conceptual distinction between personal and social optimism. An optimistic person may be expected to perceive small chances of harm both for himself (compared to others) and for society in general, leading to a prediction of a correlation between measures of unrealistic optimism and perceptions of societal risk. However, a person driven by self-enhancement motives to make advantageous comparisons of himself to others may perceive small chances of harm for himself (compared to others) but not for society in general, leading to no correlation between these factors.

Method

Participants and Procedure

A sample of 221 French undergraduate students (79 2nd year Psychology students from the University of Toulouse-II, 35 1st year Law students from the University of Toulouse-I, 43 1st year students from the National Polytechnic Institute in Toulouse, and 64 1st year students from the University Technical Institute in Toulouse) [6] completed two questionnaires consisting of the same scales as in Study 1 as well as new ones. One questionnaire comprised the MLK scale (K-R 20 = .59) followed by its performance self-placement task (reliability = .68), the LOT-R (Cronbach α = .75), and two additional scales: an unrealistic optimism scale (adapted from Weinstein, 1980) and general academic self-placement questions. The other questionnaire consisted of the Einstein scale (K-R 20 = .60) followed by its specific performance self-placement task (reliability = .52), the self-efficacy scale (Cronbach α = .76) as well as two other new scales: a Locus of Control scale (Dubois, 1985) and a general risk perception scale (adapted from Finucane, Slovic, Mertz, Flynn, & Satterfield, 2000). Two weeks separated the completion of the questionnaires. To test for practice effects on the calibration tasks, half of the participants completed the questionnaire with the MLK scale first whereas the other half began by the questionnaire containing the Einstein scale.

New Instruments

Unrealistic optimism. For the purpose of the present study, Weinstein's (1980) original comparative rating form was reduced and adapted for the French population. This scale is designed to capture to what extent people believe that they are more likely than others to experience positive events and less likely to experience negative events. Among the original 42 events, we retained 8 positive (e.g., living past 80) and 8 negative events (e.g., developing cancer), which were then intermixed. Participants were asked to indicate what the chances were that the following events would happen to them compared to other students of their university of the same sex and age. The ratings were made on 7-point scales (1 = *less chances for me than for others*, 4 = *same chances for me as for others* and 7 = *more chances for me than for others*). The 8 negative items were reverse coded before scoring. One item that considerably lowered the internal consistency of the data was removed (Cronbach α = .54).

Academic self-placement (aSP). To examine whether the WTA effect was specific or not to the calibration tasks, participants were also asked to make comparative evaluations on 6 items referring to a more general and familiar domain, that is their academic work. They thus had to indicate to what extent they thought their attention and learning abilities, exam success, general knowledge, professional plans, work strategy and ability to take notes were worse, the same or better than those of most other students of their university (Cronbach α = .68). All ratings were made on 11-point scales (0 = *worse that most other students*, 5 = *the same as other students*, 10 = *better than other students*).

Academic locus of control (aLOC). The French Locus of Control (LOC) scale of Dubois (1985), which is largely drawn from Rotter's (1966) scale, was used to assess the participants' perceived control. From the original scale, which consisted of 42 items (14 fillers, 14 items referring to social relationships and 14 referring to the academic work), we only retained items concerning the academic domain (e.g., Academic success is directly linked to my efforts). For each statement, participants had to indicate their level of agreement (on 5-point scales ranging from 1 = *strongly disagree* to 5 = *strongly agree*). A high score indicates an internal locus of control. Two items were removed since they considerably lowered the internal consistency of the data (Cronbach α =.68).

Societal risk perception. It is possible to be more optimistic or pessimistic about risks without considering oneself to be more or less vulnerable than others. For example, Satterfield et al. (2004) find that white males report uniformly low perceptions of environmental health risks compared to black males and to women. This kind of optimism or pessimism may be thought of as reflecting a difference in the base-rate probabilities of a disaster occurring (species/group optimism), regardless of whether the self is more or less likely than others in one's reference group to be afflicted (unrealistic/comparative optimism). We therefore created a 14-item scale, which was adapted from the general risk perception scale developed by Finucane et al. (2000) to assess people's perception of the risk for the whole population to experience negative events. It consisted of hazards (e.g., being a victim of an air crash) for which participants were asked to indicate the level of risk for the French population as a whole (Cronbach α = .76). All ratings were made on 4-point scales (1 = *almost no risk*, 2 = *slight risk*, 3 = *moderate risk* and 4 = *high risk*).

Results and Discussion

Miscalibration

The average proportion of surprises was higher for the MLK scale (*M* = 74.57%, *SD* = 19.55) than for the Einstein scale (*M* = 64.39%, *SD* = 21.07), t (220) = 7.25, $p < .0001$ (Cohen's $d = .50$), whatever the order of presentation used [7]. These findings replicate those of our previous study and indicate that the MLK task is more difficult than the Einstein task (see Table 1). They also suggest that the lower proportion of surprises for the Einstein scale is not due to a calibration learning effect.

In addition, we also calculated each participant's interval width score by taking each item, ranking the width of each respondent's interval, and summing the ranks across participants. As expected, a tendency to use narrow interval widths on the MLK and Einstein scales predicted miscalibration on the scale ($r = .73$, $p < .01$ and $r = .62$, $p < .01$ respectively). More importantly, interval width on the MLK scale predicted miscalibration on the Einstein scale ($r = .47$, $p < .01$) and vice versa ($r = .45$, $p < .01$). This is consistent with the view that miscalibration on the interval production task is due to an individual's general tendency to use overly narrow intervals (Soll & Klayman, 2004).

General and Specific Self-placement: Evaluations of Achievement on Academic vs. Calibration Tasks

The calibration self-placement (cSP) index showed a WTA effect which was significant for both the MLK scale (*M* = -1.30, *SD* = 1.84, t (216) = -10.51, $p < .0001$) and the Einstein scale (*M* = -1.24, *SD* = 1.63, t (217) = -11.12, $p < .0001$). Unlike the previous study, there was no difference in the present study in the strength of this effect in the two scales (t (214) = .52, $p = .60$). Participants' academic self-placements however did not significantly differ from the midpoint of 5 (*M* = 4.98, *SD* = .94). Contrary to their self-placements on the calibration tasks, they thus judged themselves on the whole as no better nor worse than other students on more usual academic tasks.

Correlation Analyses: Positive illusions and miscalibration

Results obtained with scales shared by both studies will first be described, then followed by the results taking into account the new variables used in the present study. As seen in Table 3, most of the correlation patterns obtained in Study 1 were replicated. First, the calibration scores were positively correlated ($r = .47$, $p < .01$) showing moderate test-retest consistency, but were not significantly associated with either self-efficacy or dispositional optimism (LOT-R) scores. Second, the self-placement scores specific to the calibration tasks (cSP) were again positively correlated ($r = .60$, $p < .01$) and were also significantly positively associated with the LOT-R scores ($r = .21$, $p < .01$ for cSP specific to the MLK scale and $r = .14$, $p < .05$ for cSP specific to the Einstein scale). Third, the positive relationship between the LOT-R and self-efficacy scores was also replicated ($r = .49$, $p < .01$). However, some differences appeared. Contrary to Study 1, both calibration scores were negatively related to cSP scores specific to the MLK scale ($r = -.24$, $p < .01$ for the MLK scale and $r = -.22$, $p < .01$ for the Einstein scale) as well as to cSP scores specific to the Einstein scale ($r = -.13$, $p < .07$ for the MLK scale and $r = -.19$, $p < .01$ for the Einstein scale). These results suggest that the more participants were miscalibrated, the more they reported that they had performed worse than most other students. In addition, the correlations between self-efficacy and self-placement slightly differed between studies. In Study 1, the positive correlation between self-efficacy and self-placement reached significance for cSP scores specific to the MLK scale. In Study 2 on the other hand, we observed a significant positive correlation with cSP scores specific to the Einstein scale ($r = .14$, $p < .05$).

Analysis revealed that none of the new scales - academic self-placement, unrealistic optimism and aLOC - were significantly correlated with the miscalibration scores. In contrast, all measures were positively correlated with each other, except the aLOC score, for which correlations with unrealistic optimism and self-efficacy scores did not reach significance. Unrealistic optimism and the academic self-placement (aSP) scores were both positively correlated with all the other indices, that is with LOT-R, self-efficacy, and the self-placement scores specific to the calibration tasks (cSP). The correlation pattern for the aLOC score was quite similar except that it did not correlate with the MLK cSP score and correlated negatively with the Einstein cSP score.

Finally, the societal risk perception score showed negative correlations with the cSP score specific to the Einstein scale ($r = -.20$, $p < .01$), unrealistic optimism scores ($r = -.17$, $p < .05$) and dispositional optimism scores ($r = -.15$, $p < .05$). Interestingly, these last two correlations suggest that those who are optimistic about societal risks are actually pessimistic about their chances (relative to others) of escaping harm. These negative correlations underscore the importance of distinguishing optimism about societal risk from optimism about personal risks. They suggest that the optimism scores reflect self-enhancement motives (the belief that one will fare better than others) rather than a generalized Panglossian belief that all will turn out for the best for everyone (including oneself). It seems that the lack of self-reference in perceptions of societal risks makes these beliefs independent of positive illusions motivated by self-enhancement.

Finally, we constructed a composite “positive illusions” score by summing locus of control, dispositional optimism, academic self-placement, self-efficacy and unrealistic optimism scores. The correlations of the composite positive illusion measure with both the MLK and Einstein calibration scores were both insignificant (r =.01 & r =.08, $N = 221$, both *ns*), confirming that there is no overall relation between positive illusions and miscalibration on the interval production task.

Conclusions to Study 2

In sum, our second study confirms that judgmental overconfidence as assessed through miscalibration on the interval production task is independent of positive illusions such as the better-than-average effect, optimism and beliefs in high personal control. While self-related beliefs such as the better-than-average effect, optimism, unrealistic optimism and beliefs in high personal control share considerable variance, they are either independent of or negatively correlated with perceptions of societal risks. These results suggest that overconfidence may take up to three forms. The first form is judgmental overconfidence as measured by the interval production task. The second form involves self-enhancement biases that inflate perceptions of the self's capacity to influence events and chances of doing better than others. The third form concerns optimism with respect to societal risks, which do not measure the self's chances compared to others of avoiding catastrophes.

Study 3: Do Positive Illusions Predict Overconfidence on Probability Evaluation Tasks?

Although Studies 1 and 2 have shown that positive illusions do not predict overconfidence as measured by confidence intervals, it is possible that they will predict confidence as measured by probability evaluation. For example, Moore and Healy (2008) suggest that the probability evaluation technique conflates overprecision and overestimation, which may be influenced by positive illusions about the self. Self-enhancement biases may lead to inflation of miscalibration scores on the probability evaluation task, as this task explicitly requires participants to express a subjective confidence level, and thus express something about their own feeling of confidence. Interval production tasks focus participants' attention on properties of the external world (in particular the range of a distribution of a set), and thus may not activate self-related motivational biases that affect the way participants set confidence intervals. We therefore assess whether there is a correlation between our 'positive illusion' indicators and miscalibration measured on both the interval production and probability evaluation techniques.

This 3rd study thus enables a test of whether the same individuals who are the most overconfident on interval production are also the most overconfident on probability evaluation. While Budescu and Du (2007) find considerable consistency across methods using stock predictions using interval production and probability evaluation techniques, to our knowledge, no study has yet addressed the question of whether the same individuals who are the most overconfident on interval production are also the most overconfident on probability evaluation. We address this question in our final study, using general knowledge questions, a domain in which interval production techniques reliably yield overall higher overconfidence scores than probability evaluation techniques.

Method

Participants and Procedure

Ninety seven French undergraduate psychology students from the University of Toulouse-II completed two questionnaires, administered two weeks apart at the end of their classes. For subsequent analyses, four participants providing missing data on the interval task were excluded from the sample. The paired questionnaires were designed to enable comparisons between performance on an interval production (IP) task and a probability evaluation (PE) task. We use the same questions on the same group of participants in a counterbalanced design. The 'positive illusion' measures used in Study 2 were also included.

The first questionnaire comprised of unrealistic optimism (one item was removed to reach a modest but acceptable reliability; Cronbach $\alpha = .54$) and dispositional optimism (Cronbach $\alpha = .82$) measures, one of the calibration tasks (either IP or PE), a self-placement task specific to participants' performance on the calibration task, and an academic self-placement task (Cronbach $\alpha = .74$). The second questionnaire - its counterpart - included self-efficacy (Cronbach $\alpha = .88$) and locus of control (Cronbach $\alpha = .67$) scales, the second calibration task (either PE or IP) and a societal risk measure (Cronbach $\alpha = .81$).

Order of presentation and type of content was varied to control for order and practice effects. Full Einstein and MLK versions comprising 20 questions (See Appendices 1 and 2) were alternatively presented to participants, who either began with an IP or a PE version of the scales. Thus, all in all, 50 participants completed an MLK IP scale coupled with an Einstein PE scale and the remaining 43 participants completed an Einstein IP scale coupled with a MLK PE scale. Furthermore, these two groups were split into two sub-groups with one group completing the IP scale first (n = 20 & n = 20) and the other group completing the PE scale first (n = 30 & n = 23). As the order of presentation had no significant effect, this factor was ignored in the analyses presented below.

New Instruments

Calibration tasks. We used the MLK and Einstein scales to assess participant's ability to produce well calibrated responses and we added 10 general knowledge questions to each scale (see Appendices 1 & 2). In addition, a new calibration task was designed to complete the lengthier version of the MLK and Einstein interval production (IP) tasks. This new calibration task consisted of a content-equivalent Probability Evaluation (PE) task. For the PE task, we did not require participants to produce an upper and lower limit in response to each question. Instead, respondents were presented with a possible answer (e.g.: "Was Einstein older than 69 when he died?") and asked to (a) evaluate it by indicating whether they thought it was true or false ("*yes*" or "*no*") and (b) estimate the probability that their choice of response would be correct on a 6-point scale ranging from "uncertain" (*50%*) to "certain" (*100%*). Each possible answer was generated at random (using a function provided in Excel), within a range defined by the mean upper and lower limits produced by Study 2 participants**.** The internal consistencies for the IP scales was lower (MLK: α = .67, Einstein: .58) than those of the PE scales (MLK: α = .88, Einstein: .85). For both scales, we computed an *overconfidence rate* by subtracting the percentage of correct responses from the percentage of *expected hits*. For IP scales, the percentage of *expected hits* had been predefined by the 90% probability interval (i.e. a theoretical hit rate of 90%, given 18 expected correct from a 20-item scale).

For PE scales, we calculated an expected hit rate by calculating the average of their confidence estimates. This was defined as the percentage of correct responses expected, given the probabilities assigned by each participant. Each participant's expected hit rate was calculated by adding the probabilities assigned by the participant and dividing this sum by the number of questions posed (19 questions for the Einstein version of the PE scale and 20 questions for remaining scales, one of the supplementary questions in the Einstein PE scale having been eliminated because of the ambiguity of the word "billion").[8] For example, on the 20-item scale, if participant A chose a 60% confidence level for her responses to ten items, and a 80% confidence level for the other ten items, we would score this as an overall expected hit rate of 70% (14 hits expected out of 20). If participant B produced 5 confidence estimates of 60%, 5 judgments of 80% and 10 judgments of 90%, we calculated an overall expected hit rate of 80% (16 hits expected out of 20), and so on. Thus if participants A and B both got twelve responses right on the multiple choice items, participant A would be scored as having two misses (14-12), and participant B as having four (16-12).

Finally, interval width was also computed as before for the IP scales by calculating the difference between the upper and lower limit of each interval produced and attributing a rank to the difference for each variable. We expected narrow interval width to be correlated with high degrees of confidence on the probability evaluation task, as both indicate high expected precision for her judgment by a participant.

Results and Discussion

Miscalibration

Given our previous results, we expected the MLK scale to be more difficult than the Einstein scale. This difference was marginally significant for the PE scales, as there was a slight tendency for overconfidence rates to be higher for the MLK PE scale (M = 15.05%, SD = 11.98, n = 43) than for the Einstein PE scale (M = 10.62%, SD = 17.10, n = 50), t (91) = -1.422, $p < .10$, one-tailed, Cohen's d = -.30 (see Table 1). The overconfidence rate was also higher for the MLK IP scale (M = 69.50%, SD = 14.08, n = 50) than for the Einstein IP scale (M = 66.39%, SD = 13.20, n = 43), but this effect did not reach significance (t (91) = 1.091, p = .13, one-tailed, Cohen's d = 0.23).

Overall, the average overconfidence rate was much higher for the IP scales (M = 68.06%, SD = 13.69, N = 93) than for the PE scales (M = 12.67%, SD = 15.04, N = 93), t (92) = -27.665, $p < .0001$. The size of this effect was very large (Cohen's d = -5.77), thus replicating earlier findings comparing these formats (e.g. Klayman et al., 1998).

General and Specific Self-placement: Evaluations of Achievement on Academic vs. Calibration Tasks

Overall participants tended to evaluate their academic performances as slightly above average as academic self-placement significantly differed from the midpoint of 5 (M = 5.23, SD = 1.07, t (92) = 2.049, $p < .05$). However, most participants reported for both calibration self-placement scales that they had performed *worse* than most other students, thus producing a significant worse-than-average (WTA) effect (M = -1.09, SD = 1.41, t (92) = -7.420, $p < .0001$ for the IP scales and M = -.84, SD = 1.26, t (92) = -6.408, $p < .0001$ for the PE scales). Participants also generally tended to express a stronger WTA effect after having completed the IP scales than after having completed the PE scales (t (92) = -1.848, $p < .05$, one-tailed). This pattern is again consistent with the dissociation predicted by Larrick et al. (2007): the task that induced the highest level of overconfidence in judgement also led to a stronger level of underconfidence in self (WTA).

Correlation Analyses: Positive illusions and miscalibration on the IP and PE tasks

Table 4 shows the correlation matrix obtained for this sample. Overall, we obtained a similar pattern of correlations to Study 2, with numerous correlations between the different "positive illusion" measures. The calibration self-placement measures correlated positively between themselves ($r = .54$, $p < .01$) as well as with unrealistic optimism, dispositional optimism and academic self-placement ($p < .05$). Dispositional optimism correlated positively with unrealistic optimism ($r = .40$, $p < .01$) and academic self-placement ($r = .24$, $p < .05$), unrealistic optimism correlated with academic self-placement ($r = .46$, $p < .01$) and locus of control ($r = .28$, $p < .01$). No significant correlations were found between perceived societal risk and the "positive illusion" measures.

We also found significant correlations between some "positive illusion" and calibration measures. There was a negative correlation between academic self-placement and the overconfidence rate for the IP scale ($r = -.27$, $p < .01$). This correlation goes against the direction that would be predicted if the better-than-average (BTA) effect would measure the same construct as judgmental overconfidence. We also found a positive correlation between dispositional optimism and the expected hit rate for the PE scale ($r = .23$, $p < .05$) but no correlation between dispositional optimism and overconfidence rates. This suggests that an optimistic outlook tends to increase participants' tendency to express confidence in their responses without increasing the number of errors they make.

The IP and PE calibration scores were not significantly correlated ($r = .10$, *ns*), suggesting that these two scales measure distinct constructs. However, IP interval width scores correlated negatively with expected hit rate for the PE task ($r = -.29$, $p < .01$). This means that those participants who tended to produce smaller intervals on the IP task also expressed greater confidence in their responses during the PE task. This result suggests that the IP and PE tasks share significant variance and thus measure a common construct that would seem to correspond to "overprecision" (Moore & Healy, 2008).

Do Positive Illusions Predict Probability Evaluations Independently of Overprecision?

Various measures of positive illusions correlated positively with the expected hit rate for the PE task. These included the calibration self-placement score specific to the PE task ($r = .31$, $p < .01$) and unrealistic optimism and dispositional optimism ($r = .25$, $p < .05$. and $r = .23$, $p < .05$). This raises the possibility that while positive illusions do not influence interval width estimates, they may influence probability evaluations by leading participants to inflate their expressed confidence levels. As in Study 2, we tested this hypothesis more specifically through a composite "positive illusions" score. Consistent with our analysis, this score correlated significantly with the PE overconfidence rate ($r = .21$, $p < .05$) but not the IP overconfidence rate ($r = -.07$, *ns*).

In addition, results of an equation in which expected hits on PE was regressed on positive illusions and IP interval width revealed that the two variables contributed to predicting significant and independent variance in the expected hits score ($F(2, 90) = 10.204$, $p < .0001$, Adjusted $R^2 = .167$, Interval width: $\beta = -.30$ $p < .005$, Positive illusions: $\beta = .31$, $p < .005$). This means that the PE estimates are in part predicted by overprecision but also by positive illusions.

Conclusions to Study 3

In this 3rd study, we show that while disposition to positive illusions does not predict miscalibration on the interval production task, it does predict miscalibration on the probability evaluation task. This difference may emerge due to the explicit demand to express confidence through choosing probability levels in the evaluation task, as opposed to the interval production task, which taps confidence levels in a more implicit manner. However, it is striking that positive illusions do not predict degree of overconfidence on the interval production task, despite the higher overall level of overconfidence on this task (for further discussion of the relation between self-esteem and format dependence, see Juslin, Winman, & Hanson, 2007; Winman, Hanson, & Juslin, 2004;).

Miscalibration scores on the interval production and probability evaluation tasks are only weakly (and non-significantly) correlated. This suggests that these two methods of assessing calibration tap different judgment processes (Teigen & Jorgensen, 2005). The finding that the interval width on the IP task predicts theoretical hit rate on the PE task however, shows that overprecision in the form of overly narrow confidence intervals predicts miscalibration on both the IP and PE tasks. This suggests that the two tasks measure a common underlying component of judgmental overconfidence understood as overprecision - overestimation of the precision of one's judgments. Last, we find that positive illusions contribute independently of overprecision to inflating confidence judgments on the PE task. This suggests that there are two different sources of miscalibration on the probability evaluation task – overprecision and positive illusions.

General Discussion

Our research has produced four major sets of findings. First, people show systematic individual differences in their tendency to experience the positive illusions identified by Taylor and Brown (1988), namely the better-than-average effect, illusion of control, and unrealistic (comparative) optimism. This is indexed by the numerous intercorrelations between our measures of disposition to experience positive illusions (Calibration self-placement, academic self-placement or BTA effect, self-efficacy, academic locus of control, dispositional optimism, comparative optimism). In addition, the self-enhancing nature of positive illusions is underscored by the finding that optimism in perception of personal risks is *not* correlated with optimism in perception of societal risks.

Second, miscalibration on the interval production task can be regarded as a stable personal trait. Studies 1 and 2 both showed that overconfidence on the interval production task has test-retest stability across two-week intervals, using different sets of questions, and thus can be considered to be a stable individual difference characteristic. Although no significant correlations were observed between miscalibration scores assessed by the IP and PE methods (collected at intervals of at least one week), other measures were significantly correlated across the two tasks (e.g. interval width and chosen confidence levels), suggesting the existence of a stable underlying construct of overprecision.

Third, all three studies showed that disposition to experience positive illusions does not predict miscalibration on the interval production task. This kind of overprecision corresponds to the model of judgmental overconfidence proposed by Biais et al. (2005), who argued that underestimation of population variability may lead to overestimation of the diagnosticity of a market signal (or, in the language of economics, underestimation of conditional uncertainty). In line with this analysis, Biais et al. found that participants who showed high overconfidence on an interval production task were most likely to lose money in an experimental financial market in situations of high signal ambiguity, where the stated market price had low diagnosticity and was least likely to reflect the true value of the asset being traded. This kind of overprecision may result from cognitive biases, as proposed by Juslin, Winman & Hanson's (2007) naïve sampling model of overconfidence which proposes that the overconfidence observed using the interval production technique is due to failure to a) recognize that the naïve sample is not representative of the target population, and b) correct for this, leading to systematic underestimation of the true population variance (see also Moore and Healy (2008)'s analysis of overprecision in terms of overly restricted subjective probability distributions).

Finally, Study 3 shows that disposition to positive illusions does predict miscalibration on the probability evaluation task. This suggests that miscalibration on the interval production and probability evaluation tasks have different sources, and in particular that miscalibration on the probability evaluation task is influenced by motivational biases. Moreover, the finding that the width of intervals independently predicts miscalibration on the probability evaluation task is in line with Moore and Healy's proposal that the latter task confounds overprecision and overestimation of the quality of one's performance (2008, p. 503). Thus, our findings suggest that the interval production task measures judgmental overconfidence (Moore and Healy's overprecision) in a "pure" way, whereas the probability evaluation task measures both judgmental overconfidence and tendency to experience positive illusions. It seems important to measure self-enhancement bias with our positive illusion measures, as Larrick et al. (2007) failed to find any relation between various personality scales (e.g., self-esteem, need for uniqueness, narcissism, defensive pessimism, need for cognition and need for closure…) and miscalibration measured using a probability evaluation procedure.

The Behavioural Consequences of Judgmental Overconfidence vs. Positive Illusions

The distinction between judgmental overconfidence in the form of overprecision and positive illusions about the self has important implications for models of how psychological biases might influence behaviour (e.g. Odean, 1998). Both psychologists (Taylor & Brown, 1988) and economists (Bénabou & Tirole, 2003) have argued that positive illusions such as inflated self-esteem and optimism may lead individuals to attain better outcomes, for example through motivating them to work harder and persist when the going gets tough. Such positive illusions may indeed act as happily self-fulfilling prophecies: for example, Aspinwall and Taylor (1992) report longitudinal field data that students who have high self-esteem, are optimistic, and have high beliefs in their control over their life and desire for control are more likely to report being well-adjusted to college three months later. Those with high self-esteem, high desire for control and internal locus of control are more likely to show greater motivation, and to obtain high grade point averages 21 months later, controlling for intelligence (SAT scores). In similar vein, Murray and Holmes (1997) show that relationship optimism and perceptions of control predict relationship survival a year later even controlling for scores of initial relationship quality (satisfaction, trust, etc.).

While positive illusions may sometimes make people happy, they may also make them poor. Research to hand suggests that positive illusions can have negative effects on economic outcomes, through encouraging people to engage and persist in unprofitable activities (Hilton, 2007). For example, Gibson and Sanbonmatsu (2004) show that in a domain such as gambling where increased effort cannot make a difference, optimism can lead to persistence after failure and thus cause yet greater monetary losses. In an experimental market setting, Camerer and Lovallo (1999) have found that an experimental manipulation leading participants to overestimate their chances of success relative to others leads to financial losses due to excess market entry (other overconfident players enter the same market, leading to too many players chasing too little business). It seems plausible that positive illusions incite people to become more active in markets. An internet survey by Glaser and Weber (2007) for instance, found that better-than-average scores predicted trading activity on the stock market, whereas miscalibration measured by the interval production method did not.

In contrast to positive illusions, judgmental overconfidence, as measured by the interval production task has not been found to predict greater market activity. It nevertheless predicts poorer economic performance on both experimental (Biais et al., 2005; Deaves et al., 2005) and real financial markets (Bonnefon, Hilton, & Molian, 2005; Oberlechner & Osler, 2008). These findings suggest that it pays to have accurate beliefs in competitive markets, where perspicacity and accuracy in judgment may count for more than motivation and persistence. One possible mechanism to explain this relationship is that realism may facilitate successful economic performance through enabling an individual to accurately identify and invest in areas of competitive advantage. In this vein, Försterling and Morgenstern (2002) show that giving participants accurate, rather than self-enhancing, feedback for task performance in a learning phase enables them to specialize in tasks that they are good at in the subsequent test phase, and thus improve their chances of gaining a financial reward for superior performance.

Future work will therefore do well to specify when, why and how the different forms of overconfidence – whether positive illusion or judgmental overconfidence - will enhance or impede personal well-being and/or economic performance. It should measure these hypothesised constructs in appropriate ways, and where possible provide empirical tests of the causal paths implied by theoretical hypotheses.

References

Alicke, M. & Govorlin, O. (2005). The better-than-average effect. In M.D. Alicke, D.A. Dunning & J. Krueger (Eds.). *The self in social judgment* (pp. 85-106). New York: The Psychology Press.

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 294-305). Cambridge: Cambridge University Press.

Aspinwall, L. G., & Taylor, S. E. (1992). Modeling cognitive adaptation: A longitudinal investigation of the impact of individual differences and coping on college adjustment and performance. *Journal of Personality and Social Psychology, 63,* 989-1003.

Barber, B., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance, 55,* 773-806.

Barber, B., & Odean, T. (2001). Boys will be boys: Gender, overconfidence and common stocks investments. *Quarterly Journal of Economics, 116,* 261-292.

Barber, B., & Odean, T. (2002). Online investors: Do the slow die first? *Review of Financial Studies, 15,* 455-487.

Bénabou, R., & Tirole, J. (2003). Self-knowledge and self-regulation: An economic approach. In I. Brocas & J. D. Carrillo (Eds.), *The psychology of economic decisions: Volume 1: Rationality and well-being.* Oxford: Oxford University Press.

Biais, B., Hilton, D., Pouget, S., & Mazurier, K. (2005). Judgmental overconfidence, self-monitoring and trading performance in an experimental financial market. *Review of Economic Studies, 72,* 297-312.

Biais, B., & Weber, M. (2005). Hindsight bias and trading performance. Paper presented at the 3rd Toulouse Workshop on Judgment and Decision Making and Experimental Economics.

Bonnefon, J-F., Hilton, D.J., & Molian, D.J. (2005). A portrait of the unsuccessful entrepreneur as miscalibrated thinker: Profit decreasing ventures are run by overconfident owners. Unpublished manuscript: University of Toulouse.

Bruine de Bruin, W., Parker, A.M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 41*, 938-956.

Budescu, D., & Du, N. (2007). The coherence and consistency of investors' probability judgments. *Management Science, 53,* 1731-1744.

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review, 89,* 306-318.

Deaves, R., Lüders, E., & Luo, G.Y. (2005). An experimental test of overconfidence and gender on trading activity. Working paper. DeGoote School of Business, McMaster University.

Dubois, N. (1985). Une échelle française de "locus of control". [A French LOC scale]. *Revue de Psychologie Appliquée, 35,* 215-233.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519-527.

Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2000). Gender, race, and perceived risk: the "white male" effect. *Health, Risk & Society, 2,* 159-172.

Försterling, F., & Morgenstern, M. (2002). Accuracy of self-assessment and task performance: Does it pay to know the truth? *Journal of Educational Psychology*, *94,* 576-585.

Gibson, B., & Sanbonmatsu, D.M. (2004). Optimism, pessimism and gambling: The downside of optimism. *Personality and Social Psychology Bulletin, 30,* 149-160.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Glaser, M., Langer, T., & Weber, M. (2005). Overconfidence of professionals and lay men : Individual differences within and between tasks. Working Paper : University of Mannheim.

Glaser, M., Nöth, M., & Weber, M. (2004). Behavioral Finance. In D.J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 527-54). Malden, MA, US: Blackwell Publishing.

Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *Geneva Risk and Insurance Review, 32,* 1-36.

Griffin, D., & Brenner, L. (2004). Probability judgment calibration. In D.J. Koehler N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 177-198). Malden, MA, US: Blackwell Publishing.

Hilton, D. J. (2007). Overconfidence, trading and entrepreneurship: Cognitive and cultural processes in risk-taking. In R. Topol & B. Walliser (Eds.). *Contributions to economic analysis: Cognitive economics: New Trends* (pp. 225-235). Elsevier Press.

Hoorens, V., & Buunk, B. (1993). Social comparison of health risks : Locus of control, the person-positivity bias and unrealistic optimism. *Journal of Applied Social Psychology, 23,* 291-302.

Jonsson, A-C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain and gender. *Personality and Individual Differences*, *34*, 559-574.

Juslin, P., Winman, A., & Hansson, P. (2007). The *Naïve* Intuitive Statistician: A Naïve Sampling Model of Intuitive Confidence Intervals. *Psychological Review, 114,* 678-803.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Klayman, J., & Burt, R. S. (1998). *Individual differences in confidence and experiences in social networks.* (Unpublished manuscript, Center for Decision Research, Graduate School of Business, University of Chicago.)

Klayman, J., Soll, J. B., Gonzales-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*, 216-247.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77,* 221-232.

Kullback, S. (1959). *Information Theory and Statistics*. Wiley : New York.

Langer, E. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32,* 311-328.

Langer, E., & 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*, *32*, 951-955.

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

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Cambridge, England: Cambridge University Press.

Macko, A., & Tyska, T. (2006). Entrepreneurship and risk taking. Paper presented at the IAREP/SABE conference, Paris, July 2006.

McKenna, F. (1993). It won't happen to me: Unrealistic optimism of illusion of control? *British Journal of Psychology, 84,* 39-50.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502-17.

Murray, S. L., & Holmes, J. G. (1997). A leap of faith ? Positive illusions in romantic relationships. *Personality and Social Psychology Bulletin*, *23,* 586-604.

Oberlechner, T. & Osler, C. (2008). Overconfidence in currency markets. Unpublished manuscript. Webster University.

Odean, T. (1998). Volume, volatility, price and profit when all traders are above average. *Journal of Finance, 53*, 1887-1934.

Parker, A., & Fischhoff, B. (2002). Decision-Making Competence. *Advances in Consumer Research, 29,* 434-437.

Parker, A., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual differences approach. *Journal of Behavioral Decision-Making, 18,* 1-27.

Rotter, J. (1966). Generalized expectations for internal versus external control of reinforcement. *Psychological Monographs, 80,* 1-28.

Russo, J., & Schoemaker, P. J. H. (1992). Managing over-confidence. *Sloan Management Review*, *33*, 7-17.

Satterfield, T. A., Mertz, C. K., & Slovic, P. (2004). Discrimination, vulnerability and justice in the face of risk. *Risk Analysis, 24,* 115-129.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the life orientation test. *Journal of Personality and Social Psychology, 67*, 1063-1078.

Schwarzer, R., & Jerusalem, M. (1995). Self-efficacy measurement: Generalized Self-Efficacy Scale (GSES). In J. Weinman, S. Wright & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, UK: NFER-Nelson.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M.P. Zanna (Eds.). *Advances in Experimental Social Psychology, 29*, 209-269. New York: Academic Press.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30,* 290-314.

Stanovich, K. E., & West, R. F. (1998). Individual differences and rational thought. *Journal of Experimental Psychology: General, 127,* 161-188.

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

Taylor, S., & Brown, J. (1988). Illusion and well being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193-210.

Teigen, K. H., & Jörgensen, M. (2005). When 90% confidence intervals are only 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology, 19*, 455-475.

Weinstein, N. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*, 806-820.

Weinstein, N. (1982). Unrealistic optimism about susceptibility to health problems. *Journal of Behavioral Medicine, 5,* 441-460.

Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: How to cure overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1167-1175.

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy–informativeness tradeoff. *Journal of Experimental Psychology: General*, *124,* 424-432.

Notes

[1] Ratings for both Dispositional optimism and Self-efficacy were made on a seven-point scale (1 = *strongly disagree* ; 7 = *strongly agree*) for Psychology students but on a five-point scale (1 = *strongly disagree* ; 5 = *strongly agree*) for the two other populations (i.e., managerial economics students and *Grande Ecole* students in management). To control for this variation in the scaling, Z-scores were computed on both variables and then used for all the following analyses.

[2] In addition to the surprise rates, informativeness scores (which is indicative of the intervals' width) as well as accuracy scores (indicating to what extent the midpoints of the intervals are close to the true value) were also computed for both miscalibration scales following the procedure of Yaniv and Foster (1995). For all studies reported here, findings on these scores indicated that lower rates of surprises were due to wider intervals and/or midpoints closer to the true answer.

[3] Consistent with the hypothesis that higher academic ability is associated with lower miscalibration, the ESC students obtained better calibration scores on both scales than the other two groups. This difference did not result in different relationships between miscalibration scores and positive illusions. A Fisher's Z transformation of the correlation coefficients (Kullback, 1959) confirmed that the correlations were equal in all the samples.

[4] Comparing across students samples, only ESC students did not display the WTA effect but rather reported being neither better nor worse than average.

[5] For all the present studies, correlation analyses led to exactly the same results as regression analyses (where the surprise index was regressed on its corresponding calibration self-placement and all positive illusion variables measured in each Study). Consequently, only results of correlation analyses (two-tailed) were reported.

6 Contrary to Study 1, the different students samples were not used to control for differences in academic abilities (which had no impact on correlations between overconfidence and positive illusions), but simply to get a diverse population, not limited to psychology students.

7 The average proportion of surprises was always higher for the MLK scale than for the Einstein scale, whether the MLK scale was completed first (M = 78.25%, SD = 14.89 for the MLK scale, M = 65.09%, SD = 22.19 for the Einstein scale, t (113) = 7.32, $p < .0001$, Cohen's d = .70) or after the Einstein scale (M = 70.65%, SD = 22.95 for the MLK scale, M = 63.64%, SD = 19.88 for the Einstein scale, t (106) = 3.27, $p < .001$, Cohen's d = .33). These findings.

8 "Is the number of hectares of forest in the world inferior to 2 billion?"

Acknowledgments

Our research on miscalibration owes much to discussions with Bruno Biais, Sébastien Pouget and Jim Shanteau. We thank the editor and three anonymous referees for the comments on an earlier draft of this paper, and Bernard Gaffié, Gayannée Kédia, Julia Kogan, Florence Loose, Stéphane Perrissol and David Stolin for help with the collection of data.

Table 1

Overconfidence Rates and Self-placement Score in each of the Three Studies

	Study 1		Study 2		Study 3	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*
Overconfidence Rates						
Production Interval MLK Task	78.14	15.13	74.57	19.55	69.50	14.08
Production Interval Einstein Task	62.41	19.82	64.39	21.07	66.39	13.20
Probability Evaluation MLK Task	--	--	--	--	15.05	11.98
Probability Evaluation Einstein Task	--	--	--	--	10.62	17.10
Self-placement on Calibration Tasks						
Production Interval MLK Task	-1.11	1.64	-1.30	1.84	-1.36	1.55
Production Interval Einstein Task	-.77	1.85	-1.24	1.63	-.77	1.17
Probability Evaluation MLK Task	--	--	--	--	-.58	1.16
Probability Evaluation Einstein Task	--	--	--	--	-1.06	1.31
Academic Self-placement	--	--	4.98	.94	5.17	1.04

Table 2

Pearson Correlation Matrix for Study 1

	1	2	3	4	5	6
1. Percentage of surprises on MLK scale	--					
2. Percentage of surprises on Einstein scale	.37**	--				
3. Calibration self-placement (cSP) specific to MLK scale	-.04	-.06	--			
4. Calibration self-placement (cSP) specific to Einstein scale	-.02	-.10	.56**	--		
5. Dispositional Optimism (LOT-R)	-.09	-.08	.23**	.22**	--	
6. Self-efficacy	-.04	.05	.20**	.10	.28**	--

Note. ** $p < .01$.

Table 3

Pearson Correlation Matrix for Study 2

	1	2	3	4	5	6	7	8	9	10	11	12
1. Percentage of surprises on MLK scale	--											
2. Interval width for MLK scale	-.73**	--										
3. Percentage of surprises on Einstein scale	.47**	-.51**	--									
4. Interval width for Einstein scale	-.45**	.62**	-.65**	--								
5. cSP for MLK scale	-.24**	.23**	-.22**	.13	--							
6. cSP for Einstein scale	-.13=	.11	-.19**	.04	.60**	--						
7. Academic self-placement	.06	-.02	.00	-.05	.31**	.25**	--					
8. Unrealistic optimism	.02	.08	.06	.00	.19**	.12	.36**	--				
9. Dispositional optimism (LOT-R)	-.02	.05	.10	-.06	.21**	.14*	.30**	.40**	--			
10. Self-efficacy	.04	.00	.10	-.04	.12	.14*	.33**	.30**	.49**	--		
11. Academic locus of control	-.06	.07	-.02	-.00	.00	-.14*	.15**	.08	.13=	.12	--	
12. Societal risk perception	.12	-.16*	.10	-.04	-.11	-.20**	-.09	-.17*	-.15*	-.12	-.01	--

Note. = $p < .07$. *$p < .05$. **$p < .01$., two-tailed.

Table 4

Pearson Correlation Matrix for Study 3

	1	2	3	4	5	6	7	8	9	10	11	12
1. Overconfidence rate for IP scale	--											
2. Interval width for IP scale	-.63**	--										
3. Overconfidence rate for PE scale	.10	-.20	--									
4. Theoretical hit rate for the PE scale	.16	-.29**	.66**	--								
5. cSP for IP scale	-.20	.12	.12	.13	--							
6. cSP for PE scale	-.14	.06	.15	.31**	.54**	--						
7. Academic self-placement	-.27**	.10	.14	.13	.27*	.37**	--					
8. Unrealistic optimism	-.10	-.02	.14	.25*	.34**	.45**	.46**	--				
9. Dispositional optimism (LOT-R)	.04	.02	.17	.23*	.36**	.32**	.24*	.40**	--			
10. Self-efficacy	.15	.03	-.00	.12	.13	.03	-.04	-.02	.36**	--		
11. Academic locus of control	-.04	-.11	.18	.15	.04	-.04	.18	.28**	.10	-.06	--	
12. Societal risk perception	.10	-.08	.17	-.05	-.08	-.03	-.07	-.10	-.14	.04	-.15	--

Note. $^{=}p < .07$. $*p < .05$. $**p < .01$. Two-tailed.

Appendix 1: Items used in "MLK" IP scale

	Low	High
Martin Luther King's age at death.		
Length of the Nile River (in miles).		
Number of countries that are members of OPEC.		
Number of books in the Old Testament.		
Weight of an empty Boeing 747 (kgs.).		
Year in which J.S. Bach was born.		
Gestation period (in days) of an Asian elephant.		
Diameter of the moon (in miles).		
Air distance from London to Tokyo.		
Deepest known point in the Oceans (in ft.).		
Population of Thailand		
Number of Landers in Germany		
Surface area of Switzerland		
Number of countries and territories in the world		
Number of provinces in Canada		
Number of countries which have designated French as their official language		
Population of New York		
Number of time zones in Asia		
Altitude of the highest summit of the Andes (in meters)		
Life expectancy in China in 2006		

Appendix 2: Items used in "Einstein" IP scale

	Low	High
Einstein's age at death		
Distance from earth to moon		
Number of countries that are members of NATO		
Number of French inhabitants in 2002		
Maximal length of a whale		
Year in which the Ariane rocket project began		
Year in which W. A. Mozart was born		
Height of the Eiffel Tower		
Surface of the Amazonian forest		
Altitude of Mont Blanc		
Number of countries and territories of the Asian continent		
Population of Japan		
Life expectancy in India in 2002		
Population density in Saudi Arabia (per km^2)		
Number of hectares of forest in the world		
Tourists in France in 2002		
Women head of state or government (either president or prime minister or chancellor) in 2006		
Population of Tokyo		
Number of languages spoken in the world		
Number of seas surrounding Asia		