

Overconfidence in absolute and relative performance: The regression hypothesis and Bayesian updating*

Daniela Grieco¹ & Robin M. Hogarth²

¹Università Bocconi, Milano, Italy

²ICREA & Universitat Pompeu Fabra, Barcelona, Spain

daniela.grieco@unibocconi.it

robin.hogarth@upf.edu

July 11, 2008

* The authors' names are listed alphabetically. They thank Carlos Trujillo, Julian Rode, and Enric Soria for help in conducting the experiments. They have also benefited from the helpful comments of Michele Bernasconi, Natalia Karelaia, Franco Malerba, Don Moore, Rosemarie Nagel, Mercè Roca, Joachim Vosgerau, and Lorenzo Zirulia. This research was financed partially by grants from the Spanish Ministerio de Educación y Ciencia SEJ2006-14098 (Hogarth). For correspondence, please contact Robin M. Hogarth at Universitat Pompeu Fabra, Department of Economics & Business, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain. Tel: + 34 93 542 2561.

Abstract

Studies have found that people are overconfident in estimation involving difficult tasks but underconfident in easy tasks. Conversely, they are overconfident in placing themselves in easy tasks but underconfident in hard tasks. Moore and Healy (2008) explain these findings by a regression hypothesis that we test in four experiments. The hypothesis successfully predicts estimation behavior. However, for placement (involving both easy and hard tasks) we find no overconfidence, regression effects due to low and high anchor points, and extreme underconfidence (when people choose between multiple alternatives). We further test a Bayesian model of reactions to signals about absolute performance when success depends on relative performance and find behavior consistent with this model. Finally, since placement judgments are important in many competitive settings, we stress the need for more research to identify their determinants.

JEL codes: D8, L0, M1

PsycINFO codes: 2340, 3040

Keywords: Overconfidence, estimation, placement, “better-than-average” effects.

Overconfidence is a central topic of research in judgment and decision making as well as behavioral economics (see, e.g., Plous, 1993; Camerer & Lovallo, 1999). In an important contribution, Moore and Healy (2008) recently clarified much of this extensive research by identifying different paradigms as well as the conceptual meaning of the term “overconfidence” itself. In particular, Moore and Healy distinguished three ways in which overconfidence is used.

First, people can be overconfident in estimating their ability to do something. For example, a person might overestimate his ability to run a Marathon within a certain time. Moore and Healy call this overestimation. Interestingly, overestimation is not universal. A robust finding is that people tend to overestimate their own skill on hard tasks but underestimate it on easy tasks (Burson, Larrick, & Klayman, 2006; Moore & Cain, 2007).

Second, a person might express overconfidence in ability relative to others; for example, the belief that one can run the Marathon faster than, say, 80% of a specific population. Moore and Healy call this overplacement. It is also referred to as the “better-than-average” effect whereby people judge their abilities in familiar domains such as driving as being superior to that of the “average” person (Svenson, 1981). At the same time, however, there is a tendency to judge oneself as below average in unfamiliar and, therefore, hard tasks such as juggling (Kruger, 1999). Also, as Hoelzl and Rustichini (2005) demonstrate, this type of overconfidence may be moderated when people are required to make incentive-compatible choices as opposed to expressing opinions.

Third, people can be overconfident when estimating future uncertainty; for example, when providing confidence intervals for forecasts of, say, sales that

subsequently turn out to be too narrow (see, e.g., Alpert & Raiffa, 1982; Klayman et al., 1999). Moore and Healy call this overprecision.

Although much needs to be done to achieve a better understanding of overprecision (Juslin, Winman, & Hansson, 2007), this paper concentrates on overestimation and overplacement and, in particular, considers the relation between them. This is an especially important consideration in competitive environments where success depends on rank in a population (i.e., how able relative to others) but where the decision maker may only have information about absolute level. As examples, consider entrepreneurs facing novel business opportunities, firms confronting research and development expenditures, the structure of many career choices, or even students considering which new courses to take.

Insert Figure 1 about here

Interestingly, Moore and Healy (2008) suggest that systematic errors in estimation and placement might arise from random errors in judgment. To understand why, assume that in estimating performance a person's judgments are subject to random error and consider the types of errors that are likely to be observed across a range of tasks that vary from hard to easy. This can be depicted by the relation between estimated and true performance shown in Figure 1 – the heavy line. When tasks are difficult, error implies that people overestimate their performance. When tasks are easy, error implies that people underestimate their performance (see also Dawes & Mulford, 1996; Erev, Wallsten, & Budescu, 1994).

What happens when people judge the performance of others? Assume, first, that these judgments are also subject to error and that people generally have less information about others than themselves. Thus, there will be more error associated with judgments of others than of oneself, and the relation between estimated and true performance for others will be flatter than that for oneself – see dotted line in Figure 1. Comparing the two lines, it can be seen that, for difficult tasks, people will underplace themselves relative to others; for easy tasks, they will overplace themselves.

This “regression” hypothesis makes strong and simple predictions. When people overestimate their ability in a specific, difficult task, they should also underplace themselves. Thus, for example, a track athlete who overestimates her performance in a particular difficult event would be expected to underplace herself relative to competitors in the same event (assuming, of course, that she does not have excellent information about her competitors). Conversely, when people underestimate their ability in a specific simple task, they should also overplace themselves. Using the athlete example again, overestimation for a simple event should be accompanied by underplacement.

The main purpose of this paper is to test the regression hypothesis. Is the asymmetry between estimation and placement observed when people face the same tasks? Moreover, if the asymmetry does not exist, what moderates patterns that are observed?

A related issue considers what impact information relevant to estimation has on decisions concerning placement. Can knowledge about one be informative about the other and, if so, do people make such inferences appropriately? For example, if you

know your score on a test, can you use this information to make a decision that depends on placement (i.e., when assessing your score relative to others)?

Our goal is to investigate these two issues experimentally. The paper is organized as follows. In the next section we introduce the experimental paradigm used to test whether people exhibit over- or underplacement and/or over- and under-estimation in the same task (a general knowledge quiz). The paradigm also permits investigation of how people update judgments of their relative level of performance after receiving information about their absolute level. Moreover, we specify a Bayesian model of this process as a behavioral benchmark. This is followed by presenting the results of four experiments.

In brief, our first two experiments demonstrate that even if people are over- or underconfident in the sense of estimation, their choices reveal no bias in terms of placement. In addition, people generally update their judgments about placement in accordance with the Bayesian model. In our third experiment, we extend our investigation by considering placement decisions that are tailored to individual performance. In the fourth experiment, participants were required to rank order four alternatives instead of choosing between two. In doing so, they exhibited underplacement irrespective of whether they were also over- or underconfident in the sense of estimation. We conclude by discussing the implications of our findings.

The experimental paradigm

Our experimental paradigm has seven basic features. First, participants answer a multiple-choice test where their remuneration depends on how many questions they

answer correctly. Second, they guess their scores (thereby allowing us to estimate over- or underconfidence in terms of estimation).

Third, participants choose between six pairs of 2-outcome gambles each of which involves monetary gains and losses. In one gamble in each pair, the probability of winning is 0.50. In the other gamble, this probability is unspecified but participants are told that it reflects their relative test score. Specifically, the probability of winning is equal to the percentile of the score in the group (e.g., a score at the 0.80 percentile translates into a 0.80 probability). Clearly, if participants believe that their score exceeds (falls short of) the 0.50 fractile, they should choose the gamble with the unspecified (0.50) probabilities. Participants' choices therefore provide behavioral tests for over- or underplacement at the group level.

Fourth, participants are given feedback in the form of their absolute test scores.

Fifth, participants are required to choose again between the same six gambles. Since they now know their absolute test scores, we can observe how this affects their placement (i.e., how knowledge of absolute level affects judgments of relative levels).

Sixth, participants are remunerated by the sum of (a) a small, fixed show-up fee, (b) according to the number of questions answered correctly in the quiz, and (c) the outcome associated with one of their twelve choices selected at random.

The seventh feature is that there are two versions of the quiz, one "hard" and one "easy" that are used to induce over- and underestimation, respectively, in two groups in each experiment.

Placement based on knowledge of one's absolute score. At first thought, one might think that knowledge about the absolute score on the test should have no impact on

a person's assessment of relative performance. To support this idea, consider how students graded "on a curve" should interpret their absolute scores on a test. A student, it is argued, should pay no attention to her absolute score since she is ignorant of the distribution of scores of the other students (cf., Windschitl, Kruger, & Simms, 2003¹).

However, a deeper analysis of this problem leads to a different conclusion. (See also, Moore & Cain, 2007; Moore & Healy, 2008). Assume that the student characterizes the scores of the population with whom comparisons will be made by the random variable, X , and that what matters is where her score lies relative to the mean of this distribution. Assume further that she has a prior distribution over μ , the mean of X , denoted by $p(\mu)$.

Now, the student learns her own score, x . This can be characterized by the likelihood function $p(x|\mu, \sigma_o^2)$ where σ_o^2 is the (assumed known) variance of the distribution. Using Bayes' theorem, the signal or score received by the student can be used to update the prior distribution over μ , that is

$$p(\mu|x, \sigma_o^2) \propto p(\mu) \cdot p(x|\mu, \sigma_o^2) \quad (1)$$

where " \propto " means "is proportional to." If the prior and likelihood functions involve normal distributions, it can be shown that the mean of the posterior distribution (the left hand side of Equation 1) is

$$E\mu = \alpha x + (1-\alpha)\mu \quad (2)$$

where $\alpha = \{\sigma_a^2/(\sigma_a^2 + \sigma_o^2)\}$ is the ratio of the variance of the prior distribution (σ_a^2) to the sum of the variances of the prior and likelihood (see, e.g., Zellner, 1971).

¹ However, we note that these researchers have in mind a somewhat different paradigm where the information about the test score (e.g., easy conditions) is common to all participants.

This is clearly an “as if” model that involves various technical assumptions. However, its qualitative implications are that the student’s best updated estimate of the mean score in the population is a weighted average of her prior “guess” and her own score. Thus, her best estimate of the mean will always lie between these two quantities, that is, be bigger (smaller) than the prior when her own score is bigger (smaller) but not as big (small) as her own score. In particular, note that to the extent that there is less variation in X (or σ_o^2 is small), more weight should be given to own score.

Finally, in one of our experiments we test what happens when participants’ probabilities of success do not depend on their own performance but rather on that of another participant selected at random. How should a participant take account of feedback concerning her own performance in this case? Here, we note that although the participant is aware of her own performance, she does not know the performance of the randomly selected participant. However, following the logic given above, she can update her estimate of mean performance and use that as the expected performance of a randomly selected participant. But since the performance of the randomly selected participant is just as likely to be above or below the updated median, the participant making the decision has no reason to change from what she did prior to receiving feedback about her own particular performance.

Experimental evidence

Experiment 1

Design. We sought to answer two issues. The first was whether participants who were over- or underconfident in terms of estimation would take decisions that implied

over- or underconfidence in terms of placement. The second was whether participants' choices after receiving feedback would be in accordance with the Bayesian model outlined above. The design of the experiment followed our experimental paradigm.

To address the first issue, we constituted two groups of participants at random. One had to choose between two possible answers for each question of the general knowledge quiz (the “easy” condition); the other had to choose between five possible answers (the “hard” condition). Both groups faced the same time limit of 45 seconds to answer each question. Based on results in the literature (see, e.g., Burson et al., 2006), we hypothesized that the group in the hard (easy) condition would be overconfident (underconfident) in terms of estimation.

Procedure. The experiment was conducted in several phases (see also instructions in Appendix A). First, participants took a test involving 20 general knowledge and logic questions in a multiple choice format. After completing the test, participants were required to estimate the number of questions they had answered correctly. (Their remuneration depended on the number of correctly answered questions.)

Second, participants were given the option of leaving the experiment or to continue in a task that would involve choosing between gambles where they could actually lose money.² They were informed that, at the end of the experiment, their remuneration would depend on playing out the consequences of one of their chosen gambles selected at random.

² The decision to allow participants to leave the experiment was guided by ethical considerations concerning the possibility of participants losing money.

Third, the participants remaining in the experiment faced a series of six choices between gambles. We refer to these subsequently as the “pre-feedback” choices and describe the gambles below.

Fourth, participants received feedback in the form of their actual (absolute) test scores.

Fifth, the participants chose between the six gambles for a second time albeit in a different order. We refer to these subsequently as the “post-feedback” choices.

Sixth, participants faced the consequences of playing out one of their choices that was determined randomly and were remunerated accordingly. Participants also completed a post-experimental questionnaire that inquired, inter alia, how they evaluated their scores on the test relative to their fellow participants.

Choice tasks. Participants were faced with a series of choices between two gambles. One of these provided a 50% chance of winning money and a 50% chance of losing money where the expected value was 1€ The other gamble also involved sums to be won or lost but the probabilities were not specified. Participants were told that their individual-specific probabilities depended on their relative performance on the test.

Each choice could therefore be characterized by a “0.50/0.50” gamble to win or lose ($w: l$) versus unspecified chances to win or lose ($w': l'$). Thus, for example, the choice between a 0.50/0.50 gamble paying (3€-1€) and an unspecified probability gamble paying (3€-1€) can be described by the notation “3:-1 vs. 3:-1” (i.e., 0.50/0.50 chances on the left, unspecified probabilities on the right). As noted above, we maintained the expected value of all six gambles with known probabilities equal to 1€

however, we varied the amounts involved (from 3€-1€ to 5€-3€) and whether outcomes were symmetric or asymmetric, e.g., “3:-1 vs. 3:-1” or “3:-1 vs. 5:-3.”

Participants. Participants were recruited through notices on the campus of Universitat Pompeu Fabra and the experiment was conducted in the Leex laboratory using the z-Tree software (Fischbacher, 1999). They were remunerated by a show-up fee, how well they answered the test, and the outcome of the randomly selected gamble. In all four experiments reported here we followed these same procedures concerning participants.

In this experiment, participation involved four groups of fifteen persons where two groups received the easy test and two the hard test. The mean age of participants was 20.5 (little variance) with almost as many males as females. On average, participants earned 7.03 €

Results of Experiment 1. The left hand panel of Table 1 reports – separately for the sub-groups taking the hard and easy tests – the mean test scores, mean estimated test scores, and mean levels of overconfidence. On average, participants who took the hard test were overconfident (2.40), whereas those who took the easy test were underconfident (-1.85). The difference between these means is significant ($t=4.86$, $p<.001$) and consistent with previous findings in the literature (see above).

As noted earlier, both before and after receiving feedback, participants made six choices between gambles with, on the one hand, “0.50/0.50” probabilities and, on the other hand, unspecified probabilities. To develop reliable measures of preferences for the unspecified options, we coded each choice in binary fashion (unspecified probabilities = 1, “0.50/0.50” probabilities = 0) and summed these indicators to create an “uncertainty

score” by participant. By definition, the scores can range from 0 to 6 (the larger the score, the more often the participants chose gambles with unspecified probabilities). Consequently, a score of 3 indicates indifference between the gambles involving “0.50/0.50” and unspecified probabilities.

Insert Tables 1 and 2 about here

In Appendix B (top panel), we provide details of all the gambles and the proportions of participants preferring the gambles with unspecified probabilities. This information is presented both for the hard and easy sub-groups and choices made before and after receiving feedback. We further provide the corresponding mean uncertainty scores that are also summarized in the top panel of Table 2 (together with standard errors).

For Experiment 1, the top panel of Table 2 shows – for the pre-feedback choices – that the mean uncertainty scores of the hard and easy sub-groups (3.16 and 2.63) are neither significantly different from each other ($t = 1.05$, $p = .299$) nor different from 3.00 that represents indifference between the gambles with “0.50/0.50” and unspecified probabilities ($t = -0.48$ and -0.72 for the hard and easy sub-groups, respectively, ns). In other words, there is no over- or underplacement.

Second, since participants in the hard (easy) condition received on average scores that were lower (higher) than their expectations, the Bayesian model predicts that their post-feedback decisions will involve choosing the unspecified gambles at a lower (higher) rate than the pre-feedback decisions. This is exactly what the data show. In the hard sub-group, the mean uncertainty score decreases from 3.16 to 1.72 ($t = 4.04$,

$p < .001$). In the easy sub-group, it increases from 2.63 to 3.78 ($t = -2.86$, $p = .008$). This interaction is further supported by the fact that the beta coefficient of the regression of the difference between pre- and post-feedback uncertainty scores on the easy/hard condition (coded as a dummy variable) is significant ($t = 4.79$, $p < .001$).

Discussion of Experiment 1. Experiment 1 provides further evidence of the so-called “hard-easy” effect in estimation (Juslin, 1994; Juslin, Winman, & Olsson, 2000). Participants facing the hard test overestimated their ability; those facing the easy test underestimated their ability. At the same time, neither group was over- or underconfident in their pre-feedback choices in the sense of over- or underplacement. In other words, there was no relation between over- or underestimation and over- or underplacement.

Second, changes in the participants’ pattern of choices after receiving feedback was consistent with the Bayesian model. That is, those in the hard sub-group reduced their choices with unspecified probabilities; those in the easy sub-group increased them.

In the post-experimental questionnaire participants were asked to assess their skill in answering the general knowledge questionnaire relative to their peers. They were given five options from “much worse than others” to “much better than others” with a mid-point of “similar to others.” A large majority (76%) checked this latter category. Participants were also asked to estimate in quantitative terms their relative position in the distribution of scores. Their mean judgment implied an overall probability of success of 0.58. Thus, when explicitly asked participants stated that their competence was similar to their peers. This response was clearly consistent with their pre-feedback choices. Experiment 2 was designed to provide further tests of this point.

Experiment 2

Design and procedure. We employed the same basic paradigm (gambles and test) as in Experiment 1 but with three exceptions. First, we changed the order in which participants made their choices and answered the test questionnaires.³ Thus, in Experiment 2, the ordering of activities was as follows: (1) instructions; (2) participants were offered the chance to leave the experiment if they did not wish to face the possibility of losses; (3) choices between gambles; (4) test questionnaire; (5) assessment of own score on test; (6) choices between gambles; (7) final questionnaire; and (8) payment based on randomly selected choice.

Second, the change in the order of tasks meant that participants did not receive feedback on their absolute test scores before making the second set of choices. Instead, one can argue that taking the test was a form of imperfect feedback (given that they chose before taking the test).

Third, whereas the unspecified probabilities of half of the participants depended on their percentile scores in the test (as in Experiment 1), the probabilities for the others depended on the percentile score of a randomly selected participant. In other words, half of the participants faced the same situation as those in Experiment 1 (and were informed as to how the probabilities had been calculated); the other half was informed that the unspecified probabilities were equal to the percentile score of one of their colleagues chosen at random. If – on average – participants are neither over- nor underconfident in their ability relative to others, there should be no difference between those choosing on

³ The decision to change the order of tasks was in response to the possibility that the results of Experiment 1 might be due to order effects. Specifically, in related work Camerer and Lovallo (1999) used the order adopted in Experiment 2.

the basis of their own scores and those choosing on the basis of randomly selected fellow participants (i.e., no over- or underplacement).

This time, however, the Bayesian model predicts a difference in how participants should react to feedback. Relative to pre-feedback responses, participants choosing based on their own performance should shift their post-feedback responses in the manner observed in Experiment 1. However, those choosing on the basis of the performance of a randomly selected fellow participant would have no reason to change their pre-feedback responses.

As in Experiment 1, half of the participants received the hard version of the test questionnaire and half received the easy version. In summary, participants in Experiment 2 were assigned at random to four (= 2 x 2) groups composed of two conditions of unspecified probabilities (based on own ability or that of a random other) and two levels of test difficulty (hard or easy).

Participants. There were 57 participants with an average age of 21.6 years; 40% were male. On average, participants earned 6.98€

Results of Experiment 2. The middle panel of Table 1 shows that participants exhibit the “hard-easy” effect for the estimation task whether they make decisions based on their own score or that of a randomly selected colleague. The hard-easy differences are significant in both sub-groups ($t = 4.35$ and $t = 5.13$, $p < .001$).

Proportions of choices favoring the unspecified probabilities made by participants in each of the four experimental conditions – pre- and post-feedback – are detailed in the middle section of Appendix B. The upper middle section of Table 2 summarizes the mean uncertainty scores.

These data show, first, for the pre-feedback choices there are no differences between the hard and easy sub-groups in both the “own score” and “random other” conditions (2.93 vs. 3.06, $t = -0.22$, $p = 0.828$; and 3.31 vs. 2.69, $t = 0.95$, $p = 0.352$). Moreover, there are no differences overall between the own score and random other conditions (3.00 vs. 3.00, $t = 0.00$, $p = 0.999$) which, of course, are also not significantly different from 3.00. In other words, these data provide no evidence of over- or underplacement whether participants are in the hard or easy sub-groups, on the one hand, or choose based on their own score or that of a randomly selected colleague.

Second, as in Experiment 1, the mean uncertainty scores for the post-feedback choices are smaller than their pre-feedback counterparts in the hard sub-group but larger in the easy sub-group. However unlike Experiment 1, with one exception these differences are not statistically significant (2.93 vs. 2.60, $t = 0.64$; 3.06 vs. 3.13, $t = -0.22$; 3.31 vs. 3.08, $t = 0.76$; but 2.69 vs. 3.46, $t = -2.99$, $p = .011$).

Testing for interactions by regressing the difference between pre- and post-feedback scores on, first, treatment (own score vs. random other) and condition (hard/easy) reveals no significant beta coefficients ($t = 1.12$ and 1.84 , respectively, ns). However, repeating the same regression on condition (hard/easy) but within treatments, the coefficient for hard/easy is not significant for own score ($t = 0.68$) but is significant in the random other treatment ($t = 2.52$, $p = .019$).

Interpreting these results within the Bayesian logic outlined above, we note that participants in the own score condition exhibited changes pre- to post-feedback that were consistent with the Bayesian model albeit statistically insignificant (see also further below). Participants in the random other condition, on the other hand, exhibited

statistically significant changes that violated the “no change” prediction of the Bayesian model.

Discussion of Experiment 2. Experiment 2 replicates the important result of Experiment 1, namely: over- or underestimation due to hard and easy tests is not accompanied by overconfidence in the form of over- or underplacement. Moreover, this finding is replicated even when the order of tasks has been changed (i.e., of taking the test and the first set of choices) and when participants use the probabilities of a randomly assigned other to make their choices as opposed to their own probabilities.

Experiment 2 does not replicate the results of Experiment 1 with respect to feedback for the own score sub-group. However, whereas the feedback in Experiment 1 was definitive – actual absolute test scores – in Experiment 2 it was imperfect, simply taking the test. Thus, if there were to be any effect, one would expect it to be much weaker.

As to the random other sub-group, the data indicated significant pre- to post-feedback shifts that were similar to those exhibited in Experiment 1. However, this time the Bayesian model predicted no change. In correctly judging themselves, on average, as similar to a random other, it would seem that our participants “over-generalized” this thinking when reacting to the feedback of actually taking the test.

Although participants in Experiment 2 were offered the chance to leave the experiment prior to taking gambles, none elected to do so. We speculate this may have occurred because the first set of choices was made before participants had answered the quiz that determined their probabilities of success.

Behavior in terms of placement was consistent with statements made in the post-experimental questionnaire. Seventy-four percent of participants stated that their skill in answering the test was “similar to others.” Moreover, in evaluating their relative position in the distribution of scores, the overall mean percentile estimate was 0.45.

In Experiments 1 and 2 placement decisions were essentially made against a single reference point of 0.50 or the median of the performance distributions. Experiment 3 therefore sought to extend the experimental paradigm by exploring the effects of different reference points or fractiles of the performance distribution.

Experiment 3

Design, procedure, and choice tasks. Experiment 3 was exactly like Experiment 1 with one exception. When participants faced choices between known and unspecified probabilities, the known probabilities were not all stated to be “0.50/0.50” (“win/loss”). Instead, these were tailored for each participant so that they were equal to their individual unspecified probabilities which, as in Experiment 1, depended on how well they had answered the quiz compared to their fellow participants. Thus, for example, a participant whose test score was at the 0.65 fractile of the distribution would be faced with a choice between known probabilities of “0.65/0.35” and unspecified probabilities. (Participants were not informed that the known probabilities were a function of their own performance.)

Of particular interest is how participants act when faced with high ($p > 0.50$) as opposed to low ($p \leq 0.50$) known probabilities. Assume, for example, that a participant starts with a prior that she is no different from the median peer. Imagine that, on taking

the test, she feels she does well and that her performance is better than most. From a rational perspective, she should combine this signal with her prior thereby assessing her ability as being above her prior but below her signal. Thus, if faced with gambles involving high known and unspecified probabilities, she should be “regressive” and choose the gamble with known probabilities. By parallel logic, participants facing gambles with low known probabilities should choose the gambles with unspecified probabilities.

Participants. There were 58 participants with an average age of 20.8 years; 46% were male. On average, participants earned 6.88 €

Results of Experiment 3. Five of the 58 participants did not wish to continue beyond the quiz to the choices. Thus, we were left with 53 participants, 27 in the hard and 26 in the easy conditions. As shown in Table 1, participants in the hard conditions were, on average, overconfident in the sense of estimation and participants in the easy condition underconfident (3.56 vs. -1.50, $t = 7.00$, $p < .001$). This replicates the results of Experiments 1 and 2.

Table 2 (lower middle section) reports results in terms of participants’ mean pre- and post-feedback uncertainty scores broken down by the hard and easy conditions. In addition to showing results for the whole sample (total), we also show separately the results of those participants who faced known probabilities greater than 0.5 ($p > 0.50$) and those who faced probabilities less than or equal to 0.5 ($p \leq 0.50$).

The “total” results are similar to Experiment 1. At the pre-feedback stage, the means of the two sub-groups (3.26 and 3.00) are neither significantly different from each other ($t = 0.43$, ns) nor from 3.00 ($t = 0.60$ and 0.01 , respectively, ns). That is, on average

there is no over- or underplacement in either the hard or easy sub-groups. At the same time, the placement decisions following feedback are in accordance with the Bayesian model and supported by statistical analysis. Specifically, the regression of the difference between pre- and post-feedback scores on the hard/easy condition reveals a significant coefficient for the latter ($t= 2.90$, $p=.005$).

On the other hand, there is a striking distinction between participants who faced known probabilities larger and smaller than 0.50. Participants for whom $p>0.50$ chose predominantly the known probabilities as witnessed by their low uncertainty scores (1.62 and 1.69 for the hard and easy sub-groups, respectively). Both means were significantly different from 3.00 ($t=3.07$ and 2.57 , $p<.01$) but not from each other ($t=0.11$, $p=.911$).

Participants for whom $p \leq 0.50$ chose predominantly the unspecified probabilities as witnessed by their high uncertainty scores (4.79 and 4.31 for the hard and easy sub-groups, respectively). Both means were significantly different from 3.00 ($t=4.59$ and 2.78 , $p<.01$) but not from each other ($t= 0.78$, $p=.911$).

Participants were also differentially sensitive to feedback. For $p > 0.50$, there was no effect on uncertainty scores of the interaction captured by the regression of the difference between pre- and post-feedback scores on the hard easy/condition ($t = 0.87$, ns). However, for those for whom $p \leq 0.50$, the coefficient of the analogous regression was significant ($t=3.30$, $p=.003$).

Discussion of Experiment 3. The results of Experiment 3 add to the evidence on the hard-easy effect for estimation: overestimation for the hard condition and underestimation for the easy condition. *At the aggregate level*, the data also replicate

Experiment 1 in terms of no over- or underplacement and participants' responses are consistent with the Bayesian model following feedback.

However, the picture is quite different for participants who faced known probabilities greater or smaller than 0.50. The former showed a strong tendency to avoid gambles with unspecified probabilities whereas the latter predominantly chose them. In other words, those facing probabilities greater than 0.50 exhibited underplacement and those facing probabilities below 0.50 exhibited overplacement. This differential behavior of participants facing low ($p \leq 0.50$) and high ($p > 0.50$) known probabilities is, of course, consistent with the arguments of “regressive” reasoning outlined above. Moreover, these tendencies were observed irrespective of whether participants were over- or underconfident in the sense of estimation.

As is well known, contextual effects have frequently been shown to influence the expression of preference (Lichtenstein & Slovic, 2006). What then might be the effect of different frames when people make judgments concerning placement? This was the purpose of Experiment 4.

Experiment 4

Design, procedure, and choice tasks. Experiment 4 was exactly like Experiment 1 with one exception. Participants did not simply choose between gambles that, on one hand, had known “0.50/0.50” probabilities and, on the other, unspecified probabilities that depended on how well they answered the quiz compared to their fellow participants. Instead, participants had to rank order (by preference) the following four outcomes: (1) A “sure thing” of 0.75€ (2) The gamble with “0.50/0.50” probabilities (as in Experiment

1); (3) The gamble with unspecified probabilities where no information was provided about the probabilities (so-called “ambiguity,” Ellsberg, 1961); and (4) The gamble with unspecified probabilities where, as in Experiment 1, the probabilities depended on how well participants had answered the quiz compared to their fellow participants.

From a formal viewpoint, we note that the comparison between the second and fourth alternatives implies identical choices to those examined in Experiments 1 and 2. Thus, consistency would lead one to expect no over- or underconfidence in terms of placement.

Participants. There were 64 participants with an average age of 20.6 years; 44% were male. On average, participants earned 7.93€

Results of Experiment 4. The right hand panel of Table 1 presents the results concerning estimation. As in Experiments 1, 2, and 3, participants in the hard sub-group are overconfident (mean of 1.77) whereas those in the easy sub-group are underconfident (mean of -1.70). Moreover, these means are significantly different from each other ($t=3.94$, $p<.001$).

The bottom sections of Appendix B and Table 2 report the data relevant to placement. Unlike Experiments 1, 2, and 3, however, we do not report the results of direct binary choices between gambles involving known and unspecified probabilities but infer these decisions from participants’ rank orderings of the four choice alternatives enumerated above. There are two striking results.

First, the data exhibit considerable underplacement, i.e., underconfidence. For the pre-feedback choices, there are no differences in mean uncertainty scores between the hard and easy sub-groups (1.20 vs. 1.59, $t = 0.77$, $p = 0.446$) but these are both

significantly smaller than the 3.00 that would imply no over- or underplacement ($t = 5.00$ and 4.03 , $p < .001$).

Second, although participants are not acting as though their average performance is equal to the mean of the group at the pre-feedback stage, they do adjust their choice strategies subsequently in accordance with the direction of their feedback. In the hard sub-group, this decreases from a mean uncertainty score of 1.20 to 1.10 ($t = 0.40$, $p = 0.693$); whereas in the easy sub-group this increases from 1.59 to 2.85 ($t = -3.83$, $p < .001$).⁴ Interestingly, the post-feedback choices of the easy sub-group reveal no over- or underplacement (2.85 is not significantly different from 3.00, $t = 0.39$). As might be expected, testing this interaction by regressing the difference between pre- and post-feedback scores on the hard/easy condition reveals a significant beta coefficient ($t = 3.22$, $p = .002$).

Finally, unlike Experiment 1, there was no differential self-selection into the experimental task. All of the participants elected to do the choice task after completing the quiz. We have no explanation for the differences between the two experiments in this respect.

Discussion of Experiment 4. The key – and surprising – result of Experiment 4 is the strong underplacement that is apparently induced by the rank ordering of choice alternatives as opposed to direct choices between gambles involving known and unspecified probabilities (as in Experiments 1, 2, and 3). These results are even more surprising in that a large majority of participants (73%) checked the mid-point of “similar to others” when asked in the post-experimental questionnaire to assess their skill in

⁴ Whereas the result is statistically significant in the easy sub-group, this is not the case in the hard sub-group. One reason for the latter could be a floor effect, i.e., there is little room to be even more “uncertainty averse” as one moves from the pre-feedback to the post-feedback condition.

answering the general knowledge questionnaire relative to their peers. Also the mean estimate of their percentiles in the distribution of scores was 0.47. In other words, when explicitly asked, participants assessed themselves as being, on average, no different than their colleagues.

The results of this experiment can be viewed from two perspectives. One is the relation between estimation and placement; the other is the effect of numbers of alternatives on revealed preferences. As to the former, the results suggest that the relation between tasks involving estimation and placement are subject to framing effects. Although interesting, this considerably complicates making any general statements.

As to the latter, the finding of greater uncertainty avoidance (or “risk aversion”) can be related to two literatures. One concerns findings in the animal literature that certain birds – common starlings – show less risk aversion when faced with binary choices as opposed to multiple alternatives (Schuck-Paim & Kacelnik, 2007). The explanatory hypothesis is that most animal choices are binary in nature, e.g., eat versus do not eat. They do not involve choosing one of many alternatives, e.g., which of k (>2) alternative foods to eat. When confronted by many alternatives, animals are confused and either delay choice or seek safer options.

The second literature relates to the so-called “paradox of choice” (Iyengar & Lepper, 2000; Schwartz, 2004). This refers to the fact that the satisfaction people derive from choice does not necessarily increase with the number of alternatives they face (if anything, it is an inverted U-shaped function, Reutskaja & Hogarth, 2008). This has important implications in applied settings. For example, when faced with many alternatives for investing in pension funds, people have been found to delay choice, even

when this reduces their financial returns (Iyengar, Jiang, & Huberman, 2004). In general, this literature emphasizes the dysfunctional aspects of having “too many” choice alternatives that can result in dissatisfaction, excessive risk aversion, and procrastination. The relation between our results and these findings clearly demands further investigation.

General Discussion

Summary. In all of our four experiments participants exhibited over- and underconfidence in the sense of estimation, based on responses to hard and easy tasks respectively. This replicated results in the literature (see e.g., Juslin et al., 2000) and was consistent with the regression hypothesis proposed by Moore and Healy (2008). The picture, however, was not so clear concerning over- and underconfidence in the sense of placement and is summarized in Table 3.

Insert Table 3 about here

The upper panel (A) of Table 3 reports the results of regressing pre-feedback uncertainty scores for all four experiments on the different experimental manipulations. Consider Model 1 in the first column. The constant term estimates the mean uncertainty score across all experiments independent of their separate effects. The estimate of 3.90 (standard error of 0.63) is not significantly different from 3.00 that represents no overall effect of over- or underplacement.

Second, there is no effect due to whether participants faced unspecified probabilities based on their own performance or that of a randomly selected other (Self vs. Other, Experiment 2).

Third, there was a strong effect when participants faced “0.50/0.50” or “specifically tailored” probabilities (Fixed vs. Variable Probabilities, Experiment 3).

Fourth, the framing effect of expressing preferences in terms of binary choices as opposed to rank ordering preferences was also significant (Experiment 4).

Fifth, there was no effect overall if participants had taken the hard or easy versions of the quiz. In other words, at an aggregate level, the previous effects were independent of whether participants were over- or underconfident in the sense of estimation.

Finally, in Model 2 (the second column) we consider individual confidence scores (estimated vs. actual quiz scores) as an additional explanatory variable. At the individual level, overconfidence (in the sense of estimation) does increase uncertainty scores significantly but marginally (explaining only an additional 2% of variance).

The lower panel (B) of Table 3 provides a similar analysis for reactions to feedback. In Model 1, the difference between pre- and post-feedback uncertainty scores is regressed on the different experimental conditions. As can be seen, the effect for “hard vs. easy” (that captures changes consistent with the Bayesian hypothesis) is highly significant ($t = 6.48$). This effect, of course, is over and above smaller effects for “Fixed vs. Variable Probabilities” and “Binary Choice vs. Rank Order.” The fact that there is no significant effect for “Self vs. Other” supports our earlier analysis that although participants in the “own score” condition changed their responses in accordance with the

Bayesian hypothesis, those in the “random other” also changed but should not have done so (Experiment 2). Finally, as in panel A, we include the effect of individual confidence in Model 2 (second column). This has a strong effect, accounting for some 10% of additional variance, and documents stronger reactions to feedback by those who were revealed to be most overconfident (in the sense of estimation).

Implications. The main result of our work centers on the difference between overconfidence in the senses of estimation and placement. Whereas the former can be relatively easily explained in terms of the regression hypothesis (Moore & Healy, 2008), this is not true of the latter.

For predicting placement, an important assumption of the regression hypothesis is that the person knows more about her performance than that of others. However, this may not be the only way in which people think about the problem of placement. For example, in our task it may be cognitively simpler for a person just to assume that her performance is, on average, no different from the mean of her peers. Indeed, this is consistent with the results of Experiments 1 and 2 where there was no over- or underplacement and even if results depended on the performance of a randomly selected other.

In Experiment 3 participants were confronted with known probabilities that differed from 0.50 and thus could not simply compare themselves with the median peer. Interestingly, what we observed was regressive behavior. When faced with a choice between gambles characterized by low known probabilities, on the one hand, and unspecified probabilities, on the other hand, there was a strong tendency to choose the latter. Similarly, choices between gambles involving high known probabilities versus unspecified probabilities led to choosing the former.

Experiment 4 demonstrated extreme underplacement when participants' choices were inferred from the rank ordering of four alternatives. As noted above, this framing effect cannot be explained by the regression hypothesis and suggests broader questions such as the effects of numbers of choice alternatives on attitudes toward risk.

In many applied settings people do receive “signals” about their ability prior to entering competitive situations. Consider, for example, students who can assess their abilities based on past examinations or athletes who can review past performance prior to competition. Our experimental paradigm modeled these kinds of situations by observing how participants took account of feedback after the first set of decisions. In general, our participants acted in accordance with a Bayesian updating model except when decisions involved the performance of a randomly chosen colleague. We suspect that, for this decision, reasoning in the shoes of another involved too much complexity. This hypothesis should be investigated in future work.

Finally, we note that many studies have reported significant overconfidence in the area of entrepreneurial entry decisions (see, e.g., Cooper, Wu, & Dunkelberg, 1988; Bernardo & Welch, 2001; Koellinger, Minniti, & Schade, 2007). Moreover, there is increasing awareness of different types of overconfidence and that it is particularly misplacement that leads to erroneous decisions (Wu & Knott, 2006). In this paper we have not studied the effects of overconfidence in the sense of precision and thus do not comment on its importance relative to the other types of overconfidence. However, we suspect that much more can be gained in the future by concentrating efforts to identify the role of factors than determine placement as opposed to estimation.⁵

⁵ For an intriguing analysis of the meaning of overconfidence in the sense of placement, see Benoît and Dubra (2008).

References

- Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky, A. (eds), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press
- Bernardo, A.E., & Welch, I. (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy*, 10 (3), 301-330.
- Benoît, J.-P., & Dubra, J. (2008). *Overconfidence?* Working paper, London Business School.
- Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, 90(1), 60–77.
- Camerer, C. F., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306-318.
- Cooper, A.C., Woo C.J., & Dunkelberg, W.C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing*, 2, 97-108.
- Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment, or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes*, 65, 201-211.
- Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, 75(4), 643-669.

- 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.
- Fischbacher, U. (1999). *z-Tree 1.1.0.: Experimenter's manual*. University of Zurich, Institute for Empirical Research in Economics, <http://www.iew.unizh.ch/ztree/index.php>
- Hoelzl, E., & Rustichini, A. (2005). Overconfident: Do you put your money on it? *Economic Journal*, *115* (April), 305-318.
- Iyengar, S. S., Jiang, W., & Huberman, G. (2004). How much choice is too much? Contributions to 401 (k) retirement plans. In O. S. Mitchell, & S. Utkus (Eds.), *Pension design and structure: New lessons from behavioral finance* (pp. 83-96). Oxford: Oxford University Press.
- Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995-1006.
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, *57*, 226-246.
- Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review*, *114*(3), 678-703.

- Juslin, P., Winman, A. & Olsson, H. (2000). Naïve empiricism and dogmatism in confidence research. A critical examination of the hard-easy effect. *Psychological Review*, *107*, 384-396.
- Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what and whom you ask. *Organizational Behavior and Human Decision Processes*, *79* (3), 216-247.
- Koellinger, P., Minniti, M., & Schade, Ch. (2007). “I think I can, I think I can”: Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology*, *28*, 502–527.
- 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* (2), 221-232.
- 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*, *103*, 197-213.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115* (2), 502-517.
- Plous, S. (1993). *The psychology of judgment and decision making*. Philadelphia, PA: Temple University Press.
- Reutskaja, E., & Hogarth, R. M. (2008). *Satisfaction in choice as a function of the number of alternatives: When “goods satiate” but “bads escalate.”* Working paper, Universitat Pompeu Fabra, Barcelona.

- Schwartz, B. (2004). *The paradox of choice: Why more is less*. New York: Eco/HarperCollins Publishers.
- Schuck-Paim, C., & Kacelnik, A. (2007). Choice processes in multialternative decision making. *Behavioral Ecology*, *18*, 541-540.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*, 143-148.
- 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*, *85*(3), 389-408.
- Wu, B., & Knott, A. M. (2006). Entrepreneurial risk and market entry. *Management Science*, *52*, 1315-1330.
- Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York, NY: Wiley.

Table 1: Overconfidence (overestimation) for hard and easy tests

	<u>Experiment 1</u>		<u>Own score</u>		<u>Experiment 2</u>		<u>Random other</u>		<u>Experiment 3</u>		<u>Experiment 4</u>	
	<u>Hard</u>	<u>Easy</u>	<u>Hard</u>	<u>Easy</u>	<u>Hard</u>	<u>Easy</u>	<u>Hard</u>	<u>Easy</u>	<u>Hard</u>	<u>Easy</u>	<u>Hard</u>	<u>Easy</u>
	(n=25)	(n=27)	(n=15)	(n=16)	(n=13)	(n=13)	(n=27)	(n=26)	(n=30)	(n=34)		
Mean test score	8.36	13.52	6.67	14.19	6.23	14.38	7.37	13.92	8.13	13.62		
Mean of estimated test scores	10.76	11.67	9.67	12.75	10.54	11.92	10.93	11.42	9.90	11.92		
Overconfidence ¹	2.40	-1.85	3.00	-1.44	4.31	-2.46	3.56	-2.50	1.77	-1.70		
t-statistic: Hard vs. Easy	4.86		4.35		5.13		7.00		3.84			
	p< .001		p< .001		p< .001		p< .001		p< .001			

¹ Positive values indicate overconfidence (overestimation); negative values indicate underconfidence (underestimation)

Table 2. Uncertainty scores for Experiments 1 - 4: Means/<standard errors>

		<u>Sub-groups</u>		<u>Total</u>
		<u>Hard</u>	<u>Easy</u>	
<u>Experiment 1</u>				
	Pre-feedback	3.16 <0.37>	2.63 <0.35>	2.88 <0.25>
	Post-feedback	1.72 <0.35>	3.78 <0.33>	2.79 <0.29>
<u>Experiment 2</u>				
	Own score			
	Pre-feedback	2.93 <0.25>	3.06 <0.52>	3.00 <0.29>
	Post-feedback	2.60 <0.51>	3.13 <0.46>	2.87 <0.34>
	Random other			
	Pre-feedback	3.31 <0.51>	2.69 <0.40>	3.00 <0.32>
	Post-feedback	3.08 <0.56>	3.46 <0.46>	3.27 <0.36>
<u>Experiment 3</u>				
Total:	Pre-feedback	3.26 <0.43>	3.00 <0.43>	3.13 <0.30>
	Post-feedback	2.11 <0.45>	3.62 <0.40>	2.85 <0.32>
p > 0.50:	Pre-feedback	1.62 <0.45>	1.69 <0.51>	1.65 <0.33>
	Post-feedback	1.62 <0.55>	2.38 <0.55>	2.00 <0.39>
p ≤ 0.50:	Pre-feedback	4.79 <0.39>	4.31 <0.47>	4.56 <0.30>
	Post-feedback	2.57 <0.71>	4.85 <0.32>	3.67 <0.45>
<u>Experiment 4</u>				
	Pre-feedback	1.20 <0.36>	1.59 <0.35>	1.41 <0.20>
	Post-feedback	1.10 <0.33>	2.85 <0.38>	2.03 <0.18>

Table 3. Regressions involving all experiments

A. Dependent variable: Uncertainty score

	<u>Model 1</u>		<u>Model 2</u>	
	<u>Coefficient</u>	<u>t-ratio</u>	<u>Coefficient</u>	<u>t-ratio</u>
Constant	3.90	6.23	3.89	6.27
Self vs. Other	-0.03	-0.08	0.01	0.02
Fixed vs. Variable Probabilities	-5.05	-5.98	-4.69	-5.53
Binary Choice vs. Rank Order ¹	1.56	5.75	1.52	5.68
Hard vs. Easy ²	0.13	0.54	-0.29	-0.98
Confidence	x	x	<u>0.09</u>	2.38
	R ²	0.25	0.27	
	n	226	226	

B. Dependent variable: Difference between pre- and post-feedback uncertainty score

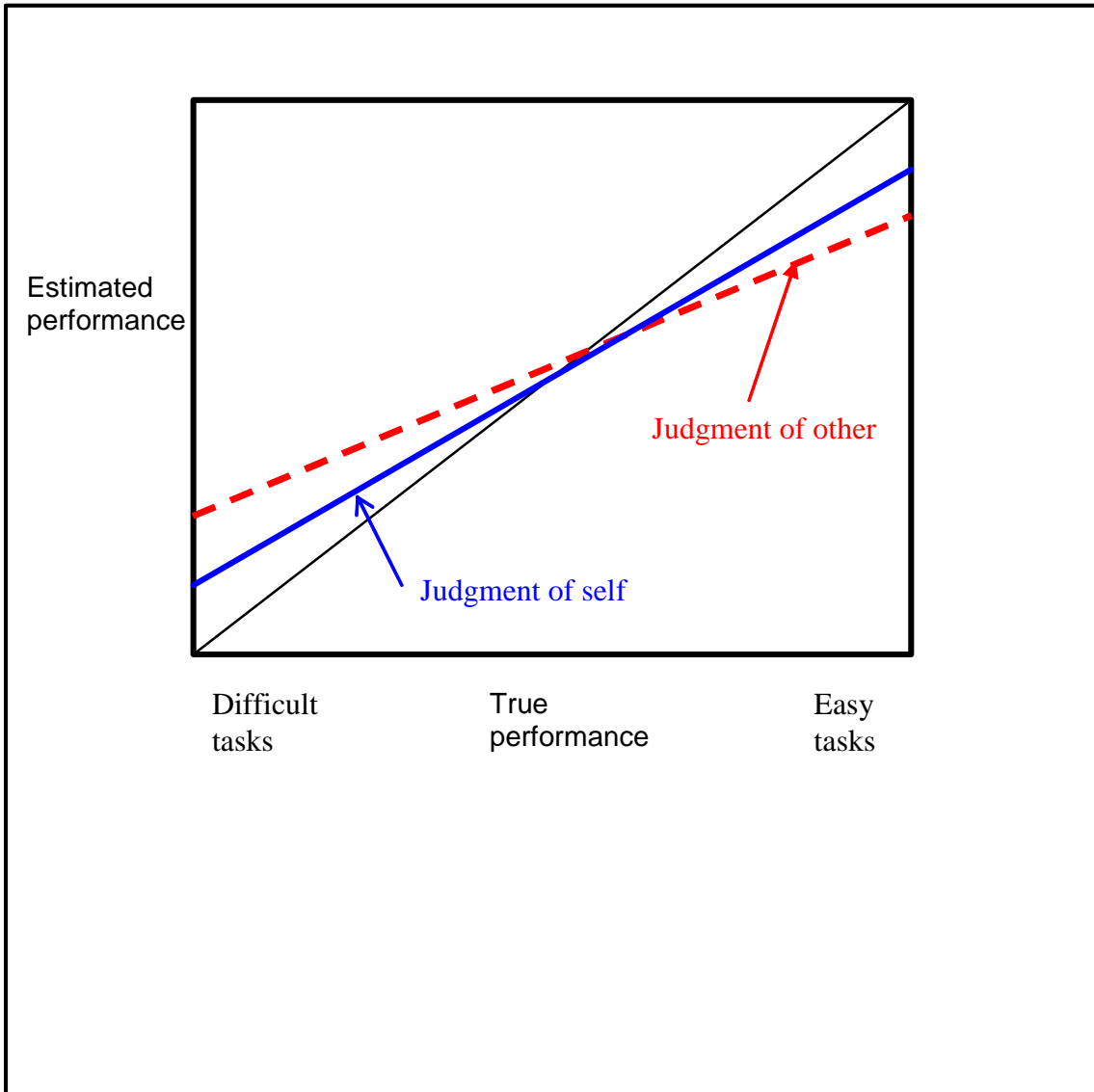
	<u>Model 1</u>		<u>Model 2</u>	
	<u>Coefficient</u>	<u>t-ratio</u>	<u>Coefficient</u>	<u>t-ratio</u>
Constant	-0.82	-1.29	-0.86	-1.44
Self vs. Other	0.44	1.14	0.53	1.46
Fixed vs Variable probabilities	<u>-1.96</u>	-2.28	-1.15	-1.41
Binary Choice vs. Rank Order ¹	<u>0.75</u>	2.72	<u>0.67</u>	2.59
Hard vs. Easy ²	1.57	6.48	<u>0.63</u>	2.22
Confidence	x	x	0.20	5.59
	R ²	0.20	0.30	
	n	226	226	

Note: Coefficients significant at $p < .001$ are in **bold**; those underlined are significant at $p < .05$.

¹ Binary Choice = 1; Rank Order = 0.

² Hard = 1; Easy = 0.

Figure 1. Judgments of self and others across task complexity



Appendix A – Experimental instructions for Experiment 1

(Translated from Spanish)

Thank you for participating in this experiment about decision making behavior that is part of a research project. What you earn will depend on your skill as well as the skill of your peers.

Please follow the instructions carefully. You have the opportunity to gain more than the 3 € that are already assured by your participation in the experiment. As from now until the end of the experiment you are not allowed to talk amongst yourselves. Raise your hand if you have a question and one of the instructors will attend to you. Please, do not ask in a loud voice. Thank you. The rules are the same for all participants.

Experimental procedure

The experiment consists of 3 phases.

Phase 1 consists of a set of 20 general knowledge questions. Each question is independent of the others. For each question, you have to decide between A and B.* Your earnings at the end of the experiment depend on the accuracy of your responses. You will earn 0.25 € for each question that you answer correctly.

Phase 2 is optional. You can

- claim all the money that you have earned up to this point (that is, the initial 3 € and what you earned from the questionnaire) and go directly to Phase 3 or
- modify your earnings by way of considering 12 pairs of gambles (2 groups of 6 pairs) and for each pair choosing one of two possible options. You have to decide between Option 1 and Option 2 for the 12 pairs of gambles. The gambles give you the chance to gain or lose certain sums of money. Each gamble involves different amounts of euros: you can gain between 3 and 5 euros and lose between 1 and 3 euros. The chances of winning or losing in these gambles can be the same for all the participants or depend on your own skill – compared to the skill of your peers – in the questionnaire of Phase 1.

* These are the instructions for the “easy” experimental condition. In the “hard” condition, participants had to choose one of five possible answers.

After you have chosen between each pair of gambles, one of the gambles that you have chosen will be selected and you will face the consequences of that gamble. The money depending on the outcome of this gamble will be added to (if you win) or subtracted from (if you lose) your earnings from Phase 1 (that is the initial 3 € plus your earnings from the questionnaire). Thus, when you choose each gamble, remember that it can be the one that is selected and that your final earnings can depend on its outcome!

In Phase 3, we ask you to complete a general questionnaire.

Phase 1

This phase consists of 20 multiple-choice questions. For all the questions you have to decide between A and B. (See last footnote.*) Your earnings depend on the correctness of your responses: you will receive 0.25 € for each correct answer. You have a maximum of 45 seconds to give each response: if you indicate no choice between A and B, your answer will be considered wrong.

At the end of this questionnaire, we will ask you to estimate the number of questions you answered correctly.

Phase 2

This phase is optional. You can decide between:

(a) keeping all the money you have earned up until now (that is the initial 3€ plus your earnings from the questionnaire) and go directly to Phase 3; or

(b) participate in a game where you can influence your own earnings by way of choosing between two gambles, called Option 1 and Option 2, for 12 pairs of gambles. The gambles give you the possibility of winning between 3 and 5 euros and losing between 1 and 3 euros. Your probability of winning or losing in these gambles can be the same as the probabilities for your peers (Option 1) or can depend on how well you answered the questionnaire in Phase 1 relative to your peers (Option 2). Your probability of winning will be equal to your percentile. Specifically, if you were at the 90th percentile of the distribution of points, your probability of winning would be 0.90; if you were at the 50th percentile of the distribution of points, your probability of winning would

be 0.50 (that is at the middle); if you were at the 30th percentile of the distribution of points, your probability of winning would be 0.30; and so on.

Remember that your percentile indicates the percentage of participants who obtained fewer points than you!

All the participants have the same opportunity.

If you choose (a), go to Phase 3.

If you choose (b), you have to decide between Option 1 and Option 2 for each pair of gambles. At the end, one of the gambles that you have chosen will be drawn at random and your final earnings will depend on the outcome of that gamble.

Phase 3

In the third phase we ask you to complete a general questionnaire. One of the instructors will give it to you.

When you have completed the general questionnaire of Phase 3, we will give you the sum that you have earned in Phases 1 and 2 plus the fixed fee of 3€

Thank you for your participation!

Appendix B. Proportions of participants preferring gambles with unspecified probabilities: Experiments 1-4

	n	3:-1 vs. 3:-1	3:-1 vs. 4:-2	3:-1 vs. 5:-3	4:-2 vs. 4:-2	4:-2 vs. 5:-3	5:-3 vs. 5:-3	Mean	Uncertainty score (mean)
<u>Experiment 1</u>									
Pre-feedback									
Hard sub-group	25	0.56	0.40	0.48	0.72	0.44	0.56	0.53	3.16
Easy sub-group	27	0.52	0.33	0.41	0.52	0.41	0.44	0.44	2.63
Total	52	0.54	0.37	0.44	0.62	0.42	0.50	0.48	2.88
Post-feedback									
Hard sub-group	25	0.43	0.43	0.47	0.45	0.40	0.40	0.43	1.72
Easy sub-group	27	0.58	0.52	0.57	0.49	0.54	0.56	0.54	3.78
Total	52	0.51	0.47	0.52	0.47	0.47	0.48	0.49	2.79
<u>Experiment 2</u>									
Own score									
Pre-feedback									
Hard sub-group	15	0.47	0.40	0.60	0.40	0.47	0.60	0.49	2.93
Easy sub-group	16	0.44	0.50	0.63	0.50	0.63	0.38	0.51	3.06
Total	31	0.45	0.45	0.61	0.45	0.55	0.48	0.50	3.00
Post-feedback									
Hard sub-group	15	0.49	0.47	0.51	0.47	0.44	0.49	0.48	2.60
Easy sub-group	16	0.54	0.50	0.52	0.50	0.46	0.52	0.51	3.13
Total	31	0.52	0.48	0.52	0.48	0.45	0.51	0.49	2.87
Random other									
Pre-feedback									
Hard sub-group	13	0.46	0.54	0.69	0.31	0.69	0.62	0.55	3.31
Easy sub-group	13	0.54	0.23	0.38	0.62	0.46	0.46	0.45	2.69
Total	26	0.50	0.38	0.54	0.46	0.58	0.54	0.50	3.00
Post-feedback									
Hard sub-group	13	0.46	0.56	0.54	0.49	0.51	0.46	0.50	3.08
Easy sub-group	13	0.49	0.56	0.51	0.54	0.46	0.60	0.53	3.46
Total	26	0.47	0.56	0.53	0.51	0.49	0.53	0.52	3.27
<u>Experiment 3</u>									
Total:									
Pre-feedback									
Hard sub-group	27	0.59	0.56	0.48	0.52	0.56	0.56	0.55	3.26
Easy sub-group	26	0.46	0.46	0.35	0.69	0.50	0.54	0.50	3.00
Total	53	0.53	0.51	0.42	0.60	0.53	0.55	0.52	3.13
Post-feedback									
Hard sub-group	27	0.37	0.41	0.37	0.33	0.30	0.33	0.35	2.11
Easy sub-group	26	0.62	0.50	0.50	0.62	0.69	0.69	0.60	3.62
Total	53	0.49	0.45	0.43	0.47	0.49	0.51	0.47	2.85
p>0.50:									
Pre-feedback									
Hard sub-group	13	0.31	0.31	0.39	0.08	0.39	0.15	0.27	1.62
Easy sub-group	13	0.46	0.15	0.08	0.46	0.15	0.39	0.28	1.69
Total	26	0.39	0.23	0.23	0.27	0.27	0.27	0.28	1.65
Post-feedback									
Hard sub-group	13	0.31	0.39	0.39	0.15	0.23	0.15	0.27	1.62
Easy sub-group	13	0.46	0.31	0.39	0.31	0.46	0.46	0.40	2.38
Total	26	0.39	0.35	0.39	0.23	0.35	0.31	0.34	2.00
p≤0.50									
Pre-feedback									
Hard sub-group	14	0.86	0.79	0.57	0.93	0.71	0.93	0.80	4.79
Easy sub-group	13	0.46	0.77	0.62	0.92	0.85	0.69	0.72	4.31
Total	27	0.67	0.78	0.59	0.93	0.78	0.82	0.76	4.56
Post-feedback									
Hard sub-group	14	0.43	0.43	0.36	0.50	0.36	0.50	0.43	2.57
Easy sub-group	13	0.77	0.69	0.62	0.92	0.92	0.92	0.81	4.85
Total	27	0.59	0.56	0.48	0.70	0.63	0.70	0.61	3.67
<u>Experiment 4</u>									
Pre-feedback									
Hard sub-group	30	0.23	0.24	0.17	0.24	0.20	0.13	0.20	1.20
Easy sub-group	34	0.21	0.21	0.29	0.35	0.21	0.33	0.27	1.59
Total	64	0.22	0.22	0.23	0.30	0.21	0.23	0.24	1.41
Post-feedback									
Hard sub-group	30	0.20	0.14	0.13	0.24	0.20	0.20	0.19	1.10
Easy sub-group	34	0.50	0.53	0.38	0.50	0.44	0.50	0.48	2.85
Total	64	0.36	0.34	0.26	0.38	0.33	0.36	0.34	2.03