Research Project: Learning to be less miscalibrated

by Marie-Pierre Dargnies and Guillaume Hollard

In this experiment, we investigate how people can be taught to be less miscalibrated. Miscalibration is one of the several forms taken by overconfidence. It is usually measured by asking subjects 90% confidence intervals for a set of questions. A perfectly calibrated subject's intervals should contain the correct answer 90% of the time. It is actually found that all people are miscalibrated. Glaser, Langer and Weber [2005] found that the correct answer falls outside the confidence intervals provided by the subjects about 70% of the time exhibiting a high level of miscalibration. People are, to a certain extent, aware that they are affected by the miscalibraton bias. Indeed, when asked how many of the correct answers to the ten questions they have provided confidence intervals for they think fall outside of the intervals they have given, their average answer is 4,18. Nevertheless, the actual number of surprises exceeds their prediction, indicating that miscalibration is not only due to the fact that people do not understand what a 90% confidence interval is.

Being miscalibrated has economic consequences. People exhibiting a high level of miscalibration earn less profits on experimental financial markets. It is all the more interesting that traders are found to be even more miscalibrated than lay men. Biais, Hilton, Mazurier and Pouget [2005] in an experiment mimicking a financial market in a laboratory find evidence that the more miscalibrated people are the more they are prone to suffer from winner's curse.

Knowing that miscalibration is correlated with the performance of traders on a financial market, it would be interesting to find out, first whether people can learn to be less miscalibrated and second, whether this learning improves their performance on the financial market.

Our experiment aims at finding out whether people can be taught to be less miscalibrated. In the following sections we will present two experimental protocols conceived to reduce subjects' miscalibration and some preliminary results. In the future, it would be interesting to investigate the economic consequences of such a training.

1 First attempt: making miscalibration costly

The experimental subjects were separated in two groups. The subjects of the first group attended a training session aiming at reducing their miscalibration bias before they were asked to provide 90% confidence intervals for a set of ten questions. The second group was a control group, the subjects just answered the same set of ten questions with 90% confidence intervals allowing to check whether the training was successful in reducing miscalibration in the first group.

1.1 Experimental design

The experiment took place at the laboratory of experimental economics of Paris 1 in january 2007. 58 subjects participated in the experiment. We used a between treatment: 21 undergraduate students in mathematical economics went through the training session before they completed the 90% calibration task while the control group was composed of 37 students in various majors who just completed the 90% calibration task.

1.1.1 The training session

In the training session, the participants were asked to answer a set of ten questions by providing intervals. The set of questions used in the training session as well as the one used in the next step were the ones used in Biais and al's experiment. In this training session, the subjects were provided for each question with a reference interval they can be 100% sure that the correct answer belongs to. They had to give an interval included in the reference interval. The payoff rule was the following, the payoffs are expressed in experimental currency (ECU).

- If the interval they provided did not contain the correct answer to the question, they earned nothing for this question.
- If the interval provided contained the correct answer to the question, the narrower the interval provided, the higher the payoff according to the following formula:

$$\text{Payoff} = 100 \times \left[1 - \frac{\text{Length of the interval provided by the subject}}{\text{Length of the reference interval}}\right]$$

According to this formula, the payoff is maximal and equal to 100 when the interval provided by the subject is a unique value, this value being the right answer to the question. The payoff is equal to zero if the subject provides the reference interval and then takes no risk at all. There is therefore a trade-off between the risk taking and the amount of ECU a subject could get if the correct answer fell inside of his interval.

After they had answered a question, the subjects got a feedback providing them with their payoff for the question they had just answered. They could easily infer from this feedback whether they had taken too much risk.

People being miscalibrated, we expected them to realize it as they saw that the correct answer fell outside of their interval much more often than they expected it which resulted in a loss of money. As a result, we expected they would give wider intervals for the next questions.

1.1.2 90% confidence intervals

In the next step, the subjects that have participated to the training as well as the subjects of the control group were asked to give 90% confidence intervals for a set of ten questions. There was no feedback between the questions. We expected that the individuals who attended the training would be less miscalibrated than the subjects of the control group.

1.1.3 Third step: Questions concerning the experiment

After they had answered the ten questions, all the subjects were asked questions concerning the 90% confidence intervals they had given in the previous step. For instance, they were asked how many correct answers they thought fell inside of the intervals they had given, how many correct answers they thought fell inside of the intervals given by the average subject, whether they thought they had taken more risk than the average subject... This allows us to check whether the subjects exhibit an other form of overconfidence, namely better than average effect for the miscalibration task.

1.2 Preliminary results

Contrary to what we expected, we do not find that the subjects who went through the training session are better calibrated than those who did not. Actually and weirdly enough, subjects who went through the training appear to be even ***more*** miscalibrated even if the difference is not significative.

A lot less than 9 correct answers out of 10 fall inside the 90% confidence intervals for both the subjects who did and did not go through the training. The evaluation of the hit rate is higher for the sujects from the control group but the difference is not significative. While there is no significative difference between the evaluation of one's hit rate and the evaluation of the average subject's hit rate fo the subjects of the control group, there is a significative difference for the subjects who went through the training who anticipate a higher hit rate for themselves.

We can wonder why we find such surprising results.

2 Second attempt: step-by-step calibration

The subjects were split into two groups. Subjects from one group had to answer to a set of ten questions by providing first 50% and then 90% confidence intervals while subjects from the control group just had to provide 90% intervals for the same set of ten questions. The subjects from a third group had to provide 50% confidence intervals in order to allow us to check there was no initial difference in miscalibration at the 50% level between the group we wanted to cure and other subjects. We intended to compare the miscalibration at the 90% calibration task from the group who had to provide both 50% and 90% intervals and the group who just gave 90% intervals.

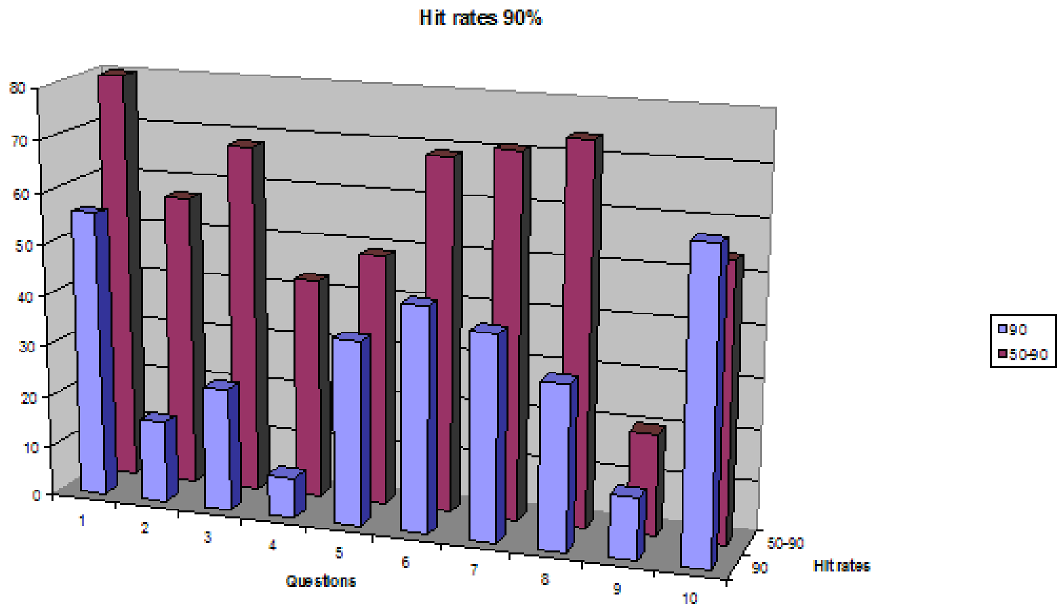
2.1 Experimental design

The experiment took place at the university of Marne-La-Vallée in february 2007. 86 students in economics participated in the experiment. 26 students provided 50% confidence intervals, 25 students provided 90% confidence intervals and 35 students provided both 50% and 90% confidence intervals all for the same set of ten questions.

In addition to providing confidence intervals, subjects were asked to estimate how many correct answers would fall inside their confidence intervals.

2.2 Preliminary results

It appears very clearly that the subjects who gave both 50% and 90% confidence intervals did a lot better at the 90% calibration task than the subjects who just gave 90% intervals. Not only are their hit rates significatively higher, the intervals they provided are also much wider.



The performances at the 50% calibration task from the subjects who provided 50% and 90% intervals and the subjects who only provided 50% intervals are not significatively different (about 3 correct answers out of ten fall inside their intervals). While we would expect that people anticipate that 9 correct answers out of ten fall inside their 90% intervals, they actually expect a much lower hit rate. Interestingly enough, subjects who did the step-by-step calibration (who did better at the 90% calibration task) expect a higher hit rate (7,2) than those who did not (5,6). However, for the 50% calibration task, subjects expect that about 5 correct answers out of 10 will fall in their 50% intervals (4,5 for the subjects who just provided 50% intervals and 4,3 for those who proided first 50% intervals and then 90% intervals). Maybe it is the case that people better understand what is a 50% confidence interval. The gap between the evaluation and the performance is smaller for the 50% calibration task (1,8 and 0,9 respectively for the subjects who just provided 50% intervals and for those who provided both 50% and 90% intervals) than for the 90% calibration task (respectively 2,4 and 1,4). For the 90% calibration task, this gap is smaller for the subjects who had first given 50% confidence intervals: the evaluations of the hit rates and the hit rates are both higher for subjects who did the step-by-step calibration but the difference in hit rates between the two groups exceeds the difference in evaluation. About 5,9 correct answers out of 10 fall inside the 90% intervals from the subjects who first provided 50% intervals and their evaluation is 7,2 corresponding respectively to a hit rate 2,6 points and an evaluation 1,6 points higher than the corresponding values of the control group. Both gaps between the hit rates and the evaluation are significative.

3 What we intend to do next

Most past works about miscalibration have focused on calibration at high levels, 90% generally and report a wide-spread miscalibration bias. However, when we asked for 50% confidence intervals, we observe that miscalibration is still present but to a lower extent. We also note that people expect that almost 5 correct answers out of 10 will fall inside of their 50% confidence intervals. This leads us to investigate calibration at lower levels. Furthermore, we wish to investigate the differences in miscalibration at these different levels between people who are destined to become traders and the rest of the population.

In these purposes, we intend for a set of ten questions to ask both students in finance and in various majors to give their best guess of the correct answers and confidence intervals at 1%, 10%, 50%, 90%, 99%. They will also be asked to evaluate their hit rate and the average subject's hit rate. This should allow us to find out whether there exists some differences in calibration and in evaluation between the different levels. It could for instance be the case that the miscalibration is reversed at low levels and that more than 10% of correct answers fall inside the 10% confidence intervals provided. It is also possible that people expect a hit rate higher than 10% or that their actual hit rate excceds their expectation. We would be interested in studying the differences both in hit rates and in expectations between people who are destined to work in finance whose choices and expectations in the movement of financial assets will matter and lay men for whom miscalibration is less likely to have an influence.