Self-Confidence in Information Processing: Experimental Methods and Evidence

Sidak Yntiso*

NYU Abu Dhabi, Class of 2014
sg210@nyu.edu

Abstract

Do individuals interpret ambiguous feedback optimistically or pessimistically when their self-esteem is at stake? Ertac (2011) finds that generally people overweigh negative performance feedback, which leads to overly-pessimistic beliefs about ability. If these results are valid and generalizable, more studies are necessary to study the effects and extent of pessimism. Towards this end, I have replicated Ertac's experiment with a number of differences in methodology (performance groups, signal and payment). In confirmation of her findings, after participants received an accurate yet negative signal, there was a significant pessimistic effect in belief updating. This paper also compares belief-accuracy across two types of belief elicitation mechanisms — the quadratic scoring rule and the crossover mechanism. There is no significant difference in belief accuracy between the two mechanisms.

Keywords: *self-confidence, belief updating, elicitation mechanisms*

I. Introduction

MANY psychology and recent economics studies show that human beings exhibit positive confidence biases when making decisions about their self-image. People are too confident about their own abilities, too confident about their control and too optimistic about the future. These biases could exist to promote happiness and contentment, the ability to care about others, and the ability to be productive and creative (Taylor and Brown 1988). Factors influencing the extent of these optimistic biases in any given trait include, the degree of desirability, the perceived probability, previous experience with the trait, the degree of controllability or the existence of representativeness heuristic (Kahneman and Tversk, 1972), and stereotyping (Weinstein 1980).

Other studies have shown that these biases have real world consequences in market entry, stock trading and investment (e.g. Camerer and Lavollo 1999; DellaVigna and Malmendier 2006; Odean 1999; Malmendier and Tate 2008). There are also a number of economic implications of my study for optimal feedback designs in organizations and schools (Ertac 2006). Institutions should seek to relate performance feedback in a way that does not induce pessimism or adversely affect morale.

What is meant by optimism or overconfidence, and pessimism or underconfidence? Many studies have defined optimism as when more than x percent of the population believes they are in the top y percent of population (x>y) in performance on a task (Benoit and Dubra 2011). This paper differs from those studies in that this experiment required the complete belief distribution of all possible states, not just a participant's estimate of their rank. The complete belief distribution must average out to the true distribution, which if weighted towards high ability is evidence of overconfidence. Another definition of overconfidence, as in Mobius et al. (2011), is a persistent difference between responsiveness and positive and negative signals. Eil and Rao (2011) similarly define overconfidence as a positive asymmetry in beliefs caused by ignoring strong negative signals. This form of optimism is not incompatible with Ertac's (2011) definition of pessimism. With this form of pessimism, subjects are pessimistic when they overweigh the importance of an uninformative and unfavorable signal and therefore assign high likelihoods to being in lower performance groups.

This senior thesis tests whether Ertac's pessimism is replicable despite several differences in experimental methodology. If pessimism can be shown, this will feed into a more complicated picture of self-confidence biases in information processing. Additionally, this paper compares the effectiveness of two belief elicitation tools.

*I would like to thank Professor Yaw Nyarko for his advice and support. Thanks are also due to Professors Jonathon Rogers and Rebecca Morton for help running the experiments. Finally, thank you to the faculty and staff at the NYUAD Division of Social Sciences for your valuable support.

II. Literature Review

The studies most closely related to this one follow the same basic methodology, though they differ in signal structure, task and belief elicitation (Mobius et al. 2011; Eil and Rao 2011; Ertac 2011; Grossman and Owens 2011; Charness et al. 2011). First, subjects complete an ego-dependent task i.e. a task for which they care about their performance (as opposed to an ego-independent task, for instance, your beliefs about a random number). Subjects are made to report their beliefs about their performance. Then subjects receive noisy or biased, positive or negative feedback and are asked to report their beliefs again. These papers provide compelling evidence that there are significant differences in belief updating in ego-independent and ego-dependent tasks. Therefore, it is unnecessary to include an ego-independent control task in this current study.

Mobius et al. (2011) tracked the evolution of beliefs about relative performance on an IQ test and the evolution of beliefs about performance on an ego-independent task (a robot's IQ score). The authors required subjects to report a single number, representing their belief about the probability of being in the top half of performance on the IQ test. From this, they computed subjects' responsiveness to positive and negative signals and found that the difference between the two was positive and significant (subjects over-weigh positive information). Furthermore, the posterior beliefs for 224 subjects who received two positive and two negative signals (and hence no new information at all) were significantly higher than the priors. This positive asymmetric bias was insignificant when the same subjects with the same priors observed the same information flow about a robot's IQ. The authors concluded that positive asymmetry leads to overconfidence; they also presented evidence of persistence and conservatism.

Eil and Rao (2011) used broader tasks (beauty and intelligence) as ego-dependent tasks and a randomly selected number from one to ten as an ego-independent task. They used a ten-step ranking procedure, with the beauty task subjectively graded by fellow subjects. The reported signal was a message: "You are ranked higher" or "You are ranked lower," than another randomly selected participant in the study. Beliefs were incentivized with the quadratic scoring rule.

Eil and Rao (2011) found that the direction of feedback affected self-confidence biases. Namely, when given positive feedback, inference more closely resembles Bayes rule in accuracy and precision, with an optimistic bias. But when given negative feedback, individuals tend to ignore feedback, have more erratic updating and are more information adverse. This asymmetry was not the case in the ego-independent control task (Eil and Rao 2011).

Finally, Ertac (2011) used a three-level belief elicitation mechanism where subjects can be either in the top 20 percent, middle 60 percent, or bottom 20 percent categories. She used simple addition and GRE verbal tests as the ego-dependent tasks. She also used the quadratic scoring rule to compensate belief estimates. Subjects were in one of two sessions — where they receive either a top/not top or a bottom/not bottom signal. The top/not top signal tells individuals whether they were in the top portion or not. The bottom/not bottom signal works analogously. If subjects receive a top signal, there is no ambiguity. If they receive a not top signal, they can interpret this as meaning they are in bottom 20 percent or middle 60 percent.

She confirms that subjects deviate more from Bayesian benchmarks in ego-dependent tasks. However, she also finds general evidence of pessimism — overweighing negative feedback. She shows that after receiving a signal stating that they were not in the top portion, subjects were less likely to believe they were in the middle and more likely to believe they were in the bottom portion than a Bayesian agent would. The average difference between reported beliefs of being in the middle category and the Bayesian posteriors of being in this category was negative and significant. The not top signal does not warrant this overly pessimistic change of belief. Furthermore, this bias is not significant after receiving the Not Bottom signal. Interestingly, female subjects displayed greater pessimistic deviations then men did. On the other hand, self-confident subjects (subjects with high priors) had a self-serving bias (reacting to positive signals more optimistically than rational).

The innovative top/not top signal structure may have captured how people may be pessimistic processers in certain situations. However, whether using the quadratic scoring rule is justified in these kinds of experiments (see Section 3.2). Also, the ego-dependent tasks (addition and GRE verbal questions) are very specific. Arguably, this pessimism effect may not exist in more salient ego-dependent tasks. Finally, this paper differs from Ertac's in its use of five performance categories. With five categories, once two have been removed, subjects can display more accurate beliefs by placing themselves in one of three remaining categories.

The second set of results in this study has to deal with belief incentivization. There are many elicitation tools used in experimental economics — Becker-DeGroot-Marschak (Becker et al. 1964), the scorings rule, the crossover mechanism and others (Karni 2009). With so many more or less mathematically reliable

(monotonicity, salience and dominance are necessary conditions (Smith 1976)), which tool should experimenters use? Researchers cannot ignore belief incentivization altogether because evidence shows that having incentives increases the accuracy of stated beliefs (Gachter and Renner 2006; Grether 1992). Therefore, it is necessary to compare the existing tools by how often they are associated with accurate belief reporting.

III. Experimental Designs and Procedures

I. Experimental Designs

The experiment was conducted through March and April 2014, in the Social Science Experimental Laboratory at NYU Abu Dhabi with university funding for payoffs. A total of sixty subjects were successfully recruited. The subjects completed the experiments in groups of 5 or 10 on computers using z-tree.

First, subjects reported their beliefs about being in each of 5 categories of relative ability on the exam (each category representing 20 percent of the population respectively). Then there was one practice round consisting of a difficult question and another belief stage as well as a short quiz on instructions. The subjects reported their beliefs again (pre-exam priors). The full exam consisted of 15 Mensa level questions (Grosswirth et al. 1999). These questions are difficult enough that subjects probably have a vested interest in how they do and they have been used in the literature (Grossman and Owens 2011). Then subjects reported their beliefs again (post-exam priors).

In the feedback phase, subjects received a signal stating whether they were in the top two (Not Bottom Signal) or not in the top two groups (Not Top Signal). They were made to report their beliefs about being in each of the five categories again. Finally in the payment stage, each subject's belief accuracy was rewarded by either the crossover mechanism as used in Mobius et al. (2011) or the quadratic scoring rule.

II. Belief Elicitation

The quadratic scoring rule was first introduced by Brier (1950). It is now the most popular elicitation mechanism in experimental economics (Hollard et al., 2010). In the experiment, the quadratic scoring rule was used:

$$1 + 20 * p(x) - w, \tag{1}$$

where *p(x)* is the likelihood expressed for the actual value of X (the true performance category) and, w is the sum squares of the likelihoods attached to each belief category.

There are two problems with the use of this rule. Firstly, the quadratic scoring rule is strictly only incentive compatible with risk neutral subjects. Offerman et al. (2009) discuss the risk problem of using proper scoring rules and propose a solution. Ertac (2011) used data from the ego-independent task to identify risk neutral subjects and was able to replicate her findings in this subgroup.

Secondly, the quadratic scoring rule provides the opportunity for risk averse subjects to hedge their stated beliefs against adverse outcomes. For example, subjects may perform poorly and state low beliefs. It is noteworthy that Blanco et al. (2008) found no evidence of this kind of hedging when comparing a hedging-prone elicitation design with a hedging-proof elicitation design in the prisoner's dilemma setting.

The quadratic scoring rule and three examples employing it were included in the instructions read out loud before the experiment. Participants were also made to apply the rule to a hypothetical case during the preliminary quiz to ensure they understood its reliability. Finally, participants were also informed that "you *will get paid the most when you honestly report your best guesses about the likelihood of each of the different possible outcomes*," several times throughout the experiment.

The crossover mechanism (Mobius et al. 2011) is another reliable elicitation mechanism. In this study, I have extended its use to five quintiles. With the crossover mechanism, participants are asked for a value of x such that they are indifferent between:

a. Receiving 20 AED (roughly 6 USD) with certainty if they are in each category

b. Receiving 20 AED with probability x, x $\in$ (0, 0.01, 0.02, . . . , 0.99, 1)

Then computer program randomly selects y,

a. If y>x, receive 20 AED with probability y

b. If y<x, receive 20 AED only if you are in the top category.

The crossover mechanism solves the two problems mentioned above. It is strictly incentive compatible with non-standard preferences such as Prospect theory and only assumes monotonicity. It also garners belief payoffs that are strictly increasing with performance (Mobius et al. 2011). However, the crossover mechanism cannot be meaningfully extended to belief elicitation across more than two incompatible categories or events. The problem in doing so is that the reported beliefs are

no longer probabilities; for example, subjects can attach 100 percent likelihood to incompatible events like being in both the bottom and top quintiles. Despite these interpretative concerns, the findings in the belief accuracy section (Section 5.8) indicate that these probabilities are as accurate or more accurate than the quadratic scoring rule in predicting subjects' actual performance quintile.

One solution to this problem of analysis is to divide the reported belief for each category by the sum of beliefs for every category and so produce an average measure of likelihood that could be understood as a probability. It is unclear whether this measure actually corresponds to subjects' beliefs. Therefore, I have dropped the crossover mechanism elicited beliefs from analysis for self-confidence biases.

III. Hypothesis

The main experimental evidence hypothesis tested is: (a) Subjects have correct initial beliefs i.e. there is no overconfidence or under confidence. (b) Subjects update their beliefs after viewing a signal based on Bayes' rule. If they observe a not-bottom signal:

$$P_T^{'} = \frac{P_T}{P_T + P_{NT}} \tag{2}$$

$$P_{NT}^{'} = \frac{P_{NT}}{P_T + P_{NT}} \tag{3}$$

$$P_M^{'} = P_N^{'}B = P_B^{'} = 0 \tag{4}$$

If they observe a not-top signal:

$$P_M^{'} = \frac{P_M}{P_M + P_{NB} + P_B} \tag{5}$$

$$P_{NB}^{'} = \frac{P_{NB}}{P_M + P_{NB} + P_B} \tag{6}$$

$$P_B^{'} = \frac{P_B}{P_M + P_{NB} + P_B} \tag{7}$$

$$P_{T^-}^{'} = P_{NT}^{'} = 0 \tag{8}$$

where $P^{'}{}_T$, $P^{'}{}_{NT}$, $P^{'}{}_M$, $P^{'}{}_{NB}$ and $P^{'}{}_B$ denote the posterior probability of being in the top, next to top, middle, next to bottom and bottom categories (P_T,P_{NT}, P_M, P_{NB}, P_B denote the prior probabilities).

Receiving feedback generates neither overconfidence nor underconfidence. The explicit measure of interest is the average difference between the subjective posterior and Bayesian posterior of being in the middle category.

IV. Results

I. Task Performance

Across all sessions, there were 60 participants — 32 male and 28 female (see Table 3). The average score on the quiz was 8 out of a possible 15 points. There was no significant difference in performance between male subjects and female subjects (p=0.16). The results for the quadratic scoring rule sessions are presented first. Then the two types of sessions are compared for elicited belief accuracy.

II. Evolution of a-priori beliefs

Subjects were required to present their beliefs immediately before and after taking the quiz to control for the possibility that taking the quiz had better informed participants of their performance. This was not the case, there was no significant change in participants' beliefs except for beliefs about being in the next to bottom state (p = 0.028) (see Table 4). From this point on, I will be referring to the post-exam priors.

III. Belief Accuracy

The first result is that subjects do not, on average, place themselves across the five quintile groups equally. Table 5 summarizes the breakdown of actual belief categories against perceived categories. As a group, all subjects slightly over-weigh the probability of being in the next to top and middle categories and under-weigh the remaining categories. The Actual Top performance participants accurately place the highest weight on being Top, but this may reflect either self-awareness or overconfidence. On average, most performance groups assign the highest prior belief to their actual category, which is evidence of self-awareness. Finally, the Actual Bottom scorers consider themselves only slightly more likely to be in the bottom quintile as in the top quintile, which is some indication of initial overconfidence.

IV. Information Processing

Table 1 presents the subjective posteriors and Bayesian posteriors for each belief category. The significant point is the bias measure, the difference between the two. After receiving a negative (Not Top, NT) signal, participants under-weigh the likelihood of being in the middle category by 26 percent as compared to a Bayesian. This is significant evidence of pessimism, similar to findings in Ertac (2011).

Table 1: *Biases in Information Processing*

	Bottom	Not Bottom	Middle	Not Top	Top	N
			Subjective Posterior			
Total	0.13	0.23	0.20	0.18	0.26	50
Not-Top Signal	0.22	0.38	0.33	0.05	0.02	30
	(0.024)	(0.037)	(0.033)	(0.022)	(0.015)	
Not Bottom Signal	0.00	0.00	0.00	0.38	0.62	20
	(0.000)	(0.000)	(0.000)	(0.043)	(0.043)	
			Bayesian Posterior			
NT Bayesian posterior	0.154	0.261	0.585			
NB Bayesian posterior				0.543	0.457	
			Bias			
NT Signal	0.063	0.122	**−0.260**	0.051	0.024	30
NB Signal	0.000	0.000	0.000	−0.162	0.161	20

Participants over-select into the bottom and next to bottom categories 6 percent and 12 percent of the time respectively. It is significiant that subjects also over-select into irrelevant top categories. This indicates that some subjects did not understand the experiment completely. However, the magnitude of the observed pessimism about the middle category, as compared to the other lower categories, suggests that pessismism not misunderstanding was more important (see Section 5.5 for a discussion of possible cognitive limitations).

After receiving a positive (Not Bottom, NB) signal, participants under-weigh the likelihood of being in the next to top category 16 percent of the time and overweigh the likelihood of being in the top quintile 16 percent of the time. This is evidence of overconfidence in information processing.

V. Cognitive Limitations

Mobius et al. (2011) point out that patterns in information processing could easily arise from cognitive limitations, such as inattentiveness, laziness or misunderstanding, rather than motivated behavior. Presumbably, if these causes were the case, more able subjects (as determined by scores) would make less mistakes on average. If we use the average score (payoff) as a measure of ability, we find no evidence of this hypothesis (Table 6). The size of the individual bias (difference between submitted and Bayesian posterior) is not significantly correlated with the average score. This gives credence to interpreting these patterns as motivated behavior. Furthermore, pessimism was not an effect in Ertac's ego-indepdent control case.

VI. Comparative Belief Accuracy

Eliciting subjective beliefs assumes that humans have a true belief about these categories. However, subjects may not consciously understand objective probability theory (or implicitly use it), so showing that one tool is more closely associated with Bayesian inference than another does not prove its effectiveness (Hollard et al. 2010). Instead, the most effective mechanism must be associated with the greatest accuracy of elicited beliefs in predicting future outcomes.

Across all subjects, two belief accuracy measures were constructed. The first measures the relationship between the performance category the subject viewed as most likely and her actual performance category (Ertac 2011). For example, if a subject reported a 40 percent likelihood of being in the middle category and that was her highest reported probability, then her perceived perfomance is the middle quintile.

As you can see in Table 2, both the crossover mechanism and the quadratic scoring rule invoked accurate perceived performance 30 percent of the time. But the quadratic scoring rule is associated with more overconfident priors. Across all subjects, initial overconfidence occurred 36.7 percent of the time, which provides further evidence for initial overconfidence.

Another measure for comparing elicitation mechanisms is the calibration index (Hollard et al. 2010). Calibrated stated beliefs minimize the difference between the predicted and actual success state. I have used the following measure of calibration:

Table 2: *Comparative Belief Accuracy. Note: CM is the Crossover Mechanism and WSR is the Quadratic Scoring Rule*

	CM	QSR	Total	Difference (QSR−CM)	Prob. (QSR = CM)
Overconfident	30% (0.483)	38% (0.516)	36.7% (0.486)	8% (0.169)	0.6386
Accurate	30% (0.483)	30% (0.516)	30% (0.463)	0% (0.1462)	1
Underconfident	40% (0.49)	32.00% (0.4712)	33.33% (0.475)	−8.00% (0.169)	0.638
Calibration Index	0.444 (0.125)	0.375 (0.112)	0.386 (0.116)	0.69 (0.039)	0.0873
N	10	50	60		

$$\text{calibration index: } \frac{1}{n}\sum_{j=1}^{n}\sqrt{\frac{1}{5}\sum_{i=1}^{5}(p_i - x_i)^2} \qquad (9)$$

For n individuals, p_i is the stated probability for the i^{th} performance group, x_{i} is an indicator variable that takes value of 1 only if the individual is in that actual performance group.

Therefore, a perfectly calibrated mechanism has a calibration index of 0 (Hollard et al. 2010). A higher value of the calibration index implies less belief accuracy.

The crossover mechanism is associated with a slightly lower calibration index than the quadratic scoring rule (significant at a 90 percent level). In unison, these results about belief accuracy indicate that neither the crossover mechanism nor the quadratic scoring rule is much more effective than the other.

V. Discussion

I have assumed that if the other reasons subjects incorrectly process information (unexplained pessimistic-looking heuristics, biases and cognitive limitations) exist, they should be correlated with ability for participants in the middle category. Even if this is so, it may be that all middle quintile participants do not have the ability to ensure managing these biases. Therefore, showing that the pessimistic bias is not correlated with ability for the middle quintile participants is not sufficient to show that cognitive limitations may not play a part in information processing. The cognitive limitations problem could be corrected by taking an ego-independent control task and comparing the evolution of beliefs across the two tasks (Mobius et al. 2010).

Another possible limitation of the current study is the error caused by the sample size of only 60 students. This can be corrected with further, larger studies employing different tasks.

VI. Conclusion

In conclusion, this paper first directly tests the Ertac (2011) findings that negative yet uninformative feedback is processed over-pessimistically. Uniquely, this paper has introduced a new ego-dependent task for which participants have more ego invested, five ability categories and a different signal structure to the Ertac (2011) methodology. The average difference between the subjective posterior and Bayesian posterior of being in the middle state was significantly negative, confirming Ertac's findings. In this case, individuals exhibit negative self-image biases. This paper also shows evidence of initial optimism and optimism in updating after participants receive a positive signal. These findings necessitate more study and a more complex and multi-faceted understanding of human processing in ego-dependent activity.

Secondly, this paper tests the salience of the quadratic scoring rule and the crossover mechanism against each other. There is little evidence to support the use of one tool over the other. This implies that the quadratic scoring rule can be used as in this study or in Ertac (2011), despite the shortcomings mentioned above. The scoring rule may be preferable in these studies because it can be applied across more than two events or outcomes.

References

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak. 1964. "Measuring Utility by a Single-

response Sequential Method." *Behavioral Science* 9(3): 226−32.

Benoit, Jean-Pierre, and Juan Dubra. 2007. "Overconfidence?" Working Paper 6017, MPRA.

Blanco, Mariana, Dirk Engelmann, and Alexander K. Koch. 2008."Belief Elicitation in Experiments: Is There a Hedging Problem?" Discussion Paper 3517, Institute for the Study of Labor.

Brier, Glenn W. 1950."Verification Of Forecasts Expressed In Terms Of Probability." *Monthly Weather Review* 78(1): 1−3.

Camerer, Colin, and Dan Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review* 89(1): 306−18.

Della Vigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* w 96(3): 694−719.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven 2011. "Self-Confidence and Strategic Deterrence." Working Paper 11-151/1, Tinbergen Institute Discussion Papers.

Eil, David, and Justin M. Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3(2): 114−38.

Ertac, Seda. 2011. "Does Self-relevance Affect Information Processing? Experimental Evidence on the Response to Performance and Non-performance Feedback." *Journal of Economic Behavior & Organization* 80(3): 532−45.

Ertac, Seda. 2005. "Social Comparisons and Optimal Information Revelation: Theory and Experiments.." PhD diss., University of California, Los Angeles.

Gachter, Simon, and Elke Renner. 2006. "The Effects of (Incentivized) Belief Elicitation in Public-Goods Experiments." Working Paper 2006-16, CeDEx.

Grether, David M. 1992. "Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence." *Journal of Economic Behavior & Organization* 17(1): 31−57.

Grossman, Zachary, and David Owens. 2011. "An Unlucky Feeling: Persistent Overestimation of Absolute Performance with Noisy Feedback." Working Paper 02−11, Department of Economics at UC Santa Barbara.

Grosswirth, Marvin, Abbie F. Salny, and Alan Stillson. 1999. *Match Wits with Mensa: The Complete Quiz Book (Mensa Genius Quiz)*. Cambridge, MA: De Capo Press.

Hollard, Guillaume I., SÃľbastien Massoni, and Jean-Christophe Vergnaud. 2010. "Subjective Beliefs Formation and Elicitation Rules: Experimental Evidence." Working Paper, UniversitÃľ Paris1 PanthÃľon-Sorbonne.

Kahneman, Daniel, and Amos Tversky. 1972. ""Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3(3): 430−54.

Karni, Edi. 2009. "A Mechanism for Eliciting Probabilities." *Econometrica* 77(2): 603−06.

Klar, Yechiel, and Eilath E. Giladi. 1997. "No One in My Group Can Be below the Group's Average: A Robust Positivity Bias in Favor of Anonymous Peers." *Journal of Personality and Social Psychology* 73(5): 885−901.

Malmendier, Ulrike, and Geoffrey Tate. 2008. "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction?." *Journal of Financial Economics* 89(1): 20−43.

MÃűbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2011. "Managing Self-Confidence: Theory and Experimental Evidence." Working Paper 11-14, Federal Reserve Bank of Boston.

Odean, Terrance. 1999. "Do Investors Trade Too Much?." *American Economic Review* 89(5): 1279−298.

Offerman, Theo, Joep Sonnemans, Gijs Van De Kuilen, and Peter P. Wakker. 2009. "A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes." *Review of Economic Studies* 76: 1461−489.

Smith, Vernon L. 1976. "Experimental Economics: Induced Value Theory." *The American Economic Review* 66(2): 274−79.

Taylor, Shelley E., and Jonathon D. Brown. 1988. "Illusion and Well-being: A Social Psychological Perspective on Mental Health." *Psychological Bulletin* 103(2): 193-210.

Weinstein, Neil D. 1980. "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology* 39(5): 806−20.

VII. Appendix A

Table 3: *Data*

	Sessions 1−2	Sessions 3−11
Belief Elicitation	Crossover Mechanism	Quadratic Scoring Rule
No. of Performance Categories	5	5
No. of Subjects	10	50
No. of Observations	30	150
Average Score	7.86	9.4

Table 4: *Difference between pre-exam and post-exam prior beliefs; Note: Interpret significance levels as * (p<0.10), ** (p<0.05), *** (p<0.01)*

	Pre-exam prior	Post-exam prior	Mean Difference	Prob (Pre−exam =Post-exam prior)	N
Bottom	0.17	0.12	−0.013	0.323	50
	(0.052)	(0.25)	(0.046)		
Next to Bottom	0.14	0.18	−0.005	0.689	50
	(0.021)	(0.036)	(0.013)		
Middle	0.21	0.31	−0.012	0.690	50
	(0.044)	(0.109)	(0.031)		
Next to Top	0.32	0.35	−0.019	0.371	50
	(0.049)	(0.048)	(0.021)		
Top	0.22	0.17	0.05	0.122	50
	(0.031)	(0.027)	(0.032)		

Table 5: *Breakdown of Subjective Priors by Performance Group*

	Bottom	Not Bottom	Middle	Not Top	Top
Actual Top	0.13	0.06	0.18	0.32	0.39
	(0.023)	(0.023)	(0.055)	(0.048)	(0.095)
Actual Next to Top	0.08	0.13	0.26	0.35	0.18
	(0.033)	(0.035)	(0.043)	(0.048)	(0.042)
Actual Middle	0.17	0.19	0.19	0.26	0.07
	(0.073)	(0.053)	(0.109)	(0.097)	(0.022)
Actual Next to Bottom	0.18	0.24	0.28	0.19	0.11
	(0.036)	(0.028)	(0.028)	(0.033)	(0.020)
Actual Bottom	0.17	0.25	0.25	0.25	0.09
	(0.052)	(0.049)	(0.051)	(0.075)	(0.028)
Total	0.13	0.17	0.25	0.27	0.17
N	50	50	50	50	50

Table 6: *Cognitive limitations and gender by bias measure; Note: Bias = Subjective Bayesian posterior of the middle quintile*

	Bias Coefficient	T	P > t	R−sq	N
Ability (Score)	0.0174 (0.0239)	0.475	0.6393	0.023	29
Male	0.0712 (0.1026)	0.60	0.493	0.018	29

Table 7: *Differences between men and women beliefs by actual group*

	Subjective Priors			Subjective Posteriors		
	Difference	Prob (M=F)	N	Difference	Prob. (M=F)	N
Actual Bottom	.015 (0.109)	0.2114	3	0.03 (0.106)	0.7674	7
Actual Next to Bottom	.072 (0.056)	0.2342	4	−0.06 (0.078)	0.4787	6
Actual Middle	−0.01 (0.273)	0.9813	6	0 (0.211)	1	2
Actual Next to Top	−0.13 (0.093)	0.2011	6	−0.03 (0.108)	0.7554	4
Actual Top	−0.30 (0.139)	0.1565	7	−0.17 (0.156)	0.3209	3

Table 8: *Biases in Information Processing with Overconfidence*

	Subjective Posterior for Middle Category					
	Self-Confident Participants	Non-Confident Participants	Difference	T	P > t	N
Not-Top Signal	0.34 (0.061)	0.37 (0.059)	−0.03 (0.0864)	−0.39	0.69	25
Not-Top Bayesian posterior	0.457	0.460				
	Bias					
Bias Measure (for middle category)	−0.119	−0.089				

VIII. Appendix B: Crossover Mechanism Instructions

Instructions

Welcome to the Social Science Experimental Laboratory. This is an experiment on the economics of decision-making. Your earnings will depend partly on your performance, partly on your decisions, partly on the decisions of others and partly on random chance. A 25 AED show-up fee will be added to your earnings, for your participation. You will be paid privately, in cash, at the end of the experiment. You are under no obligation to tell anyone how much you have earned in the experiment.

The experiment will consist of two parts, and the second part will consist of three related steps. We will now explain the procedures of the first part of the experiment. You will also have a chance to practice on the computer, to help you get familiar with the computer program. Please raise your hand if you have any questions throughout the experiment, and the experimenter will assist you.

1. Instructions for Part 1 — Belief Self-Report Rounds

After you take the quiz, the program will rank the scores of all the people who took the quiz and calculate your percentile rank. For instance, if you scored better than 25% of the people who take the quiz, your percentile rank is 25%. Based on your percentile rank you will be put into one of five performance quintile groups âĂŞ the bottom 20%, the next to bottom 20%, the middle 20%, the next to top 20% and the top 20% of relative performance.

a. Crossover Mechanism Rule (optional to read)

Please report what value of x makes you indifferent between

a) Receiving 20 AED with certainty IF you are in the top category

b) Receiving 20 AED with probability x,x $\in$ 0, 0.01, 0.02 . . . , 0.09, 1

After the quiz, the experimenter will know whether you are in the top category or not. For example, you have a 20% belief of being in the top. Then you would have to report x is 20% for indifference between the two. If you reported less, say x=10% then you would strictly prefer option (a).
To decide payoffs, the experimenter randomly chooses a y between 0 and 1.

(a) if y>x, you receive 20 AED with probability y.

(b) if y<x you receive 20 AED only if you are in the top category.

Naturally, the percent chances that you submit need not add up to 100%.

b. Narrative Form (required to read)

Mohammed T

Mohammed T is, on average, the best quiz-taking robot type. There are exactly 100 clones of this robot. Each Mohammed T clone has a different chance of scoring in the top quintile and the chances are given by:

1. Mohammed T 1 has a 1% chance of scoring in the top quintile.

2. Mohammed T 2 has a 2% chance of scoring in the top quintile.

3. . . . etc. . .

4. Mohammed, T 100 has a 100% chance of scoring in the top quintile.

In general, Mohammed T X has an X% chance of scoring in the top quintile. You will be assigned one of the Mohammed T clones by the computer program randomly. You and this Mohammed T are in a team. For payoffs, the computer program will use either your score or the score of the Mohammed T on your team. Whichever score it uses, you will earn 20 AED if that score is in the top quintile of all scores.

But you will have to help the experimenter decide whether to use your score or the robot's score as your team's score. To do this, you will have to guess which Mohammed T you are most like. Which of the 100 Mohammed T clones do you think is as likely as you are to score in the top quintile?

The computer program will pick whichever member of your team is more likely to score in the top quintile. For example, suppose that you say you are as good as Mohammed T 60. If the robot on your team is Mohammed T 34, then your payoff will be based on your score, since you are more likely to score in the top quintile. In this case, you will earn 20 AED only if you are in the top quintile. So if you say you are likely to be in the top quintile (similar to Mohammed T 60), you had better be in the top quintile.

But if the actual robot on your team is Mohammed 97, then the program will use the robot's score, since the robot is more likely to score in the top quintile. In this case, you will earn 20 AED with probability Y (here, Y=97%).

The bottom line is that you are most likely to win 20 AED if you are as accurate as possible when you estimate your probability of scoring in the top quintile and pick the corresponding Mohammed T.

MohammedNT

You face the same problem for each of the other four quintile groups. You will be paired off with one of a hundred Mohammed NT robot clones.

Mohammed NT 1 has a 1% chance of scoring in the top quintile.

Mohammed NT 2 has a 2% chance of scoring in the top quintile.

. . . etc . . .

Mohammed NT 100 has a 100% chance of scoring in the second to top quintile. Mohammed NT X has an X% chance of scoring the second to top quintile. You can help the experimenter pick a Mohammed NT by reporting which Mohammed NT you are most like. Which of the 100 Mohammed NT clones do you think is as likely as you are to score in the second to top quintile? Again, your teamâĂŹs score will depend on either your performance or the performance of the Mohammed NT you have been paired off with. So you can maximize your payoff by correctly estimating your probability of being in the next to top quintile and picking an appropriate Mohammed NT.

Mohammed M

Mohammed M X has an X% chance of scoring the middle quintile. Which of the 100 Mohammed M clones do you think is as likely as you are to score in the middle quintile?

Mohammed NB

Mohammed NB X has an X% chance of scoring the second to bottom quintile. Which of the 100 Mohammed M clones do you think is as likely as you are to score in the next to bottom quintile?

Mohammed B

Mohammed B X has an X% chance of scoring the bottom quintile. Which of the 100 Mohammed B clones do you think is as likely as you are to score in the bottom quintile?

2. Instructions for part 2—Performance Rounds In the second part of the experiment, you will be working on tasks that require ability and effort. In each round, you will be presented with a set of questions to work on within a fixed amount of time. Part of your earnings will come from your performance in these questions. For each question that you answer correctly within this time frame, you will earn 4 AED.

3. Part 3 — Beliefs (same as above)

4. Part 4 — Signal Structure

In this section, you will receive a signal of your relative performance. Interpret the signal as follows:

1. Not Bottom: You are on or above the 60th percentile. You are not in any of the bottom three categories. Please restrict your responses to the top two categories (Second to Top 20% or Top 20%).
2. Not Top Signal: You are on or below the 60th percentile. You are not in either of the top two categories. Please restrict your responses to the bottom three categories (bottom 20%, second to bottom 20% or middle 20%).

5. Part 5 — Beliefs (same as above)

6. Part 6 — Payoffs

Your total earnings in a given round in this first part of the experiment will be the sum of the show-up fee (25 AED), the earnings you get from task performance (4 AED*number of correct questions), and the payoffs you get from the average of the initial, second and third assessment of your performance.

Experiment Begins

We will now start the computer program and run three unpaid practice rounds for you to get more familiar with the stages of the experiment and the software. The task that you will face in these practice rounds will be different than the tasks that you will face in the real rounds.[1] You will also face more questions and have more time in the actual paid rounds. These practice rounds are just for you to familiarize yourself with the procedures. Before the paid rounds start, you will see a short quiz on your screen to test your understanding of the instructions.[2] The program will not proceed unless you answer all questions correctly. Please raise your hand if you are having trouble with any of the questions in the quiz.

IX. Appendix C: Quadratic Scoring Rule Instructions

Instructions

Welcome to the Social Science Experimental Laboratory. This is an experiment on the economics of decision-making. Your earnings will depend partly on your performance, partly on the performance of others, partly on your decisions and partly on random chance. In any case, a 30 AED show-up fee will be added to your earnings, for your participation. You will be paid privately, in cash, at the end of the experiment. You are under no obligation to tell anyone how much you have earned in the experiment. The experiment will consist of two sections - one short 15 question quiz and three belief reporting stages. You will also have a chance to practice on the computer, to help you become familiar with the computer program. Please raise your hand if you have any questions throughout the experiment, and the experimenter will assist you. We will now examine the procedures for the first section of the experiment.

1. Instructions for Part 1 — Performance Rounds

In this part of the experiment, you will be working on tasks that require ability and effort. You will be presented with a set of questions to work on within a minute each. Part of your earnings will come from your performance in these questions. For each question that you answer correctly within this time frame, you will earn 4 AED.

2. Instructions for Part 2 — Belief Self-Report Rounds

After you take the quiz, the program will rank the scores of all the people who took the quiz and calculate your percentile rank. For instance, if you scored better than 25% of the people who take the quiz, your percentile rank is 25%. Based on your percentile rank your score will be put into one of five performance quintile groups — the bottom 20%, the next to bottom 20%, the middle 20%, the next to top 20% and the top 20% of relative performance. The belief reporting rounds require that you answer the following questions:

1. What is the likelihood that your score ranks in the lowest quintile?

2. What is the likelihood that your score ranks in the second to lowest quintile?

3. What is the likelihood that your score ranks in the middle quintile?

4. What is the likelihood that your score ranks in the second to top quintile?

5. What is the likelihood that your score ranks in the top quintile?

The sum of these likelihoods must necessarily equal 100% and you will not be allowed to proceed unless this is so. You will earn the most payoffs if you honestly report your best guess about each likelihood. This is because your belief payoffs are

$$1 + 20 * p(x) - w, \tag{10}$$

[1] Subjects are given simple questions to understand the experment

[2] Subjects are given instructions-related questions

where: X is your actual performance category, as per your score; *p(x)* is the likelihood that you expressed for X; w is the sum squares of the likelihood that you attached to each belief category is the likelihood expressed for the actual value of X (the true performance category) and, w is the sum squares of the likelihoods attached to each belief category.
The range of your payoffs is from 0 AED to 20 AED for each guess.

Example:

As an example, imagine that, at the end of the experiment, you are in the middle category. If you had expressed that you were 100% certain that you were in the middle category:

- X = middle
- $p(x)$=1.0
- $w = 0^2 + 0^2 + 1^2 + 0^2 + 0^2 = 1$
- Earnings from estimation = $\$(1 + 20 * 1.0 - 1) = 20.00$ AED

If you had expressed that you were 100% certain that your score was in the top quintile:

- $p(x)$=0
- $w = 0^2 + 0^2 + 0^2 + 0^2 + 1^2 = 1$
- Earnings from estimation = $\$(1 + 20 * 1.0 - 1) = 20.00$ AED

If you expressed likelihoods of 10%, 10%, 30%, 10% and 40% for the bottom, second to bottom, middle, second to top and top categories , respectively:

- $p(x)$=.30
- $w = 0.1^2 + 0.1^2 + 0.3^2 + 0.1^2 + 0.4^2 = 0.28$
- Earnings from estimation = $\$(1 + 20 * (.30) - .28) = 6.72$ AED

2. Instructions for Signal Structure

After the second belief reporting stage, you will receive a signal of your relative performance. Interpret the signal as follows:

Not Bottom: You are on or above the 60th percentile. You are not in any of the bottom three categories. Please restrict your responses to the top two categories (Second to Top 20% or Top 20%).

Not Top Signal: You are on or below the 60th percentile. You are not in either of the top two categories. Please restrict your responses to the bottom three categories (bottom 20%, second to bottom 20% or middle 20%).

3. Instructions for Payoffs

Your total earnings in a given round in this first part of the experiment will be the sum of the show-up fee (30 AED), the earnings you get from task performance (4 AED*number of correct questions), and the payoffs you get from the average of the initial, second and third assessment of your performance.

Experiment Begins

We will now start the computer program and one unpaid practice round for you to get more familiar with the stages of the experiment and the software. The task that you will face in this practice round will be similar to the tasks that you will face in the real rounds. You will also face more questions and have more time in the actual paid rounds. This practice round are just for you to familiarize yourself with the procedures. Before the paid rounds start, you will also see a short quiz on your screen to test your understanding of the instructions. The program will not proceed unless you answer all questions correctly. Please raise your hand if you are having trouble with any of the questions in the quiz