Does Self-Relevance Affect Information Processing? Experimental Evidence on the Response to Performance and Non-Performance Feedback

Seda Ertac [a,*]

[a] Department of Economics, Koc University

Abstract

In many settings, individuals are confronted with decision problems that involve information relevant to their self-image. This paper uses an experiment to explore whether the self-relevance of information influences information processing. The experiment implements two information processing tasks that are identical from a theoretical perspective, but differ in the type of information provided: performance feedback versus information within the context of a purely statistical updating problem. The results suggest that information processing differs significantly across self-relevant and self-irrelevant contexts. In the self-relevant context, except in cases where initial self-confidence is high, subjects overweigh unfavorable performance feedback, leading to overly pessimistic beliefs. This is in contrast to the corresponding self-irrelevant setup, where departures from Bayes' rule do not follow a consistent pattern in terms of direction, and are smaller in magnitude. In addition, I find that women may interpret positive feedback more conservatively than men, leading to more pessimistic posteriors.

JEL Classification: C91; D83.

Keywords: Experiments, beliefs, performance feedback, information processing, self-confidence, gender.

1. Introduction

In many economic decisions involving dynamic uncertainty, individuals are faced with the task of revising their beliefs in response to new information. These posterior beliefs then determine the decisions that they make. The main assumption of economic theory about the processing of information is that agents are Bayesian, that is, they use Bayes' rule in updating their prior beliefs when faced with new information. Although economic theory makes no distinctions here with respect to the nature of the information, individuals' response to information may crucially depend on the type of information in question. For example, feedback received in organizational or academic settings (the outcome of a project or an investment decision, or performance in a task that requires skill) usually affects one's self-image and morale, and is ego-relevant. On the other hand, some types of information collected or received before making decisions can be irrelevant to one's self-image, e.g., a disinterested manager evaluating an employee's work or a recruiter making a judgment about the success of a potential job candidate.

It is a well-known fact that judgment biases such as overconfidence may come into play when the decision at hand involves a trait that is important to the individual's self-image. This has been shown to have consequences for economic decisions such as market entry, stock trading, and investment (e.g. Camerer and Lovallo (1999), Gervais and Odean (2001), Hoelzl and Rustichini (2005)). Similarly, it is natural to conjecture that the way in which individuals process information may be different when information is ego-relevant, from both the case of information that is unrelated to the self, and the predictions of economic theory. For instance, individuals might use such information in a self-serving way, downplaying negative signals or signals that conflict with their initial beliefs.[1]

Moreover, since many economic settings involve the processing of self-relevant feedback, addressing the evolution of self-assessments in a dynamic framework where there is inflow of information is crucial from an economic standpoint. For example, from an organizational design point of view, it is important to know what agents' response to feedback about their performance would be, for the optimal design of interim performance feedback policies in

---

[1] Indeed, there is considerable literature in psychology on motivated reasoning, ego-protection, and ego-enhancement (see Baumeister (1998) for a review).

firms as well as grade revelation policies in schooling (Ertac (2006), Azmat and Iriberri (2010)).

This paper experimentally investigates how individuals process information, in the context of both information that is relevant to the self (performance feedback in tasks that require ability and effort), and information that is self-irrelevant (within the context of a purely objective statistical updating task). The primary goal of the paper is to identify the effects of the information being personal on its use. The basic design involves eliciting subjects' beliefs about their relative performance after they complete a task, then giving them incomplete information about their relative performance, and eliciting beliefs once again after this information is received. Information processing in this performance-related setting is contrasted with a case where individuals face an updating problem that is theoretically equivalent, but does not have a personal aspect. By maintaining the exact statistical structure of the updating task and changing only the self-relevance of the information, I am able to study whether or not any observed departures in behavior from theoretical predictions are due to the nature of the information, as opposed to general errors in updating that would be common to any type of information.

The main insight the study yields is that whether or not information is self-relevant matters for information processing. In general, subjects frequently deviate from Bayesian benchmarks, which is in line with earlier studies on Bayesian updating. Both the magnitude and the direction of these deviations, however, are significantly different across the performance and non-performance contexts. While the direction of the deviations is quite symmetric in the updating context irrelevant to the self, there is evidence of pessimism on average, especially in interpreting negative feedback, when the feedback is related to one's own performance. Information processing is also heterogeneous in the self-relevant context and correlates with how positive initial beliefs are: self-confident subjects process information more optimistically than they should, exhibiting self-serving tendencies. The magnitude of the biases is also larger in the self-relevant case, suggesting that individuals are less Bayesian in contexts that are related to themselves. In addition, I find some evidence for a task-dependent gender difference in information processing: in the verbal task (GRE verbal questions), which both genders regarded as highly difficult, women commit significantly more pessimistic deviations from Bayesian benchmarks than men in response to positive feedback. This suggests that at least in some contexts, pessimistic beliefs may be more persistent in women.

The paper also investigates whether the difference in updating can be explained by the subjectivity of the prior beliefs in the performance context (as opposed to objective priors in the non-performance context). In order to account for this, I use a non-performance updating task with subjective priors and compare it with the performance context. This treatment provides data on information processing under subjective priors/ambiguity, which has not been studied with incentivized beliefs in the literature before.[2]

The paper provides a direct comparison of the use of personal versus non-personal information, and is related to a long line of work in decision sciences on information processing, which has documented several heuristics and biases in judgment.[3] An important and relevant example is the self-serving attribution bias (e.g. Miller and Ross (1975)), which may manifest itself in a tendency to interpret ambiguous information in a positive light. This has been studied in economic contexts such as bargaining (Babcock and Loewenstein (1997)) or the decisions of mutual fund managers (Choi and Lou (2008)). In the economics literature, there are several studies that test experimentally for confirmation bias, which refers to the tendency to interpret new evidence to fit existing beliefs (Dave and Wolfe (2004), Jones (2008), Jones and Sugden (2001)). Dave and Wolfe (2004) and Ertac (2006), in line with some earlier psychology studies (e.g. Slovic and Lichtenstein (1971)), find evidence of conservatism in information processing, which refers to a tendency to update beliefs too little in response to new information. Charness and Levin (2005) study the link between Bayesian updating and reinforcement.[4]

The papers closest to the current paper in motivation and design are Clark and Friesen (2009), Eil and Rao (2011) and Mobius et al. (2010). Clark and Friesen (2009) also study self-assessments in a performance context by eliciting beliefs, but they analyze how point predictions about future task performance are revised after completing the task, rather than explicitly comparing Bayesian versus submitted posteriors. Mobius et al. (2010), on the other hand, find evidence of conservatism and asymmetry (stronger reaction to positive than negative feedback) in response to feedback on IQ, and also find that men are significantly less conservative than women. Similarly, Eil and Rao (2011) study updating in response to information about beauty and IQ, and find that individuals incorporate positive information

---

[2] See Halevy (2007) for an experimental study that analyzes attitudes toward ambiguity and compound lotteries.
[3] Rabin (1998) provides a review of some of these biases as they relate to economic decision-making
[4] In addition to the experimental studies, there are also several theoretical studies that focus on departures from Bayesian behavior in information processing (e.g. Epstein et al. (2005)). Most related to the current paper are Koszegi (2001, 2006), which analyze the effects of "ego utility" on information collection and decision-making.

4

more into their beliefs than negative information. Although motivated by similar questions, the current study differs from their work in signal structure, task, and belief elicitation mechanism.[5]

The organization of the paper is as follows: Section 2 describes the experimental design, procedures, and hypotheses. Section 3 presents and discusses the results on self-assessments and information processing. Section 4 concludes.

2. Experimental Design and Procedures

2.1. Design

The main treatment variable in the experiment is the self-relevance of information, which varies within-subject. In the self-relevant treatment, participants work on tasks that require ability and effort, and the information they are given is about their relative performance. In the self-irrelevant treatment there is no performance task, and the information that participants are asked to process is about a state of the world that has no relevance for self-image. Accurate beliefs about the true state are rewarded, in all but two sessions run to study the effects of incentivizing beliefs. In the self-relevant setting, the true state is the subject's place (top, middle, or bottom) in the performance distribution for the task that was just completed. In the self-irrelevant setting, the true state is one of three potential outcomes and is randomly picked by the computer according to a pre-determined probability distribution. Within these two main settings, three types of updating task are used: (1) Performance-related updating task, (2) Non-performance updating task with objective priors, (3) Non-performance updating task with subjective priors/ambiguity. These updating tasks are explained below.

2.1.1. Performance-Related Updating Task

In the performance (self-relevant) treatment, subjects work on a task for a specified amount of time, and are compensated using a piece-rate for each question they answer correctly. They are initially not given any information on their performance, and are asked to state their beliefs about their likelihood of being in the top 20%, middle 60%, or bottom 20% of the performance distribution. These estimates are compensated using a quadratic scoring rule.

---

[5] Eil and Rao (2010) elicit beliefs on 10 possible states of nature with a quadratic scoring rule and give perfect information on rank comparison with another subject, whereas Mobius et al. (2011) elicit beliefs on two states through a cutoff mechanism based on selecting own performance or a robot's performance for payment, and use a signal structure that has 75% accuracy. Neither of the papers implements an ambiguity treatment.

To study how self-relevant information is processed, participants are given feedback after submitting their initial beliefs about their relative performance. Depending on session, the information received is either a "top/not top (T/NT)" signal that tells the subject whether they were in the top portion or the distribution or not, or a "bottom/not bottom (B/NB)" signal that tells whether they were in the bottom or not. If they receive information that they were in the top (or in the bottom, depending on signal type) there is no remaining uncertainty, that is, feedback is truthful. After receiving the feedback, subjects are asked to submit their beliefs about being in the top/middle/bottom again. This second estimation is also compensated in the same fashion as the first, using the scoring rule. The total payoff in a performance round, therefore, is the sum of the payoffs from performance, from prior beliefs, and from posterior beliefs.

2.1.2. Non-Performance Updating Tasks

 Non-Performance Updating with Objective Priors:

In the non-performance (self-irrelevant) treatments, there is no task performance stage. Instead of being linked to performance, the states "top", "middle" and "bottom" are merely outcomes that have certain prior probabilities of being picked by the computer. At the beginning of each round, subjects are shown these prior probabilities, and then are asked to assign likelihoods to each of the three states being drawn in that round. As in the performance rounds, these beliefs are compensated using a quadratic scoring rule.[6] After this initial estimation, subjects are given information on whether or not the state drawn was top or not (or bottom or not, depending on signal type). After this feedback stage, subjects are asked once again to assign probabilities for the three states, and these posteriors are incentivized in the same way.

A novel aspect of the design is that for each subject, it replicates the exact same updating problem from a theoretical perspective, with and without a performance task. The way this is achieved is as follows: the computer records the probabilities that were submitted by each participant in the performance rounds, and uses exactly those probabilities as the prior probabilities of each state being picked in the non-performance rounds. In other words, for every performance round, there is a corresponding non-performance round where the subject faces the same prior probabilities for the three states. The difference of these rounds is only

---

[6] Notice that assuming risk-neutrality, it is optimal for subjects to just submit the objective prior probabilities in this first estimation.

that the context is not related to performance. This allows a comparison across the cases of self-relevant and self-irrelevant information, when the statistical nature of the problem is the same and economic theory makes the exact same behavioral predictions.

Non-Performance Updating with Subjective Priors:

In this third type of updating task, individuals are asked to state and update their beliefs as part of a non-performance updating task again. However, now only one of the three objective prior probabilities is revealed to the subject. That is, in each round, only one of the true priors for the three states (top, middle and bottom) is revealed, whereas no information about the other two probabilities is given. The subject then submits her own prior beliefs for all three states, receives information, and updates her beliefs. This creates an environment with ambiguity that is designed to mimic the one faced by the subjects in performance estimation. The goal is to have at least a partial control for the fact that beliefs are inherently subjective in the performance rounds whereas they are objective in the non-performance rounds, which could potentially have an effect on updating that is separate from ego-relevance.

2.2. Procedures

The experiments were conducted at the California Social Science Experimental Laboratory (CASSEL) at UCLA, using undergraduate students as subjects.[7] In total, 230 subjects participated, in a total of 19 sessions.[8] Sessions lasted roughly from an hour and fifteen minutes to an hour and a half, with payments averaging around $20. After the instructions were read, participants played three practice rounds, and then completed a short quiz that tested their comprehension. Subjects were aware that there would be multiple parts to the experiment, but did not know about the nature of the different updating tasks until those treatments started. That is, instructions for the different updating tasks were read sequentially. There was also one practice round and a quiz before each new type of updating problem. A survey was given at the end of the experiment, with questions on the difficulty of the tasks, clarity of the instructions etc. (see Appendix). The exchange rate used in the experiment was 0.025.

2.2.1. Tasks and Performance

---

[7] The experiment was computerized using the Java-based Multistage software, by Raj Advani and Walter Yuan.
[8] Additionally, one pilot session was run to test subjects' understanding of the instructions and the workings of the computer program.

The first task used in the performance part of the experiment was an algebra task. In these rounds, subjects were given 2 minutes to solve 16 addition problems, which involved adding five two-digit numbers. The second task used in the experiment was a verbal task. Subjects were given multiple-choice questions from GRE verbal tests, which involved analogies and antonyms.[9] All subjects who participated in the experiment completed four rounds of the addition task and four rounds of the verbal task. Earnings from performance were calculated according to a piece-rate, with each correct answer being worth 50 points. The computer then ranked everyone's performance (in terms of number of questions correctly solved), and determined the top, middle, and bottom performance groups.[10] Since each group in the experiment consisted of 10 subjects, the top 20% corresponded to the top two performers, the bottom 20% to the worst two performers, and the middle 60% to the six subjects in-between.

2.2.2. Belief Elicitation

Beliefs were elicited using a quadratic scoring rule, which rewards accurate beliefs (see Selten (1998) and Offerman et al. (2009) for a discussion of the use of this procedure). Subjects are asked to submit percent chances for being in the top, middle, or bottom of the performance distribution in the performance rounds, and for the state top, middle or bottom being picked by the computer in the non-performance rounds. The payoff (π) from these beliefs depends on the true state, and is increasing in the probability assigned to it. The exact formula used in the experiment is given by:

$$\pi_i(p_i, p_{-i}) = 50 + 100(p_i) - 50(p_i^2 + \sum p_{-i}^2)$$

where p_i denotes the probability assigned to the true state, i, which can be either top, middle or bottom.[11] The software used in the experiment was designed in such a way that subjects could see their potential payoffs from estimation on their screen as they entered probabilities in the relevant boxes.

In addition, two sessions were conducted where beliefs were not compensated at all and only performance was compensated, in order to study the effects of incentivizing beliefs. In

---

[9] These questions were taken from "GRE: Practicing to Take the General Test" (2001), and were used in the experiment with the permission of Educational Testing Service, the copyright owner.
[10] Any ties in performance were broken using a timed tie-breaker question, given to the subjects after completing the task. The subject who got the tie-breaker question correctly and in the least amount of time was ranked higher.
[11] This belief-elicitation procedure is incentive-compatible under the assumptions of risk-neutrality and expected payoff maximization.

these sessions, subjects were paid a fixed amount of money for their estimations, which did not depend on the accuracy of their submitted beliefs.[12]

2.2.3. A Description of the Sessions and Data

With the benchmark design, the first eight rounds of the experiment were performance rounds, where subjects worked on the two performance tasks. After these rounds, the second part of the experiment (the non-performance rounds) began. In some sessions, the order of performance and non-performance treatments was switched to be able to test for order effects.[13] The non-performance rounds in some of the sessions included both the objective non-performance updating task and the non-performance updating task with subjective priors/ambiguity, as explained in the design section above. A sample timeline of the sessions that have all three updating tasks is as follows:

Part I-Performance Updating Task (8 rounds)

> Subjects work on performance task, submit beliefs, receive information, update beliefs

Part II-Non-Performance Updating Task with Objective Priors (8 rounds)

> Subjects see objective priors for 3 states, submit their own priors, receive information, update beliefs

Part III-Non-Performance Updating Task with Subjective Priors (8 rounds)

> Subjects see the objective prior for only one state, submit their own priors, receive information, update beliefs

Part IV-Post-experiment survey

Notice that Part 3 was not run in all of the sessions, and the order of parts 1 and 2 depends on the type of session. Table 1 summarizes the types of tasks and the number of subjects and observations in each session.

2.3. Hypotheses

The main research questions are summarized by the following hypotheses:

---

[12] This treatment was used in two sessions with 30 subjects, and the T/NT type of signal was used.
[13] In the sessions where non-performance rounds came first, the priors given to the subjects in the non-performance updating task were taken from a previous session's performance rounds. The reason was to maintain comparability of priors in the performance and non-performance contexts, while still being able to switch the order.

1. Subjects hold "correct" initial beliefs. That is, there is no underconfidence or overconfidence, either at the individual or the group level.

2. Subjects are Bayesian: Submitted posterior probabilities for the three states after receiving information are equal to the Bayesian posteriors, which are given by:

$$p'_T = 0, \ p'_M = \frac{p_M}{p_M + p_B}, \ p'_B = \frac{p_B}{p_M + p_B}$$

if the feedback indicates that the true state is "not top", and

$$p'_T = 1, \ p'_M = \ p'_B = 0$$

if the information observed indicates that the true state is top, where p_M and p_B denote the prior probabilities for the states middle and bottom, respectively.[14]

3. Given the same priors in the performance and non-performance treatments, the posterior beliefs submitted should be the same.[15]

2. 4. Discussion

Within the above setup, both the accuracy of initial beliefs and the updating of beliefs in response to feedback can be studied. The most important feature of the design, however, is that it allows a direct comparison of information processing in performance and non-performance contexts, within person. It is quite likely that the updating behavior of individuals will be affected by some general errors, heuristics and biases, regardless of the nature of the information in question. The replication of the same updating task from a theoretical perspective, only with a different type of information, can help isolate the role that self-relevance might play in information processing.

One potential issue with the design could be that the compensation of both beliefs and performance in a given round might give rise to moral hazard. In such a case, subjects would

---

[14] For the bottom/not bottom signal type, the posterior beliefs will change accordingly.
[15] Notice that this would automatically follow if subjects are Bayesian in every context, and if not, it means that the types of updating errors are of similar direction and magnitude across the two contexts.

purposefully not solve questions in the performance part, and assign probability one to being in the bottom of the distribution, in order to reap sure payoffs from estimation. In the design, the maximum payoff that can be received from a single estimation is 100 points, and the marginal payoff to solving one question is 50 points. Examining the data, I find no evidence to suggest the presence of moral hazard: in only 1.2% of the cases do subjects get all questions wrong (less than 2 questions right in less than 5% of the cases), and 91% of the time, this is not coupled with a tendency to submit a probability of one for being at the bottom.

A similar potential issue with the interpretation of submitted beliefs arises from the possibility of subjects hedging beliefs across the two estimation stages in a given round, since both prior and posterior beliefs are compensated. Fortunately, there are multiple ways to address this issue. First, one can study the cases where subjects receive non-informative feedback (e.g. cases where they got the information that the true state was not top, but had assigned zero prior probability for top anyway). If subjects switch the state that they find most likely after such non-informative feedback (for example if they assign higher probability to middle in priors but higher probability to bottom in the posteriors), this could be evidence for hedging, in order to guarantee themselves a higher total payoff from the first and second estimations in a given round. It turns out that only 11% of the time there is a switch in the ranking of the states after receiving non-informative feedback. This suggests that there is at least no strong support for hedging. Another way of bypassing this issue in the analysis is to use the data from the non-performance task to identify risk-neutral subjects, and restrict attention to cases where risk-neutrality can be ensured. I report results from this subsample separately in what follows. It should also be noted that even if the difference between actual posteriors and Bayesian benchmarks were colored by incentives to hedge, such incentives are common to both the performance and non-performance contexts, and the main message of the paper would be intact if subjects use information differently across the two contexts.

3. Results
3.1. Initial Self-Assessments in the Performance Context

I start with some observations on task performance and beliefs in the performance rounds. On average, subjects solve 5.04 questions in the addition task, and 4.57 questions in the verbal task. The first interesting question here is how positive subjects' beliefs are, compared to their actual outcomes. To obtain an aggregate measure for this, I examine the population belief distribution for being in the top, middle and bottom. If subjects hold correct beliefs, the

average probability attributed to the top and bottom states should be 0.2, whereas it should be 0.6 for middle. Table 2 presents these probabilities for the addition task and the verbal task, across the three actual performance categories. The results show a marked difference between the two tasks, with subjects being underconfident as a group in the verbal task but not the addition task.

Another way of looking at the data is to compare the performance category that subjects view as most likely in each round, with their actual performance category. For instance, if a subject submitted a probability of 0.4 for top, 0.3 for middle and 0.3 for bottom, top would be her perceived most likely category. The results in Figure 1 show that perceived performance can be classified as correct 52% of the time, lower than actual 18% of the time, and higher than actual 30% of the time in the addition task. In the verbal task, on the other hand, it is correct 50% of the time, lower than actual 31% of the time, and higher than actual 19% of the time. Tests of proportions confirm that optimistic (pessimistic) assessments are more frequently observed in the addition (verbal) task (p=0.000 for both). [16] [17]

These findings are consistent with the studies which find that people are more likely to be overconfident in domains that they perceive to be easy (e.g. Kruger (1999), Moore and Cain (2007)), since most subjects reveal in the post-experiment survey that they found the verbal task more difficult. The result is also consistent with Clark and Friesen (2009), who find evidence of underconfidence in a verbal task. Most importantly, the marked difference in the self-confidence levels of subjects in the two tasks provides a potentially valuable environment to study how initial confidence correlates with belief-updating. I return to this point later.

3.2. Information Processing in the Performance Context

The feedback structure used in the experiment is such that in each round, two out of ten subjects learn their true performance, whereas for the remaining eight some uncertainty remains. Not surprisingly, subjects who learn that they were in the top (or bottom, depending

---

[16] Benoit and Dubra (2007) show that x% of a population may rationally report they are in the top y% of the distribution (where x>y). The data used in the analysis above do not come from subjects picking a single state, but reporting a complete probability distribution for all possible states, and the complete belief distribution should average out to the true distribution. In addition, the main goal of the paper is to analyze how departures from Bayesian information processing contribute to the evolution of beliefs in a dynamic framework with inflow of information, as in Benoit and Dubra's model, therefore the results of the paper are not subject to this critique.
[17] Notice that the data used in the above analyses includes 4 observations on each subject for each task, and learning and patterns of updating can play a role. If we restrict attention to only the initial rounds of each task where no feedback has yet been observed, we get the same pattern, whereby individuals have more positive beliefs in the addition task than the verbal task.

on the signal type) generally revise their beliefs as they should, assigning a probability of one to the correct state. In the following, I focus on the direction and strength of belief updating when subjects do not receive perfect feedback, i.e. receive either the signal *not top* (NT) or *not bottom* (NB). I restrict attention to non-trivial updating tasks, where all three states have positive prior probability, and also disregard the few instances where it is clear that the subject did not believe the feedback.[18] I pool the data across different orders of updating task (performance first or non-performance first), and across verbal and mathematical tasks.[19]

In order to understand whether subjects are Bayesian, I compare submitted and Bayesian posteriors for the state middle, for each of the two signal types (T/NT and B/NB). Table 3 (rows 1 and 2) displays the means and standard deviations of the bias (submitted minus Bayesian posteriors), and Table 4 (rows 1 and 2) lists the frequency of errors of different direction (overassigning or underassigning probability to middle), for each signal type. The data show that when they learn they are not in the top, subjects tend to attribute more probability to being in the bottom than they should. That is, on average, they update their beliefs in a more pessimistic way than Bayesian after receiving negative feedback. A Wilcoxon signed-rank test for the bias being equal to zero confirms that the posterior probability assigned to middle is significantly lower than the Bayesian posterior, which implies that the probability assigned to bottom is higher than it should be when the signal is *not top* (p=0.000).[20] This effect, however, is not significant in the case of a *not bottom* signal (p=0.35). Considering the magnitude of the deviations from Bayesian posteriors without regard to direction (the absolute value of the bias), a signed-rank test confirms that subjects display significant departures from Bayesian updating (p=0.0000). This holds for both types of signals, with no significant difference in the magnitude of errors across signal type (p=0.54).

The above analysis does not account for the potential correlation of updating patterns with initial self-confidence. Classifying instances where the individual assigns at least one-third probability to being in the top of the performance distribution as self-confident, I find that

---

[18] These are cases where the subject still places positive probability on being in the top after receiving the NT signal, or being in the bottom after receiving the NB signal. Such behavior is observed in only 36 out of 1344 cases, and could also be due to random errors.

[19] There are no significant order effects in terms of the magnitude of bias in belief-updating in the non-performance rounds. However, the absolute value of deviations from Bayesian updating in performance rounds is lower if non-performance rounds come first (p=0.02). There is no significant difference in information processing across the verbal and mathematical tasks; however, there are some interesting interactions of the tasks with gender, which are presented in Section 3.7.

[20] In this test and in similar tests that follow, observations were averaged within subjects to account for dependence.

when subjects are confident, they commit more optimistic updating errors when they get the information that they are not in the top (p=0.000, two-tailed test of proportions). In fact, the average bias in the posterior for middle is 0.09 (exhibiting a self-serving tendency) for the cases where subjects are confident whereas it is -0.04 for the not-confident, a difference of 13 percentage points that is significant at the 1% level. A Mann-Whitney test confirms that the corresponding measure of confidence does not generate a significant difference in belief updating in the non-performance rounds (p=0.38). Naturally, there is also a significant difference across the two updating contexts: For self-confident subjects, in response to an NT signal, the bias for the state middle is significantly higher in the performance updating task than in the non-performance task (p=0.01). In other words, when they receive negative feedback, confident subjects assign significantly more probability to the "good" state than they should in the self-relevant context.

A legitimate concern about the analyses reported here is risk-preferences potentially biasing the submitted beliefs of the individuals, given that the quadratic scoring rule theoretically works only under risk-neutrality. The within-subject design and the nature of the non-performance updating task fortunately allow for a test of whether this changes the main results. Comparisons of submitted and objective priors are available from the non-performance rounds, making it possible to identify subjects who behave in a consistently risk-neutral way under the quadratic scoring rule.[21] Restricting attention to this subsample and repeating the analysis yields the result (see Table 5 and Table 6, rows 1 and 2) that the posterior probability assigned to middle is lower than Bayesian in the case of the NT signal and higher than Bayesian in the NB signal, significant at the 5 and 10% levels, respectively (p=0.011 for the former and p=0.07 for the latter in a Wilcoxon test). [22] For this subsample, the average pessimism in information processing carries over to positive (NB) feedback as well.

These results suggest that the positive and negative connotations of signals matter in the performance context, and subjects tend to use information in a pessimistic fashion: when they receive information that they are not in the top, they assign a higher probability to the worse state (bottom) then they should, whereas when they receive the information that they are not in the bottom, they may assign a lower probability to the better state (top) than they should. However, this reverses to over-optimism for the self-confident subjects, which suggests that

---

[21] These are subjects who always submit the exact true probability under the quadratic scoring rule in all rounds of the non-performance updating tasks. About one in five subjects (22%) behaves in this way.
[22] Restricting attention to this subsample of subjects implicitly assumes that this sample of subjects has similar biases to others.

positive beliefs about the self (and possibly overconfidence) may be persistent because of a biased and self-serving interpretation of information.

3.3. Information Processing in the Non-Performance Context

I now turn to the analysis of how individuals use information when the information concerns an ego-irrelevant event. Table 3 (rows 3 and 4) summarizes the bias in the posterior for the state middle in the non-performance rounds, for the two signal types. In this case, it is not possible to reject that the average bias is equal to zero (p=0.35, sign-rank test), although the hypothesis that subjects are Bayesian is easily rejected (p=0.000). Notice that the directional strength of the bias is different from that observed in the performance rounds, where there was a more significant underassignment of probability to the state middle with the NT signal. A Mann-Whitney test confirms that the bias after an NT signal is significantly more negative in the performance rounds, meaning that the posterior (compared to the Bayesian benchmark) is lower (p=0.01). Similarly, a chi-square test of proportions rejects the hypothesis that the distribution of the directions of updating errors in performance and non-performance contexts is the same (p=0.015).

Another result concerns the comparison of the magnitude of deviations from Bayesian posteriors in the self-relevant and self-irrelevant contexts. The absolute value of the bias in the posterior for middle is found to be higher in the performance rounds than in the non-performance rounds, which is significant at the 1% level (Mann-Whitney test, p=0.006). This suggests that individuals are better Bayesians in the context where information is self-irrelevant.[23]

Again, it is useful to see whether and how the results change if we restrict attention to cases where subjects behave in a risk-neutral manner and submit exactly the same priors as the objective priors in the first estimation. With this subset of the data, it is still not possible to reject the bias being equal to zero, for either signal type (p=0.62 and p=0.99, respectively). Moreover, the types of deviations seem to be quite balanced (see Table 6, rows 3 and 4). Figure 2 displays updating patterns across the performance and non-performance contexts in the risk-neutral case. Another interesting result is that the absolute value of the bias is significantly lower (a roughly 3 percentage point difference) when we restrict attention to

---

[23] This difference in updating patterns also translates into statistically significant differences in earnings from beliefs in the performance vs. non-performance rounds (p=0.000)

cases where subjects behave in a risk-neutral manner (p=0.0002), which generates an even starker comparison with the level of the bias in the performance rounds.

3.4. Within-Subject Comparison of Performance and Non-Performance Contexts

In sessions where performance rounds come first, the objective prior probabilities used in the non-performance rounds are taken directly from the prior probabilities submitted by the same subject in the performance rounds. Because of this structure, the design allows for a stronger within-subject comparison between self-relevant and -irrelevant information. In this comparison, I focus only on the cases in which the submitted priors for all three states in the non-performance task are exactly the same as the priors submitted in the performance task and the feedback observed is the same as well, so that the updating problems are exactly the same.[24]

   Table 7 summarizes the bias difference in the non-performance and performance rounds within subject: When the information received is NT, the bias in the posterior for the state middle is significantly more positive in the non-performance than in the corresponding performance updating task (p=0.022, Wilcoxon sign-rank test). Since the Bayesian posterior is the same in the two updating problems, this means that after receiving the NT signal, subjects assign a higher probability to the state middle in the non-performance (self-irrelevant) updating problem.[25] In other words, bad news is processed more pessimistically in the self-relevant context. This corroborates the result suggested by the between-person analysis. The difference in posteriors across performance and non-performance contexts for the NB signal is not significant for the subsample under consideration.[26] As for the comparison of the magnitude of deviations from Bayesian updating, the absolute value of the bias is significantly higher in the performance rounds than in the non-performance rounds (Wilcoxon test, p=0.000). This result is also consistent with the across-subject comparison.

   Using the within-subject data, it is also possible to get a sense of whether individuals get better at updating over time. The current data do not show a clear downward trend in deviations from Bayesian benchmarks, in either the performance or the non-performance task. However, it should also be noted that not in all rounds of a task are subjects faced with a non-

---

[24] Notice that this amounts to focusing on instances where the quadratic scoring rule works.
[25] Reducing the sample even further to consider only subjects who *always* submit risk-neutral priors does not change the result: in fact, the average difference in the posterior becomes larger at 6.5 percentage points, with p=0.0056.
[26] It should be noted that the sample size for the NB signal is quite low when attention is restricted to risk-neutral cases.

trivial updating problem; therefore, the opportunities for learning may not be significant enough with the current design.

3.5. Information Processing in the Non-Performance Context with Subjective Priors

If subjects are Bayesian, then the non-performance and performance contexts are theoretically equivalent, because posterior beliefs determined by Bayes' rule will be equal as long as the priors and the feedback are the same. However, priors in the performance context are subjective, whereas they are objective in the non-performance context. This could affect subjects' information processing independently of whether the information is self-relevant or not. For example, ambiguity in subjects' beliefs (which could be modeled as having probability distributions over priors) could potentially skew their response to information, e.g. lead them toward putting excessive weight on the state that the feedback indicates, regardless of self-relevance.

The subjective priors/ambiguity treatment allows for an analysis of how much of the difference between the performance and non-performance contexts can be attributed to the self-relevance of information. As mentioned before, in this treatment subjects face a non-performance updating task where only one of the objective priors (the probability of top, middle or bottom) is revealed rather than all three. This modification generates a setting where the priors submitted are (partially) subjective, but the information is not self-relevant.[27]

Restricting attention to the cases where the true probability is elicited, I find that it is not possible to reject the hypothesis that the average bias is zero, just as in the non-performance updating task with no ambiguity. Tables 3 and 4 (rows 5 and 6) summarize the bias and the frequency of different types of deviations from Bayesian updating in this case. Overall, the results do not look similar to the ones in the performance context.[28] On the other hand, there is no significant difference in updating behavior in the non-performance context with and

---

[27] Another way of capturing subjectivity and ruling out ego-relevance would be updating beliefs about the performance of another person, which can be explored in future research.

[28] It is important to note that the large difference in the amount of exact Bayesian updating across the two type of signals is an artifact of the state for which the probability was "unmasked": in most non-trivial updating problems with ambiguity, subjects assign the same probability to the two unknown states, and when these states happen to be the not-top states, they know that they should assign the same posterior probabilities as well. By chance, it has happened that many of the unmasked priors were for *top* in the case of the NT signal, and subjects were able to perfectly Bayesian update in this case. This did not happen with the NB signal.

without ambiguity. These results suggest that subjectivity is not what drives the difference between the self-relevant and self-irrelevant updating tasks.

3.6. Information Processing When Beliefs Are Not Compensated

In the sessions where beliefs are not compensated, the first result to note is that there is no significant difference from belief-incentivized sessions in terms of either task performance (measured by the number of correct answers) or the distribution of initial submitted beliefs in the performance rounds (Mann-Whitney test, p=0.27 and p=0.89, respectively). The lack of a performance difference is encouraging, since it suggests that incentives to perform have not been distorted by the compensation of beliefs. Initial beliefs being similar may suggest that subjects are not prone to submitting random beliefs or overly optimistic beliefs when there is no monetary consequence to doing so. In fact, the reaction to feedback is also pessimistic, in the sense that in a majority of the cases (77%), subjects place more probability on being in the bottom than they should when they receive the NT signal. This is directionally consistent with the compensated case; moreover, the magnitude of the deviations is even larger (see Table 8, rows 1 and 2). It is not possible to reject the equality of the bias and the absolute value of the bias in the compensated and uncompensated contexts (p=0.13 and p=0.79, respectively, in a Mann-Whitney test). These results suggest that the insights obtained with compensated beliefs in the performance rounds are not an artifact of the belief elicitation method, since the bias goes in the same direction with uncompensated beliefs as well. Likewise, in the non-performance updating context (Table 8, rows 3 and 4), the direction and magnitude of the bias are not significantly different across compensated and uncompensated beliefs either (p=0.68 and p=0.33, respectively).

3.7. Self-Confidence, Information Processing and Gender

Gender differences in beliefs and behavior have been the subject of much recent research in economics. The general result is that men tend to be more self-confident than women, even when actual performance is controlled for (e.g. Niederle and Vesterlund (2007), Barber and Odean (2001)). However, there is not as much work on how gender affects information processing, either in the context of performance feedback or in objective updating tasks.[29]

Studying gender differences in information processing is important, since this could provide clues as to why there is a self-confidence difference across genders, and also whether

---

[29] A recent exception is Mobius et al. (2010).

the gap is likely to persist in a dynamic framework with inflow of information. The first gender result that can be obtained from the data is that there is a marked difference between men and women in terms of how positive their beliefs about their performance are. Without conditioning on actual performance, women are more pessimistic than men: on average, they place a higher probability than men to being in the bottom, and lower probability to being in the top in both tasks (significant at the 1% level). In terms of actual performance (being in the top/middle/bottom), there is no significant difference between genders in either task. Women also are significantly more pessimistic in the verbal than the addition task, suggesting that perceptions of difficulty may override the potential positive impact of the verbal task being viewed as part of the female domain. These results, overall, corroborate the previous findings that men tend to have a more positive view about their relative performance than women.

Although there is no significant difference in the processing of performance feedback across genders if we pool the data across tasks, an interesting result emerges if tasks and signal types are considered separately: women's response to positive feedback seems to differ from men's. In the verbal task, which is regarded as highly difficult, when women observe the signal that they are not in the bottom, they tend to place more probability on middle than they should, interpreting good news more conservatively. This is in contrast to men's behavior in the same task, who interpret this information in a more positive way (a chi-square test of the equality of the distribution of the deviations from Bayesian benchmarks across genders rejects the hypothesis, with p=0.027). This differential pessimism across genders does not exist in the addition task. Similarly, for negative feedback (the NT signal), there is no difference in updating behavior across genders, for either task. In the self-irrelevant context, updating does not differ significantly between men and women either (Kolmogorov-Smirnov test, p=0.98). In addition, in terms of the magnitude of the errors from Bayesian benchmarks, there is no significant difference between genders in either the self-relevant or the self-irrelevant setting. That is, neither gender can be labeled as "more Bayesian".

3.8. Discussion and Remarks

In light of the information processing data, there are several possible hypotheses that can be evaluated. One possible behavioral hypothesis is that subjects interpret feedback in a self-serving way, that is, more optimistically than Bayesian. This would imply a higher-than-Bayesian probability being assigned to the state middle in the case of a negative (NT) signal and a lower-than-Bayesian probability in the case of a positive (NB) signal. The data do not

yield much support for this, since on average subjects respond to negative feedback in a significantly more pessimistic way than they should. This behavior reverses, however, when one considers the most confident subjects--for them the errors go in the direction predicted by self-serving tendencies. The heterogeneous updating patterns depending on confidence can be behind some of the differences from the papers that find stronger support for self-serving interpretation of information (Eil and Rao (2011), Mobius et al. (2010)). The priors in the current paper are frequently pessimistic, creating a unique environment to study updating. Moreover, the tasks used in the current paper could also be less important for ego than the tasks used in the other studies (IQ and beauty) for most of the subjects. [30] In fact, analyzing survey responses, one finds that the self-confidence and relevance to ego are positively correlated, so the self-serving tendencies that are prevalent in more confident subjects could be due to ego-protection motives.

Another potential explanation for the pessimistic updating of beliefs could be a motive to hedge against psychological losses with monetary gains. If there is psychological disutility from being at the bottom of the performance distribution, a risk-averse individual may have an incentive to submit pessimistic beliefs to hedge against such a psychological loss with monetary gain. However, this hypothesis does not seem to hold, since we observe the same type of updating behavior in the case of uncompensated beliefs also.

Besides the main comparisons between the processing of personal and non-personal information, the data can also shed some light on a methodological question that is quite important in experimental economics. Quadratic scoring rules have been commonly used in economics experiments to elicit subjects' beliefs about states of nature, or other players' actions (e.g. Nyarko and Schotter (2002)). Since the non-performance task provides a comparison of true, objective priors with the beliefs submitted under a quadratic scoring rule, the design allows for a direct test of how well the method works in eliciting true beliefs. I find that at least in the current context, deviations from true priors are quite common (subjects submit true priors 63% of the time), and frequently involve higher-than-objective priors for the most likely state. This could partly be explained by risk-loving behavior given the low stakes used in belief elicitation.

---

[30] Subjects do care about the tasks used in the paper: 90% of the participants submit a non-zero answer to the question of how important it was for them to do better than others in the addition task (on a 0 to 10 scale, with 0 corresponding to "not important at all"), and 83% in the verbal task, with median responses being 7 and 5, respectively.

4. Conclusion

In many economic settings, the information that agents need to process is relevant to themselves, such as how they have performed in a certain task at the workplace. Understanding how individuals react to such information is valuable from an organizational design perspective, since the resulting beliefs usually affect important decisions such as effort and persistence. Laboratory evidence has commonly suggested that individuals' updating behavior may not always be Bayesian in nature. However, there is not much work that analyzes how self-relevant information is processed. This study attempts to fill this gap by presenting a set of experiments that compares information-processing in a self-relevant and a self-irrelevant context, keeping the updating problem theoretically equivalent across the two contexts.

The main conclusion to be driven is that there are significant differences in terms of both the direction and the magnitude of errors in information processing, depending on the self-relevance of the information and initial beliefs. Overall, pessimism in response to negative information is common in the self-relevant context, while self-serving errors are also observed among the very confident. Another major result is that behavior in the non-performance task departs significantly less and in a more balanced way from the Bayesian benchmark, i.e. individuals are better information processors when the information is not relevant to themselves.

The results suggest that apart from the cases where they are quite confident in their abilities to start with, individuals may end up with a more negative self-image than they should after receiving unfavorable performance feedback. This is an important result to consider, since morale-preservation is a major goal in many organizational and educational settings. The policy implication could be that especially in difficult tasks where initial beliefs are not positive, withholding of relative performance feedback could be warranted, for example in a setting where pessimistic beliefs will lead to low effort. Self-serving errors are common, however, under high initial self-confidence, which could suggest that overconfidence is likely to be persistent in the face of incoming information about performance. The larger frequency of deviations from Bayesian benchmarks for self-relevant information could also implicate the delegation of information processing to an uninterested outside party.

The current paper provides a first look into how performance and non-performance feedback are processed. Further research should explore the dynamic evolution of beliefs under different types of feedback structure (i.e. vary the content and precision of the information) and updating contexts (e.g. updating about oneself vs. updating about other people), in order to obtain a fuller picture of how different types of information are processed and to explore the reasons for the underlying causes of deviations from Bayesian benchmarks.

References

Azmat, G., Iriberri, N., 2010. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. The Journal of Public Economics 94 (7-8), 435-452.

Babcock, L., Loewenstein, G., 1997. Explaining bargaining impasse: The role of self-serving biases. Journal of Economic Perspectives 11(1), 109-126.

Barber, B.M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics 116, 261-292.

Baumeister, R. F., 1998. The Self. In: Gilbert, D.T., Fiske, S. T., Lindzey,G. (Eds.), The Handbook of Social Psychology Vol.1, Boston: McGraw-Hill, 680-740.

Benoit, J.P., Dubra, J., 2007. Overconfidence?. Mimeo.

Camerer, C., Lovallo, D., 1999. Overconfidence and excess entry: An experimental approach. American Economic Review 89(1), 306--318.

Charness, G., Levin, D., 2005. When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. American Economic Review 95 (4), 1300-1309.

Choi, D., Lou, D., 2008. A test of self-serving attribution bias: Evidence from mutual funds. Mimeo.

Clark, J., Friesen, L., 2009. Overconfidence in forecasts of own performance: An experimental study. Economic Journal 119, 229-251.

Dave, C., Wolfe, W. W., 2004. On confirmation bias and deviations from Bayesian updating. Mimeo.

Eil, D., Rao, J. M., 2011. The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. American Economic Journal: Microeconomics, in press.

Ertac, S., 2006. Social comparisons and optimal information revelation: Theory and experiments. PhD dissertation, University of California Los Angeles.

Epstein, L.G., Noor, J., Sandroni, A., 2005. Non-Bayesian Updating: A Theoretical Framework. Working paper, University of Rochester.

Gervais, S., Odean, T., 2001. Learning to be overconfident. Review of Financial Studies 14(1), 1-27.

Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: Gender differences. Quarterly Journal of Economics 118(3), 1049-1074.

GRE: Practicing to Take the General Test. 9th Edition, Educational Testing Service, 2001.

Halevy, Y., 2007. Ellsberg revisited: An experimental study. Econometrica 75(2), 503-536.

Hoelzl, E., Rustichini, A., 2005. Overconfident: Do you put your money on it? Economic Journal 115, 305-318.

Jones, M.K., 2008. Positive confirmation in rational and irrational learning. Journal of Sociological Economics 37, 1029-1046.

Jones, M., R. Sugden, 2001. Positive confirmation bias in the acquisition of information. Theory and Decision 50, 59-99.

Koszegi, B., 2001. Ego utility and information acquisition. Mimeo.

Kozsegi, B., 2006. Ego-Utility, overconfidence and task choice. Journal of the European Economic Association 4(4), 673-707.

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

Miller, D. T., Ross, M., 1975. Self-serving biases in the attribution of causality: Fact or fiction? Psychological Bulletin 82 (2), 213-225.

Mobius, M., Niederle, M., Niehaus, P., Rosenblat, T., 2010. Managing self-confidence: Theory and experimental evidence. Mimeo.

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(2), 197-213.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much?.Quarterly Journal of Economics 122(3), 1067-1101.

Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70(3), 971-1005.

Offerman, T., Sonnemans, J., Van de Kuilen, G., Wakker, P.P. (2009). A truth-serum for non-Bayesians: Correcting proper scoring rules for risk attitudes. Review of Economic Studies 76(4), 1461-1489.

Rabin, M., 1998. Psychology and economics. Journal of Economic Literature 36, 11-46.

Selten, R., 1998. Axiomatic characterization of the quadratic scoring rule. Experimental Economics 1, 43-62.

Slovic, P., Lichtenstein, S., 1971. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance 6, 649-744.

APPENDIX

Table 1: Sessions and Data

	Sessions 1-7	Sessions 8-10, 16-17	Sessions 11-14	Sessions 15, 18
Beliefs compensated	Yes	Yes	Yes	No
Signal type	Top/Not Top	Bottom/Not Bottom	Top/Not Top	Top/Not Top
Order: Performance first	Yes	Yes	No	Yes
Non-performance task w/ ambiguity	No	Yes	Yes	Yes
No. of subjects	70	60	70	30
No. of observations	1120	1440	1360	720

Table 2: Prior Beliefs in the Performance Tasks by Actual Performance Categories

	Priors, Addition Task			Priors, Verbal Task		
	Top	Middle	Bottom	Top	Middle	Bottom
Actual Top	0.488	0.452	0.06	0.344	0.533	0.123
	(0.317)	(0.277)	(0.114)	(0.313)	(0.269)	(0.183)
Actual Middle	0.243	0.602	0.155	0.156	0.59	0.253
	(0.268)	(0.248)	(0.209)	(0.216)	(0.227)	(0.249)
Actual Bottom	0.116	0.589	0.295	0.059	0.474	0.467
	(0.18)	(0.271)	(0.297)	(0.137)	(0.276)	(0.314)
All Subjects	0.267	0.569	0.164	0.174	0.556	0.27
	(0.29)	(0.27)	(0.228)	(0.244)	(0.25)	(0.275)
No. of observations	840	840	840	840	840	840

Note: Means reported, standard errors in parentheses.

Table 3: Biases in Information Processing

	Bias	Absolute Value of Bias	No. of observations
NT, Performance	-0.032	0.103	236
	(0.137)	(0.096)	
NB, Performance	-0.003	0.104	77
	(0.133)	(0.083)	
NT, Non-Performance	-0.01	0.07	301
	(0.11)	(0.08)	
NB, Non-Performance	-0.02	0.072	60
	(0.11)	(0.108)	
NT, Non-Performance, Subjective	0.003	0.066	79
	(0.13)	(0.11)	
NB, Non-Performance, Subjective	-0.018	0.11	142
	(0.13)	(0.09)	

Table 4: Updating Patterns

	Overassign to Middle	Exactly Bayesian	Underassign to Middle
NT, Performance	36.44%	5.08%	58.47%
NB, Performance	48.05 %	2.60%	49.35 %
NT, Non-Performance	40.5%	10.7%	48.8%
NB, Non-Performance	33.33 %	8.34%	58.33 %
NT, Non-Performance, Subjective	14.29%	71.43%	14.29%
NB, Non-Performance, Subjective	33.92%	5.36%	60.72%

Table 5: Biases in Information Processing with Risk-Neutrality

	Bias	Absolute Value of Bias	No. of observations
NT, Performance	-0.041	0.9	35
	(0.11)	(0.08)	
NB, Performance	0.046	0.11	34
	(0.14)	(0.09)	
NT, Non-Performance	-0.003	0.051	136
	(0.08)	(0.07)	
NB, Non-Performance	-0.009	0.04	35
	(0.11)	(0.04)	

Notes: Means reported, standard deviations in parentheses. Bias is defined as submitted minus Bayesian posteriors for the state MIDDLE.

Table 6: Updating Patterns with Risk-Neutrality

	Overassign to Middle	Exactly Bayesian	Underassign to Middle
NT, Performance	20%	11.43%	68.57%
NB, Performance	61.76%	0%	38.24%
NT, Non-Performance	41.9%	11.7%	46.4%
NB, Non-Performance	42.9 %	5.7%	51.4 %

Table 7: Bias Difference in Performance and Non-Performance Rounds

	Bias in NP minus bias in P	No. of Observations
Top/Not Top	0.049	68
	(0.123)	
Bottom/Not Bottom	0.016	31
	(0.13)	

Notes: Means reported, standard deviations in parentheses. NP=Non-performance, P=Performance. Bias is defined as submitted minus Bayesian posteriors for the state MIDDLE.

Table 8: The Effects of Incentivizing Beliefs on Information Processing

	Bias	Absolute Value of Bias	No. of observations
Beliefs Not Compensated, Performance	-0.07	0.09	56
	(0.111)	(0.09)	
Beliefs Compensated, Performance	-0.032	0.103	236
	(0.138)	(0.096)	
Beliefs Not Compensated, Non-Performance	0.001	0.039	34
	(0.06)	(0.05)	
Beliefs Compensated, Non-Performance	-0.003	0.051	136
	(0.08)	(0.065)	

Notes: Means reported, standard deviations in parentheses. Bias is defined as submitted minus Bayesian posteriors for the state MIDDLE.
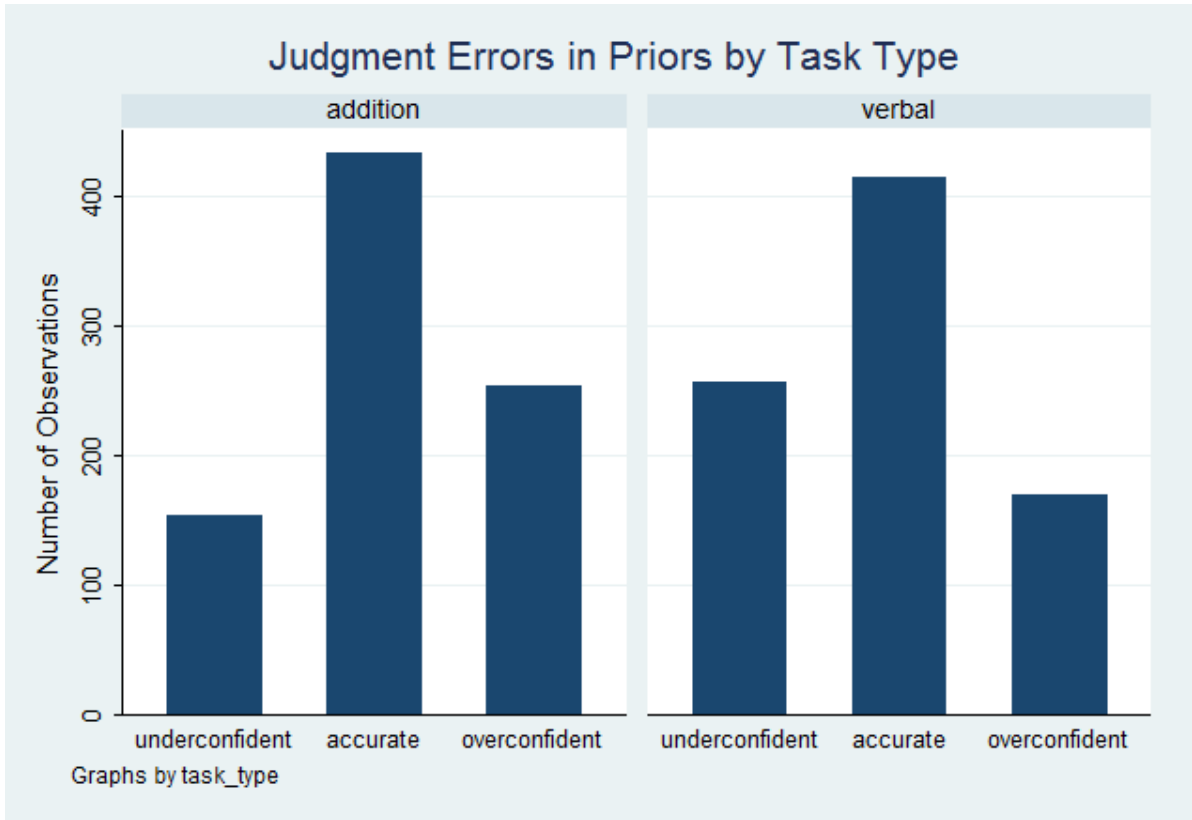
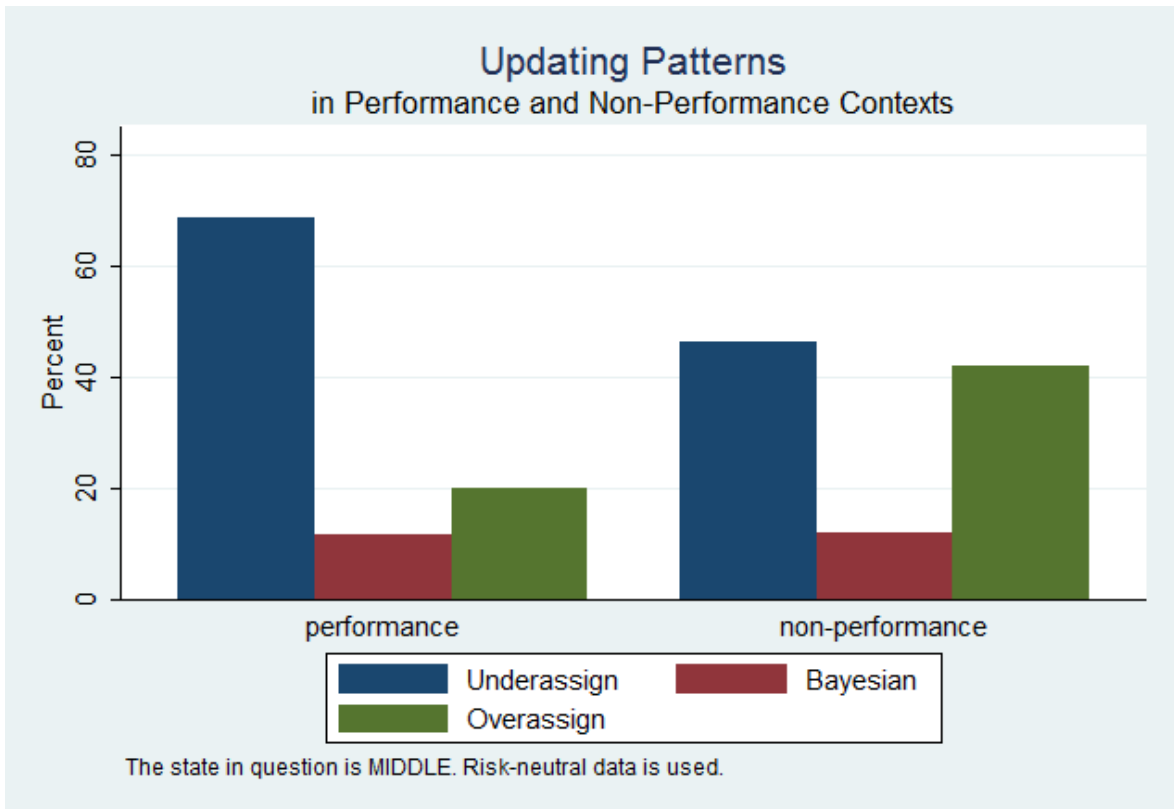**Figure 1.** Initial versus actual beliefs across task type

Figure 2. Updating patterns across performance and non-performance contexts.

Instructions [31]

Welcome to CASSEL. This is an experiment on decision-making. Your earnings will depend partly on your performance, partly on your decisions, and partly on chance. Earnings in this experiment will be denominated in "points". At the end of the experiment, the total number of points that you have earned will be converted to dollars at an exchange rate of 0.0025, which means that every 100 points is worth 25 cents. In addition, a $5 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. Now, please write your name and computer number (starting with SSEL) on the record sheets that you have been given.

The experiment will consist of three parts, and each part will consist of 8 rounds. We will now explain the procedures of the first part of the experiment. You will also have a chance to play some practice rounds 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.

Instructions for Part 1---Performance Rounds

In the first 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 50 points. [show Slide 1] Once this stage is completed, the computer will record and rank everyone's performance. A tie-breaker question will be used to resolve ties, if any. We will give you more information about the tie-breaking procedures later. Once performances are ranked, the computer will form three groups: the top 20% of the performance distribution, which corresponds to the top 2 performers of the 10 participants, the middle 60%, which corresponds to the next 6 performers of the 10 participants, and the bottom 20%, which corresponds to the worst 2 performers of the 10 participants. [show Slide 2]

You will then be asked to submit your beliefs about the likelihood of being in each performance group [show Screenshot 1]. You will see three statements on your computer screen (I was in the top 20% of the distribution, I was in the middle 60% of the distribution, I

---

[31] The instructions were read to the subjects, and were supplemented by a powerpoint presentation that included bullet-point representations of the main points, and some screenshots from the experiment (available from the author). The instructions presented here are the ones used in the sessions where the ordering is performance, non-performance with objective priors, and non-performance with ambiguity, with a feedback structure that reveals top/not top.

was in the bottom 20% of the distribution), with text boxes across from each statement. You will then be asked to submit the percent chances with which you think each statement is true. Naturally, the percent chances that you submit should add up to 100%, because your performance will be in one of these three groups.

The second part of earnings comes from the accuracy of your assessment of your performance. The payoff you are going to get from the beliefs you submit about your performance depends on what your performance actually was. The payoff rule is such that the worst thing to do in terms of payoffs would be to assign 100% chance to an incorrect statement, and the best thing to do would be to assign 100% chance to the correct statement. Your potential payoffs will vary with the specific percent chances you submit, and the correct state. As you submit your beliefs, you will be able to see on your screen the potential payoffs that you can get with this estimation, depending on what your true performance will be. The payoff rule has been designed such that you will be maximizing your ex-ante expected payoff if you are truthful about your assessment of your performance.

After you have made your initial assessment of your performance, you will receive feedback. This feedback will be in the form of revealing whether you were actually in the top 20% of the distribution or not. That is, if you were actually in the top, you will learn it. If you were not in the top, you will get a message that says: "you were not in the top, and may be in the middle or the bottom". The feedback given by the computer will always be truthful. After receiving the feedback, you will be asked to submit your beliefs about your performance again. This estimation will be compensated using the same payoff rule as the first estimation. After you have made this second estimation, the round will end.

Your total earnings in a given round in this first part of the experiment will be the sum of the earnings you get from task performance (50 points*number of correct questions), and the payoffs you get from the initial assessment of your performance and the second assessment of your performance after the feedback.

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.[32]

You will also face more questions and have more time in the actual paid rounds. These

---

[32] A number sequence task was used in the practice rounds, only for illustration purposes. That is, subjects were given questions involving the guessing of the logically following number in a sequence of numbers with a pattern. Subjects were given two questions, and only 30 seconds to work on the task in the practice rounds.

practice rounds are just for you to familiarize yourself with the procedures. Now, please pull out your dividers, and double-click on the MC (red dog) icon on your desktop.

[Go over instructions briefly once again, step by step as they actually play the stages on the computer]

[Talk about the tie-breaker question as they see it on their screen: "After the normal time for the task is up, you will see a single tie-breaker question on your screen. This question will be used by the computer to resolve any ties in the ranking and determination of the top, middle and bottom groups. If one or more participants have the same number of correct questions, the person who got the tie-breaker question right and in a shorter amount of time will be ranked higher by the computer"]

The task that you will be working on in the first four rounds of the experiment is addition. You will be given 15 addition questions at the start of each round, and will have 120 seconds to solve as many questions as you can.

[Show slide with examples]

Before the paid rounds start, you will 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.

[Paid rounds start after everyone has completed the quiz successfully, 4 rounds with the first task (addition) is played]

[Before the next 4 rounds with the verbal task starts, an example is given on the screen about the type of task they will work on in the coming 4 rounds].

Instructions for Part 2--Non-performance rounds

In this second part of the experiment, there will be no task performance stage. In these rounds, "top", "middle" and "bottom" will have nothing to do with your performance, but will just be "states of the world" that have certain chances of being chosen by the computer. In other words, in each round, the computer will pick either top, or middle, or bottom. You will be given information on how likely each state is to be picked, before the round starts. For example: if the computer has a 50% chance of picking top, 25% chance of middle, and 25% chance of bottom, you will see this on your computer screen. You will then be asked to submit your beliefs about the likelihood of different states to be picked in that round. You will be paid for the accuracy of these beliefs, in the same fashion as in the performance rounds.

After you make this initial estimation, you will be given some information about the state actually picked by the computer. The computer will tell you whether the chosen state was top,

or not top. After receiving this information you will be asked to submit your beliefs again about the likelihood of the three states being the chosen state. This second assessment is also compensated in the same way as the first. We will run 8 rounds with this configuration.

Instructions for Part 3--Non-performance rounds with ambiguity

We will now start the third and last part of the experiment. In this part of the experiment, just like the previous part, there will be no task performance stage. Again, in these rounds, "top", "middle" and "bottom" will just be states of the world that have certain chances of being chosen by the computer. You will again be given some information on how likely it is for certain states to be picked, before the round starts. However, you will see the likelihood of only one of the three states. That is, you will know the percent chance with which one of the three states will be picked, but will not know the likelihood of the other two states. You will then be asked to submit your beliefs about the likelihood for all three states of being picked in that round. You will be paid for the accuracy of these beliefs, in the same fashion as in the previous rounds.

After you make this initial estimation, you will again be given some information about the state actually picked by the computer. The computer will tell you whether the chosen state was top, or not top. After receiving this information you will be asked to submit your beliefs again about the likelihood of the three states being the chosen state. This second assessment is also compensated in the same way as the first. We will run 8 rounds with this configuration. [After the 8 rounds] The decision part of the experiment is now over. You will now see a survey on your screen. Please fill in your answers in the text boxes. Thank you very much for your participation.

Post-Experiment Survey Questions

1. Which of the two tasks (algebra or verbal) do you think is more reflective of a person's overall ability or intelligence?
2. Which of the two tasks did you find more difficult?
3. Which of the two tasks did you enjoy more?
4. If/when you received favorable feedback about your performance (the message "you were in the top"), did it affect your morale? Please answer on a scale of --10 to 10: -10=extremely negatively, 0=not at all, 10=extremely positively. If not applicable, please write N/A.
5. If/when you received favorable feedback about your performance, did it affect your morale? Please answer on a scale of --10 to 10: -10=extremely negatively, 0=not at all, 10=extremely

positively. If not applicable, please write N/A.

6. If/when you received unfavorable feedback about your performance (the message "you were not in the top"), did it affect your morale? Please answer on a scale of --10 to 10: -10=extremely negatively, 0=not at all, 10=extremely positively. If not applicable, please write N/A.

7. If/when you received unfavorable information about your performance, do you think it affected your performance in the subsequent rounds? Please answer on a scale of --10 to 10: -10=extremely negatively, 0=not at all, 10=extremely positively. If not applicable, please write N/A.

8. How important was it for you to do better than others in the addition task? Please answer on a scale of 0 to 10: 0=not important at all, 10=extremely important.

9. How important was it for you to do better than others in the verbal task? Please answer on a scale of 0 to 10: 0=not important at all, 10=extremely important.

10. Please rate your understanding of the experiment on a scale of 0 to 10. How clear were the instructions?

11. What is your major?

12. What is your gender?