Measuring Responsiveness to Feedback as a Personal Trait

Thomas Buser[1]
Leonie Gerhards[2]
Joël J. van der Weele[1]

[1] Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute, the Netherlands;
[2] University of Hamburg, Germany.

Measuring responsiveness to feedback as a personal trait

Thomas Buser[*]   Leonie Gerhards[◇]   Joël van der Weele[×]

June 2, 2016

Abstract

People typically update their beliefs about their own abilities too little in response to feedback, a phenomenon known as "conservatism", and some studies suggest that they overweight good relative to bad signals ("asymmetry"). We measure individual conservatism and asymmetry in three tasks that test different cognitive skills, and study entry into a winner-takes-all competition based on similar skills. We show that individual differences in feedback responsiveness explain an important part of the variation in confidence and competition entry decisions. Conservatism is correlated across tasks and predicts competition entry both by influencing beliefs and independently of beliefs, suggesting it can be considered a personal trait. Subjects tend to be more conservative in tasks that they see as more ego-relevant and women are more conservative than men. Asymmetry is less stable across tasks, but predicts competition entry by increasing self-confidence.

---

[*]University of Amsterdam, Tinbergen Institute. Contact: t.buser@uva.nl

[◇]University of Hamburg. Contact: leonie.gerhards@wiso.uni-hamburg.de

[×]University of Amsterdam, CREED, Tinbergen Institute. Contact: vdweele@uva.nl

1   Introduction

The ability and willingness to take into account feedback on individual performance can profoundly influence life decisions. For instance, it determines the formation of under- or overconfident beliefs that are associated with personal costs. Reflecting the importance of the topic, there is a substantial literature in psychology and economics on Bayesian updating and the role of feedback in belief formation, which we review in more detail below. This literature shows that people are generally "conservative", i.e. they are less responsive to noisy feedback than Bayesian theory prescribes. Möbius et al. (2014, henceforth MNNR) also suggest that people update in an "asymmetric" way about their own ability, placing more weight on positive than on negative feedback.

In our experiment, we investigate whether feedback responsiveness can be considered a personal trait that affects economic decisions. We measure how participants update their beliefs about their relative performance on three cognitive tasks. The tasks require three distinct cognitive capabilities, namely verbal skills, calculation skills and pattern recognition. The feedback after each task consists of six consecutive noisy signals about the likelihood that they scored in the top half of their reference group. Our feedback structure is inspired by MNNR, but we elicit substantially more observations per individual. This allows us to construct individual measures of conservatism and asymmetry for each subject, defined as the individual deviation from the average updates of subjects with similar priors and feedback content.

Using these measures, we generate a number of new insights. First, we investigate whether conservatism and asymmetry can be considered personal traits, using within-subject variation to assess the consistency of our measures across tasks. From this analysis, conservatism appears as a stable trait while asymmetry does not. A second contribution is to show the relation between our measures and post-feedback confidence and competitiveness. Variation in feedback responsiveness between individuals explains a substantial share of post-feedback beliefs, controlling for the content of the feedback. Specifically, conservatism raises beliefs after receiving many bad signals and lowers them after receiving many good signals. Asymmetry raises beliefs across the board.

We study competitiveness in a final task, composed from exercises similar to each of the previous tasks. Subjects choose whether they want to get paid on the basis of an individual piece rate, or on the basis of a winner-takes-all competition against another subject. A number of studies, reviewed below, show that the choice to enter such a competition with another subject is predictive of career choices outside of the lab. We find that conservatism predicts entry both through influencing final beliefs and independently of beliefs. Asymmetry also predicts entry by raising final beliefs. Thus, being more conservative and asymmetric is good for high-performing subjects with an expected gain from competition, and bad for the remaining subjects.

Our third contribution relates to recent theoretical and experimental studies that explore the link between updating behavior and the ego-relevance of the task. Existing evidence on this link

is based on between-subject comparisons. We exploit within-subject variation in ego relevance which stems from the fact that student participants from different majors attach different ego-relevance to each task. We find that ego-relevance leads subjects to update more conservatively but not more asymmetrically.

Finally, we use both within and between-individual variation to replicate previous estimates of conservatism and asymmetry relative to a Bayesian benchmark. We confirm that subjects are on average strongly conservative. We do not find clear evidence that subjects are asymmetric on average, although asymmetry does occur through an increase in updating "mistakes" after a bad signal.

Our findings suggest that individual conservatism and, to a lesser degree, asymmetry, is an important driver of self-confidence and competitiveness. In the conclusion, we provide a range of domains in which we expect these attributes to affect peoples' decisions and discuss how our study can help to reduce the negative effects of faulty updating.

2 Literature

A sizable literature in psychology on belief updating has identified that people are generally "conservative", i.e. they are less responsive to noisy feedback than Bayesian theory prescribes (Slovic and Lichtenstein, 1971; Fischhoff and Beyth-Marom, 1983). More recent evidence shows that when feedback is relevant to the ego or identity of experimental participants, they tend to update differently (Möbius et al., 2014; Eil and Rao, 2011; Ertac, 2011; Grossman and Owens, 2012). These studies provide a link between updating behavior and overconfidence, as well as to a large literature on self-serving or ego biases in information processes (e.g. Kunda, 1990).

More specifically, MNNR find evidence for "asymmetric" updating on ego-relevant tasks, where subjects place more weight on positive than on negative feedback. Sharot et al. (2011) find behavioral and neurological evidence that subjects update asymmetrically about the possibility of negative life events happening to them. However, it is not clear how robust these findings are. Schwardmann and Van der Weele (2016) and Coutts (2016) provide replications of the MNNR framework. While both papers find substantial conservatism in updating, they do not replicate the asymmetry result of MNNR, and even find a tendency to overweight negative rather than positive signals. The same is true for Ertac (2011), who uses a different signal structure, making her results difficult to compare directly. Kuhnen (2015) finds that subjects react more strongly to bad outcomes relative to good outcomes when these take place in a loss domain, but not when they take place in a gain domain. Thus, the degree to which people update asymmetrically is still very much an open question.

Updating biases are a potential source of overconfidence, probably the most prominent and most discussed phenomenon in the literature on belief biases. Hundreds of studies have demonstrated that people are generally overconfident about their own ability and intelligence (see Moore

and Healy (2008) for an overview). Overconfidence has been cited as a reason for over-entry into self-employment (Camerer and Lovallo (1999); Koellinger et al. (2007)) as well as the source of suboptimal financial decision making (Barber and Odean (2001); Malmendier and Tate (2008)). As a result, overconfidence is generally associated with both personal and social welfare costs.[1]

We test whether updating biases predict the decision to enter a competition with another participant. Following Niederle and Vesterlund (2007), experimental studies which measure individual willingness to compete have received a lot of attention. Their main finding is that, conditional on performance, women are a lot less likely to choose a winner-takes-all competition over a non-competitive piece rate than are men (see Croson and Gneezy (2009) and Niederle and Vesterlund (2011) for surveys, and Flory et al. (2015) for a field experiment). A growing literature confirms the external relevance of competition decisions made in the lab for predicting career choices. Buser et al. (2014) show that competing in an experiment predicts the study choices of Dutch high-school students. Other studies have found correlations with the choice of entering a highly competitive university entrance exam in China (Zhang, 2013), starting salary and industry choice of graduating MBA students (Reuben et al., 2015), as well as future salary expectations of undergraduate students (Reuben et al., 2016), the investment choices of Tanzanian entrepreneurs (Berge et al., 2015) and monthly earnings in a diverse sample of the Dutch population (Buser et al., 2015).

Closest to our paper is an early version of MNNR, in which the authors construct individual measures of conservatism and asymmetry (Möbius et al., 2007). The authors conduct a follow-up competition experiment six weeks after the main experiment using a different task. They find that conservatism is negatively correlated with choosing the competition, while asymmetry is positively but insignificantly correlated. Our results go beyond this by changing the definition of the measures, so they are less likely to conflate asymmetry and conservatism. More importantly, our dataset is much larger. While Möbius et al. (2007) record 4 updating rounds per person for 102 individuals, we have data for 18 updating rounds over three different cognitive tasks for 297 individuals. This increases the precision of the individual measures and allows us to see whether individual updating tendencies are stable across tasks.

3 Design

Our experimental design is based on MNNR. The experiment was programmed in z-Tree (Fischbacher, 2007), and run at Aarhus University, Denmark, in the spring and summer of 2015. Overall, 22 sessions took place between April and September, with each session comprising between 8 and 24 subjects. Sessions lasted on average 70 minutes, including the preparation for payments. In total, 297 students from diverse study backgrounds participated in the experiment. Each session was composed of students with the same faculty, i.e. from either social science, sci-

---

[1]Daniel Kahneman argues that overconfidence is the bias he would eliminate first if he had a magic wand. See http://www.theguardian.com/books/2015/jul/18/daniel-kahneman-books-interview.

ence or the humanities.[2] Students received a show-up fee of 40 Danish Crowns (DKK, $6.00 or €5.40).[3] Average payments during the experiment was 176 DKK with a minimum of 20 and a maximum of 980 DKK.

Subjects read all instructions explaining the experiment on their computer screens. Additionally, they received a copy of the instructions in printed form.[4] It was explained that the experiment would have four parts, one of which would be randomly selected for payment. Participants were told that the first three parts involved performance and feedback on a task as well as the elicitation of their beliefs, and that specific instructions for the last part would be displayed on the subjects' screens after the first three parts had concluded. The instructions also specified that in each task each participant would be randomly matched with 7 others, and that their performance would be compared with the participants within that group.

We then explained the belief elicitation procedure. We elicited the probability about the event that participants were in the top half of their group of 8. To incentivize truthful reporting of beliefs, we used a variation of the Becker-DeGroot-Marshak (BDM) procedure, also known as "matching probabilities" or "reservation probabilities". Participants were asked to indicate which probability p makes them indifferent between winning a monetary prize with probability p, and winning the same prize when an uncertain event E – in our experiment being in the top half – occurs. After participants indicate p, the computer draws a random probability and participants are awarded their preferred lottery for that probability. Under this mechanism, reporting the true subjective probability of E maximizes expected value, regardless of risk preferences (see Schlag et al. (2015) for a more elaborate explanation, as well as a discussion of the origins of the mechanism). We explained this procedure, and stressed the fact that truthful reporting maximizes expected earnings, using several numerical examples to demonstrate this point. This stage took about 15 minutes including several control questions about the mechanics of the belief elicitation procedure.

Subjects then were introduced to the first of three different tasks. Each task was composed of a series of puzzles, and subjects were asked to complete as many puzzles as they could within a time frame of five minutes. Their score on the task would be the number of correct answers minus one-half times the number of incorrect answers. The first task, which we will refer to as "Raven", consisted of a series of Raven matrices, where subjects have to select one out of eight options that logically completes a given pattern (subjects were told that "this exercise is

---

[2]In total, 101 social science students took part, of whom 51 males, including students in Economics and Business administration, International Business, Public Policy, Innovation Management, Economics and Management, Sustainability, and Law. In total 97 science students participated, of whom 55 male, including students from Physics, Health technology, Mathematics, Engineering, Geology, (Molecular) Biology, IT Engineering, Chemistry. Finally, 99 students from the humanities took part, among whom 30 males, from study backgrounds like European Studies, International Relations, Human Security, Journalism, Japan studies, Languages (mainly English, Spanish, Italian, German), Culture and Linguistics.

[3]At the time of the experiment, the exchange rate of 1 DKK was $0.15 or €0.135.

[4]All instructions and screen shots of the experimental program can be downloaded from `https://www.dropbox.com/s/hmv3uq2vfwbxs85/AllInstructtionsAndScreens.pdf?dl=0`.

designed to measure your general intelligence (IQ)"). In the second task, which we will refer to as "Anagram", subjects were asked to formulate an anagram of a word displayed on the screen, before moving to the next word (subjects were told that "this exercise is designed to measure your ability for languages"). In the third task, which we will refer to as "Matrix", subjects were shown a 3×3 matrix filled with numbers between 0 and 10, with two decimal places. The task was to select the two numbers that added up to 10 (subjects were told that "this exercise is designed to measure your mathematical ability").[5]

The order of tasks was counterbalanced between sessions, in order to account for effects of depletion or boredom. The details for each task were explained only after the previous task had been completed. Subjects earned 8 DKK for each correct answer and lost 4 DKK for each incorrect answer. We explained to them that their payment could not fall below 0.

After each task, we elicited subjective beliefs about a subject's relative performance. Specifically, we asked participants for their belief that they were in the top half using the BDM procedure described above. After participants submitted their initial beliefs, we gave them a sequence of noisy but informative feedback signals about their performance. Participants were told that the computer would show them either a red or a black ball. The ball was drawn from one of two virtual urns, each containing 10 balls of different colors. If their performance was actually in the top half of their group, the ball would come from an urn with 7 black balls and 3 red balls. If their performance was not in the top half, the ball would come from an urn with 7 red balls and 3 black balls. Thus, a black ball constituted "good news" about their performance, a red ball "bad news". After subjects observed the ball, they reported their belief about being in top half for a second time. This process was repeated five more times, resulting in six updating measurements for each participant for each task, and 18 belief updates overall. The prize at stake in the belief elicitation task was 10 DKK in each round of belief elicitation.

After the third task, subjects were informed about the rules of the fourth and final task, which consisted of the same kind of puzzles as the previous three tasks, mixed in equal proportions. Before performing this task, subjects were offered a choice between two payment systems, as in Niederle and Vesterlund's (2007) "Task 3". The first option consisted of a piece-rate scheme, where the payment depended on their score in a linear way (12 DKK for a correct answer, -6 DKK for an incorrect one). The second option was to enter into a competition, where their score was compared to that of some randomly chosen other participant. If their score exceeded that of their matched partner, they would receive a payment of 24 DKK for each correct answer, and -12 DKK for each incorrect one. If their matched partner scored higher, they would receive a payment of zero. In this round there was no belief elicitation.

After the competition choice, subjects were asked to fill out a (non-incentivized) question-

---

[5]We also implemented two different levels of difficulty for each task. The aim was to generate an additional source of diversity in confidence levels that could be used in our estimation procedures. As it turned out, task difficulty did not significantly affect initial confidence levels, so we pool the data from all levels of difficulty in our analysis. During the sessions subjects were always compared to other subjects who performed the task at the same difficulty level.

naire. Among other things, we asked how relevant participants thought the skills tested in each of the three tasks were for success in their field of study. We will use the answers to these questions as an individual measure of ego relevance of the tasks. Subjects also completed a narcissism scale, based on Konrath et al. (2014), and answered several questions related to their competitiveness, risk taking, and a range of activities which require confidence like playing sports or music on a high level.

4 Measuring individual responsiveness to feedback

Our measurement of feedback responsiveness will focus on asymmetry and conservatism. In the literature, conservatism refers to the tendency to update beliefs *too little* in response to new information relative to the Bayesian benchmark. Asymmetry refers to the difference in response to good and bad signals, for a given prior belief. Asymmetry is positive if a subject updates more in the positive than in the negative direction.

Measuring asymmetry and conservatism at the individual rather than the aggregate level introduces some complications. For example, at the aggregate level, one can estimate asymmetry by comparing different individuals with similar priors, who observe opposing signals. At the individual level this is not possible, as we cannot confront the same individual with different signals without changing beliefs between the two cases. In addition, while evaluating updates relative to the Bayesian benchmark is straightforward in an aggregate analysis, it is problematic for individuals with relatively high or low priors. In this case, the rational Bayesian update is small, so small absolute differences in updates translate into huge relative differences and therefore huge differences in measurement, making individual estimates very unstable.

To circumvent these issues, our individual measures of feedback responsiveness are based on absolute deviations from the average update for people with the same prior and the same signal, rather than the Bayesian benchmark. Apart from changing the scale of the measure, this shift in benchmark affects terminology; the Bayesian/rational benchmark allows the labeling of deviations as "biases", ours only allows us to talk about "relative" conservatism and asymmetry. In Section 7 we replicate earlier aggregate results using a Bayesian benchmark.

Deviations are calculated on the basis of the residuals of the following regression model, which is run separately for positive and negative signals.[6]

$$\Delta\mu_{int} = \alpha + \beta_1\mu_{in,t-1} + \beta_2\mu_{in,t-1}^2 + \gamma_1 1_1 + \gamma_2 1_2 + ... + \gamma_{10} 1_{10} + \epsilon_{int} \qquad (1)$$

Here $\Delta\mu_{int} := \mu_{int} - \mu_{in,t-1}$ is the update by individual i in feedback round t and task n and $1_1, 1_2...1_{10}$ represent dummies indicating that $0 \leq \mu_{in,t-1} < 0.1, 0.1 \leq \mu_{in,t-1} < 0.2,..., 0.9 \leq$

---

[6]Our individual measures of conservatism and asymmetry are similar to the ones developed by Möbius et al. (2007). One difference is that we use a more flexible function to approximate average updating behavior. More importantly, we give equal weight to positive and negative updates which avoids conflating asymmetry and conservatism for subjects with an unequal amount of positive and negative signals (see below).

$\mu_{in,t-1} \leq 1$ respectively. These dummies introduce an additional (piecewise) flexibility to our predicted average updates. The residuals of this regression thus measure individual deviations from the average updating for either positive or negative signals, conditional on the prior of each individual.

For each individual i and for each round t and task n, let regression residuals from (1) be denoted by r_{int}. Individual asymmetry in task n is then defined as

$$A_{in} := \frac{1}{N_{in}^-} \sum_{t=1}^{6} 1_{(s_{int}=L)} * r_{int} + \frac{1}{N_{in}^+} \sum_{t=1}^{6} 1_{(s_{int}=H)} * r_{int}, \tag{2}$$

where N_{in}^+ and N_{in}^- are the observed number of positive and negative signals respectively. Thus, A_{in} is the sum of the average residual after a positive and the average residual after a negative signal. It is positive if an individual updates upwards more than the average person. Note that this measure is only defined for people who receive at least one positive and at least one negative signal on a given task, which means that we lose a little over 10% of our observations.

To obtain an overall individual measure for asymmetry we calculate an analogous measure across all 3 tasks, spanning 18 updating decisions.

$$A_i := \frac{1}{N_i^-} \sum_{t=1}^{18} 1_{(s_{it}=L)} * r_{it} + \frac{1}{N_i^+} \sum_{t=1}^{18} 1_{(s_{it}=H)} * r_{it}, \tag{3}$$

Correspondingly, task-specific conservatism is defined as

$$C_{in} := \frac{1}{N_{in}^-} \sum_{t=1}^{6} 1_{(s_{int}=L)} * r_{int} - \frac{1}{N_{in}^+} \sum_{t=1}^{6} 1_{(s_{int}=H)} * r_{int}. \tag{4}$$

In words, C_{in} is the average residual after a negative signal minus the average residual after a positive update. Thus, C_{in} is positive if an individual updates upward less than average after a positive signal and updates downward less than average after a negative signal. We only calculate this measure for people who receive at least one positive and at least one negative signal as it is impossible to distinguish conservatism from asymmetry for those with only positive or only negative signals. To obtain an overall individual measure of conservatism we calculate an analogous measure across all 3 tasks, spanning 18 updating decisions.

$$C_i := \frac{1}{N_i^-} \sum_{t=1}^{18} 1_{(s_{it}=L)} * r_{it} - \frac{1}{N_i^+} \sum_{t=1}^{18} 1_{(s_{it}=H)} * r_{it}. \tag{5}$$

Our measurements do not weight the relative importance of updates following negative and positive signals by the amount of such signals. Doing so would lead to a conflation of our two measures. For example, a subject who is relatively conservative and receives more positive than negative signals would have a negative bias in asymmetry, as the downward residuals after a

8

positive signal would be overweighted relative to the upward residuals after a negative signal.

Finally, updates in the wrong direction pose a problem for the computation of our relative measurements. An update of the wrong sign has a potentially large impact on our measures, as it is likely to result in a large residual. However, as it seems likely that such updates at least partly reflect "mistakes", this may unduly influence our measures. To mitigate this effect, we treat wrongly signed updates as zero updates in the calculation of our individual measures. The statistical results go through if we use the original "mistaken" update instead.

4.1   Noise and updating mistakes

In the absence of knowledge of the true beliefs, the quality of our data is difficult to assess. However, one indicator is the amount of "mistakes" that are made in updating. Table 1 shows the number of wrongly signed updates in each task. Per task and round, subjects update in the wrong direction in around 10% of the cases (usually between 25 and 35 out of 297 observations).

One reason why we do not fully exclude updating mistakes from the statistical analysis is that they display an asymmetric pattern. The bottom two rows of Table 1 show that the proportion of mistakes roughly doubles when the signal is negative. This difference is significant at the 1%-level for each of the tasks.[7] This implies that wrong updates are not purely noise, but seem to be partly driven by self-serving motives. The fraction of mistake updates in Table 1 is about the same as that found in MNNR, and four percentage points higher than in Schwardmann and Van der Weele (2016), studies that use the same belief elicitation mechanism and a similar feedback mechanism.[8]

Table 1 demonstrates the advantages of eliciting feedback responsiveness in more than one task. While one needs a substantial number of observations per individual to distinguish between conservatism and asymmetry, there clearly is a limit to the amount of signals one can elicit about a single event. First, after multiple signals in the same direction, subjects will hit the "boundaries" of the probability scales. This is clear from the final three columns of Table 1 that show the number of subjects who report either 100% or 0% confidence. There is a marked uptick in numbers in the final two updating rounds. In principle, one could reduce the likelihood ratio of the signal in order to avoid this. However, a second problem is that Table 1 shows that the number of wrongly signed updates increases by more than 100% in the final updating round of each task. Whatever the reason for this pattern (perhaps subjects got bored), it implies that eliciting more than 5 updates on the same event is problematic. Thus, based on our findings, it seems that introducing several events – or tasks, as in the present study – about which to elicit probabilities is necessary to gather a sufficient amount of data.

---

[7]P-values are from regressions of a binary indicator of having made an updating mistake on a dummy indicating that the signal was negative. Standard errors were clustered at the individual level.

[8]In the instructions, Schwardmann and Van der Weele (2016) use a choice list with automatic implementation of a multiple switching point for the BDM design, rather than the slightly more abstract method of eliciting a reservation probability. This may explain the lower rate of wrong updates. Schwardmann and Van der Weele (2016) find the same asymmetry in wrong updates when it comes to positive and negative signals.

Round	Wrongly signed updates			Boundary cases		
	A	M	R	A	M	R
1	36	30	32	7	8	5
2	25	23	25	10	13	7
3	29	38	25	14	12	15
4	25	28	24	19	19	19
5	26	38	34	25	27	30
6	83	85	77	31	36	35
Proportion +	7.69	8.78	7.10			
Proportion −	17.37	18.23	17.21			

Table 1: Overview of wrongly signed updates (Columns 2-4) and subjects who report either 100% or 0% confidence (Columns 5-7) by task. A stands for "Anagram", M stands for "Matrices" and R stands for "Raven". The last two rows show the proportion of wrongly signed updates after positive (+) and negative (−) signals.

Finally, an important question is whether our measures are affected by the *order* in which positive and negative signals are received. We address this issue in Section 7, where we show that the path dependence of signals does not have a strong influence on updating behavior.

5 Results: feedback responsiveness

5.1 Initial beliefs and descriptive statistics

Figure 1 shows the distributions of initial and final beliefs (that is, the beliefs the subjects held about being in the top half before the first and after the sixth round of feedback, respectively) over all tasks. Mean initial beliefs are 54% (s.d. 0.13), indicating a modest amount of overconfidence, as only 50% can actually be in the top half. Average beliefs in the final round are roughly the same as in the initial round (55%), but the standard deviation increases to 0.19.

To understand whether beliefs have become better calibrated we run regressions of initial and final beliefs in each task on binary variables indicating the actual performance rank of subjects. In the initial round, we find that ranks explain more of the variation in beliefs in the Matrix task ($R^2 = 0.30$) than in the Anagram task ($R^2 = 0.21$) or the Raven task ($R^2 = 0.14$). In each task, we find that the R^2 of the model increases between 7 and 9 percentage points between the first and last round of belief elicitation. Thus, on average feedback does indeed succeed in providing a tighter fit between actual performance and beliefs over time.

5.2 Consistency of feedback responsiveness across tasks.

An important motivating question for our research is whether feedback responsiveness can be considered a trait of the individual. To answer this question, we look at the consistency of conservatism and asymmetry across tasks. Table 2 displays pairwise correlations between our
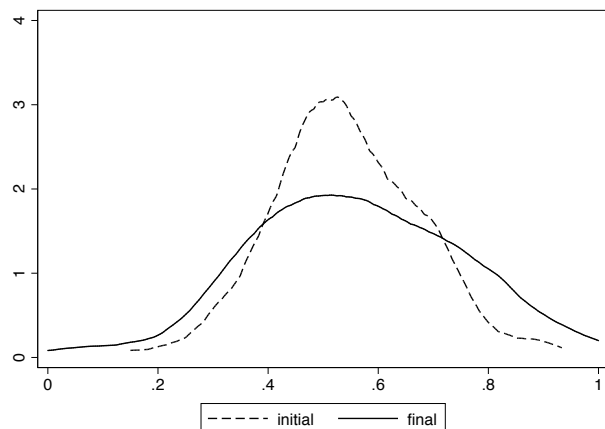
Figure 1: Density plots of initial and final belief distributions.

measures over tasks. For conservatism, we find highly significant correlations in the range of $0.14 - 0.36$. For asymmetry, correlations are smaller, and the only significant correlation is that between asymmetry in the matrix and anagram tasks. This latter result is puzzling, as these tasks are very different from each other and are seen as relevant by different people.[9]

	C(M)	C(R)		A(M)	A(R)
C(A)	0.220***	0.356***	A(A)	0.145**	-0.011
C(M)		0.142**	A(M)		0.082

Table 2: Spearman's pairwise correlations of measures over task. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. A stands for "Anagram", M stands for "Matrices" and R stands for "Raven".

Summary 1 *For a given individual, conservatism displays robust correlation over tasks, whereas asymmetry does not.*

5.3 Impact of feedback responsiveness on final beliefs

To better understand the importance of individual differences in feedback responsiveness, we look at their effect on the beliefs in the final round of each task. As the impact of conservatism

---

[9]As we show below, individuals who attach more relevance to a task tend to be more conservative in updating their beliefs about their performance in that task. However, the estimated correlations of conservatism across tasks are not due to correlations of relevance across tasks. If we regress individual conservatism in each task on individual relevance and correlate the residuals of this regression across tasks, the estimated correlations are very similar to the ones reported in Table 2

and asymmetry depends on received feedback, we run a linear regression of the form

$$\mu_{in} = \alpha + \beta_0^C * C_{in} + \beta_0^A * A_{in} + \sum_{s=1}^{5} \beta_s 1_{(s_{in}^+=s)} + \sum_{s=1}^{5} \beta_s^C 1_{(s_{in}^+=s)} * C_{in} + \sum_{s=1}^{5} \beta_s^A 1_{(s_{in}^+=s)} * A_{in} + \varepsilon_{in},$$

(6)

where μ_{in} is the final belief after the last round of feedback, and $1_{(s_{in}^+=1)}, 1_{(s_{in}^+=2)}, ..., 1_{(s_{in}^+=5)}$ represent dummies taking a value of 1 if subject i got the corresponding number of positive signals in task n. A_{in} and C_{in} are defined as in (2) and (4) above.

The left panel of Figure 2 shows the effect of an increase of one standard deviation in conservatism, separately for the amount of s positive signals received, i.e. $\beta_0^C + \beta_s^C$. The data confirm that conservatism raises final beliefs for people who receive many bad signals and lowers them for people who receive many good signals, cushioning the impact of new information. The right panel of Figure 2 shows a similar graph for the effect of a standard deviation increase in asymmetry, i.e. $\beta_0^A + \beta_s^A$. The impact of asymmetry is to raise final beliefs for any combination of signals. The effect is highest when signals are mixed, as the absolute size of the belief updates, and hence the effect of asymmetry, tend to be larger in this case.
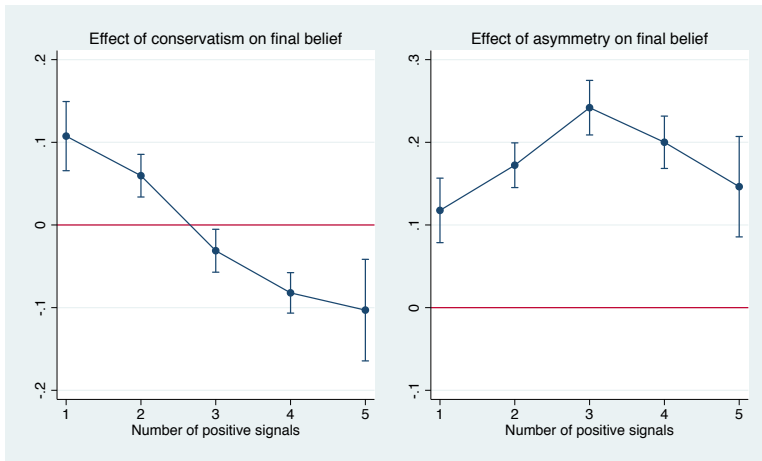

Figure 2: The impact of conservatism and asymmetry on final beliefs after the last updating round, split by the number of signals.

The direction of the effects shown in Figure 2 is implied by our definitions, and should not be surprising. More interesting is the size of the effects of both conservatism and asymmetry. For subjects with unbalanced signals, conservatism will matter most. Specifically, an increase of one standard deviation in conservatism raises final beliefs by 10 percentage points for subjects who received 1 positive and 5 negative signals, and lowers them by about the same amount for people who receive 5 positive and 1 negative signals. By contrast, asymmetry is most important for people who saw a more balanced pattern of signals. A one-standard deviation increase in

asymmetry leads to an average increase in post-feedback beliefs of over 20 percentage points for a person who saw 3 good and 3 bad signals, and therefore should not have adjusted beliefs at all.

In each individual task, the standard deviation of final beliefs is about 30 percentage points, implying that for any realization of the signal structure, variation in feedback responsiveness explains a substantial part of the variation in final beliefs. In fact, the adjusted R^2 of our regression model in (6) is 67%, which falls to 46% when we drop the responsiveness measures and their interactions from the model. Thus our responsiveness measures explain about an additional 21 percentage points of total variation in final beliefs after controlling for signals. It is interesting to ask whether this is due to feedback responsiveness explaining variations in final beliefs across individuals, within individuals across tasks or both. To answer this question, we run the same regression including individual fixed effects. The within-subjects (across-tasks) R^2 is 0.68 with and 0.50 without our responsiveness measures, while the between-subjects R^2 is 0.65 with and 0.43 without our responsiveness measures. This demonstrates that there is meaningful individual heterogeneity in responsiveness to feedback and that conservatism and asymmetry are important determinants of individual differences in belief updating and confidence.

Summary 2 *When observing unbalanced feedback with many more positive than negative signals or vice versa, a one standard deviation change in conservatism changes final beliefs by up to 10 percentage points, while asymmetry matters relatively less. For balanced feedback with similar amounts of both positive and negative signals, a one standard deviation change in asymmetry changes final beliefs by about 20 percentage points. Controlling for feedback content, conservatism and asymmetry jointly explain an additional 21 percentage points of the between-subjects variation in final beliefs after controlling for signals.*

5.4 Ego-relevance and gender effects

We now turn to an analysis of heterogeneity in feedback responsiveness related to task and subject characteristics. In turn, we discuss the role of ego-relevance and gender.

The effect of ego-relevance. Past research suggests that the ego-relevance of a task changes belief updating, and can trigger or increase asymmetry and conservatism (see Section 2), indicating that responsiveness to feedback is motivated by the goal of protecting a persons' self-image. Furthermore, Grossman and Owens (2012) show that ego-relevance also leads to initial overconfidence in the form of higher priors.

The variation in study background in our experimental sample allows us to study this directly, as it creates variation in the ego-relevance of the different experimental tasks. We measured the relevance of each task with a questionnaire item.[10] We conjecture that participants who

---

[10]For the Raven test, relevance was measured as the answer to the following question, on a Likert scale from 1

attach higher relevance to a particular task will be more confident and show greater degrees of asymmetry in that task. Furthermore, if subjects are more confident in tasks that they consider more relevant, they would have an ego motivation to be more conservative as well in order to protect any ego utility they derive from such confidence. To see whether these conjectures are borne out in the data, we first investigate whether relevance affected subjects' beliefs about their own relative performance before they received any feedback. To this end, we regress initial beliefs in each task on the questionnaire measure of relevance. We also include a gender dummy in these regressions, which is discussed below.

	(1)	(2)
Female	-0.050***	-0.030**
	(0.015)	(0.014)
Relevance	0.029***	0.021***
	(0.004)	(0.004)
Scores & ranks		✓
Relevance	0.030***	0.021***
	(0.004)	(0.003)
Scores & ranks		✓
Individual fixed effects	✓	✓
N	891	891

Table 3: OLS regressions of initial beliefs on task relevance and gender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the individual level.

The results, reported in Table 3, show that relevance has a highly significant effect on initial beliefs. Heterogeneity in scores can only explain part of this effect, as we show in Column (2) where we control for scores and normalized within-session ranks. The lower part of the table shows that the effect of relevance on initial beliefs is robust to the introduction of individual fixed effects. This implies that the effect stems from within-subject variation in relevance across tasks. That is, the same individual is more confident in tasks that measure skills that are more ego-relevant.

This result is consistent with the idea that confidence is ego-motivated: participants who think a task is more relevant to the kind of intelligence they need for their chosen career path, are more likely to rate themselves above others. Alternatively, it could mean that people choose the kind of studies for which they hold high beliefs on possessing the relevant skills. Note however that the pattern cannot be explained by participants who think that their study background gives them an advantage over others, as they knew that all other participants in the sessions had the same study background.

---

to 7: "I think the pattern completion problems I solved in this experiment are indicative of the kind of intelligence needed in my field of study." Each subject answered this question three times, one time for each task. In a regression of relevance on study background, we do indeed find that students from a science background find the Raven task significantly more relevant than students from social sciences or humanities. Conversely, students from the humanities attach significantly more relevance to the Anagrams task and less to the Matrix task than either of the other groups. We also included a control for gender in these regressions, as our study background samples are not gender balanced.

	(1) Asym.	(2) Cons.	(3) Asym.	(4) Cons.	(5) Asym.	(6) Cons.
Female	-0.081	0.173**	-0.033	0.171**	-0.024	0.176**
	(0.079)	(0.084)	(0.076)	(0.083)	(0.074)	(0.082)
Relevance	0.021	0.043**	0.006	0.043**	0.001	0.039*
	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)
Scores & ranks			✓	✓	✓	✓
Initial beliefs					✓	✓
Relevance	0.023	0.040**	0.008	0.039**	0.002	0.037*
	(0.021)	(0.019)	(0.021)	(0.020)	(0.021)	(0.020)
Scores & ranks			✓	✓	✓	✓
Initial beliefs					✓	✓
Individual fixed effects	✓	✓	✓	✓	✓	✓
N	798	798	798	798	798	798

Table 4: OLS regressions of individual asymmetry (Asym.) and conservatism (Cons.) on task relevance and gender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the individual level. Each person-task combination is one observation. Regressions with asymmetry as the outcome additionally control for conservatism and vice versa.

To test the extent to which ego-relevance can explain the variation in feedback responsiveness across tasks, we regress the individual measures of asymmetry and conservatism for each task on the relevance that an individual attaches to that task. We again include gender in the regressions. The results in Table 4 show that the impact of relevance on both asymmetry and conservatism is positive. For asymmetry, the estimated coefficient is small and insignificant. For conservatism, the effect is statistically significant and, moreover, robust to controlling for scores, ranks and initial beliefs. In the regressions reported in the lower part of the table, we add individual fixed effects to compare more and less relevant tasks within subject, disregarding between subject variation. The effect is equally strong, indicating that the same subject is more conservative in tasks that measure skills which are more ego-relevant.

Combined with the positive effect of relevance on initial beliefs, the results are consistent with the idea that people deceive themselves into thinking that they are good at ego-relevant tasks and become less responsive to feedback in order to preserve these optimistic beliefs.

Summary 3 *We find that the self-reported ego relevance of the task raises initial beliefs and has a significant effect on conservatism. We do not find an effect of ego relevance on asymmetry.*

Gender effects. Earlier studies have consistently found that women are less (over)confident than men, at least in tasks that are perceived to be more masculine (see Barber and Odean (2001) for an overview of the literature). Table 3 provides corroborative evidence for this, with women being about 3 percentage points less confident about being in the top half of performers across all three tasks, after controlling for ability.

MNNR find that women also update more conservatively. This result is replicated using our individual measure in Table 4, where we a see a significant negative effect of a female dummy on

individual conservatism across the three tasks, an effect that is robust to controlling for scores and initial beliefs. We do not find a significant gender difference in asymmetry.

Summary 4 *Women are initially less confident and update more conservatively than men.*

5.5 Questionnaires

In the questionnaire following the competition, we asked several questions to see if conservatism and asymmetry correlate with other aspects of a person's personality and behavior outside the lab. Apart from the relevance of the different tasks, which we discussed above, we asked how engaged subjects are in their studies, whether they have ambitions in academia, and whether they are involved in music and sports at a competitive level. We also asked how risk-seeking, attractive and honest they thought themselves to be, and whether they considered themselves a narcissist. Loosely speaking, our hypotheses were that narcissism, as well as participation in competitive sports or music are correlated with increased asymmetry. However, we are not testing a specific model, nor do we correct for multiple hypothesis testing, so the results should be taken mostly as indicative of fruitful directions for future research.

Table A.1 in the Appendix shows coefficients of ordered probit regressions of our individual measures of conservatism and asymmetry on the questionnaire answers. There are few consistent correlations, and narcissism correlates only weakly with asymmetry after controlling for scores and initial beliefs. An interesting correlation is the one between asymmetry and the likelihood to consider an academic career and enjoyment of talking about one's studies with friends and relatives. This is consistent with the idea that asymmetry increases confidence in academic abilities which is then broadcasted to others. The reverse causality is also possible: our questionnaire items may measure the relevance of study major to our participants, which may induce increased asymmetry to protect ego utility.

6 Feedback responsiveness and competitiveness

In this section we investigate how feedback responsiveness translates to the choice to enter a competition, and whether it may lead to better or worse outcomes for the participants involved.

6.1 Predictive power of conservatism and asymmetry

The final task of the experiment was based on a mixture of matrix, anagram and raven tasks. Before they performed this final task, subjects decided between an individual piece-rate payment and entering into a competition with another subject, as described in Section 3. Posterior beliefs are likely to influence this decision. Specifically, we would expect that asymmetry raises the likelihood to enter a competition because it inflates self-confidence. The hypothesized effect of conservatism is more complex. Conservatism raises final beliefs, and supposedly competition

entry, for those who have many negative signals. However, it should depress competition entry for those who received many positive signals. In addition to this belief channel, it may be that updating behavior is correlated with unobserved personality traits that affect competitiveness.

To investigate these hypotheses, we run probit regressions of the (binary) entry decision on our measures of asymmetry and conservatism, controlling for ability. We also include gender, as it has been shown that women are less likely to enter in a competition (Niederle and Vesterlund, 2007), a finding we confirm in our regressions. The results are reported in Table 5. Column (1) controls for ability (assessed by achieved scores and performance ranks), but not beliefs, and shows that both conservatism and asymmetry have a positive effect on entry. The coefficient for asymmetry is not affected when we control for signals or initial beliefs (Columns 2-3), but virtually disappears when we control for final beliefs (Columns 4-5). This shows that asymmetry affects entry only through its effect on final beliefs rather than through a correlation with any unobserved characteristics.

	(1)	(2)	(3)	(4)	(5)
Female	-0.121**	-0.114**	-0.070	-0.089**	-0.084*
	(0.048)	(0.048)	(0.046)	(0.045)	(0.046)
Asymmetry	0.079***	0.104***	0.066***	0.019	0.039
	(0.025)	(0.029)	(0.024)	(0.025)	(0.033)
Conservatism	0.054**	0.220***	0.047**	0.052**	0.129*
	(0.024)	(0.074)	(0.023)	(0.022)	(0.075)
Conservatism x #signals		-0.018**			-0.009
		(0.008)			(0.008)
#signals		0.024*			0.005
		(0.013)			(0.013)
Scores and ranks	✓	✓	✓	✓	✓
Initial beliefs			✓	✓	✓
Final beliefs				✓	✓
N	297	297	297	297	297

Table 5: Probit regressions of competition entry on measures of feedback responsiveness. Marginal effects reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In order to better understand the belief effect of conservatism, we interact conservatism with the amount of positive signals. The estimated coefficients show that for a person with no positive signals, an increase of one standard deviation in conservatism raises the probability of entry by 22 percentage points, an effect that is larger than the gender effect. If we compare the coefficient of the interaction term with the coefficient of the number of positive signals in Column (2), we see that an increase in conservatism reduces the effect of a positive signal by about 75% (0.018/0.024). The total effect of conservatism is negative for someone with the maximum amount of 18 positive signals. The effect of more positive signals and its interaction with conservatism disappear when we control for final beliefs (Column 5), confirming that these effects indeed go through beliefs. However, there is still a large positive direct effect of conservatism. Together, Columns (4) and (5) clearly suggest that, in addition to its effect on final beliefs, there is an effect of conservatism

that may be a part of a person's personality.[11]

Contrary to our results, Möbius et al. (2007) find that conservatism is negatively correlated with tournament entry. Like us, they find that asymmetry positively predicts tournament entry. However, it is difficult to compare their results to ours. Their measures are based on only 102 individuals and four updating rounds (versus 18 in our case) and their competition experiment is based on a task which is quite different from the one used in the main experiment. They enter conservatism and asymmetry in separate regressions which ignores the fact that for subjects with an unequal amount of positive and negative signals, their measures of conservatism and asymmetry are mechanically correlated and one therefore picks up the effect of the other.

Summary 5 *Asymmetry raises the probability of competition entry by increasing final beliefs. Conservatism raises the probability of competition entry for people with many negative signals, and diminishes it for those with many positive signals. Conservatism also has an independent, positive effect on entry, suggesting it may be correlated with competitive aspects of personality.*

6.2 The cost of conservatism and asymmetry

We now address how asymmetry and conservatism affect expected earnings in the final task. To derive expected earnings, we compute the probability of winning based on the own score, and the empirical distributions of scores in the individual's session. We expect that asymmetry, which increases the chance of entry, will increase earnings for high performers for whom entry is beneficial, and lower earnings for low performers for whom it is not.

To investigate this, Table 6, we regress expected earnings on our updating measures controlling for scores and ranks. Column (1) indicates that both asymmetry and conservatism lead to higher earnings. As mentioned previously, we would expect the effect of asymmetry and (conditional on holding high initial beliefs) conservatism to positively influence earnings only for individuals with a high probability of winning the competition. This idea is explored in Column (2) where we interact asymmetry and conservatism with the average normalized rank across the three tasks.[12] As expected, the effect of both conservatism and asymmetry on expected earnings is negative for low-ranked individuals and positive for high-ranked individuals.

The positive effect of conservatism on earnings may be explained by a "strategy" to protect high (and accurate) beliefs by high performing subjects. This explanation is explored in Columns (3) and (4) of Table 6. Here we only include subjects whose average initial belief (that is, the belief they held after their performance and before the first round of feedback) across the three tasks is higher than 50 percent. Column (3) shows that the positive effect of conservatism on expected earnings is indeed higher for this subgroup and Column (4) shows that both the negative

---

[11]Controlling for the relevance that subjects attach to the three tasks does not alter any of the results.

[12]Normalized mean rank is calculated within session as (# of participants with lower score)/(# of participants-1) and therefore corresponds to the likelihood of having performed better than an arbitrary other subject from the same session.

effect of conservatism for low performers and the positive effect on high performers are larger for this subgroup, although the estimates are imprecise. Column (4) also shows that the estimated effect of asymmetry for high-performing subjects is similar to that for the entire sample, although the coefficients are slightly smaller in magnitude and less precisely estimated.

| | (1) | (2) | (3) | (4) |
		all	mean initial belief$>$0.5	
Female	-1.658	-2.819**	-2.404	-3.631*
	(1.082)	(1.281)	(1.663)	(2.010)
Asymmetry	1.208**	-3.383**	1.174	-2.351
	(0.591)	(1.556)	(0.934)	(2.502)
Conservatism	0.929*	-0.805	2.258***	-1.517
	(0.543)	(1.549)	(0.800)	(2.903)
Conservatism x mean rank		1.866		4.072
		(3.735)		(6.212)
Asymmetry x mean rank		7.904**		5.279
		(3.729)		(5.275)
Mean rank		37.126***		44.368***
		(3.545)		(5.230)
Scores and ranks	✓		✓	
N	297	297	171	171

Table 6: OLS regressions of expected earnings in task 4 on asymmetry and conservatism. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the individual level.

Summary 6 *We find that asymmetry and conservatism are beneficial for high (and confident) performers as they raise the probability of entry, which is profitable in expectation. By contrast, asymmetry and confidence are detrimental to low performers, who are better off not participating.*

7 Aggregate updating biases: a replication of MNNR

In this section, we will focus on *aggregate estimates* of asymmetry and conservatism, and reproduce the analysis in MNNR with our data. Since we are no longer looking at individual data but at average updates, we now follow MNNR in using the rational Bayesian update as a benchmark, rather than the sample average. Note that this allows us to speak about updating "biases", i.e. deviations from rational behavior on the aggregate level.

Figure 3 reproduces the main graphs in MNNR with our data.[13] Panel (a) shows the actual updates after both a positive and a negative signal as a function of prior belief, indicating the rational Bayesian update in dark bars as a benchmark. It is immediately clear that updating is

---

[13]The design of MNNR includes four updating rounds and they omit all observations from subjects who ever update in the wrong direction or hit the boundaries. To replicate these conditions faithfully, we exclude all observations for a given task for subjects who update at least once in the wrong direction or hit the boundaries of the probability space before the last belief elicitation. Moreover, we only use data from the first four updating rounds, omitting the noisy last two rounds. In our regressions, we also show results after we relax these sample selection criteria.

conservative, as subjects update too little in both directions. To investigate asymmetry, panel
(b) of Figure 3 puts the updates in the two directions next to each other. In contrast to the
results by MNNR, no clear pattern emerges. While updates after a negative update are slightly
smaller for some priors, no clear asymmetric pattern emerges. Thus, at first glance the data
do not replicate the finding of aggregate asymmetry. This corroborates null-findings in other
studies.[14]
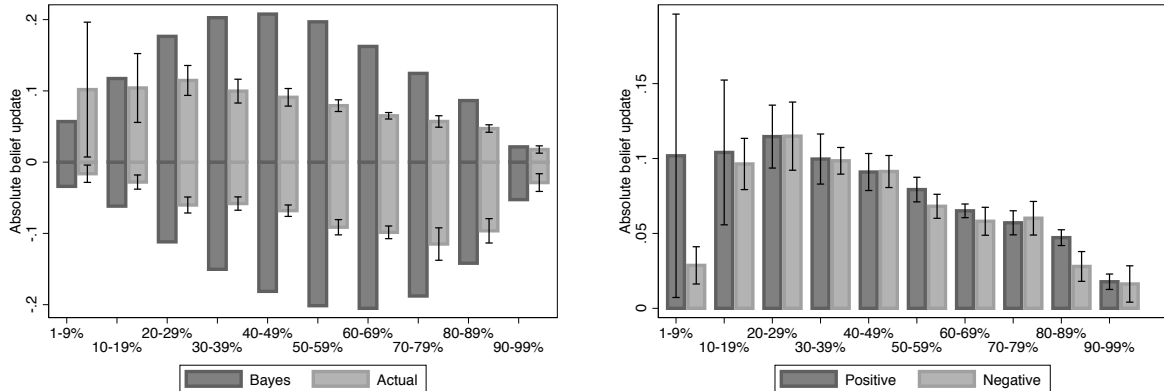
(a) Aggregate conservatism: Actual updates vs. Bayesian benchmark.

(b) Aggregate asymmetry: Average updates after positive and negative signals.

Figure 3: Overview of updating behavior, reproducing graphs in MNNR. The x-axis shows categories of prior beliefs, the y-axis shows the average size of the updates. 95% confidence intervals are included. Updating data come from round 1-4, updates in the wrong direction are excluded, replicating exactly the sample selection rules of MNNR.

To investigate these patterns further, and control more systematically for the influence of prior beliefs, following MNNR we estimate the regression model of a linearized version of Bayes' formula given by

$$logit(\mu_{int}) = \delta logit(\mu_{in,t-1}) + \beta_H 1_{(s_{int}=H)}\lambda_H + \beta_L 1_{(s_{int}=L)}\lambda_L + \epsilon_{int}. \tag{7}$$

Here, μ_{int} represents the posterior belief for person i in task $n \in \{1,2,3\}$ after signal in round $t \in \{1,2,3,4,5,6\}$, and $\mu_{in,t-1}$ represents the prior belief (i.e. the posterior belief in the previous round). Thus, our belief data have a panel structure, with variation both across individuals and over rounds/signals. λ_H is the natural log of the likelihood ratio of the signal, which in our case is $0.7/0.3 = 2.33$, and $\lambda_H = -\lambda_L$. $1_{(s_{int}=H)}$ and $1_{(s_{int}=L)}$ are indicator variables for a high and low signal respectively. The standard errors in all our regressions are clustered by individual.

From the logistic transformation of Bayes rule one can derive that $\delta, \beta_H, \beta_L = 1$ correspond

---

[14]Ertac (2011), Coutts (2016) and Schwardmann and Van der Weele (2016) even find a tendency in the opposite direction. Note that Ertac (2011) uses a different signal and event space, making it harder to compare her results to ours or MNNR's.

to perfect Bayesian updating (see MNNR for more details). Conservatism occurs if both $\beta_H < 1$ and $\beta_L < 1$, i.e. subjects place too little weight on either signal. If $\beta_H \neq \beta_L$, this implies "asymmetry", i.e. subjects place different weight on good signals compared to bad signals, where MNNR find that $\beta_H > \beta_L$ on an IQ quiz.

Table 7 presents the results. The first three columns of Table 7 show the results of all tasks pooled together, excluding all observations for a given task if the subject hits the boundaries of 0 or 1. In Column (2) we also exclude observations of subjects who update in the wrong direction one or more times, and in Column (3) we restrict the data to the first four updating rounds in each tasks, to avoid the noisy last two rounds.[15]

In each of the three columns, we see clear evidence for conservatism: both the coefficients on the positive and negative signal are very far from unity, the coefficient consistent with Bayesian updating. This implies that most subjects in our sample are indeed conservatively biased in their updating. The evidence for asymmetry is more mixed. The Wald test that both signals have the same coefficient, reported in the rows just below the coefficients, provides strong evidence for asymmetry in Column (1) only. In Columns (2) and (3) where we exclude all observations for a given task of subjects that make one or more mistakes, asymmetry is much weaker and not statistically significant. Thus, it seems that asymmetry is driven mostly by the fact that people are more likely to update in the wrong direction after a negative signal, which confirms the patterns in Table 1. This is weaker evidence for asymmetry than found in MNNR, who find a strong effect even when individuals making updating "mistakes" are excluded.

In Table 7, we also present regressions where we check whether we can replicate our findings that conservatism depends on ego relevance and that women are more conservative than men using the MNNR methodology. In Columns (4) and (5), we estimate the response to positive and negative signals separately for observations from people with above or below-median task relevance. The post-estimation Wald tests reported below the coefficients reveal no significant difference in aggregate conservatism or asymmetry when we compare the results by relevance. One reason for this could be that the (necessary) exclusion of observations from individuals who hit the boundaries likely excludes the least conservative (and most asymmetric) individuals. In Columns (6) and (7), we estimate the response to positive and negative signals separately by gender. We find the same significant gender difference in conservatism which we discovered using our individual measures.

Finally, Bayesian updating should only depend on the initial belief and the last signal, not on past signals, a property that MNNR call "sufficiency". If sufficiency is violated, then any measure of individual feedback responsiveness (including ours) will necessarily depend on the random order of signals. MNNR test for this by including lagged signals in their regression, and find that these are not significant, thus confirming sufficiency. In Appendix A, Table A.2, we

---

[15]Because we exclude all data for an individual in a given task when there is a single update in the wrong direction, and there was a strong increase in wrong updates in the last two rounds, excluding those rounds actually leads to an increase in the number of observations.

	(1)	(2)	(3)		(4)	(5)		(6)	(7)
Logit prior	0.860***	0.951***	0.948***	Logit prior	0.859***	0.948***	Logit prior	0.859***	0.948***
	(0.017)	(0.010)	(0.013)		(0.017)	(0.013)		(0.017)	(0.013)
Signal High	0.358***	0.404***	0.476***	Signal High			Signal High		
	(0.018)	(0.022)	(0.020)	Relevance low	0.357***	0.480***	Men	0.429***	0.545***
					(0.021)	(0.024)		(0.030)	(0.030)
				Relevance high	0.359***	0.471***	Women	0.302***	0.414***
					(0.026)	(0.026)		(0.020)	(0.024)
Signal Low	0.254***	0.398***	0.464***	Signal Low			Signal Low		
	(0.017)	(0.020)	(0.019)	Relevance low	0.275***	0.472***	Men	0.296***	0.514***
					(0.022)	(0.024)		(0.026)	(0.030)
				Relevance high	0.223***	0.451***	Women	0.222***	0.416***
					(0.024)	(0.027)		(0.021)	(0.023)
P (Prior=1)	0.000	0.000	0.000						
P (Asymmetry)	0.000	0.759	0.583	P (Asymmetry) (high rel)	0.000	0.533	P (Asymmetry) (Men)	0.001	0.426
				P (Asymmetry) (low rel)	0.002	0.764	P (Asymmetry) (Women)	0.001	0.923
				P(Conservatism, high vs low rel)	0.310	0.563	P(Conservatism men vs women)	0.000	0.000
				P(Asymmetry high vs low rel)	0.189	0.768	P(Asymmetry men vs women)	0.235	0.452
No boundary priors in task	✓		✓		✓	✓		✓	✓
No wrong updates in task		✓	✓		✓	✓		✓	✓
Only rounds 1-4			✓		✓	✓		✓	✓
Observations	4507	2197	2375		4507	2375		4507	2375
Subjects	288	218	272		288	272		288	272

Table 7: Regression results for model (7). All tasks are pooled. Columns (1) - (3) show robustness to sample selection criteria. Columns (4)-(5) show results of sample split by median task relevance. Columns (6)-(7) show results of sample split by gender. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

22

reproduce this exercise. We find that coefficients on past signals are positive and statistically significant. However, their impact on posteriors is smaller, by about a factor 8, than that of the last signal received. Thus, the sequence of signals has at most a small impact on the individual measurements in our data.

Summary 7 *Using the methodology from MNNR, we find strong evidence for aggregate conservatism, but not for asymmetry. However, updating "mistakes" do exhibit an asymmetric pattern.*

8 Discussion and Conclusion

Our individual measures of feedback responsiveness deliver a number of new insights. We find that differences in relative conservatism are correlated across tasks that measure different cognitive skills, indicating that they can be considered a characteristic or trait of the individual. The same thing cannot be said about asymmetry, which is not systematically correlated across tasks. We also find that individuals are more conservative, but not more asymmetric, in tasks that they see as more ego-relevant. Both measures have substantial explanatory power for post-feedback confidence and competition entry. Conservatism affects entry both through beliefs and independently, whereas relative asymmetry increases entry by biasing beliefs upward. Finally, we find that women are significantly more conservative than men.

Our study demonstrates both the strengths and limitations of our measurements of asymmetry and conservatism. Measuring updating biases is complex. There is some noise in our measures, and their elicitation is relatively time consuming. Future research could investigate if simpler or alternative measures could deliver similar or better predictive power. For example, several other incentivized belief elicitation methods exist that also control for risk aversion and have not, to our knowledge, been used in updating tasks (Offerman et al., 2009; Hossain and Okui, 2013; Schlag and Van der Weele, 2013).

Nevertheless, our results hold promise for researchers in organizational psychology and managerial economics, where feedback plays a central role. Specifically, an interesting research area would be to investigate the predictive power of these measures in the field. It would be interesting to correlate relative conservatism and asymmetry with decisions such as study choice, the decision to start a business and its probability of success, as well as a range of risky behaviors in which confidence plays a central role. In doing so, it could follow research that has linked laboratory or survey measurements of personal traits to behavior outside the lab. For instance Ashraf et al. (2006), Meier and Sprenger (2010), Almlund et al. (2011), Moffitt et al. (2011), Castillo et al. (2011), Sutter et al. (2013) and Golsteyn et al. (2014) link self-control, patience, conscientiousness and risk attitudes to outcomes in various domains such as savings, education, occupational and financial success, criminal activity and health outcomes. If such a research program is successful, it could reduce the costs of overconfidence and underconfidence to the individual and to society as a whole.

References

Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman, and Tim D. Kautz, "Personality psychology and economics," in E. Hanushek, S. Machin, and L. Woessman, eds., *Handbook of the Economics of Education*, Amsterdam: Elsevier, 2011, pp. 1–181.

Ashraf, Nava, Dean Karlan, and Wesley Yin, "Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines," *Quarterly Journal of Economics*, 2006, *121* (2), 635–672.

Barber, Brad M. and Terrance Odean, "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics*, 2001, *116* (1), 261–292.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Armando Jose Garcia Pires, and Bertil Tungodden, "Competitive in the lab, successful in the field?," *Journal of Economic Behavior and Organization*, 2015, *118*, 303–317.

Buser, Thomas, Lydia Geijtenbeek, and Erik Plug, "Do Gays Shy Away from Competition? Do Lesbians Compete Too Much?," *IZA Discussion Paper 9382*, 2015.

_ , Muriel Niederle, and Hessel Oosterbeek, "Gender, Competitiveness and Career Choices," *Quarterly Journal of Economics*, 2014, *129* (3), 1409–1447.

Camerer, Colin and Dan Lovallo, "Overconfidence and excess entry: An experimental approach," *The American Economic Review*, 1999, *89* (1), 306–318.

Castillo, Marco, Ragan Petrie, and Clarence Wardell, "Fundraising through online social networks: A field experiment on peer-to-peer solicitation," *Journal of Public Economics*, 2011, *114*, 29–35.

Coutts, Alexander, "Good News and Bad News are Still News: Experimental Evidence on Belief Updating," *Mimeo, Nova School of Business and Economics*, 2016.

Croson, Rachel and Uri Gneezy, "Gender differences in preferences," *Journal of Economic Literature*, may 2009, *47* (2), 448–474.

Eil, David and Justin M. Rao, "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 2011, *3* (2), 114–138.

Ertac, Seda, "Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback," *Journal of Economic Behavior and Organization*, 2011, *80* (3), 532–545.

Fischbacher, Urs, "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 2007, *10* (2), 171–178.

Fischhoff, Baruch and Ruth Beyth-Marom, "Hypothesis evaluation from a Bayesian perspective.," *Psychological Review*, 1983, *90* (3), 239–260.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List, "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions," *The Review of Economic Studies*, 2015, *82* (1), 122–155.

Golsteyn, Bart H., Hans Grönqvist, and Lena Lindahl, "Adolescent time preferences predict lifetime outcomes," *The Economic Journal*, 2014, *124* (580), F739–F761.

Grossman, Zachary and David Owens, "An unlucky feeling: Overconfidence and noisy feedback," *Journal of Economic Behavior and Organization*, 2012, *84* (2), 510–524.

Hossain, Tanjim and Ryo Okui, "The Binarized Scoring Rule," *The Review of Economic Studies*, 2013, *80* (3), 984–1001.

Koellinger, Philipp, Maria Minniti, and Christian Schade, ""I think I can, I think I can": Overconfidence and Entrepreneurial Behavior," *Journal of Economic Psychology*, 2007, *28* (4), 502–527.

Konrath, Sara, Brian P. Meier, and Brad J. Bushman, "Development and Validation of the Single Item Narcissism Scale (SINS)," *PLoS ONE*, 2014, *9* (8), 1–15.

Kuhnen, Camelia M., "Asymmetric Learning from Financial Information," *Journal of Finance*, 2015, *70* (5), 2029–2062.

Kunda, Ziva, "The case for motivated reasoning," *Psychological Bulletin*, 1990, *108* (3), 480–498.

Malmendier, Ulrike and Geoffrey Tate, "Who makes acquisitions? CEO overconfidence and the market's reaction," *Journal of Financial Economics*, 2008, *89* (1), 20–43.

Meier, Stephan and Charles Sprenger, "Present-biased preferences and credit card borrowing," *American Economic Journal: Applied Economics*, 2010, *2* (1), 193–210.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat, "Gender Differences in Incorporating Performance Feedback," *Mimeo, Harvard University*, 2007.

_ , _ , _ , **and** _ , "Managing Self-Confidence," *Mimeo, Stanford University*, 2014.

Moffitt, Terrie E., Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W. Roberts, and Stephen Ross, "A gradient of childhood self-control predicts health, wealth, and public safety," *Proceedings of the National Academy of Sciences*, 2011, *108* (7), 2693–2698.

Moore, Don A. and Paul J. Healy, "The trouble with overconfidence," *Psychological Review*, 2008, *115* (2), 502–17.

Niederle, Muriel and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

_ **and** _ , "Gender and Competition," *Annual Review of Economics*, 2011, *3* (1), 601–630.

Offerman, Theo, Joep Sonnemans, Gijs Van de Kuilen, and Peter Wakker, "A Truth Serum for Non-Bayesians," *Review of Economic Studies*, 2009, *76* (4), 1461–1489.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar, "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender," *Economic Journal*, 2016, *In Press*.

_ , **Paola Sapienza, and Luigi Zingales**, "Taste for competition and the gender gap among young business professionals," *NBER Working Papers Series*, 2015, *21695*.

Schlag, Karl H. and Joël J. Van der Weele, "Eliciting Probabilities, Means, Medians, Variances and Covariances without Assuming Risk Neutrality," *Theoretical Economics Letters*, 2013, *3* (1), 38–42.

_ , **James Tremewan, and Joël J. Van der Weele**, "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs," *Experimental Economics*, 2015, *18* (3), 457–490.

Schwardmann, Peter and Joël J. Van der Weele, "Deception and Self-deception," *Tinbergen Institute Discussion paper*, 2016, *012/2016*.

Sharot, Tali, Christoph W. Korn, and Raymond J. Dolan, "How unrealistic optimism is maintained in the face of reality," *Nature Neuroscience*, 2011, *14* (11), 1475–1479.

Slovic, Paul and Sarah Lichtenstein, "Comparison of Bayesian and regression approaches to the study of information processing in judgment," *Organizational Behavior and Human Performance*, 1971, *6* (6), 649–744.

Sutter, Matthias, Martin G Kocher, Daniela Rützler, and Stefan Trautmann, "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior," *American Economic Review*, 2013, *103* (1), 510–531.

Zhang, Y. Jane, "Can Experimental Economics Explain Competitive Behavior Outside the Lab?," *Working paper, Available at SSRN 2292929*, 2013.

Appendix A: Additional Tables

	C	A	C	A	C	A
Study effort	.052	.058	.028	.099	.017	.099
Importance of grades	.067	.036	.051	.062	.051	.057
Talk about studies	-.072	.15**	-.081	.157**	-.085	.153**
Consider academic career	-.056	.209***	-.034	.168***	-.021	.168***
Practice sports	-.055	.009	-.013	-.007	-.003	-.018
Sports at high level	-.026	-.055	.011	-.064	.007	-.079
Sports competition	.058	-.052	-.015	-.071	-.021	-.081
Sports duration	-.044	-.048	.005	-.074	.02	-.082
Play instrument	.085	.08	.076	.096	.089	.098
Music at high level	013	.053	.006	.071	.003	.059
Music in public	013	.053	.006	.071	.003	.059
Music duration	.08	.076	.069	.085	.084	.081
IQ rank	059	.302***	.075	.248***	.084	.205***
Grades rank	.082	.127**	.085	.099	.084	.077
Attractiveness rank	.008	.028	.03	.013	.03	-.011
Integrity rank	.01	.13**	-.017	.141**	-.018	.122**
Risk	.047	-.05	.056	-.056	.057	-.082
Competititiveness	018	.146**	.052	.099*	.045	.074
Narcicism	-.1	.146**	-.089	.121**	-.085	.112*
Gender and age			✓	✓	✓	✓
Scores			✓	✓	✓	✓
Initial beliefs					✓	✓
N	281	281	281	281	281	281

Table A.1: Coefficients of OLS regressions of individual asymmetry (A) and conservatism (C) on questionnaire score. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) Round 1	(2) Round 2	(3) Round 3
Logit prior	0.924***	0.947***	0.943***
	(0.026)	(0.021)	(0.027)
Signal High	0.497***	0.419***	0.404***
	(0.033)	(0.026)	(0.037)
Signal Low	0.456***	0.423***	0.384***
	(0.030)	(0.028)	(0.031)
Signal (t-1)	0.059***	0.076***	0.073***
	(0.023)	(0.020)	(0.026)
Signal (t-2)		0.068***	0.066***
		(0.018)	(0.024)
Signal (t-3)			-0.001
			(0.026)
No boundary priors in task	✓	✓	✓
No wrong updates in task	✓	✓	✓
Observations	600	600	575
Subjects	272	272	267

Table A.2: Results for regressions model (7), with the additional inclusion of lagged priors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.