Title

An Unlucky Feeling: Overconfidence and Noisy Feedback

Authors
Grossman, Zachary
Owens, David

Publication Date
2010-11-11

An Unlucky Feeling: Overconfidence and Noisy Feedback

Zachary Grossman* and David Owens‡

November 11, 2010

Abstract

How does overconfidence arise and how does it persist in the face of experience and feedback? In an experimental setting, we examine how individuals beliefs about their own performance on a quiz react to noisy, but unbiased feedback. In a control treatment, each participant expresses her beliefs about another participants performance, rather than her own. On average, they express accurate posteriors about others scores, but they overestimate their own score, believing themselves to have received unlucky feedback. However, this driven by overconfident priors, as opposed to biased information processing. We also find that, while feedback improves estimates about the performance on which it is based, this learning does not translate into improved estimates of related performances. This suggests that people use performance feedback to update their beliefs about their ability differently than they do to update their beliefs about their performance, which may contribute to the persistence of overconfidence.

*Department of Economics, 2127 North Hall, University of California, Santa Barbara, CA 93106

†Haverford College, Department of Economics, 370 Lancaster Ave, Haverford, PA 19041

‡We are grateful to Gary Charness and Muriel Niederle, as well as seminar participants at Swarthmore College, University of Delaware, and the ESA North American meetings for helpful comments and to Daniel Saunders for excellent research assistance. This research was partially funded by an SSRGP grant from the Institute for Social, Behavioral, and Economic Research (ISBER) at the University of California, Santa Barbara.

1 Introduction

Life experience consistently provides information about our personal attributes and our performance in a variety of settings. A restaurateur's experience provides her with information not only about the likelihood that her restaurant will succeed, but also about her ability to successfully manage and operate a restaurant business. A Bayesian's beliefs about her own skill would grow progressively more accurate and she would use this information to inform future decisions about whether and how to undertake risky ventures.

However, psychologists and economists have found that people tend to be overconfident about a broad range of qualities, (Hoelzl and Rustichini 2005, Burks, Carpenter, Goette, and Rustichini 2010, Moore and Healy 2008) even to their own financial detriment (Malmendier and Tate 2008, Camerer and Lovallo 1999). This include attributes about which people receive a multitude of feedback over the course of their lives. For example, Svenson (1981) finds that 93% of U.S. students believe that they are better than average drivers. Most people drive on a regular basis, and each trip is an opportunity to learn about one's driving ability.

How does this overconfidence arise and and how does it persist in the face of experience and feedback? One cannot answer these questions without recognizing that in any task-feedback situation there are two possible sources of overconfidence in posterior beliefs: overconfident priors and biased information processing. However, the updated beliefs emerging from one situation become the priors for the next, which makes it difficult to distinguish between these two channels and to understand how overconfidence may persist over a series of tasks. Answering these questions can provide insights into entrepreneurial and investor choice, and, in general, into the behavior of individuals taking on risk.

We present an experiment that allows us to separate the effects of overconfident priors and biased information processing in generating overconfident posterior beliefs about performance, while shedding light on how overconfidence may persist over time. We investigate how individuals' beliefs about their absolute (as opposed to relative) performance on a quiz reacts to noisy, but unbiased feedback and how experience and feedback from the quiz affect subsequent prior and posterior beliefs about performance in a second, very similar quiz.

Section 2 presents the experimental design, which has four main features that allow us to distinguishes itself from other experiments investigating overconfidence and the reaction to feedback and to gain significant insight into the origins of overconfidence, its persistence, and how people's response to information regarding their performance depends upon the way performance is measured. Participants take a ten-question quiz on logic and reasoning. In two separate conditions, participants estimate a different target score: r their own score in the *Self* condition and the score of another participant in the *Other* condition. The *Other* condition serves as a control because, while the quantity about which participants must form and state beliefs is not ego-relevant, it presents an otherwise identical task.

Second, participants estimate absolute, as opposed to relative performance. This allows us to determine the extent to which the "good news-bad news effect" found by Eil and Rao (2010) and Moebius, Niederle, Niehaus, and Rosenblat (2007) depends upon social comparisons and biases in processing information about others, as opposed to a bias merely in processing information about the self. Third, the nature of the feedback we give—a perturbation of the raw score, as opposed to a garbled binary signal—allows us to examine quantitatively how individuals attribute performance to external circumstances or luck. Finally, repeating the quiz/belief-elicitation protocol allows us to examine how information and experience carry over from one task or situation to the next.

We present the results in Section 3. We find that participants overestimate their own score—but not that of others—both before and after receiving the unbiased noisy feedback. While on average they perceive the feedback about their own score to be to be influenced by bad luck, given the overconfidence in their priors, this is largely consistent with Bayes' rule. Furthermore, we find no evidence that participants' beliefs respond conservatively to feedback about their performance. Rather, their beliefs actually react slightly *more* than warranted by Bayes' rule. Similarly, we do not find evidence they react differently to 'good news' versus 'bad news'. Thus, post-feedback overconfidence is primarily driven by overconfident priors. While experience and feedback from the first quiz improve the accuracy with which participants estimate others' scores on the second quiz, we find only minimal improvement in the accuracy of participants' estimates of their own scores.

We conclude with a discussion of these results in Section 4. Our results are consistent with recent research which has found differences in overconfidence and information processing for information that is and is not ego-relevant.[1] Charness,

[1] Not all research has found exclusive overconfidence and overoptimism. Ertac (2009) finds evidence of pessimism in interpreting feedback that is incomplete (as opposed to noisy) and Clark and Friesen (2009) find zero mean error and *underconfidence* to be more prevalent in forecasts of both relative and absolute performance on a real effort task. They also find that people have less difficulty predicting relative performance than predicting

Rustichini, and van de Ven (2010) find that people make more errors in estimating their own performance on a mental ability task than on another in which feedback relates to a more abstract question. Ertac (2009) finds that whether or not information is ego-relevant affects the way it is process. Moebius, Niederle, Niehaus, and Rosenblat (2007) find that people (especially males) react *conservatively* to news about their relative performance on trivia questions, in the sense that on average their beliefs do not shift as much as would those of a perfect Bayesian with the same expressed priors.

Recent work has also found an asymmetric response to news—participants respond much less to negative information about themselves than positive information. Eil and Rao (2010) ask subjects to estimate their ranking in both an IQ task and in a measure of individual attractiveness both before and after receiving binary feedback and find similar results, labeling the asymmetric response the "good news-bad news" effect. We find no such effects in our results, which suggests that the asymmetric response to feedback found in other studies, which all elicit beliefs about *relative* performance, may depend upon mis-attribution in social comparisons and that people's response to feedback depends upon whether the performance measure is relative or absolute.

Previous studies clearly establish the role of biased processing of performance feedback on generating overestimation in posterior beliefs about performance and this phenomenon is likely at play in many general settings. However, the simple and uniform nature of the noise-generating in our experiment may help our participants conform to Bayes' rule to a greater degree than in the aforementioned studies. In fact, our participants slightly *over-reacted* to the unbiased feedback, which—given their overconfident priors—might be expected to sharply curb posterior overconfidence. Though, like Clark and Friesen (2009), we find that updating is not sufficient to eliminate eliminate forecast errors, participants *do* seem to learn well about their quiz performance from the feedback as the degree of overestimation sharply drops after the feedback.

However, overconfidence remains consistent *across* quizzes. Overestimation in prior beliefs barely drops from the first quiz to the second, among those estimating their own score, despite the fact that participants in the second quiz have the experience and feedback from the first quiz to draw upon. Thus, they appear to fail to transfer learning from one quiz into the next. This suggests another channel through which overconfidence may persist. Performance feedback does improve

absolute performance.

the accuracy of these participants' beliefs about the relevant performance, but not necessarily about the underlying abilities that generate similar performances. Thus, not only do people process information about themselves differently than they do information about others, they use performance feedback to update their beliefs about their ability differently than they do to update their beliefs about their performance.

2 Experimental Design

The experiments took place at the Experimental and Behavioral Economics Laboratory (EBEL) at the University of California, Santa Barbara. Participants were randomly recruited from the EBEL subject pool (largely comprised of UCSB students and staff) using the online system ORSEE (Greiner 2003). Full instructions for all conditions, as well as screenshots are presented in the Appendix.[2]

We conducted twelve sessions, six for each condition. We conducted all sessions in pairs featuring the same two quizzes, alternating the order of the *Self* and *Other* condition in each pair. We also reversed the order of the quizzes in consecutive session pairs.[3] Each session lasted roughly 75 minutes and, with one exception, included 11 participants.[4] Participants earned \$12.65 on average, which includes a \$5 show-up fee and which was paid privately in cash at the end of the session. Each participant took part in only one session.

Upon arriving at the experiment, participants sat at computer terminals, were given a paper copy of the instructions, and followed along as the experimenter read them aloud. Participants then completed a protocol in which they took a quiz and were asked three times to state their beliefs about performance on the quiz. In two separate conditions participants were asked to estimate different target scores: participants in the *Self* condition were asked about their *own* score, while participants in the *Other* condition we asked about the score of an anonymous randomly selected other participant (RSOP). We elicited beliefs three times: 1) *Ex ante* or prior beliefs (b_1), before participants took the quiz; 2) *interim* beliefs (b_2), immediately after taking the quiz; and 3) *Ex post* or posterior beliefs (b_3), after receiving feedback about the performance being estimated.

[2] The software are available from the authors upon request.

[3] One block of four sessions featured an error that led some participants to be given an incorrect score for quiz S. As a result we omit from the data analysis all 39 observations for which the target score was miscalculated and 20 observations for a quiz following on in which the participant's target score was miscalculated.

[4] Session 9, which was in the *Self* condition, had only 9 participants.

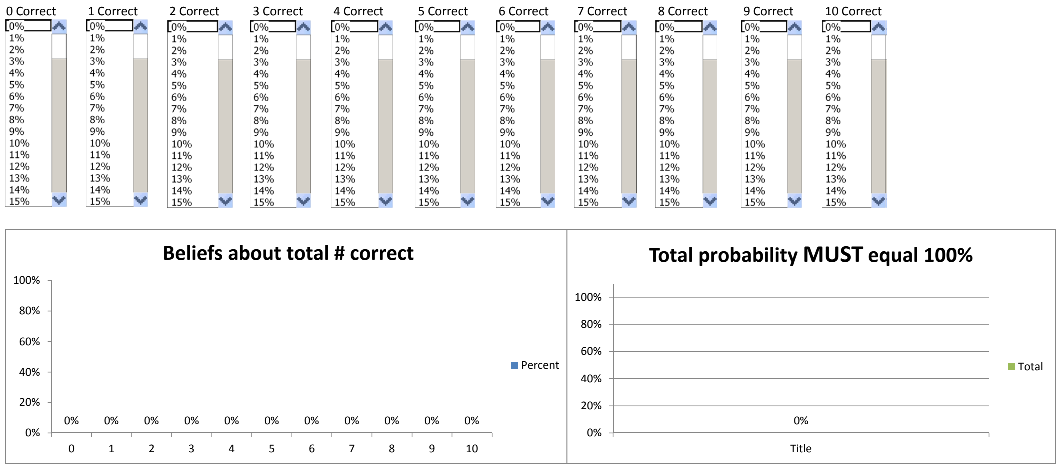
After completing this process, participants repeated it with a second quiz, which featured new questions that were drawn from the same source as the first quiz. When expressing their beliefs, participants did not view beliefs that they had expressed previously. Further, they did not view payoffs for or scores on the *first* quiz until they had completed the second. The key distinction between the first and second round of quizzes is that participants in the second round were *experienced*, having already experienced taking the first quiz and receiving noisy feedback about their first-round target score.

Quizzes The quizzes consisted of ten multiple-choice questions selected from a book of Mensa quizzes (Grosswirth, Salny, and Stillson 1999). They were presented in a Microsoft Excel spreadsheet which allowed participants to select one answer from a menu of possible answers. We constructed a total of four quizzes, two of which were used in each session.[5] To accommodate the possibility of quiz and/or order effects, we reversed the order in which the two quizzes were taken in consecutive *Self/Other* session-pairs. To incentivize effort, *for each quiz* we selected one question at random at the end of the session. Participants earned an extra $5 if their answer to this question were correct.

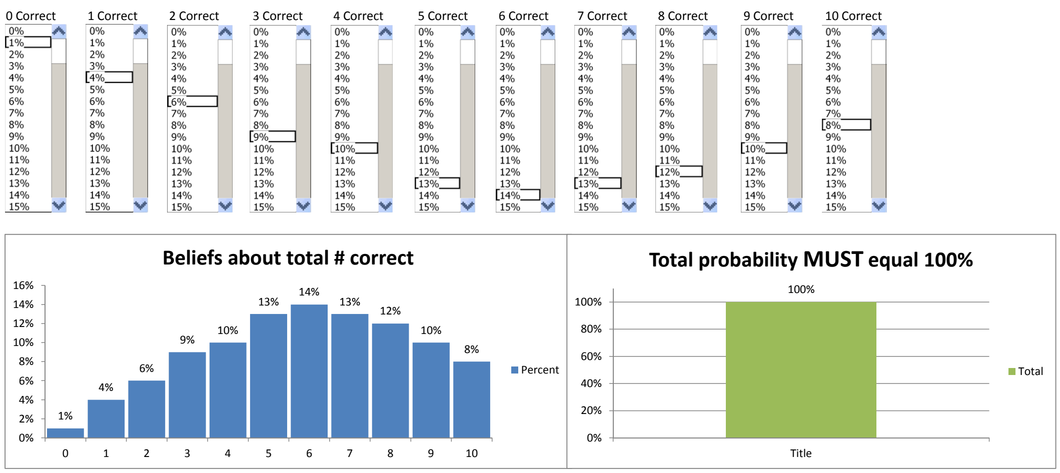
Belief Elicitation Rather than simply asking subjects to guess how many questions they answered correctly, we elicited the entire distribution of their beliefs, using an interface and incentive scheme similar to those used by Moore and Healy (2008) and Eil and Rao (2010). Figure 1 presents a screenshot of our computer interface, which was programmed using Microsoft Excel. For each raw-score outcome from zero to ten correct answers, participants select the likelihood (in percentage terms) with which they believe that outcome to have occurred. The percentages initially were set to zero for each outcome and participants made adjustments by selecting numbers from the drop-down menu corresponding to each outcome.

Two charts on the same screen were provided as a visual aids to the participants. An auto-adjusting histogram provided a visual profile of the current selection and a single bar showed the running total of percentage-points selected. The experimenter verified that the total was 100% before the participant was allowed to save the estimate and close the program. We used the same interface

[5] All quizzes, plus correct answers are provided in the Appendix.



(a) The initial appearance



(b) With entered beliefs

Figure 1: The beliefs-elicitation interface

for each elicitation, but participants did not have access to their history of belief estimates in the form of previous files.[6]

We paid participants for their probability estimates using an incentive-compatible quadratic scoring rule, for each quiz selecting one of the three estimates for payment. Specifically, for each paid elicitation, a participant was paid $1 + 5p_x - w$, where p_x is the participant's estimated likelihood for the true outcome and w is the sum of squares of the likelihoods for each of the eleven possible outcomes (zero correct through ten correct). Following Moore and Healy (2008) and Eil

[6] Participants were allowed to write on scrap paper and may have had access to their previous estimates had they recorded the estimates on the paper.

and Rao (2010), after learning the scoring-rule formula, participants were told, "This formula may appear complicated, but what it means for you is very simple: You get paid the most when you honestly report your best guesses about the likelihood of each of the different possible outcomes." Furthermore, the instructions walked participants through four detailed examples of how the payoff for the beliefs-elicitation was calculated.[7]

Feedback For each quiz, after all of the participants had entered their second score estimate, we handed each one a slip of paper (face down for privacy) with a number on it that constituted his or her feedback, f.[8] This number was generated by perturbing the target score, s^T,[9] with a noise or 'luck' term, ℓ, drawn uniformly from then set $\{-2, -1, 0, 1, 2\}$. Participants were instructed as such and the instructions emphasized the fact that the number on the paper was equally likely to be equal to $s^T - 2$, $s^T - 1$, s^T, $s^T + 1$, or $s^T + 2$.

After receiving this feedback participants entered their beliefs a third and final time. After completing this step for the first quiz, participants moved on to the second part of the session, in which they took a second quiz and entered their beliefs another three times. After completing this step for the second quiz, participants were given a statement summarizing their payment and their scores on the two quizzes, received their payment and exited.

3 Results

We begin the discussion of our results with a brief analysis of participants' performance on the quiz. Table 1 summarizes scores and feedback by condition, quiz order, and the exact quiz employed.[10] We present results at the individual quiz level. Thus, each participant is represented by two separate sets of observations.

Participants answered roughly half of the ten questions correctly, on average, and those in the *Other* treatment perform marginally better, though this difference does not approach statistical significance ($t = 1.20$, $p = 0.23$). Performance improves, in both conditions, from the first quiz to the second, and there appears to be some variance in the difficulty of the four quizzes.[11] By construction,

[7] We submit that few subjects understand the quadratic scoring rule computationally, but feel that they generally trust the experimenter when told that honestly reporting their beliefs is the best thing to do.

[8] In the instructions, we used neutral language and avoided using the word 'feedback'. Instead we called it 'imperfect information' about the score the participant was trying to estimate.

[9] The variable s is defined as a participant's own score. In the *Self* condition, $s^T = s$.

[10] See the appendix for a complete list of the questions used in each quiz.

[11] Table 8 in the appendix further breaks down scores and beliefs by quiz and by order.

Table 1: Average score (out of ten) and feedback

	Other			Self		
	s	f	N	s	f	N
Combined	5.33	5.37	99	4.91	4.93	90
	(0.25)	(0.30)	—	(0.25)	(0.28)	—
First Quiz	5.15	5.20	54	4.68	4.74	50
	(0.34)	(0.41)	—	(0.31)	(0.39)	—
Second Quiz	5.56	5.58	45	5.20	5.17	40
	(0.37)	(0.44)	—	(0.39)	(0.42)	—
Quiz A	6.90	7.10	20	5.70	5.70	20
	(0.58)	(0.62)	—	(0.59)	(0.53)	—
Quiz B	4.53	4.59	34	4.40	4.50	30
	(0.35)	(0.46)	—	(0.36)	(0.49)	—
Quiz C	4.85	4.65	20	4.65	4.65	20
	(0.61)	(0.72)	—	(0.49)	(0.64)	—
Quiz D	5.56	5.64	25	5.15	5.10	20
	(0.44)	(0.57)	—	(0.58)	(0.62)	—

Standard errors in parentheses

feedback in both conditions is unbiased—on average neither an upward nor a downward perturbation of scores—so it is not surprising that f is on average very similar to s.

The data analysis proceeds as follows: first, we examine participants' beliefs for evidence of overconfidence. Then, we explore the extent to which they find feedback on their own performance to be 'unlucky,' in spite of the fact that it is generated by a transparently unbiased process. Next, we analyze the process by which participants update their interim beliefs after receiving feedback. Finally, we evaluate the learning process between the first and second quiz in each condition.

3.1 Overestimation

Eliciting beliefs as a distribution allows us to compute Bayesian posteriors. However, we first examine participants' 'best guess' about their performance, which we define as $\bar{b}_j = \sum_{i=0}^{10} b_j^i \times i$. $\bar{b}_j$ is the mean of the belief distribution for elicitation j, where b_j^i is a participants' expressed belief in elicitation j that $s^T = i$. We focus on overconfidence manifested by $b_j - s^T$, the degree to which the mean

Table 2: Mean Overestimation

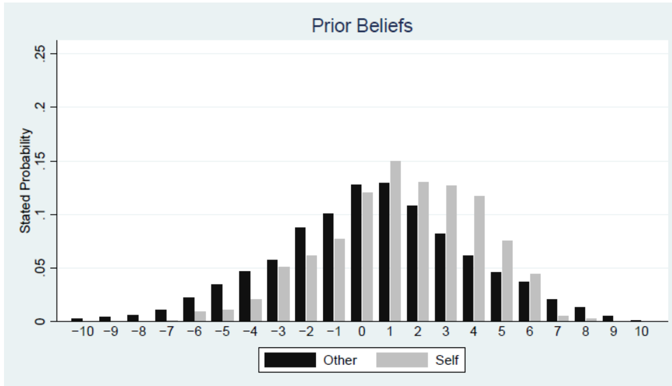
	$\bar{b}_1 - s^T$	$\bar{b}_2 - s^T$	$\bar{b}_3 - s^T$
Other	0.47	-0.23	-0.05
$N = 99$	(0.29)	(0.30)	(0.14)
Self	1.40	1.14	0.67
$N = 90$	(0.23)	(0.20)	(0.15)

Standard errors of *mean* overestimation in parentheses

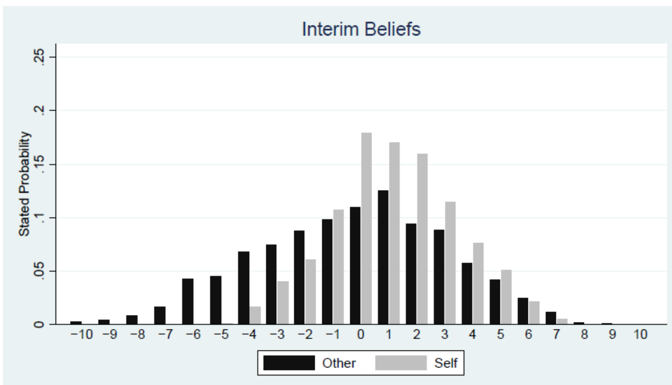
of beliefs exceeds the target score. Table 2 presents this measure by condition, separately for the prior, interim, and posterior belief elicitations.

In the *Self* condition, average overestimation is significantly greater than zero across all three elicitations ($t = 5.89$,5.62 , and 4.43, respectively, $p < .001$) and the magnitudes of overconfidence correspond to overestimation of 23%, 20% and 13% of the average score. In the *Other* condition, only prior beliefs exhibit noticeable overestimation, corresponding to 9% of quiz scores, and this overestimation is only marginally significant ($t = 1.58$, $p = .12$), with average overestimation in interim and posterior beliefs no significantly different from zero ($t = -0.77, -0.06$ and $p = .12, .72$, respectively) and, in fact, slightly negative. The difference in mean overestimation across the two conditions, as tested by a two-sample t-test, is highly significant ($p < .001$) for each of the three beliefs estimates. The decrease in overestimation from prior to interim beliefs in both conditions suggests that quizzes are more difficult than participants expect. Estimates also increase in accuracy from the interim to the posterior stage, implying that feedback improves estimation.

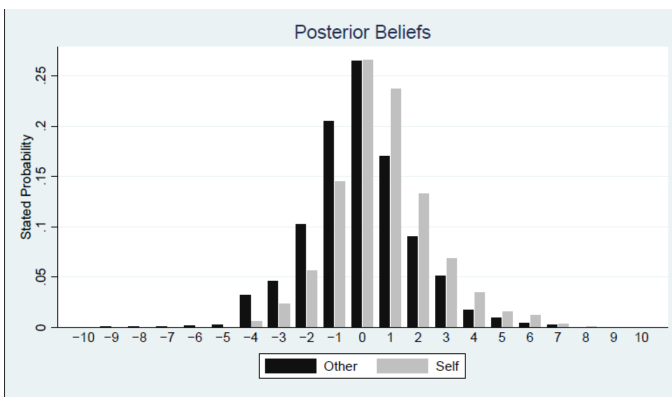
Because we elicit an entire distribution of beliefs, we may also characterize overestimation as a distribution, $b_j - s^T$, by subtracting the true target score, s^T, from each b_j. We average this distribution across all observations and display the resulting average distributions of overestimated beliefs in Figure 2, by condition and separately for each elicitation. The extent to which overestimation in the *Self* condition exceeds that in the *Other* condition is apparent in all three elicitations. On average, participants in the *Self* condition place greater weight on outcomes above the actual target score than do those in the *Other* condition, while the opposite is true for outcomes below the target score. While the participants in the *Other* condition on average do not appear to significantly overestimate the target score, both measures show that overestimation is pervasive in the *Self*



(a) Prior Overestimation: $b_1 - s^T$



(b) Interim Overestimation: $b_2 - s^T$



(c) Posterior Overestimation: $b_3 - s^T$

Figure 2: Overestimation

condition.

Table 3: Mean Estimated Luck

	Other	*Self*
$\bar{\ell}^e = f - \bar{b}_3$	0.11	-0.65
	(0.14)	(0.15)
N	99	90

Standard error of mean estimated luck in parentheses.

3.2 Estimated Luck

Next we investigate whether subjects consider noisy but unbiased feedback about their own performance to be 'unlucky.' As feedback is generated by the process $f = s^T + \ell$, the gap between feedback and posterior beliefs implicitly defines participants' estimates of ℓ. The distribution characterized by $f - b_3$ specifies the probability that subjects attribute to every possible outcome of ℓ.[12] We define this distribution $\ell^e \equiv f - b_3$ *estimated luck*. Table 3 summarizes the mean of estimated luck in each condition, which measures participants' estimate of the noise that contributed to their feedback.

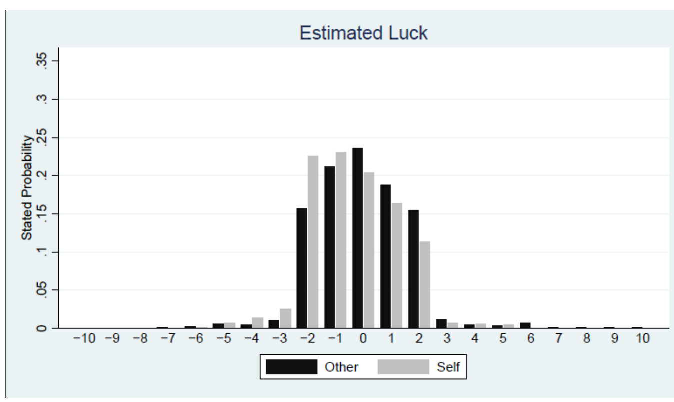


Figure 3: Estimated Luck

Table 3 shows stark differences across treatments in estimated luck. While participants in the *Other* condition express beliefs that s^T is perturbed upward by 0.11 questions, on average, those in the *Self* condition estimate that it is perturbed *downward* by 0.65 questions. A two-sample t-test deems this difference significant ($t = 2.57$, $p = .005$). Thus, participants appear to believe that feedback about their own score is unlucky, but that feedback about their colleague's score

[12]While the support of ℓ is confined to $\{-2, -1, -, 1, 2\}$, some subjects express b_3 such that ℓ^e falls outside of this range.

is relatively accurate.

Figure 3 presents the empirical distribution of $\ell = f - b_3$, aggregated across individuals. Note that for both conditions, the bulk of mass lies between -2 and 2, reflecting the fact that most participants comprehended that their actual target score had to be within 2 of the feedback they receive. The figure shows a noticeable difference in the distribution of estimated luck across the conditions. In the *Other* condition, it is distributed more-or-less symmetrically around a peak of 0, suggesting that participants believe their colleagues' feedback to be neither lucky nor unlucky. In the *Self* condition, the mass of the distribution lies in the negative range, meaning that a plurality of participants feel that their feedback under-represents their score, by close to a full point. Naturally, as the feedback is jointly determined by their performance and luck, noisy feedback allows them to overestimate their performance.

3.3 Information Processing

Above, we present evidence that participants find unbiased performance feedback to be 'unlucky', in the sense that it under-represents their actual performance. An alternative interpretation of this finding is that subjects attribute positive outcomes to their performance and negative outcomes to luck. Intuitively, this may seem to suggest that the process by which they incorporate feedback into their beliefs is biased.

We also find that participants significantly overestimate their performance across all three sets of beliefs. Importantly, overconfidence and this unlucky feeling are not independent of one another. Consider a participant with overconfident b_2, which is expressed immediately after taking the quiz, but before receiving feedback. Such a participant typically expresses b_2 with more weight on scores higher than s^T than on scores lower than s^T. If such a subject happens to receive accurate feedback ($f = s$), for example, it is rational for her to continue believe with that $s^T > f$ greater probability than $s^T < f$.[13] Such a participant would believe that f is 'unlucky,' even if she updates her beliefs in a manner perfectly consistent with Bayes' rule. In this section, we examine the extent to which the process by which participants incorporate feedback into their beliefs is consistent with Bayes' rule. Interim beliefs, b_2, are expressed before participants view f, while b_3 is expressed afterward. We calculate b_3^*, the Bayesian posterior beliefs

[13] This is especially true in our design, where f is uniformly distributed on $[s^T - 2, s^T + 2]$, but is true of any process where feedback is unbiased.

Table 4: Summary statistics for participants with 'Un-Updateable' interim beliefs

	s	s^T	f	$\bar{b}_1$	$\bar{b}_2$	$\bar{b}_3$
Other	4.50	6.94	7.33	5.69	4.20	6.78
$N = 18$	(0.52)	(0.76)	(1.00)	(0.42)	(0.43)	(0.62)
Self	3.25	3.25	2.42	6.19	6.43	4.75
$N = 12$	(0.70)	(0.70)	(0.78)	(0.54)	(0.53)	(0.64)

Notes: Standard error of means in parentheses. Participants summarized in this table expressed beliefs that place no probability on scores within 2 points of f.

following f given the expressed b_2.

'Un-Updateable' Interim Beliefs Before we compare participants' belief-updating process to their Bayesian benchmark, we must address a weakness of Bayes' rule's application to our design. A number of our subjects express interim beliefs that cannot be updated according to Bayes' rule given the feedback that they receive. Feedback f is generated by a uniform process on $[s^T - 2, s^T + 2]$. Thus, f contains conclusive evidence that $s \in [f - 2, f + 2]$. If participants express b_2 with zero probability on $[f - 2, f + 2]$, then Bayes' rule fails to provide a benchmark for b_3. Such beliefs were expressed by 18 participants in the *Other* condition, and 12 participants in the *Self* condition. Table 4 summarizes the performance and beliefs of these participants.

Table 4 contains several insights about the participants whose feedback is inconsistent with interim beliefs. First, the score of these participants is lower than the average in each condition ($p < .05$), suggesting that this subset of participants may be less intelligent on average than those who *can* update according to Bayes' rule.[14] Second, $\bar{b}_2 < s^T < f$ in the *Other* condition, with the opposite true in the *Self* condition. This is natural, as this table represents only those participants for whom f came as a surprise. Underconfident priors and a positive random shock (ℓ) both contribute to surprise in the Other condition, with the converse combining to surprise subjects in the *Self* condition. Those represented in the table are strongly predisposed to underestimation in the *Other* condition and overestimation in the *Self* condition. Of the 18 participants from the *Other* condition represented in table 4, 14 underestimate their colleague's score, according to their

[14] This could also be explained by the fact that people who perform better have less room to err above the range of outcomes consistent with the feedback received.

interim beliefs ($\bar{b}_2 < s^T$). Of the 12 from the *Self* condition, 11 overestimate their own score ($\bar{b}_2 > s^T$).

There is a marked difference in participants' responses to feedback between conditions. In the *Other* condition, $\bar{b}_3$ settles near f (and s), while subjects feel that f *wildly* under-represents s^T in the Self condition. The *average* of $\bar{\ell}^e$ for these subjects is -2.33 in the *Self* condition, which is lower than ℓ's *minimum* of -2.

Participants represented in table 4 make up a small but non-negligible (16%) percentage of the sample. The next portion of our analysis involves comparing the belief-updating process to its Bayesian benchmark for those participants for which this undertaking is possible (those *not* represented in table 4). The results of this table suggest that the participants being omitted from the next portion analysis are markedly different from those that are included. The analysis that follows should be read with the knowledge that the 16% of participants omitted are less intelligent, less responsive to information, and potentially more biased than their counterparts represented in the analysis.

Belief-updating and its Bayesian benchmark As stated above, a Bayesian updater with overconfident priors will typically consider unbiased feedback to be unlucky. As participants in our design are overconfident, believing that their feedback is unlucky does not necessarily mean that they do not update their beliefs correctly. We now examine the extent to which the 'unlucky feeling' described above is consistent with Bayes' rule.

The first column of table 5 displays $f - \bar{b}_3$, the same measure of estimated luck as does table 3. The second columns summarizes the same only for the updateable subjects, while the third column displays the Bayesian benchmark for the latter. In the *Self* treatment, mean estimated luck is significantly higher for the updateable subjects than for the un-updateable ($-0.39 > -2.33$, $p < .001$ according to a two-sample t-test), and very similar to Bayesian mean estimated luck (-0.39 vs. -0.43, $p = .71$ according to a paired two-sample t-test). In other words, while they feel unlucky, they update their beliefs correctly on average. Updating differs more with respect to the Bayesian benchmark in the *Other* condition, where participants should find their colleagues less lucky than they do.

Previous experimental research on overconfidence (Eil and Rao (2010) and Moebius, Niederle, Niehaus, and Rosenblat (2007), for example) has uncovered an ego-preserving bias in the way that participants incorporate noisy feedback into their beliefs about relative performance. Specifically, they find beliefs to be

Table 5: Average estimated luck versus Bayesian estimated luck

	$(f - \bar{b}_3)$ *(all participants)*	$(f - \bar{b}_3)$: *(updateable b_2)*	Bayesian $(f - \bar{b}_3^*)$ *(updateable b_2)*
Other	0.11	0.01	-0.29
	(0.14)	(0.12)	(0.14)
N	99	81	81
Self	-0.65	-0.39	-0.43
	(0.15)	(0.13)	(0.14)
N	90	78	78

Notes: The first column displays $\bar{\ell}^e$, the mean of (observed) estimated luck, in each condition. The second column displays the mean of estimated luck only for those subjects whose interim beliefs (b_2) could be updated according to Bayes' rule. The third column displays the mean $\bar{\ell}^e$ that would be expressed for subjects if they updated their expressed b_2 according to Bayes' rule.

more responsive to "Good News": feedback that improves one's self-image, than they are to "Bad News": feedback that harms it. Within our design, this would suggest that $\bar{b}_3$ is more responsive to feedback when f is Good News ($f > \bar{b}_2$) than when f is Bad News ($f < \bar{b}_2$).[15]

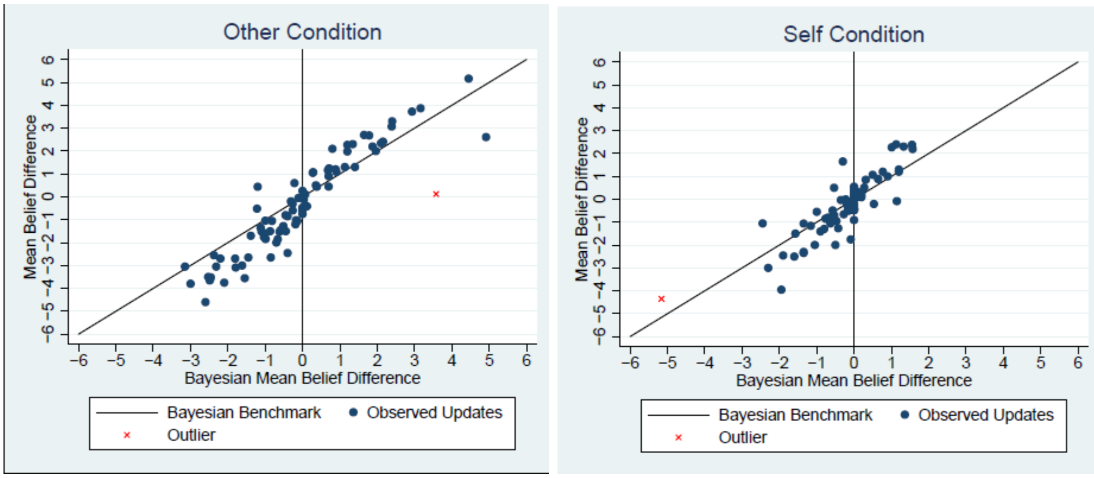
To test for such an asymmetry in the updating process, we calculate a simple measure of the responsiveness of particpants' beliefs to the information in f. A straightforward measure of beliefs' responsiveness to f is the change in mean beliefs before and after participants view f, or $(\bar{b}_3 - \bar{b}_2)$. As discussed above, Bayes' rule provides a benchmark for how much participants *should* update their beliefs: $(\bar{b}_3^* - \bar{b}_2)$. Asymmetric responsiveness would suggest that $(\bar{b}_3 - \bar{b}_2)$ is higher, relative to $(\bar{b}_3^* - \bar{b}_2)$, when $(\bar{b}_3^* - \bar{b}_2)$ is positive than when it is negative.

Figure 4 plots $(\bar{b}_3 - \bar{b}_2)$ vs. $(\bar{b}_3^* - \bar{b}_2)$. One observation is labeled an 'outlier' (marked with an 'X') in each condition, for reasons addressed below. The diagonal line $(\bar{b}_3 - \bar{b}_2) = (\bar{b}_3^* - \bar{b}_2)$ is included to facilitate a comparison of particpants' mean difference in beliefs to its Bayesian benchmark. Observations located above the diagonal represent more favorable updating than the Bayesian benchmark, and those below represent less favorable updating.

Figure 4 presents no obvious evidence of conservatism or asymmetry in the updating process, in either condition. Conservatism, or unresponsive beliefs, would cause the observations to form a flatter line than the Bayesian benchmark. Asymmetry would suggest differences in responsiveness on either side of $(\bar{b}_3^* - \bar{b}_2) = 0$.

[15] Participants for whom $f = \bar{b}_2$ are categorized neither as Bad nor Good news, and are therefore omitted from Good News-Bad News analysis.

While neither is evident in the figure, responsiveness is explored more quantitatively below.



(a) $(\bar{b}_3 - \bar{b}_2)$ vs. $(\bar{b}_3^* - \bar{b}_2)$: Other condition

(b) $(\bar{b}_3 - \bar{b}_2)$ vs. $(\bar{b}_3^* - \bar{b}_2)$: Self condition

Figure 4: Updating Mean Beliefs

Table 6 presents a series of regressions that facilitate a quantitative exploration of conservatism and asymmetry. Regression (1) presents the results of the most basic regression, where beliefs are assumed to depend only on their Bayesian benchmark. In both conditions, the coefficient on the Bayesian difference in mean beliefs is slightly greater than one, though it is only significantly greater in the *Other* condition. Moreover, regression (2) shows coefficients slightly higher, both significantly greater than one when a single outlier is dropped from each condition. With coefficients no less than one, we do not find evidence of *conservatism* in response to information. If anything, participants are on average responding slightly more strongly to information than called for by Bayes' rule. Thus, deviations from Bayesian updating do not exacerbate posterior overestimation (in either condition). Instead, the positive posterior overestimation in the *Self* condition is primarily driven by overconfident priors, for which either a Bayesian response to feedback would result in positive, albeit diminished, overconfidence or which exhibit overconfidence so strong as to render Bayes' rule inapplicable.

In order to isolate any asymmetry in the response to information with a positive versus negative valence, we include the interaction term $(\bar{b}_3^* - \bar{b}_2) \times GN$ in regressions (3), (4), and (5), where GN takes the value of one if $f > \bar{b}_2$ and zero otherwise. Intuitively, this term's coefficient will tell us whether participants respond differently to f when it represents Good News. If participants in our design exhibit an image-preserving bias in their incorporation of feedback into their beliefs, we would expect the coefficient on this interaction term to be positive and significant in the *Self* condition, but not in the *Other* condition, in which feedback is not ego-relevant.

The regression (3) interaction coefficient in the *Self* condition is positive (0.28) but not significant, which, at most, only suggests the possibility of a good news-bad news effect. However, in the *Other* condition, the coefficient (-0.39) is significantly negative, implying that subjects' beliefs are *more* responsive to negative news than to positive news about the performance of their peers. While these two observations appear to be at odds with the findings of Moebius, Niederle, Niehaus, and Rosenblat (2007) and Eil and Rao (2010), they can in fact be reconciled by noting that participants in our study estimate absolute, not relative performance, and that good news about others' absolute performance constitutes bad news about one's own relative performance.

Nonetheless, we must remark upon the extent to which our results are sensitive to the inclusion of the two observations that we classify as outliers, the same observations excluded from the column (2) regressions. Column (4) presents the regression results when these outliers are excluded. In both conditions, the coefficients on $(\bar{b}_3^*-\bar{b}_2)\times GN$ decrease in absolute value and fall far short of significance, suggesting that the effects observed in column (3) are driven disproportionately by these two data points. Thus, the appropriate conclusion about the overall behavior of our participants is that they show no asymmetry in their response to good news versus bad news.

Finally, we attempt to incorporate into our analysis of information processing the observations from the participants whose feedback was inconsistent with their interim beliefs by computing the appropriate posterior beliefs using a proxy for Bayes' rule. Where Bayes' rule does not apply, participants are assumed to put 100% probability on the possible score (given f) that is closest to the mean of her interim beliefs, $\bar{b}_2$. Thus, $b_3^p = 1$, where $p = f - 2$ if $f > \bar{b}_2$, and $p = f + 2$ if $f < \bar{b}_2$. Column (5) presents the results of the regressions that includes these additional 30 observations. The inclusion of these proxied observations reduces the fit of the model and the coefficient on the Bayesian response, $(\bar{b}_3^*-\bar{b}_2)$, while the coefficient on $(\bar{b}_3^*-\bar{b}_2)\times GN$ becomes marginally significant in the *Self* condition. Intuitively, the inclusion of these participants makes the responses in overall beliefs look significantly more conservative and somewhat more asymmetric.

Table 6: Updating Regressions

		(1)	(2)	(3)	(4)	(5)
Other	$(\bar{b}_3^* - \bar{b}_2)$	1.18	1.25	1.14	1.14	0.70
		$(0.06)^{***}$	$(0.05)^{***}$	(0.13)	(0.12)	$(0.12)^{**}$
	$(\bar{b}_3^* - \bar{b}_2) \times GN$	—	—	−0.39	−0.23	0.22
		—	—	$(0.17)^{**}$	(0.17)	(0.15)
	GN	—	—	1.13	0.97	1.39
		—	—	$(0.28)^{***}$	$(0.28)^{***}$	$(0.30)^{***}$
	Constant	−0.29	−0.24	−0.53	−0.53	−0.93
		$(0.10)^{***}$	$(0.08)^{***}$	$(0.18)^{***}$	$(0.16)^{***}$	$(0.20)^{***}$
	Outliers Dropped	—	✓	—	✓	—
	Bayes' Proxy	—	—	—	—	✓
	R^2 (adj.)	0.8304	0.8741	0.8653	0.8734	0.8693
	N	81	80	80	79	98
Self	$(\bar{b}_3^* - \bar{b}_2)$	1.13	1.27	0.94	1.11	0.74
		$(0.08)^{*}$	$(0.09)^{***}$	(0.10)	(0.14)	$(0.09)^{***}$
	$(\bar{b}_3^* - \bar{b}_2) \times GN$	—	—	0.28	0.11	0.44
		—	—	(0.24)	(0.26)	$(0.26)^{*}$
	GN	—	—	0.41	0.33	0.49
		—	—	$(0.20)^{**}$	(0.20)	$(0.26)^{*}$
	Constant	−0.01	−0.01	−0.26	−0.18	−0.33
		(0.08)	(0.08)	$(0.11)^{**}$	(0.12)	$(0.15)^{**}$
	Outliers Dropped	—	✓	—	✓	—
	Bayes' Proxy	—	—	—	—	✓
	R^2 (adj.)	0.7418	0.7334	0.7619	0.7379	0.6910
	N	78	77	76	75	88

Notes: Dependent Variable: $(\bar{b}_3 - \bar{b}_2)$. $GN = 1$ *iff* $\bar{b}_3^* > \bar{b}_2$. Three participants for whom $f = \bar{b}_2$ are not included in this analysis, as it is considered neither good nor bad news. Asterisks applied to $(\bar{b}_3^* - \bar{b}_2)$ denote significant difference from one. Asterisks on all other coefficients denote significant difference from zero.

Table 7: Overestimation By Elicitation and Quiz

	Beliefs	*Other* Overest.	*Other* MSE	*Self* Overest.	*Self* MSE
First Quiz	b_1	0.88	12.52	1.57	10.11
		(0.41)	(1.88)	(0.33)	(1.10)
	b_2	-0.19	11.42	1.18	6.00
		(0.41)	(1.65)	(0.26)	(0.79)
	b_3	-0.10	4.21	0.68	3.71
		(0.21)	(0.69)	(0.21)	(0.74)
		$N = 54$		$N = 50$	
Second Quiz	b_1	-0.02	11.07	1.18	8.26
		(0.43)	(1.47)	(0.34)	(1.19)
	b_2	-0.28	11.82	1.08	6.70
		(0.45)	(1.49)	(0.32)	(1.28)
	b_3	-0.01	2.78	0.65	3.61
		(0.17)	(0.43)	(0.22)	(0.83)
		$N = 45$		$N = 40$	

3.4 Learning Transfer

In the previous section, we show that deviations from Bayesian updating cannot account for our participants' post-feedback overestimation of their own scores, so the remaining overconfidence must primarily be due to overconfident priors. Among all studies cited above, beliefs do respond to feedback, even if the process is slow or biased. Thus, it remains a puzzle how overconfidence originates, and how it persists in the presence of repeated instances of tasks and feedback. The evolution of beliefs over the course of our brief experiment suggests a channel through which overconfidence may persists, even in the absence of bias in the processing of information about performance.

Table 7 shows the average overestimation exhibited in prior, interim and posterior beliefs, along with their mean squared errors, for the first and second quiz in each condition. Thus, the table presents data on the six sets of beliefs that participants express. Patterns in errors across the six beliefs can, at minimum, present suggestions about learning and overconfidence.

In both conditions, the accuracy of prior beliefs improves somewhat from the first to the second quiz, with average overestimation dropping from 0.88 to -0.02 and from 1.57 to 1.18 in the *Other* and *Self* conditions, respectively. Similarly, the mean squared error of the prior estimate drops from 12.52 to 11.07 in the *Other* condition and from 10.11 to 8.26 in the *Self* condition. This may improvement may be due the experience of taking the first quiz, using the beliefs-elicitation instrument, and receiving and processing feedback. It might also be a reflection of slightly higher average scores on the second quiz in both conditions.

Interim beliefs show no improvement from the first quiz to the second, with overestimation increasing (in absolute value) from -0.19 to -0.28 in the *Other* condition and not changing in the *Self* condition, and with mean squared error dropping slightly in the *Other* condition, from 12.52 to 11.07, but increasing slightly, from 6.00 to 6.70, in the *Self* condition.

Learning patterns in posterior beliefs diverge between conditions. Average overestimation remains virtually constant for both *Other* and *Self*, mean squared error in the *Other* condition improves by 1.43, from 4.21 to 2.78, while it only improves by 0.10, from 3.71 to 3.61, in the *Self* condition. The improvement in the *Other* condition is marginally significant ($p = .052$) according to a two-sample t-test, while the improvement in the *Self* condition is not statistically significant. Furthermore, because participants in the *Other* condition estimate the performance of two different people in the two quizzes, one would expect their experience with the first quiz actually to be *less* relevant for estimating performance on the second quiz than it would be for those in the *Self* condition, who estimate the same person's performance in both quizzes. While those estimating their own performance may have private information about their own ability that would dampen the impact of information and experience on beliefs relative to those in the *Other* condition, the fact that prior overestimation in the first quiz is greater in the *Self* condition suggests that this private information is not accurate or useful.

Participants improve the way that they incorporate feedback into their beliefs about others, but not about themselves. After receiving noisy feedback, participants estimate the performance of strangers more accurately than they estimate their own.

4 Discussion and Conclusion

Participants in our experiment estimate their own performance on the quiz differently than they estimate the performance of others, a task that does not put their own image at stake. While they consistently are correct on average in their estimates of others' performance and attribute no bias to the feedback they receive about others, our participants were initially overconfident about their own performance. Furthermore, after receiving performance feedback, they remain somewhat overconfident, on average overestimating their score by 0.67 points and believing that their feedback understated their score by 0.65 points. In one sense, these results are consistent with the findings of Eil and Rao (2010) and Charness, Rustichini, and van de Ven (2010), namely that overconfident posterior beliefs are more common and more severe when one's own image is at stake than when processing feedback about randomly assigned rankings and or engaging in abstract probability conditioning.

However, while those studies, along with Moebius, Niederle, Niehaus, and Rosenblat (2007), find that information-processing errors drive their results, our participants actually do a reasonably good job processing the feedback about their own performance. The post-feedback overestimation is forty percent lower than the 1.14 questions observed prior to receiving feedback and the mean squared error of the score estimates drops by 36% after feedback is given. Posterior overestimation is by and large consistent with Bayes' rule and if any bias is reflected in the data, it is one that generates a slightly stronger reaction to feedback than warranted, which would tend to diminish the level of overestimation *ex post*.

Furthermore, while Moebius, Niederle, Niehaus, and Rosenblat (2007) and Eil and Rao (2010) find that people *do* under-react to noisy feedback about their relative performance, particularly to negative feedback, we find no significant asymmetry in the response to good news versus bad news. The absence of conservatism and asymmetry in our participants' responses, however, need not be regarded as completely at odds with previous results. Rather, our results suggest a refinement of our previous understanding of how people's response to information about their performance depends upon the way it is measured. The previous studies focused exclusively on estimates of relative performance, while our participants estimate their absolute performance. This suggests that the asymmetry in response to noisy feedback found by Moebius, Niederle, Niehaus, and Rosenblat (2007) and Eil and Rao (2010) may be driven largely by mis-attribution in social comparisons.

We conclude that the incorrectly high posterior beliefs we observed in the *Self* condition and that fact that estimated luck is negative on average are almost exclusively due to overconfident prior beliefs. We grant that degree to which participants' updating is consistent with Bayes' rule is likely aided by the simple and uniform noise process and the fact that our updating analysis necessarily excludes participants who received feedback regarded *ex ante* as zero-probability, a group whose members on average overestimated their score more severely and performed more poorly on the quiz than the rest of the participants. Thus, our findings do not call into dispute the conclusions of other studies finding biased updating.

Instead, our findings refine our understanding of the mechanisms that lead overconfidence to persist in the face of repeated task-feedback experience and they highlight an important distinction that needs to be made when considering how individuals update beliefs about their performance. Our participants fail to transfer learning from context to the next. Though feedback helps them learn about their performance on one task, our participants do not appear to learn about their underlying ability, and thus exhibit almost the same degree of overestimation of their own performance prior to the second quiz as they do prior to the first quiz. Thus, our participants do appear to learn within each quiz about their own performance, but they do not appear to learn from the first quiz about the process that generates their score on the second quiz, that is, their quiz-taking ability. If the feedback that affects beliefs about performances fails to alter beliefs about the aptitudes that generate them, overconfidence with respect to these aptitudes may persist even in the face of abundant and repeated feedback. Further research might examine the degree to which this failure of learning transfer persists with multiple rounds of quizzes and feedback.

Outside the laboratory, one has a lifetime over which to accumulate consistent and repeated feedback, from which to learn about learning about one's own performance and attributes. Despite the presence of self-serving biases that hinder one's ability to interpret information relevant to one's own performance, individuals *do* learn about their performances. Overestimation undoubtedly diminishes as age and experience provide repeated learning opportunities.

However, our results suggest that, beyond biased updating of beliefs about performance, overconfidence may also persist because of the failure to interpret learning acquired in one context as relevant to another. In the same way that vague or ambiguous feedback may exacerbate overconfidence due to the former, the fact that external task-feedback situations may be much less similar than the back-to-back, identical-format quizzes featured in this experiment likely exacerbates overconfidence due to the latter. It may not always be clear how relevant experience from one context is for the formation of beliefs relating to another context. More research is needed to clarify how overconfidence arises and persists outside the laboratory.

References

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini (2010): "Overconfidence is a Social Signaling Bias," IZA Discussion Papers 4840, Institute for the Study of Labor (IZA).

Camerer, C., and D. Lovallo (1999): "Overconfidence and Excess Entry," *American Economic Review*, 89(1), 306 – 318.

Charness, G., A. Rustichini, and J. van de Ven (2010): "Overconfidence, Subjectivity, and Self-Deception," Discussion paper, UC Santa Barbara, working paper.

Clark, J., and L. Friesen (2009): "Overconfidence in Forecasts of Own Performance: An Experimental Study," *Economic Journal*, 119(534), 229–251.

Eil, D., and J. Rao (2010): "The Good News-Bad News Effect: Asymmetric Processing of Objective Information About Yourself," *American Economic Journal: Microeconomics*.

Ertac, S. (2009): "Does Self-Relevance Affect Information Processing? Experimental Evidence on the Resonse to Performance and Non-Performance Feedback," working paper, Koc University.

Greiner, B. (2003): "An Online Recruitment System for Economic Experiments," *Forschung und wissenschaftliches Rechnen*, 63, 79–93.

Grosswirth, M., A. F. Salny, and A. Stillson (1999): *Match Wits With Mensa: The Complete Quiz Book (Mensa Genius Quiz)*. Da Capo Press, 1 edn.

Hoelzl, E., and A. Rustichini (2005): "Overconfident: Do You Put Your Money On It?," *Economic Journal*, 115(503), 305 – 318.

Malmendier, U., and G. Tate (2008): "Who Makes Acquisitions? CEO Overconfidence and the Market's Reactions," *Journal of Financial Economics*, 89(1), 20 – 43.

Moebius, M., M. Niederle, P. Niehaus, and T. Rosenblat (2007): "Gender Difference in Incorporating Performance Feedback," Discussion paper, working paper.

Moore, D. A., and P. J. Healy (2008): "The Trouble With Overconfidence," *Psychological Review*, 115(2), 502 – 517.

A Supplemental Tables and Figures

		Other Treatment				
		Obs.	s	$\bar{b}_1$	$\bar{b}_2$	$\bar{b}_3$
Quiz A	**First Quiz**	10	5.40 (0.82)	5.87 (0.22)	5.54 (0.40)	5.33 (0.46)
	Second Quiz	10	8.40 (0.50)	5.11 (0.42)	4.98 (0.45)	7.73 (0.56)
Quiz B	**First Quiz**	22	4.27 (0.49)	6.29 (0.31)	4.41 (0.32)	4.42 (0.42)
	Second Quiz	12	5.00 (0.43)	6.26 (0.31)	5.37 (0.37)	5.09 (0.49)
Quiz C	**First Quiz**	10	5.70 (1.01)	5.35 (0.38)	4.75 (0.41)	4.98 (0.96)
	Second Quiz	10	4.00 (0.61)	5.22 (0.37)	5.15 (0.33)	4.19 (0.59)
Quiz D	**First Quiz**	12	6.42 (0.50)	6.56 (0.17)	5.98 (0.34)	6.34 (0.49)
	Second Quiz	13	4.77 (0.64)	4.94 (0.36)	5.04 (0.49)	4.95 (0.64)
				Self Treatment		
		Obs.	s	$\bar{b}_1$	$\bar{b}_2$	$\bar{b}_3$
Quiz A	**First Quiz**	10	6.40 (0.72)	6.84 (0.56)	6.77 (0.71)	6.74 (0.68)
	Second Quiz	10	5.00 (0.91)	5.17 (0.83)	5.68 (0.91)	5.28 (0.91)
Quiz B	**First Quiz**	20	4.00 (0.42)	6.47 (0.28)	5.48 (0.34)	4.74 (0.41)
	Second Quiz	10	5.20 (0.63)	7.35 (0.19)	5.76 (0.39)	5.81 (0.50)
Quiz C	**First Quiz**	10	4.10 (0.67)	5.97 (0.77)	5.52 (0.88)	5.09 (0.86)
	Second Quiz	10	5.20 (0.71)	6.49 (0.58)	6.82 (0.54)	6.00 (0.69)
Quiz D	**First Quiz**	10	6.20 (0.70)	6.82 (0.32)	7.38 (0.26)	6.81 (0.46)
	Second Quiz	10	4.10 (0.82)	5.20 (0.59)	5.56 (0.47)	5.02 (0.59)

Table 8: Performance and Belief by Quiz

B Instructions

[Not for publication]

This appendix presents the instructions for both the *Self* and *Other* conditions.

Self Condition

The Quiz Experiment: Instructions

Introduction:

Thank you for participating in this experiment. Please follow along carefully as the experimenter reads through these instructions. If you have any questions at any point, please raise your hand.

This is an experiment in the economics of decision-making. A research foundation has provided funds for conducting this research.

For your participation, you will be paid privately and in cash at the end of this session. Your earnings will depend partly on your decisions, and partly on chance. If you follow the instructions and make careful decisions, you may earn a considerable amount of money.

You will receive $5 as a participation fee (simply for showing up on time). Details of how you will make decisions and gain subsequent earnings will be provided below.

Summary of the experimental procedure:

The experiment has two parts. Each part has five steps:

1. Before taking a 10 question quiz, you will estimate the number of questions, out of 10, than you will answer correctly.
2. You will take the 10 question quiz.
3. You will again estimate the number of questions, out of 10, that you answered correctly. *For the remainder of the instructions, we will refer to this number as* ***'X'.***
4. You will receive some information about the value of ***X***.
5. You will estimate ***X*** again.

The two parts are exactly the same, except that you will take a different quiz in each part. After you are done with both parts, you will learn the value of ***X*** for each part and then receive your earnings.

You can earn money both by answering questions correctly and by accurately estimating ***X***. The details of how your earnings are determined are explained below.

Detailed experimental procedure:

On the computer in front of you, a folder is open containing 4 files. Please do not touch the computer until you are instructed to do so.

1. ***Estimating X:***

Your estimation of ***X*** will be more involved than simply stating which value is the most likely. You will be asked to indicate how likely you think that ***X*** is equal to each number from 0 to 10.

When the experimenter tells you to do so, you will open the first file by double-clicking on the file with the mouse. This file has a spreadsheet that you will use to give us your estimate of ***X***. The spreadsheet has 11 dropdown menus, each corresponding to an integer between 0 and 10. In the menu below each number, you will select the likelihood with which you believe ***X*** is equal to that number. Beneath the drop-down menus are two charts that will help you. On the left is a chart titled *'**% Likelihood of total # correctly answered**.'* This will give you a graphical representation of the likelihoods that you have entered. On the right is a chart titled "***Total likelihood MUST add up to 100%".*** This chart adds up all of the likelihoods that you have selected. Please ensure that the total is 100%. Once you have finished making your selections, the experimenter will verify that your total adds up to 100%, and then instruct you on how to save and close the file. *Please do not close the file until you are instructed to do so.*

2. ***The Quiz:***

   When the experimenter tells you to do so, you will open the second file by double-clicking on the file with the mouse. This file contains a 10-question quiz. From the time the experimenter tells you to open the file, you will have exactly 10 minutes to answer the questions on the quiz. When the experimenter tells you that your time is up, please stop working on the quiz, and do not touch your computer's mouse or keyboard until given further instructions. While you are taking the quiz, you may use the back of these instructions pages for scratch work. We will not collect these pages or look at them.

3. ***Estimating X again:***
   After you complete the quiz you will estimate ***X*** a second time. The experimenter will ask you to open the third file. This file is identical to the one you used to estimate ***X*** the first time. Once again, you will enter the likelihood with which you believe that ***X*** is equal to each integer between 0 and 10, in a manner identical to that described in step 1, above.

4. ***Receiving information about X:***
   After you have made your second estimation on the value of ***X***, the experimenter will hand you a slip of paper with a number written on it. This number, ***'Y'***, is related to ***X*.** However, we call it *imperfect* information because ***Y*** will be partly determined by chance, and may or may not be equal to ***X***. Instead, ***Y*** is a number that we get when we add a randomly determined number to ***X***. Specifically, ***Y*** is equal to ***X*-2, *X*-1, *X*, *X*+1 or *X*+2,** each with equal probability.

5. ***Estimating X a final time:***

Finally, the experimenter will ask you to open the fourth file, which is identical to the files you used to estimate ***X*** the other times. Once again, you will enter the likelihood with which you believe that ***X*** is equal to each integer between 0 and 10, in a manner identical to that described in step 2, above.

6. ***Your earnings:***
   You can earn money in two main ways during the course of this experiment:

   1. Answering questions correctly
   At the end of the experiment, *for each part*, 1 of the 10 questions will be selected at random and you will be paid an extra $5 if you answered it correctly.

   2. Accurately Estimating Performance
   For each part, exactly one of the three estimates that you made will be randomly selected for payment. You will be paid ($1 + 5*p_X - w$) dollars for each of your two estimations, where:
   i. p_X is the likelihood that you expressed for the actual value of ***X***, and
   ii. w is the sum squares of the likelihood that you attached to each of the 11 scores.

   <u>This formula may appear complicated, but what it means for you is very simple</u>: *You get paid the most when you honestly report your best guesses about the likelihood of each of the different possible outcomes.* The range of your payoffs is from $0 to $5 for each set of guesses. Since we select two of these estimates for payment (one for each part), you could earn up to a theoretical maximum of $10 for two perfect guesses (but that would be very difficult to do).

   Example:
   As an example, imagine that, at the end of the experiment, you learn that ***X***=5.

   a. If you had expressed that you were 100% certain that ***X***=5:
      - $p_X = 1.0$
      - $w = 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 = 1$
      - Earnings from estimation = $(1 + 5*1.0 - 1) = $5.00

   b. If you had expressed that you were 100% certain that ***X***=4:
      - $p_X = 0$
      - $w = 0^2 + 0^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 = 1$
      - Earnings from estimation = $(1 + 5*0 - 1) = $0.00

   c. If you expressed likelihoods of 10%, 20%, 30% and 40% for ***X***=5, ***X***=6, ***X***=7 and ***X***=8, respectively:
      - $p_X = .10$
      - $w = .1^2 + .2^2 + .3^2 + .4^2 = .30$
      - Earnings from estimation = $(1 + 5*.10 - .30) = $1.20

d. If you expressed likelihoods of 10% and 90% for ***X***=5 and ***X***=6 respectively:
   - p_X = .10
   - $w = .1^2 + .9^2 = .82$
   - Earnings from estimation = $(1 + 5*.10 - .82) = $.78

To summarize: there are two parts. ***X*** is the number of questions that you answered correctly on the quiz. In each part, you will estimate ***X*** three times: before you take the quiz, after you take the quiz, and after you receive some information about ***X***. You will be paid $5 + your earnings from the first quiz + your earnings from one of the estimates of ***X*** from part 1 + your earnings from the second quiz + your earnings from one of the estimates of ***X*** from part 2.

Once the experimenter has answered any questions you may have, we will begin part 1.

Other Condition

The Quiz Experiment: Instructions

Introduction:

Thank you for participating in this experiment. Please follow along carefully as the experimenter reads through these instructions. If you have any questions, please don't hesitate to raise your hand.

This is an experiment in the economics of decision-making. A research foundation has provided funds for conducting this research.

For your participation, you will be paid privately and in cash at the end of this session. Your earnings will depend partly on your decisions, partly on the decisions of others, and partly on chance. If you follow the instructions and make careful decisions, you may earn a considerable amount of money.

You will receive $5 as a participation fee (simply for showing up on time). Details of how you will make decisions and gain subsequent earnings will be provided below.

Summary of the experimental procedure:

The experiment has two parts. Each part has six steps:

1. You will be randomly paired with one of the other participants in the room, whom we will call your 'colleague.' You will not learn which of the other participants is your colleague.
2. Before taking a 10 question quiz, you will estimate the number of questions, out of 10, that your colleague will answer correctly.
3. You will take the 10 question quiz. Everyone in the room will take the same quiz.
4. You will again estimate the number of questions, out of 10, that your colleague answered correctly. *For the remainder of the instructions, we will refer to this number as '**X**'.*
5. You will receive some information about the value of ***X***.
6. You will estimate ***X*** again.

The two parts are exactly the same, except that you will take a different quiz in each part. After you are done with both parts, you will learn the value of ***X*** for each part and then receive your earnings.

You can earn money both by answering questions correctly and by accurately estimating ***X***. The details of how your earnings are determined are explained below.

Detailed experimental procedure:

On the computer in front of you, a folder is open containing 4 files. Please do not touch the computer until you are instructed to do so.

1. ***Estimating X:***
   Your estimation of ***X*** will be more involved than simply stating which value is the most likely. You will be asked to indicate how likely you think that ***X*** is equal to each number from 0 to 10.

   When the experimenter tells you to do so, you will open the first file by double-clicking on the file with the mouse. This file has a spreadsheet that you will use to give us your estimate of ***X***. The spreadsheet has 11 dropdown menus, each corresponding to an integer between 0 and 10. In the menu below each number, you will select the likelihood with which you believe ***X*** is equal to that number. Beneath the drop-down menus are two charts that will help you. On the left is a chart titled *'**% Likelihood of total # correctly answered**.'* This will give you a graphical representation of the likelihoods that you have entered. On the right is a chart titled "***Total likelihood MUST add up to 100%".*** This chart adds up all of the likelihoods that you have selected. Please ensure that the total is 100%. Once you have finished making your selections, the experimenter will verify that your total adds up to 100%, and then instruct you on how to save and close the file. *Please do not close the file until you are instructed to do so.*

2. ***The Quiz:***

   When the experimenter tells you to do so, you will open the second file by double-clicking on the file with the mouse. This file contains a 10-question quiz. From the time the experimenter tells you to open the file, you will have exactly 10 minutes to answer the questions on the quiz. When the experimenter tells you that your time is up, please stop working on the quiz, and do not touch your computer's mouse or keyboard until given further instructions. While you are taking the quiz, you may use the back of these instructions pages for scratch work. We will not collect these pages or look at them.

3. ***Estimating X again:***
   After you complete the quiz you will estimate ***X*** a second time. The experimenter will ask you to open the third file. This file is identical to the one you used to estimate ***X*** the first time. Once again, you will enter the likelihood with which you believe that ***X*** is equal to each integer between 0 and 10, in a manner identical to that described in step 1, above.

4. ***Receiving information about X:***
   After you have made your second estimation on the value of ***X***, the experimenter will hand you a slip of paper with a number written on it. This number, ***'Y'***, is related to ***X***. However, we call it *imperfect* information because ***Y*** will be partly determined by chance, and may or may not be equal to ***X***. Instead, ***Y*** is a number

that we get when we add a randomly determined number to ***X***. Specifically, ***Y*** is equal to ***X*-2, *X*-1, *X*, *X*+1 or *X*+2,** each with equal probability.

5. ***Estimating X a final time:***
   Finally, the experimenter will ask you to open the fourth file, which is identical to the files you used to estimate ***X*** the other times. Once again, you will enter the likelihood with which you believe that ***X*** is equal to each integer between 0 and 10, in a manner identical to that described in step 2, above.

6. ***Your earnings:***
   You can earn money in two main ways during the course of this experiment:

   1. Answering questions correctly
   At the end of the experiment, *for each part*, 1 of the 10 questions will be selected at random and you will be paid an extra $5 if you answered it correctly.

   2. Accurately Estimating Performance
   For each part, exactly one of the three estimates that you made will be randomly selected for payment. You will be paid ($1 + 5*p_X - w$) dollars for each of your two estimations, where:
   i. p_X is the likelihood that you expressed for the actual value of ***X***, and
   ii. w is the sum squares of the likelihood that you attached to each of the 11 scores.

   This formula may appear complicated, but what it means for you is very simple: *You get paid the most when you honestly report your best guesses about the likelihood of each of the different possible outcomes.* The range of your payoffs is from $0 to $5 for each set of guesses. Since we select two of these estimates for payment (one for each part), you could earn up to a theoretical maximum of $10 for two perfect guesses (but that would be very difficult to do).

   Example:
   As an example, imagine that, at the end of the experiment, you learn that ***X***=5.

   a. If you had expressed that you were 100% certain that ***X***=5:
      - $p_X = 1.0$
      - $w = 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 = 1$
      - Earnings from estimation = $(1 + 5*1.0 - 1) = $5.00

   b. If you had expressed that you were 100% certain that ***X***=4:
      - $p_X = 0$
      - $w = 0^2 + 0^2 + 0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 = 1$
      - Earnings from estimation = $(1 + 5*0 - 1) = $0.00

c. If you expressed likelihoods of 10%, 20%, 30% and 40% for ***X***=5, ***X***=6, ***X***=7 and ***X***=8, respectively:
- p_X = .10
- $w = .1^2 + .2^2 + .3^2 + .4^2 = .30$
- Earnings from estimation = $(1 + 5*.10 - .30) = $1.20

d. If you expressed likelihoods of 10% and 90% for ***X***=5 and ***X***=6 respectively:
- p_X = .10
- $w = .1^2 + .9^2 = .82$
- Earnings from estimation = $(1 + 5*.10 - .82) = $.78

To summarize: there are two parts. ***X*** is the number of questions that your colleague answered correctly on the quiz. In each part, you will estimate ***X*** three times: before you take the quiz, after you take the quiz, and after you receive some information about ***X***. You will be paid $5 + your earnings from the first quiz + your earnings from one of the estimates of ***X*** from part 1 + your earnings from the second quiz + your earnings from one of the estimates of ***X*** from part 2.

Once the experimenter has answered any questions you may have, we will begin part 1.

C The Quizzes

[Not for publication]

This appendix presents a table containing the solutions to the quiz questions followed by the four quizzes.

Table 9: The quiz solutions

Question\Quiz	A	B	C	D
1	40 mph	3	20	12 stamps
2	$120	26	17	$0.05
3	5	CPOEHNANGE	208	dirty, squalid, mean, despicable
4	15	18 days	CPOEHNANGE	7
5	66.66%	2.5 hours	18 days	8
6	9:00 AM	sleight of hand, conjuring tricks	sleight of hand, conjuring tricks	240
7	21:00	208	3 rungs	6
8	Radish	20	yes	12
9	6	20	$252	11,5
10	New York	64	Liar	25

Quiz A

Quiz

#1	10 mph 20 mph 30 mph 40 mph 50 mph	A man goes to visit his friend thirty miles away. He doesn't mind speeding, so he travels at 60 miles per hour and arrives in half an hour. On the way back, however, he has a little trouble with his car, and it takes him an hour to reach home. What is his average speed for the round trip?
#2	$90 $95 $100 $105 $110 $115 $120	A man went into a jewelry store and bought a $75 chain, giving the clerk a $100 bill. He returned a few moments later and bought a new watch, giving the clerk a $20 bill and receiving $5 in change. Later, the bank told the store that both the $100 bill and the $20 bill were counterfeit. Ignoring markup, overhead, cost of merchandise, etc., how much money did the store lose?
#3	3 4 5 6 7 8	You have twenty-four socks in a drawer, six each of brown, black, white, and red. How many socks must you take out of the drawer, without looking, to be sure you have a matched pair (of any color)?
#4	1 3 5 7 11 13 15	Which number in the following series of numbers is least like the others? **1, 3, 5, 7, 11, 13, 15**
#5	33.33% 40% 50% 60% 66.66%	The price of an article is cut 40% for a sale. When the sale is over, the store owner wants to bring the price back up to the original selling price. What percentage of the sale price must be added to that sale price to bring the price back up to the original selling price?

#6

- 8:00 AM
- 8:30AM
- 9:00AM
- 9:30AM
- 10:00AM

A girl decides to take a long walk in the country and visit a friend on tohe way. She walks at a steady pace of 2.5 miles per hour. She spends 4 hours walking over to her friend's hosue; she has a cup of coffee and a sandwich and talks to her friend, all of which occupies an hour, and then her friend runs her home in the car, over some rough road, at 20 miles per hour. She gets home at 2:30 PM. When did she leave her house?

#7

- 18:30
- 20:00
- 21:00
- 22:00
- 23:30

If it were two hours later, it would be half as long until midnight as it would be if it were an hour later. What time is it now?

#8

- Banana
- Radish
- Strawberry
- Peach
- Lettuce

Pear is to apple as potato is to?

#9

- 1
- 4
- 6
- 10
- Can't Answer

A trader buys coffee for $1,200 and sells it for $1,500. For each bag of coffee he earns a profit of $50. How many bags of coffee did he have?

#10

- Canberra
- New York
- Vienna
- Madrid

From these four cities, pick the odd one out.

Quiz B

Quiz

#1	1 2 3 4 5	If a jet has a value of 1, and a plane has a value of 2, what is the value of a Concorde?
#2	21 22 23 24 25 26	How many minutes is it before six o'clock if fifty minutes ago it was four times as many minutes past three o'clock?
#3	CGHICOA TTOOORN IMMIA CPOEHNANGE	Which of the following scrambled words is the "odd man out" when the words are unscrambled? **CGHICOA, TTOOORN, IMMIA, CPOEHNANGE**
#4	16 days 17 days 18 days 19 days 20 days	A snail is climbing out of a well. The well is twenty feet deep. Every day the snail climbs up three feet and every night he slips back two feet. How many days will it take him to get out of the well?
#5	1 hour 1.5 hours 2 hours 2.5 hours 3 hours	Your doctor gives you six pills and tells you to take one every half hour. How long does it take you to use up all of the pills?
#6	a kind of imported handgun the Italian name for the town of Leghorn sleight of hand, conjuring tricks a former prime minister of France the title of a famous poem by Keats	Definition: Legedermain
#7	24 120 208	If *diaphanous* and *sheer* do not have the same meaning, cross out all the 9's in the line below. If they do, cross out all the 6's. If *slough* and *cough* are pronounced the same, multiply the number of 4's by 6. If not, add up all of the non-crossed-out numbers and multiply by 4. **9, 4, 6, 4, 9, 4, 6, 9, 4, 6, 9**

#8	10¢ 20¢ 30¢ 40¢ 50¢	Under certain special circumstances, a peach costs 20¢, a banana costs 30¢, and a grapefruit costs 40¢. How much will a pear cost under the same circumstances?
#9	10 11 19 20 21	If you count from 1 to 100, how many 7's will you pass on the way?
#10	36 45 46 64 99	Following the pattern shown in the number sequence below, what is the missing number? **1,8,27,?,125,216**

Quiz C

Quiz

#1	10 11 19 20 21	If you count from 1 to 100, how many 7's will you pass on the way?
#2	14 15 16 17 18	Which number comes next in this series of number? **2, 3, 5, 7, 11, 13, ?**
#3	24 120 208	If *diaphanous* and *sheer* do not have the same meaning, cross out all the 9's in the line below. If they do, cross out all the 6's. If *slough* and *cough* are pronounced the same, multiply the number of 4's by 6. If not, add up all of the non-crossed-out numbers and multiply by 4. **9, 4, 6, 4, 9, 4, 6, 9, 4, 6, 9**
#4	CGHICOA TTOOORN IMMIA CPOEHNANGE	Which of the following scrambled words is the "odd man out" when the words are unscrambled? **CGHICOA, TTOOORN, IMMIA, CPOEHNANGE**
#5	16 days 17 days 18 days 19 days 20 days	A snail is climbing out of a well. The well is twenty feet deep. Every day the snail climbs up three feet and every night he slips back two feet. How many days will it take him to get out of the well?
#6	a kind of imported handgun the Italian name for the town of Leghorn sleight of hand, conjuring tricks a former prime minister of France the title of a famous poem by Keats	Definition: Legedermain
#7	3 rungs 8 rungs 14 rungs 20 rungs 25 rungs	A man moors his boat in a harbor at high tide. A ladder is fastened to the boat, with three rungs showing. The rungs are twelve inches apart. At low tide the water level sinks twenty feet. How many rungs of the ladder are showing now?

#8	yes no	Does Canada have a fourth of July?
#9	$240 $244 $252 $258 $264	A man bets $24 and gets back his original bet and $48 additional. He spends 25 percent of his winnings at a restaurant to celebrate, and 50 percent of his winnings to buy a present for his wife because he was so late, and his salary was $240, from which he made his original bet. How much money does he have left when he finally arrives home?
#10	Liar Truthteller	You are at a meeting at which there are only liars and truthtellers. A woman comes up to you and says that the chairman of the meeting told her he was a liar. Is she a liar or a truthteller?

Quiz D

Quiz

#1

6 stamps
12 stamps
18 stamps
24 stamps

If there are twelve one-cent stamps in a dozen, how many two-cent stamps are there in a dozen?

#2

$0.05
$0.10
$0.50
$1.00
$1.05

You have two pencils, a good one and a cheap one. The good one cost $1.00 more than the cheap one. You spend $1.10 for both. How much did the cheap one cost?

#3

dirty, squalid, mean, despicable
deaf, hard of hearing
a type of acid
a thing of small value, a trifle
a scab

Definition: Sordid

#4

2
5
7
9
12

A horse salesman went to a horse auction with a certain number of horses. To his first customer, he sold half his horses, plus half a horse. To his second, he sold half of what he had left, plus half a horse. To the last he sold half of what he had left, plus half a horse. Each customer received whole horses, however, not half horses, and he had no horses left. How many horses did he start out with?

#5

2
3
4
5
6
7
8

Complete the series by adding the next number: **0, 0, 1, 2, 2, 4, 3, 6, 4, __**

#6

238
239
240
241
242

How many odd pages are there in a book 479 pages long?

#7

1
4
6
10
Can't Answer

A trader buys coffee for $1,200 and sells it for $1,500. For each bag of coffee he earns a profit of $50. How many bags of coffee did he have?

#8

- 4
- 9
- 10
- 12
- 15

Four years ago, Jane was twice as old as Sam. Four years from now, Sam will be 3/4 of Jane's age. How old is Jane now?

#9

- 11, 5
- 10, 5
- 10, 4
- 11, 6

Continue the following number series with the group of numbers below which best continues the series. **1, 10, 3, 9, 5, 8, 7, 7, 9, 6, ?, ?**

#10

- 23
- 25
- 28
- 30
- 32

Look at the drawing. The numbers alongside each column and row are the total values of the symbols within each column and row. What should replace the question mark?