9-2007

Bayesian Overconfidence

Paul J. Healy
Ohio State University

Don A. Moore
Carnegie Mellon University, dmoore@cmu.edu

Bayesian Overconfidence*

Paul J. Healy[†]
Department of Economics
The Ohio State University

Don A. Moore[‡]
Tepper School of Business
Carnegie Mellon University

September 24, 2007

Abstract

We study three distinct measures of overconfidence: (1) *overestimation* of one's performance, (2) *overplacement* of one's performance relative to others, and (3) *overprecision* in one's belief about private signals. A new set of experiments verifies a strong negative link between overestimation and overprecision that depends crucially on task difficulty (the 'hard-easy' effect). We present a simple Bayesian model in which agents are uncertain about the underlying task difficulty. This model correctly predicts the observed regularities. Thus, we capture several observed patterns of overconfidence without assuming any implicit behavioral biases.

*The authors wish to acknowledge the financial support of National Science Foundation Grant SES-0451736 as well as a CART grant from the Tepper School of Business at Carnegie Mellon University. We are indebted to Colin Camerer, Juan Dubra, Robin Hogarth, George Loewenstein, Ted O'Donoghue, Joel Sobel, Roberto Weber, and three anonymous reviewers for the 2007 Academy of Management Meeting Best Papers Proceedings. Zach Burns and the Pittsburgh Experimental Economics Laboratory provided invaluable assistance with data collection. The most current version of this paper is available online at http://healy.econ.ohio-state.edu.

†1945 North High Street, Columbus, OH 43210. Phone: (614) 247-8876. Email: healy.52@osu.edu.

‡5000 Forbes Avenue, Pittsburgh, PA 15213. Phone: (412) 268-5968 Email: dmoore@cmu.edu.

1 Introduction

It is often assumed that overconfident beliefs are 'irrational' because they are statistically incorrect. For example, Ola Svenson (1981) found that 93 percent of a sample of American drivers reported themselves to be more skillful at driving than the median American driver. Although it is extremely unlikely that these beliefs are correct for this sample, is it necessarily the case that some of these individuals hold irrational beliefs? Are people's beliefs naturally biased towards believing themselves better than others, or are these 'biased' beliefs the rational outcome of some Bayesian inference procedure?

In this paper, we present a very simple model in which agents participate in a task of unknown difficulty. After performing the task, each agent is asked to estimate her own performance and the performance of a randomly-selected other participant. Assuming the agent's prior beliefs are well-behaved, we show that when the task is unexpectedly easy, she will believe that she has out-performed her peers but will simultaneously underestimate her own performance. When the task is unexpectedly difficult, she will believe her performance was worse than her peers but will also overestimate her own performance.

Critical to this result is the distinction between overconfidence about one's ranking relative to others and overconfidence about one's score relative to the true score. Much of the previous literature muddles these distinct concepts, so we clearly define three types of overconfidence: overestimation (overestimating one's own performance), overplacement (overestimating one's rank relative to others), and overprecision (beliefs that have greater precision than is warranted by the data). The Bayesian model predicts overplacement and underestimation after unexpectedly easy tasks and underplacement and overestimation after unexpectedly difficult tasks.

The intuition behind the predictions of the model is straightforward: experiencing an unexpectedly good outcome implies that the task was somewhat easier than expected, but also that you performed somewhat better than average. Thus, you predict that you have outperformed your competitors. As an example, suppose every manager in a particular emerging industry agrees that the expected marginal cost of a new product is $10 per unit. After production begins, however, each firm privately observes an actual marginal cost ranging from $7 to $9 – well below the common prior expectation. Each manager might conclude that their lower-than-expected cost was partly due to an incorrect prior estimate of $10, but also partly due to his own firm's better-than-average ability at producing the product. Thus, it is possible that all managers simultaneously exhibit overplacement, believing that its costs are lower than the average competitor. Had the actual costs been higher than the prior estimate (ranging from $11 to $13, for example), the result would reverse and all managers might exhibit underplacement, believing that its costs are higher than average. Thus, we can generally conclude that overplacement is more likely after unexpectedly easy tasks and underplacement is more likely after unexpectedly difficult tasks.

Now suppose that firms build a prototype product before opening their production lines, and the cost of the prototype serves as an unbiased signal of the true production cost. If the range of prototype costs is lower than expected, managers might rationally conclude that their true production cost will lie somewhere between the prior estimate ($10) and the observed prototype cost. Given any firm whose true production cost is $8, we would expect that, on average, the prototype cost is $8 but the firm's expectation about its true production cost is higher – perhaps $9. An outsider who observes that true production costs ($8) are lower than the prior estimate ($10) will also observe that firms are, on average, overestimating their costs (at $9). The result reverses for higher-than-expected true production costs. In general, agents are more

likely to underestimate their ability when actual performance is better than previously expected and are more likely to overestimate their ability when performance is worse than expected.

To examine these predictions, we abstract away from the particulars of the competitive environment and directly examine individuals' beliefs about their own performances and the performances of others in an experimental setting where subjects participate in a sequence of trivia quizzes.[1] Our experimental results verify that overconfidence is closely linked to task difficulty; after easy tasks, subjects exhibit overplacement and underestimation. After difficult tasks, subjects exhibit underplacement and overestimation. We therefore confirm the predicted negative relationship between these two types of overconfidence. These results complement the results of Justin Kruger (1999), who finds systematic underplacement on difficult tasks, and Don A. Moore and Tai Gyu Kim (2003) or Don A. Moore and Deborah A. Small (forthcoming), who document similar connections between overestimation, overplacement, and task difficulty.

From the perspective of our model, overconfidence is a 'bias' only in the sense that observed beliefs do not match the actual distribution of outcomes. Agents' beliefs are consistent with Bayes's rule and are optimal given the information available – the only way for agents to improve the accuracy of their beliefs is with more information. The 'bias' occurs only because agents are not endowed with complete information about the task's difficulty. We therefore think of overconfidence as a 'statistical bias' caused by incomplete information rather than a 'behavioral bias' caused by errors in judgement or psychological factors. In fact, no behavioral bias is needed to generate

[1] It may be that competition exacerbates any overconfidence bias (and may introduce other biases). We study beliefs in the absence of competition as a first step in understanding the basic nature of overconfidence. Results from other studies that do incorporate competition (such as those cited above) can then be used to paint a more complete picture of the overconfidence phenomenon.

predictions in line with the statistical biases we observe in our experiment.

One notable difference between this and most other models of overconfidence is that this model does not assume overplacement or overestimation before agents experience the task; the managers' $10 estimate may be perfectly accurate (on average) given the initial information available. Only after experiencing the task and observing new information do overplacement and overestimation emerge. This is consistent with our experimental data, where overplacement and overestimation are observed after performing a task, but not before.

Many authors have shown that the presence of overconfidence leads to meaningful economic consequences, such as excess market entry (see, e.g., Richard Roll, (1986); Colin Camerer and Dan Lovallo, (1999); James G. March and Zur Shapira, (1987)), increased volatility in financial markets (e.g., Terrance Odean, (1998) and (1999); Kent D. Daniel, David Hirshleifer and Avanidhar Subrahmanyam, (2001)), or excessive investment in capital (e.g., Ulrike Malmendier and Geoffrey Tate, (2005)). We do not dispute the importance of the statistical bias of overconfidence; our model implies only that behavioral biases are not necessary to explain its existence.

The theoretical environment is described in the following section. In Section 3 we discuss the design of our experiments. The results appear in Section 4. Previous literature is discussed in Section 5 and the paper concludes with Section 6.

2 Bayesian Overconfidence

In this section we formally demonstrate how Bayesian inference can generate the observed patterns of overconfidence and underconfidence. In what follows, upper-case variables represent random variables and lower-case variables represent particular realizations of the corresponding random variable. We assume that each agent i per-

forms an identical task in isolation and receives a numerical score, denoted x_i, which quantifies her performance. Each agent believes that each score x_i is a realization of X_i, and X_i is determined by

$$X_i = S + L_i, \tag{1}$$

where S is the overall expected score across agents (or, the *simplicity* of the task) with $E[S] = \mu$ and L_i is a mean-zero idiosyncratic component that determines the difference between i's score and the overall average.[2] As a simple mnemonic, we refer to L_i as agent i's *luck*, but, depending on the application, it may include a variety of other idiosyncratic components, such as i's unknown task-specific ability level.[3] Assuming agent i has well-defined prior beliefs about the distributions of S and L_i, she can update those beliefs upon observing her realized score x_i. If agent i also has well-defined priors on L_j for some other agent $j \neq i$, her beliefs about X_j will also change as she observes x_i and updates her belief about S. In this way agent i may exhibit overplacement or underplacement with respect to others' scores after she observes her own score.

2.1 Overplacement

Under the above assumptions, i's prior expectations of her own score and the score of another agent are $E[X_i] = E[X_j] = \mu$. After performing the task and observing x_i, she updates her beliefs about S and L_i. Since she does not observe x_j for any $j \neq i$, her beliefs about L_j remain unchanged, so that $E[X_j|x_i] = E[S|x_i]$. We say that i exhibits *overplacement* if $E[X_j|x_i] < x_i$ and *underplacement* if $E[X_j|x_i] > x_i$. Note that overplacement (and underplacement) is a property of posterior beliefs, with no necessary connection to actual scores; for example, it is possible for all agents

[2] For simplicity, we assume throughout that all random variables have well-defined means.
[3] We discuss the case where $E[L_i]$ is non-zero in Subsection 2.4.2.

to simultaneously exhibit overplacement, even though some do and some do not outperform their peers.

Suppose an agent i who has never encountered the task before receives a score higher than expected ($x_i > \mu$). She might infer that her high score was due to good luck ($l_i > 0$) or a simpler-than-expected task ($s > \mu$). If she attributes her high score entirely to the task's simplicity (*i.e.*, she believes $E[S|x_i] = x_i$), then she will exhibit no overplacement because task simplicity affects all agents equally. If instead she attributes her high score at least partially to her own luck (*i.e.*, she believes $E[S|x_i] < x_i$), then she will exhibit overplacement since $E[X_j|x_i] = E[S|x_i] < x_i$. Similarly, if $x_i < \mu$ and she attributes her low score at least partially to luck, then she will exhibit underplacement. Thus, we expect overplacement after unexpectedly easy tasks and underplacement after unexpectedly difficult tasks.[4]

Whether or not we observe $E[X_j|x_i] < x_i$ when $x_i > \mu$ and $E[X_j|x_i] > x_i$ when $x_i < \mu$ depends on the belief distributions over S and L_i. If, for example, beliefs over S are uniformly distributed (and beliefs over L_i are not), then $E[X_j|x_i] = E[S|x_i] = x_i$ for all x_i and no overplacement or underplacement is observed. If the belief distributions are such that

$$E[X_j|x_i] = E[S|x_i] = \alpha\mu + (1-\alpha)x_i$$

for some $\alpha > 0$, then we must observe the required pattern of overplacement for every x_i.[5] The following examples highlight cases where $E[S|x_i]$ takes this particular form.

[4]Since this theory includes no behavioral biases, it also applies in situations where some agent k observes x_i but not x_j. Thus, we predict that people also tend to exhibit the predicted patterns overconfidence about their *friend's* performance when the friend's performance is observable but others' performances are not.

[5]Chambers and Healy (2007) explore general conditions on belief distributions that guarantee posterior expectations of the form $E[S|x_i] = \alpha\mu + (1-\alpha)x_i$ for $\alpha \in [0,1]$ and, more generally, for $\alpha \leq 1$. For example, symmetry and quasiconcavity of the densities is sufficient for the result with $\alpha \in [0,1]$.

Example 1 Suppose that i believes that $S \sim \mathcal{N}(\mu, \sigma_S^2)$ (S is distributed according to a normal distribution with mean μ and variance σ_S^2) and $L_j \sim \mathcal{N}(0, \sigma_L^2)$ for each j (including i). By Bayes's Rule, $E[X_j|x_i] = E[S|x_i] = \alpha\mu + (1-\alpha)x_i$, where $\alpha = \sigma_L^2/(\sigma_S^2 + \sigma_L^2)$.[6] □

This example is in fact a special case of a more general theorem due to Persi Diaconis & Donald Ylvisaker (1979), who show that if the distribution of X_i given S is in the exponential family, then $E[S|x_i]$ lies between μ and x_i if and only if the prior on S is conjugate. Thus, we expect the predicted pattern of overplacement when X_i given S has a normal distribution with a normal prior on S, an exponential distribution with a gamma prior, a Pareto distribution with a Pareto prior, a poisson distribution with a gamma prior, a geometric distribution with a beta prior, and, as in the next example, when X_i has a binomial distribution with a beta prior.

Example 2 In our experimental environment, subjects complete a sequence of 10-question quizzes. Suppose subjects believe their scores to be binomially distributed with parameter p (meaning they expect to get each question correct with probability p).[7] If p is an unknown parameter distributed according to a beta distribution with parameters β_1 and β_2 (so that $\mu = 10\ \beta_1/(\beta_1 + \beta_2)$) then

$$E[X_j|x_i] = \left(\frac{\beta_1 + \beta_2}{\beta_1 + \beta_2 + 10}\right)\mu + \left(1 - \frac{\beta_1 + \beta_2}{\beta_1 + \beta_2 + 10}\right)x_i.$$

Since the posterior mean for X_j lies between the prior mean and the observation of x_i, we predict the same pattern of overplacement as in example 1.[8] □

[6] This is a familiar property of normal distributions; see Berger (1980, p.127-8) for a derivation.

[7] This example assumes that subjects believe their success on each question is independent of success on all other questions *for a given* p. When p is unknown, however, independence fails since success on one question provides information about the probability of success on other questions.

[8] For a derivation of $E[X_j|x_i]$, see George Casella and Roger L. Berger (2002, p. 325).

Not all beliefs on S and X_i generate this pattern of overplacement. As noted, a uniform prior on S results in $E[S|x_i] = x_i$, so agent i expects others to do exactly as well as she. If the prior is highly bimodal then $E[X_j|x_i]$ might 'overshoot' x_i. As an extreme example, if an agent's prior belief is that S can only equal -8 or 8 and that L_i can only range from -2 to $+2$, then if she observes $x_i = 6$, she knows that $s = 8$ and so $x_j \in [6, 10]$, or $x_j \geq x_i$. If the belief on L_i is highly bimodal, the posterior might move in the *opposite* direction as the observed score. To see this, suppose now that L_i equals -8 or 8, each with probability one half, and that S ranges from -2 to 2. Now observing $x_i = 6$ leads the agent to conclude that $E[X_j|x_i] = s = -2$. In other words, receiving a high score causes the agent to believe others will do *worse* than previously expected.[9]

2.2 Overestimation

In many situations, the act of performing a task does not perfectly reveal one's performance. A competitor in a judged competition such as figure skating typically has a strong signal of her performance immediately after competing, but will not know her score with certainty until the judges tally her scores and will not know her ranking relative to others until the competition is complete. We refer to the time before the individual competes as the *ex-ante* phase, where beliefs are determined entirely by priors. The time after the individual competes but before she learns her score is the *interim* phase, where her information is based only on an imperfect signal of her performance. Upon learning her own score, she enters the *ex-post* phase, where she is fully informed about her own score, but knows nothing yet of her competitors' performances. After learning her competitors' scores, the ex-post phase is complete

[9]For a continuous example, suppose the density function on S is $(1 - |x|/3)/3$ over $[-3, 3]$ and L_i takes values of -2 or 2, each with probability one half. If $x_i \in (-2, 2)$ then $E[X_j|x_i] = -x_i$, so the agent expects others' scores to be exactly *opposite* of her own.

and she has full knowledge of the outcomes.

We model the interim phase by assuming that each agent i observes a noisy signal of her true score, y_i, which is believed to be a realization of the random variable $Y_i = x_i + E_i$, where E_i is a mean-zero error term. Since we assume $X_i = S + L_i$, an agent who observes a draw y_i makes inferences about S, L_i, and E_i. For example, a high value of y_i may lead her to conclude that S is relatively high, but that L_i and E_i were positive as well. In this case, she will expect that she did better than average (because L_i is positive), but not as well as her signal indicated (because E_i is positive). If her signal is in fact accurate, then she has underestimated her actual performance. Formally, we say that i exhibits *overestimation* if $E[X_i|y_i] > x_i$ and *underestimation* if $E[X_i|y_i] < x_i$.

In practice, we cannot observe agents' private signals, but we can observe the the resulting distribution of $X_i|y_i$. The drawback of this approach is that this posterior distribution depends crucially on the unobservable signal, so that random noise in the draws of the signals will translate into noise in the observed posteriors on X_i. To avoid these difficulties, we integrate across all possible signals (or, average across all elicited beliefs) to calculate the expected value of $E[X_i|y_i]$ when the true score (x_i) is known. Formally, we can calculate $E_{Y_i}[E[X_i|Y_i]|x_i]$ for any x_i and compare it against x_i. If this expected score is greater than x_i, we conclude that agents exhibit *overestimation in expectation*. If it is less than x_i, agents exhibit *underestimation in expectation*.

The following example shows how the results for overestimation can move in the opposite direction as those for overplacement; when a task is easier than expected, agents exhibit overplacement and *under*estimation. When a task is more difficult than expected, agents exhibit underplacement and overestimation. These predictions are summarized in Table 1.

Example 1 (Continued) Let $Z_j = L_j + E_j$. If we assume that E_j is also normally distributed with mean zero and variance σ_E^2, then $Z_j \sim \mathcal{N}(0, \sigma_L^2 + \sigma_E^2)$. Since $Y_i = S + Z_i$, we can apply Bayes's Rule to see that $E[S|y_i] = \hat{\alpha}\mu + (1 - \hat{\alpha})y_i$, where $\hat{\alpha} = (\sigma_L^2 + \sigma_E^2)/(\sigma_S^2 + \sigma_L^2 + \sigma_E^2)$. Since $E[X_j|y_i] = E[S|y_i]$, it follows that i's expectation of j's score continues to lie between her prior expectation (μ) and her private signal (y_i). Her expectation of her *own* score differs, however, because her signal also contains information about her own luck variable (L_i). Formally, since $Y_i = X_i + E_i$, $E[X_i|y_i] = \bar{\alpha}\mu + (1 - \bar{\alpha})y_i$, where $\bar{\alpha} = \sigma_E^2/(\sigma_S^2 + \sigma_L^2 + \sigma_E^2)$. Here, i's expectation about her own score also lies between her prior expectation (μ) and her private signal (y_i), but since $\bar{\alpha} < \hat{\alpha}$, we have that either

$$y_i < E[X_i|y_i] < E[X_j|y_i] < \mu$$

or

$$\mu < E[X_j|y_i] < E[X_i|y_i] < y_i.$$

In other words, i displays overplacement after high signals and underplacement after low signals.

To evaluate overestimation, note that, for this example, $E_{Y_i}[E[X_i|Y_i]|x_i] = \bar{\alpha}\mu + (1 - \bar{\alpha})E[Y_i|x_i]$, which equals $\bar{\alpha}\mu + (1 - \bar{\alpha})x_i$. Thus, agents' expected reports of $E[X_i|y_i]$ lie strictly between μ and x_i. If $\mu < x_i$, we observe underestimation in expectation, and if $x_i < \mu$, we observe overestimation in expectation. $\square$

As with overplacement, the result that overestimation depends on the realization of task simplicity does not obtain with every combination of prior beliefs. With a uniform prior over X_i, for example, agents will fully update their expected score to the realized signal regardless of the actual simplicity of the task, generating no over-

or underestimation.

2.3 Overprecision

If an agent's beliefs about her own score has higher precision (lower variance) than her actual distribution of scores (taking a frequentist's viewpoint), we say that she exhibits *overprecision*. Since our model of agents' inferences operates only on subjective beliefs without assuming those beliefs are empirically accurate, the presence of overprecision will not qualitatively affect the above results on overplacement and overestimation; however, the *level* of precision will affect the magnitudes of these effects. For example, consider an agent who exhibits excessive precision in her estimates of her own score. If this overprecision on X_i stems from overprecision in her prior over S, then she attributes her high score more to her own performance ('luck') and less to the task's simplicity, relative to someone with well-calibrated beliefs. Thus, she perceives less correlation between her own score and the scores of others, exacerbating the overplacement phenomenon when the task is easier than expected. On the other hand, if her overprecision is due to lower-than warranted variance in L_i, she will perceive more correlation than actually exists and her overplacement will be mitigated. In general, overprecision in S increases the magnitude of underestimation and overplacement after easy tasks and the magnitude of overestimation and underplacement after difficult tasks, while overprecision in L_i reduces these magnitudes.

In practice, we only observe beliefs over scores. Since we can not differentiate overprecision in S and overprecision in L_i, we cannot use the theoretical links between overprecision and the other types of overconfidence to validate or falsify this model; we can only measure and describe the observed patterns of overprecision and how it correlates with overplacement and overestimation.

2.4 Modifications and Extensions

The above theoretical model is, by design, simple and unsophisticated. Several embellishments could be added to make the model fit better certain real-world environments. The main thrust of our argument remains unchanged, however, if these modifications do not alter the conclusion that $E[X_j|x_i]$ and $E[X_i|y_i]$ lie between μ and x_i (at least, in expectation). In this section we detail three possible changes to the model and argue that these changes do not (necessarily) invalidate the main conclusions of the simple model.

2.4.1 Multiple-Dimensional Signals

Some tasks, such as exams, involve some uncertainty about the exact nature of the task that is revealed while the task is performed. According to the model above, a student taking an exam receives only a signal of how well she performed and can only make inferences about the test's true difficulty from that one signal. In some situations it may be appropriate to model the student as receiving a second signal that is directly related to the test difficulty. For example, discovering that a final exam's questions were taken from various homework problems assigned throughout the course may lead a student to increase her estimate of the average score for the entire class, regardless of she felt about her own performance.

We can model this possibility by assuming agents receive two signals while performing the task: an unbiased signal y_i of their actual performance, and a second unbiased signal r_i of the task difficulty, where r_i is a realization of $R_i = S + Q_i$ with $E[Q_i] = 0$. As before, we take expectations over the value of the signal since it is not observable to the experimenter. Thus, we calculate $E_{Y_i,R_i}[E[S|Y_i, R_i]|x_i, s]$. The following example demonstrates how the inclusion of this second signal does not qual-

itatively alter the above analysis. Intuitively, we expect that r_i equals μ on average when prior beliefs are unbiased. Adding a second signal that 'pulls' the posterior means for X_i and X_j toward μ on average does not change the prediction that these posterior means lie between μ and x_i; thus, the magnitude of the overplacement and overestimation effects may change, but the direction of the effects would not.

Example 1 (Continued) If $R_i = S + Q_i$ with $Q_i \sim \mathcal{N}(0, \sigma_L^2 + \sigma_E^2)$ then

$$E[S|r_i, y_i] = \frac{(\sigma_L^2 + \sigma_E^2)/2}{\sigma_S^2 + (\sigma_L^2 + \sigma_E^2)/2}\,\mu + \frac{\sigma_S^2}{\sigma_S^2 + (\sigma_L^2 + \sigma_E^2)/2}\left(\frac{r_i + y_i}{2}\right).$$

When taking the expectation of this expression over Y_i and R_i, we simply replace y_i with x_i and r_i with s. If prior beliefs are unbiased ($\mu = s$ on average) then the expected posterior mean of X_j is

$$\left(\frac{\sigma_L^2 + \sigma_E^2 + \sigma_S^2}{\sigma_L^2 + \sigma_E^2 + 2\sigma_S^2}\right)\mu + \left(\frac{\sigma_S^2}{\sigma_L^2 + \sigma_E^2 + 2\sigma_S^2}\right)x_i.$$

Thus, i's expectation of j's score lies (on average) between her prior mean μ and her actual score, leading to overplacement in expectation after high scores and underplacement in expectation after low scores.

2.4.2 Ability and Prior Overconfidence

The simple model does not incorporate the possibility of prior overconfidence since it is not needed to explain the observed hard-easy effect. In many situations, however, it may be more appropriate to assume agents' prior beliefs about their idiosyncratic component (L_i) is not mean-zero, perhaps due to past experience or to a true behavioral bias. Such beliefs could be incorporated by assuming $X_i = S + L_i + A_i$, where A_i represents i's prior ability level. This would be equivalent to $X_i = S + \hat{L}_i + E[A_i]$,

where $E[A_i]$ is the mean of A_i and $\hat{L}_i$ has a mean of zero. The only changes in the analysis of this model (relative to the case where $A_i = 0$) are that the 'luck' term may now have a larger variance (perhaps affecting the magnitude of overplacement and overestimation) and that the values of $E[S|x_i]$ and $E[S|y_i]$ (and, thus, $E[X_j|x_i]$ and $E[X_j|y_i]$) are shifted by $E[A_i]$. In other words, the effect of prior overconfidence is simply added to the results of the basic model; subjects' overplacement is *increased* after unexpectedly easy tasks and *reduced* after unexpectedly difficult tasks.

2.4.3 Non-Bayesian Overconfidence

The above mathematical arguments make generous application of Bayes's Rule, but the results may also hold for agents whose updating process is non-Bayesian. The only necessary component of the theory is the relative ordering of the prior mean, posterior expectation, and the observed score or signal. If a non-Bayesian subject exhibits the same ordering, then the resulting patterns of overprecision and overestimation will be the same as under Bayes's Rule. For example, suppose an agent has normally-distributed priors, as in Example 1, but her updating rule results in a posterior of the form $E[X_j|x_i] = \hat{\alpha}\mu + (1-\hat{\alpha})x_i$, where the actual weight $\hat{\alpha}$ differs from the value α required by Bayes's Rule. As long as $\hat{\alpha}$ remains between zero and one, the predictions of this section are essentially unchanged; only the magnitudes of overplacement and overestimation will change. Thus, the predictions apply to Bayesians and non-Bayesians alike, so long as the posterior mean lands between the prior mean and the observed score or signal.

An important consequence of this observation is that our experiment is not a direct test of the details of the model; instead, the experiment tests the *predictions* of the model. We cannot reject any other model that generates these same predictions. While this limits the power of our experiments, it also means that our ability to

generate patterns of overconfidence using a theoretical model devoid of behavioral biases is quite robust; the result obtains as long as agents have incomplete information about task difficulty and form posterior expectations in between prior expectations and received signals. In other words, the overconfidence result stems from the basic intuition of statistical inference rather than from various technical details of the model.

3 Experimental Design

Eighty-two undergraduate student participants were recruited from Carnegie Mellon University. Each participated on a computer terminal in the Center for Behavioral Decision Research laboratory. The experiment consisted of 18 rounds in which each participant completed a 10-item trivia quiz and reported various beliefs about their score and the scores of others.[10]

The timing of each round is broken into three phases. In the *ex-ante* phase, subjects know nothing of the content or difficulty of the upcoming quiz. After taking the quiz, subjects enter the *interim* phase in which they have experienced the quiz but do not yet know the correct answers, their score, or the scores of any other participants. In the *ex-post* phase, subjects have seen the correct answers and know their own score, but do not know the scores of others.

In each of the three phases, subjects are asked to submit a belief distribution about their own score on the quiz and a second distribution about the score of a randomly-selected previous participant (hereafter, RSPP). Subjects are not given any information about the RSPP, other than the fact that the RSPP completed the

[10]The quiz questions and answers are available in the supplemental appendix, along with the mean, median, and variance of the scores for each quiz. To experience the computerized experimental environment, visit [URL omitted for anonymity] and log in using Participant ID 0000.

same 18 quizzes at some prior date.[11] Each probability distribution consisted of eleven probabilities, one for each of the possible scores (zero through ten). Subjects were shown eleven moveable horizontal bars to represent these eleven probabilities, and could 'drag' each bar to represent the desired probability distribution.[12] Once a subject is satisfied with a particular distribution, she clicks a button to submit the reported distribution.

Specifically, the timing is as follows: Subjects in the *ex-ante* phase report a distribution for their own score, followed by a distribution for the score of the RSPP. They then complete the 10-item trivia quiz. In the *interim* phase (before learning their own score), subjects again report a distribution for their own score and a distribution for the RSPP's score. Subjects are then shown the correct answers and grade their own quizzes.[13] Finally, in the *ex-post* phase, each reports a distribution for the RSPP's score.[14]

Each subject in each period earns a payment for her performance on the trivia quiz of $\$25r$, where r is her percentile rank on the quiz relative to all previous participants, plus payments for each of the five reported distributions according to a quadratic scoring rule.[15] If subject i reports a distribution for her score of $\hat{p}_i = (\hat{p}_i(0), \hat{p}_i(1), \ldots, \hat{p}_i(10))$ and earns an actual score of x_i on the quiz, then her

[11]In early sessions, data from pilot sessions was used as the source for the RSPP. Subjects were not explicitly informed about the pool of subjects from which the RSPP was drawn.

[12]Initial bar positions were randomly set each period. Moving one bar caused the ten other bars to adjust proportionally to their current length such that the sum of the bars was continuously equal to 100 percent. Subjects proceeded at their own pace and could spend as much time as needed adjusting these bars.

[13]Since we did not verify subjects' actual scores during the experiment, they could have incorrectly reported their actual earned score. Upon checking the quizzes after the experiment, we found no such instances of blatant misreports and very few isolated instances of 'questionable' (misspelled or incomplete) answers being counted as correct. We did not remove these data from our analysis.

[14]In early sessions, data from pilot sessions was used to draw the score of the RSPP. Subjects were not given any information about the pool of past subjects from which the other score was drawn, including the number of such subjects.

[15]For the sake of computing the percentile rank r, participants were counted as having scored better than half and worse than half of those who had obtained the same score.

payment for the report is

$$1 + 2\hat{p}_i(x_i) - \sum_{k=0}^{10} \hat{p}_i(k)^2.$$

An identical formula is used for reports about the distribution of the RSPP's score. This quadratic scoring rule pays between zero and two dollars per report and induces risk-neutral expected utility maximizers to reveal their beliefs truthfully (see, e.g., Selten (1998)). Subjects were paid the sum of their earnings across the 18 rounds after the completion of the experiment.

In this setting, a subject can manipulate her quiz performance to increase the accuracy of her reported distributions. For example, a subject could intentionally score zero on the quiz and predict a score of zero with certainty in both reports about her own score. Given that subjects in practice were earning an average \$12.18 on the quiz and \$2.39 on the two reports about their own score, and scoring zero on the quiz would earn an average of \$2.54 on the quiz and \$4.00 on the two reports, only the most pessimistic subjects would find such a manipulation profitable. In practice, we do not observe these types of obvious manipulations with significant frequency.[16]

The 18 quizzes span six topics, each at three difficulty levels. The assignment of quizzes to the three difficulty levels (easy, medium, and difficult) was based on previous experience with the questions and on the subjective assessment of the authors.[17] The six topics were geography, movies, music, history, sports, and science. The 18 quizzes were randomly assigned to six blocks of three rounds each such that each block had one quiz of each difficulty level. The three difficulty levels were randomly ordered within each block, and the order in which each subject encountered the six blocks was randomized. This design allows for a relatively uniform distribution of

[16] Conditions on the beliefs necessary for manipulation to be profitable are explored in a working paper version of this manuscript, available from the authors upon request.

[17] Result 1 verifies that 'easy' quizzes were in fact easy, 'medium' quizzes were intermediate, and 'difficult' quizzes were difficult.

quiz difficulty levels across the 18 rounds while making it difficult for a subject to predict the difficulty or subject matter of an upcoming quiz.[18]

4 Results

For each subject in each round we observe five probability distributions: The subject's ex-ante and interim beliefs about her own score, and her ex-ante, interim, and ex-post beliefs about the score of the RSPP. We report the expected values of these distributions (averaged across all players and periods) for each quiz difficulty level in Table 2. Actual score averages appear in the table under the *ex-post* phase.

Before testing measures of overconfidence, we must first verify that our experimental design correctly incentives subjects to reveal the data needed to compare the results to the predictions of the theory. If subjects are manipulating their quiz performance to increase the accuracy of their predictions, for example, stated beliefs will reflect expectations about manipulations, not true abilities, and the Bayesian updating model would not apply. Although small manipulations in performance would be difficult to detect, large manipulations are fairly obvious. Scores on easy quizzes averaged 8.86 (out of 10) with a standard deviation of 2.17, so a subject scoring zero or one is most likely 'sandbagging' the quiz to make her performance more predictable. Of the 492 easy quizzes, we observe only 11 scores of zero or one.[19] Although these apparent manipulations exist, they constitute only about 2 percent of the easy quiz

[18]It is true that a difficult quiz is the least likely to appear immediately after a difficult quiz, for example. Since the ordering of difficulty levels within each block is randomized, there will be no systematic effect on the results unless there is an interaction between quiz difficulty and within-block ordering effects. Note that if blocks were not used then subjects might encounter a string of difficult quizzes after which they could become nearly certain that no additional difficult quizzes remain.

[19]These 11 low scores are due to 9 different subjects. Alternatively, we can check for manipulations by looking for extreme but correct ex-ante predictions. Subjects correctly reported an ex-ante expected score of zero or one in 21 out of 1476 quizzes and correctly reported an ex-ante expected score of nine or ten in 2 of 1476 quizzes. The latter can occur because subjects self-grade their quizzes and the accuracy of their grading is only checked after the experiment ends.

data. Since these data would likely weaken the fit with the model predictions, we do not discard them in our analyses.

4.1 The Four Main Results

We now demonstrate four main results using the data: First, the quizzes are well calibrated in the sense that subjects score higher (and correctly believe they score higher) on easy quizzes and score lower (and correctly believe they score lower) on difficult quizzes. Second, subjects do not enter the experiment with significant levels of overplacement. Third, subjects exhibit overplacement on easy quizzes and underplacement on difficult quizzes. Fourth, subjects exhibit underestimation on easy quizzes and overestimation on difficult quizzes. The first two results verify that the experimental setting is appropriate and the last two results verify the predictions of the theory.

These results are each demonstrated by regressions whose estimates and standard errors appear in Table 3. In each regression, an appropriate dependant variable is regressed against a full set of dummy variables indicating easy, medium, and difficult quizzes. Each regression was also run including dummy variables for block effects and all interactions between blocks and difficulty levels, but fewer than five percent of these block and interaction estimates are significant at the five percent level, so we omit them from subsequent analysis.[20] Since blocks act as a proxy for time effects such as experience or learning, we can also conclude that overall performance and performances within each difficulty level are all stable across the 18 periods.[21]

The first two regressions in Table 3 give the following result.

[20] The full regressions appear in the supplemental appendix. The significant block and interaction coefficients are Block 1×Difficult and Block 5×Easy in the regression of Score$-E^2$(Other), and Block 1×Difficult in the regression of E^1(Self)$-$Score.

[21] We treat each elicited belief distribution as an independent observation. Where possible, we also verify our results with non-parametric tests.

Result 1 *Scores are high on easy quizzes, low on difficult quizzes, and slightly above the overall average on medium quizzes. Subjects correctly perceive these differences immediately after taking the quiz.*

This result is important in verifying that the three difficulty levels produce significantly different scores. If all quizzes produced similar scores, then the experiment would not provide a powerful test of the hard-easy effect predicted by the theory. It is clear from column 2 of Table 3 that in fact scores vary greatly by difficulty level. The average score across all quizzes is 5.16, while scores on easy quizzes are 8.86 points on average and the average score on difficult quizzes is 0.69. The average score on medium quizzes is 5.93, meaning that medium quizzes tend to be closer in performance to easy quizzes than difficult quizzes. These differences are all highly significant.[22] The median and mode are both 10 for easy quiz scores, 0 for difficult quiz scores, and 7 for medium quiz scores.

The regression in column 3 of Table 3 can be used to verify that subjects correctly perceive the differences in difficulty after taking the quiz. Subjects' expectations of their own score are 3.29 points higher than the overall average of 5.36 after an easy quiz and 3.86 points lower after a difficult quiz. These shifts are highly significant. Note also that the shifts in beliefs are slightly smaller than the shifts in actual scores. This is predicted by the model; after taking an easy quiz and receiving a positive (unbiased) signal about her own score, a Bayesian subject expects that her score is not as high as the received signal because high signals are more likely to contain positive errors. A symmetric argument applies to difficult quizzes.

Result 2 *Subjects do not exhibit significant overplacement before taking quizzes.*

[22] Large-sample Mann-Whitney tests also confirm that the distribution of scores on medium quizzes is significantly different from that of difficult quizzes (z-stat $= 22.83$), and that scores on easy quizzes are significantly different from those on medium tests ($z = 17.08$).

A key difference between the theoretical model outlined above and the notion of overplacement as an intrinsic bias is that the above model assumes people are not necessarily overconfident *a priori*, but become overconfident after positive experiences with the task at hand. Comparing first-period ex-ante expected scores of self versus the RSPP, subjects on average report a higher expected score for themselves, though the difference is small and insignificant (Wilcoxon signed rank test p-value of 0.343). Of 82 subjects, 45 reported higher expectations about their own score than the RSPP. A simple binomial test cannot reject the hypothesis that subjects are as likely to exhibit overplacement as underplacement (p-value of 0.16). A regression of ex-ante overplacement on quiz difficulty (column 4 of Table 3) reveals no significant overplacement or underplacement for any difficulty level. Estimates are all insignificant if the same regression is run using lagged dummy variables, implying that subjects also do not exhibit significant overplacement before period t after taking an easy quiz in period $t-1$, for example.

Result 3 *Subjects exhibit overplacement after easy quizzes and underplacement after difficult quizzes. This is true whether or not subjects actually scored better than the randomly-selected previous participant.*

The remaining three regressions from Table 3 (columns 5, 6, and 7) test the predictions of Table 1. We examine two measures of overplacement: interim overplacement and *ex-post* overplacement. In the first measure, subjects are uncertain about their own scores; in the second, they are not. The regression in column 5 indicates that subjects exhibit significant overplacement in the interim phase after an easy quiz and significant underplacement after a difficult or medium quiz. Specifically, subjects expect to out-perform the RSPP by an average of 0.39 points after an easy quiz, but expect to be out-performed by an average of 1.45 points after difficult quizzes. The

result is similar in the ex-post phase; subjects exhibit overplacement by an average of 0.37 points after easy quizzes and underplacement by an average of 1.66 points after difficult quizzes.

Recall that scores on medium quizzes (as well as the associated interim and ex-post expectations) are significantly greater than the overall average of 5.16. Since these quizzes are 'slightly easy', we should expect to observe some degree of overplacement in the interim and ex-post phases. According to Table 3, the *opposite* result obtains: subjects exhibit slight underplacement on medium quizzes at both the interim phase (by 0.22 points) and ex-post phase (by 0.28 points). In terms of the percentage of subjects exhibiting overplacement, however, there is no significant pattern in the data and overplacement occurs roughly as frequently as underplacement (see column 7 of Table 4). This indicates that the slight underplacement on medium quizzes stems from the fact that, in practice, the magnitude of underplacement is larger than the magnitude of overplacement. This might occur, for example, if subjects' prior beliefs about the quiz difficulty are not symmetrically distributed.[23]

In the theory of Section 2, the link between quiz difficulty and overplacement stems from the assumption that the posterior expectation of others' scores ($E[X_j|x_i]$) lies between the prior mean (μ) and the realized score (x_i). In practice, this 'betweenness' condition is satisfied in 64.8 percent of quizzes.[24] This condition is stronger than necessary; the predicted pattern for overplacement also obtains if $E[X_j|x_i] < x_i$ when $x_i > \mu$ and $E[X_j|x_i] > x_i$ when $x_i < \mu$. This weaker sufficient condition is satisfied

[23]Subjects in an extensive pilot study who were asked to compare their score against the *median* score of the previous participants (rather than a randomly-selected previous participant) exhibited overplacement after taking the same medium quizzes (by a significant 0.26 points in the interim phase and an insignificant 0.12 points in the ex-post phase). Qualitatively, all other results were the same between the two studies, suggesting that the results for medium quizzes are not particularly robust.

[24]This assumes μ is the subjects' prior expectation of their own score. Using subjects' prior expectation of the RSPP's score, betweenness is satisfied in 71.1 percent of the quizzes.

in 80.1 percent of quizzes in our data.[25]

Assuming $E[X_j|x_i]$ is in fact a convex combination of the prior mean and the realized score, a simple linear regression of the elicited values of $E[X_j|x_i]$ against the observed values of x_i (with the constraint that $E[X_j|\mu] = \mu$, where $\mu = 5.364$ is the average prior expectation of one's own score) provides an estimate for the best-fitting parameters in the theory. A simple least-squares regression indicates that $E[X_j|x_i] = 0.387\mu + (1 - 0.387)x_i$ with a standard error of 0.012 on the coefficient. This line is plotted against the data in Figure I. Using the beta-binomial specification of Example 2, the resulting beta coefficients are $(\beta_1, \beta_2) = (3.39, 2.93)$, indicating a nearly symmetric prior over p with a mean of 0.5364 (since $\mu = 5.364$) and a skewness of only -0.095.[26]

The definition of overplacement used in this paper requires only that subjects believe they will score higher than the RSPP; it does not require that this belief be incorrect. In Table 4, we separate the interim and *ex-post* data into those observations in which subjects actually did out-perform the RSPP and those in which they did not, allowing us to examine whether overplacement is generally consistent with actual outcomes. In the interim phase after easy quizzes, for example, 44.5 percent of subjects expected that they had outperformed the RSPP (with a tolerance of $\pm 1/2$ since expectations are real-valued and actual scores are integer-valued). Of those subjects, only 35.6 percent were correct. Looking across all easy and difficult quizzes and both the interim and ex-post phases, the expected ranking predicted by Result 3 (overplacement for easy quizzes, underplacement for difficult quizzes) is the modal ranking. In each of those six cases, no more than 41.1 percent of the subjects had correct rankings of their expectations. Thus, the predicted overplace-

[25] Using subjects' prior expectation of the RSPP's score as μ, the number increases to 84.4 percent.
[26] Clearly, a more accurate model would allow for parameter heterogeneity across subjects.

ment/underplacement pattern is the modal observation even though these beliefs are inaccurate the majority of the time.

Result 4 *Subjects exhibit underestimation after easy quizzes and overestimation after difficult quizzes.*

Our measure of overestimation is the difference between a subjects' expected score in the interim phase (after taking the quiz) and their actual score. The final regression from Table 3 confirms that, on average, subjects underestimate the score by 0.22 points after easy quizzes and overestimate their score by 0.81 points after difficult quizzes. Overestimation on medium quizzes is essentially absent.

Recall from Section 2 that the standard predictions may fail to hold when prior beliefs are excessively bimodal. In fact, scores on the quizzes are highly bimodal, with nearly half of all observed scores equal to zero or ten. As we discuss in the following Section, however, subjects' beliefs do not exhibit this extreme bimodality, even in later periods where subjects have experienced up to seventeen prior quizzes. In other words, the fit between the data and the predictions was apparently improved by the fact that subjects failed to recognize this bimodality.

4.2 Overprecision

Recall that overprecision occurs when agents' belief distributions have lower variance than the distribution of actual outcomes. We consider overprecision about one's own score and about others' scores at the *ex-ante* stage and overprecision about others' scores at the interim and *ex-post* stages. Any measurement of interim or *ex-post* overprecision about one's own score faces the problem that beliefs are conditional on private information (Y_i in our model), which cannot be observed. Thus, we cannot construct the appropriate empirical distribution that conditions on this information

against which the reported beliefs should be compared.

The distribution of actual scores across all difficulty levels is highly bimodal, with 49.8% of all quiz scores equalling zero or ten.[27] On average, the combined weight subjects assigned to these two scores in their ex-ante distributions is 16.4% for their own score and 13.3% for the scores of others, both of which are less than the 18.2% weight assigned by a uniform distribution. By the final period, these average combined weights increase to only 24.7% and 23.1% for own and others' scores, respectively – still significantly below the true distribution.[28] Since subjects fail to recognize the bimodality of scores, their reported ex-ante variances (across all periods) are lower than the true variance by an average of 10.60 for their own scores and 9.99 for the scores of others.[29]

The overprecision of subjects' interim estimates of others' scores can be tested by comparing the variance of reported beliefs to the distribution of actual scores on that particular quiz.[30] On average, the actual variance is 2.00 units larger than the variance of the reported beliefs. After subjects learn their own scores, this difference increases insignificantly to 2.18.[31] Both are significantly greater than zero, indicating the presence of overprecision in interim beliefs about others' scores.

Recall from Section 2.3 that overprecision should be correlated with the magnitudes of overplacement and overestimation, though the sign of this correlation cannot

[27]This bimodality raises the issue of 'floor' and 'ceiling' effects in our design, where beliefs are simply truncated by the maximum and minimum possible score. Such an explanation would only partially explain our results, and Moore and Small (in press) see similar patterns of results without any upper or lower bounds on performance.

[28]Since we are examining *ex-ante* beliefs measured before the quiz questions were revealed, these distributions are not conditioning on quiz difficulty.

[29]Wilcoxon tests verify that these levels of overprecision are significantly different from each other ($p < 0.001$) and greater than zero ($p < 0.001$). These differences in variances drop to 9.16 and 8.67 in the final period, which are both significantly positive ($p < 0.001$) but not significantly different ($p = 0.298$).

[30]Since these are interim beliefs, they are conditional on quiz difficulty.

[31]The p-value of the Wilcoxon-Mann-Whitney test is 0.149.

be predicted without knowing the underlying distributions for S and L_i. We explore this correlation in the data using Spearman rank-order correlation coefficients. From Table 5, the correlations between ex-ante overprecision and interim overplacement are all significantly negative, and the correlations between ex-ante overprecision and interim overestimation are all positive and insignificant.[32] The first result suggests that subjects' overprecision stems from overprecision in the idiosyncratic component of their score ('luck') rather than in the common component of their score ('simplicity'). Since subjects' private signals are comprised of the quiz difficulty, their individual luck, and the signal error, the second result indicates that the overprecision in luck must be offset by underprecision in the signal errors. In other words, these results suggest that subjects are overestimating the correlation between their scores and the scores of others, but underestimating the quality of their private signals about their own scores.

5 Previous Literature

A variety of other papers explore overconfidence under differing assumptions. Overconfidence has previously been modeled as stemming from other judgement biases (Matthew Rabin and Joel Schrag (1999)) or from rational choice when beliefs are flexible and overconfidence affects other interactions and motivations (e.g., Roland Benabou and Jean Tirole (2002); Juan D. Carillo and Thomas Mariotti (2000); or Oliver Compte and Andrew Postlewaite (2004)). Other research (March and Shapira (1987); Odean (1998); Daniel, Hirshleifer & Subrahmanyam (2001); Malmendier & Tate (2005)) has demonstrated how overconfidence can lead to meaningful economic consequences. Our paper differs from these in that we assume no biases in judgement

[32] Here, ex-ante overprecision is the variance of the distribution of all scores minus the variance of the subject's ex-ante distribution of her own score.

nor do we presume that overconfidence is beneficial in either the task at hand or future interactions; instead, we show how overconfidence (and underconfidence) can arise from Bayesian inference about others' performances after observing a signal of one's own performance.

Eric Van den Steen (2004) (hereafter VdS) provides a Bayesian model of overplacement driven by assuming heterogeneous priors. In his setup, agents choose their most-preferred action from a set $A = \{a_1, a_2, \ldots a_n\}$. Actions will either succeed or fail and each person i believes each action a_n will succeed with probability p_n^i. Suppose person 1 believes a_1 has the highest probability of success and person 2 believes a_2 has the highest probability of success. Clearly, person 1 picks a_1 and person 2 picks a_2. If the two agents' priors are independent (meaning person i learns nothing by observing that person j chose $a_j \neq a_i$), then each person will conclude that the other made an inferior choice. Thus, each will exhibit overplacement.

Luis Santos-Pinto and Joel Sobel (2005) (hereafter SP&S) provide a model that generalizes VdS in which agents choose the optimal skills to acquire in order to maximize their overall ability. The basic intuition of their model is captured by the following (simplified) example: Suppose person 1 and person 2 are trying to maximize different functions, denoted $f_1(x)$ and $f_2(x)$, respectively. Think of x as a vector of skills and $f_i(x)$ as i's perception of his ability level at some task, given skills x.[33] Clearly, the optimal choices (x_1^* and x_2^*) are likely to differ between the two agents, in which case we should expect that $f_1(x_1^*) > f_1(x_2^*)$ and $f_2(x_1^*) < f_2(x_2^*)$. Thus, if person 1 evaluates person 2's choice using f_1, person 1 will conclude that he has made the better choice and therefore has the higher ability. In this way, both agents can exhibit overplacement.[34]

[33] To represent the VdS model as a special case of the SP&S environment, think of x as an n-vector with $x_k = 1$ if action a_k is chosen and zero otherwise, and let $f_i(x) = \sum_{k=1}^{n} x_k\, p_k^i$.

[34] The full SP&S model is significantly more complex; agents aim to maximize $f(x, \lambda_i)$ subject

In both the VdS and SP&S models, overplacement stems from agents using their own objective function to compare their choices against the choices of others. By contrast, our model could be described as a reduced-form version of a model in which agents share a common objective function, but its parameters are unknown and agents' beliefs about the parameters may vary. Letting 'easy' tasks be those for which the unknown parameters have low variance and 'difficult' tasks be those for which the parameters have high variance, an agent who observes an unexpectedly high value of the objective function will believe that the task was more likely easy (thus, everyone's beliefs about the parameters are likely to have less error) but also that her own prior beliefs about the parameters were likely more accurate than others (if only by luck). She therefore concludes that others will also score well, but not as well as she.

A second difference between VdS and SP&S and the current model is that the former imply prior overplacement but the latter does not. Our experiments reveal no significant prior overplacement. Although the results of Camerer and Lovallo (1999) (henceforth C&L) are frequently cited as evidence in favor of overconfidence, they too find no strong evidence for *prior* overconfidence. In their setting, subjects choose whether or not to enter a market in which entrants' profitability depends on their assigned 'rankings'. They study three treatments: in the first, rankings are randomly chosen after the entry decisions are made. In the second and third, entrants are ranked based on their performance in a trivia quiz. Subjects in the third treatment are told before choosing to participate that their payoff will depend on their score on a trivia quiz, while subjects in the second treatment are not. Thus, the third

to $x \in A(I_i)$ where λ_i and I_i are individual-specific parameters drawn from a known distribution. Person i then compares $f(x_i^*, \lambda_i)$ against $f(x_j^*, \lambda_i)$ and concludes that x_i^* was a (weakly) better choice than x_j^*. SP&S then derive conditions under which the fraction of individuals who believe they are in the top p-cile of ability levels in the population is (weakly) greater than p. This condition defines overplacement at the population level.

treatment introduces a self-selection bias that is likely to favor higher scores on the quizzes.

C&L find that subjects over-enter (relative to the risk-neutral equilibrium prediction) in the self-selection treatment, apparently because they fail to recognize that their competitors also self-selected into the experiment.[35] Comparing the first two treatments reveals that subjects enter more frequently when rankings are quiz-based, but the aggregate level of entry is at or below the risk-neutral equilibrium prediction in both cases.[36] Thus, the C&L study highlights the role of competition (and beliefs about one's competitors) in generating overconfidence but the lack of prior overconfidence in the absence of the self-selection bias is consistent with our observations.

Although we do not address the role of competition in the current study, the particular information structure we assume has been used extensively in game theoretic models. In some cases, the results are indicative of the kind of 'rational overplacement' we describe. For example, Carl Shapiro (1986) assumes that firms in an oligopoly market have constant marginal costs and that each firm's marginal cost is drawn from a common prior distribution with imperfect positive correlation between firms. With little or no correlation, a firm with unexpectedly low marginal costs will conclude that its competitors' marginal costs will not be as low as its own. In the symmetric Bayes-Nash equilibrium of the game, this low-cost firm will produce a fairly large quantity because it perceives a significant cost advantage ('overplacement'). If, on the other hand, costs are highly correlated, the low-cost firm believes other firms are likely to have similarly low costs, and so its equilibrium output is reduced.[37] In other

[35]An alternative explanation is that those subjects who self-select into the experiment are those who have had unexpectedly high scores on previous trivia quizzes and consequently increased their expectation of their own trivia-quiz ability (see Section 2.4.2).

[36]Lower entry rates in the randomly-assigned rankings treatment may be attributable to risk-averse subjects (correctly) believing that payoffs in the quiz-based rankings treatment have a lower variance than in the randomly-assigned rankings treatment.

[37]To see this from Shapiro's paper, simply compare the Bayesian equilibrium with imperfect

words, low-cost firms produce more (and high-cost firms produce less) in exactly those situations where they exhibit a greater degree of overplacement (or underplacement).

6 Discussion

This paper accomplishes two goals. First, we demonstrate that patterns of overconfidence can be generated theoretically without assuming any behavioral biases. This is achieved through a simple model of incomplete information regarding task difficulty. Second, we carefully distinguish between the various forms of overconfidence and clearly show their relationship in an experimental study. The observed relationship is exactly that predicted by the simple theoretical model.

There have been a number of recent economic models that have attempted to explain how rational Bayesian agents could display overconfidence (for example, Benabou and Tirole (2002); Ronit Bodner and Drazen Prelec, (2003); Rabin and Schrag (1999); Van den Steen (2004)). Although overconfidence has been widely observed, none of these models can parsimoniously account for the evidence from the present experiment because they do not predict the systematic *under*confidence observed in these results.

We believe that the tendency for studies to find examples of overconfidence more frequently than underconfidence is an artifact of methodology. For example, several studies confound overestimation with overprecision (Joseph W. Alba and J. Wesley Hutchinson, 2000, Baruch Fischhoff *et al.*, 1977), making it impossible to determine the degree to which each is responsible for the result. These studies measure people's confidence about the correctness of individual answers. At the individual item level it is impossible to distinguish overestimation of one's absolute performance from ex-

correlation ($\rho < 1$) to the Bayesian equilibrium with perfect correlation ($\rho = 1$), which is the Nash equilibrium of the standard game with complete information.

cessive precision in the accuracy of private beliefs — since beliefs are binomial, means and variances are closely related. These results, therefore, cannot provide unambiguous evidence for the existence of systematic overestimation. Our results lead us to speculate that these prior results may be more attributable to overprecision than to overestimation.

Additionally, overplacement and overestimation have not occurred in the same studies. Those studies in which people overestimate their absolute performance the most have tended to focus on contexts in which performance is low and success is rare (Peter Juslin *et al.*, (2000); Malmendier and Tate (2005); Neil D. Weinstein, (1980)). Those studies in which people overplace their relative ranking the most have tended to focus on contexts in which performance is high and success is likely (College Board, (1977); Kruger (1999); David M. Messick *et al.*, (1985); Ola Svenson, (1981)).

Although our model is Bayesian, there is ample reason to question whether people actually make judgments according to Bayes's Rule. Under some circumstances, people appear to neglect priors (such as base rates), overweighing recent evidence (David M. Grether, (1980) and (1990)). Under other circumstances, people appear too conservative, overweighing priors and neglecting useful new evidence (Ward Edwards, (1968); Richard D. McKelvey and Talbot Page, (1990)). Which of these errors people commit depends on the order and form in which they acquire information (Robin M. Hogarth and Hillel J. Einhorn, (1992); Gary L. Wells, (1992)). What is important for our purposes here, however, is that although people are imperfect Bayesians, they rarely abandon Bayesian logic completely. All that is necessary for our predictions to hold is that people's estimates of others lie between their priors and their beliefs about themselves. That is, that their beliefs are a convex combination of their priors and their new information. Both the results from our experiment and from other research bear this assumption out.

References

Benabou, R. and Tirole, J.: 2002, Self-confidence and personal motivation, *Quarterly Journal of Economics* **117**(3), 871–915.

Berger, J. O.: 1980, *Statistical Decision Theory and Bayesian Analysis, 2nd ed.*, Springer-Verlag, New York.

Board, C.: 1977, *Student descriptive questionnaire*, Educational Testing Service, Princeton, NJ.

Bodner, R. and Prelec, D.: 2003, Self-signaling and diagnostic utility in everyday decision making, *in* I. Brocas and J. D. Carillo (eds), *The psychology of economic decisions, volume 1: Rationality and well-being*, Oxford University Press, Oxford, pp. 105–123.

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

Carillo, J. D. and Mariotti, T.: 2000, Strategic ignorance as a self-disciplining device, *The Review of Economic Studies* **67**(3), 529–544.

Casella, G. and Berger, R. L.: 2002, *Statistical Inference*, second edn, Duxbury Press, Pacific Grove, PA.

Chambers, C. P. and Healy, P. J.: 2007, Updating toward the signal. http://healy.econ.ohio-state.edu.

Compte, O. and Postlewaite, A.: 2004, Confidence-enhanced performance, *American Economic Review* **94**(5), 1536–57.

Daniel, K. D., Hirshleifer, D. and Subrahmanyam, A.: 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* **56**(3), 921–65.

Diaconis, P. and Ylvisaker, D.: 1979, Conjugate priors for exponential families, *The Annals of Statistics* **7**(2), 269–281.

Edwards, W.: 1968, Conservatism in human information processing, *in* B. Kleinmuntz (ed.), *Formal representation of human judgment*, Wiley, New York, pp. 17–52.

Grether, D. M.: 1980, Bayes' rule as a descriptive model: The representative heuristic, *Quarterly Journal of Economics* **95**, 537–557.

Grether, D. M.: 1990, Testing bayes rule and the represetativeness heuristic: Some experimental evidence, *Journal of Economic Behavior and Organization* **17**, 31–57.

Hogarth, R. M. and Einhorn, H. J.: 1992, Order effects in belief updating: The belief-adjustment model, *Cognitive Psychology* **24**, 1–55.

Juslin, P., Winman, A. and Olsson, H.: 2000, Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect, *Psychological Review* **107**(2), 384–396.

Kruger, J.: 1999, Lake wobegon be gone! the 'below-average effect' and the egocentric nature of comparative ability judgments, *Journal of Personality and Social Psychology* **77**(2), 221–232.

Malmendier, U. and Tate, G.: 2005, Ceo overconfidence and corporate investment, *Journal of Finance* **60**(6), 2661–2700.

March, J. G. and Shapira, Z.: 1987, Managerial perspectives on risk and risk taking, *Management Science* **33**(11), 1404–18.

McKelvey, R. D. and Page, T.: 1990, Public and private information: An experimental study of information pooling, *Econometrica* **58**, 1321–1339.

Messick, D. M., Bloom, S., Boldizar, J. P. and Samuelson, C. D.: 1985, Why we are fairer than others, *Journal of Experimental Social Psychology* **21**, 480–500.

Moore, D. A. and Kim, T. G.: 2003, Myopic social prediction and the solo comparison effect, *Journal of Personality and Social Psychology* **85**(6), 1121–1135.

Moore, D. A. and Small, D. A.: forthcoming, Error and bias in comparative social judgment: On being both better and worse than we think we are, *Journal of Personality and Social Psychology* .

Moore, D. A. and Small, D. A.: in press, When it's rational for the majority to believe that they are better than average, *in* J. I. Krueger (ed.), *Rationality and Social Responsibility: Essays in Honor of Robyn M. Dawes*, Erlbaum, Mahwah, NJ.

Odean, T.: 1998, Volume, volatility, price, and profit when all traders are above average, *Journal of Finance* **53**(6), 1887–1934.

Odean, T.: 1999, Do investors trade too much?, *American Economic Review* **89**(5), 1279–1298.

Rabin, M. and Schrag, J. L.: 1999, First impressions matter: A model of confirmatory bias, *Quarterly Journal of Economics* **114**(1), 37–82.

Roll, R.: 1986, The hubris hypothesis of corporate takeovers, *Journal of Business* **59**, 197–216.

Santos-Pinto, L. and Sobel, J.: 2005, A model of positive self-image in subjective assessments, *American Economic Review* **95**, 1386–1402.

Selten, R.: 1998, Axiomatic characterization of the quadratic scoring rule, *Experimental Economics* **1**(1), 43–61.

Shapiro, C.: 1986, Exchange of cost information in oligopoly, *The Review of Economic Studies* **53**(3), 433–446.

Svenson, O.: 1981, Are we less risky and more skillful than our fellow drivers?, *Acta Psychologica* **47**, 143–151.

Van den Steen, E.: 2004, Rational overoptimism, *American Economic Review* **94**(4), 1141–1151.

Weinstein, N. D.: 1980, Unrealistic optimism about future life events, *Journal of Personality and Social Psychology* **39**(5), 806–820.

Wells, G. L.: 1992, Naked statistical evidence of liability: Is subjective probability enough?, *Journal of Personality and Social Psychology* **62**(5), 739–752.

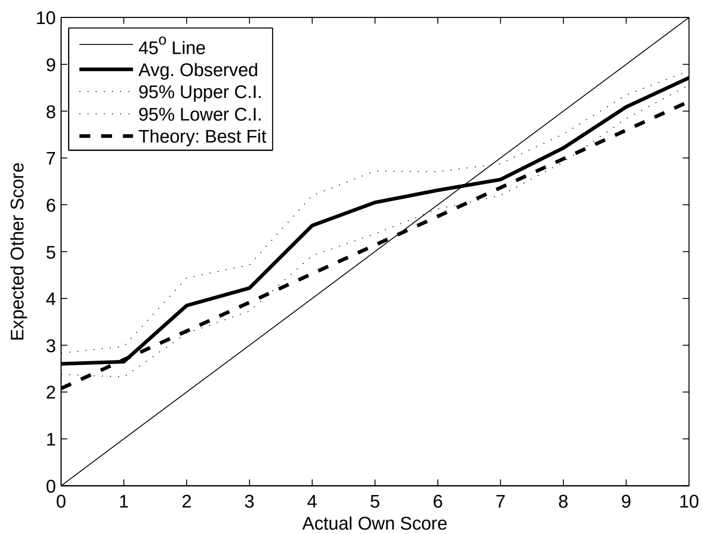

Figure I: Average reported expectation of others' scores versus own score.

Task Difficulty	Relative Performance	Absolute Performance
Easy	Overplacement	Underestimation
Difficult	Underplacement	Overestimation

Table 1: The key predictions of the Bayesian model.

			Block						
Difficulty	Phase	Distribution	1	2	3	4	5	6	Overall
Easy	Ex-Ante	Own Score	4.737	5.072	5.301	5.129	5.301	5.412	5.159
			(0.20)	(0.18)	(0.15)	(0.17)	(0.17)	(0.19)	(0.07)
		Other's Score	4.794	5.167	5.192	5.206	5.337	5.204	5.150
			(0.16)	(0.14)	(0.14)	(0.12)	(0.12)	(0.10)	(0.05)
	Interim	Own Score	8.407	8.921	8.808	8.677	8.395	8.657	8.644
			(0.23)	(0.18)	(0.25)	(0.26)	(0.27)	(0.24)	(0.10)
		Other's Score	8.018	8.286	8.300	8.426	8.219	8.306	8.259
			(0.17)	(0.16)	(0.18)	(0.16)	(0.18)	(0.19)	(0.07)
	Ex-Post	Actual Score	8.805	9.122	9.098	8.781	8.500	8.878	8.864
			(0.23)	(0.17)	(0.23)	(0.26)	(0.28)	(0.25)	(0.10)
		Other's Score	8.186	8.496	8.679	8.572	8.515	8.525	8.495
			(0.18)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.07)
Medium	Ex-Ante	Own Score	5.197	5.413	5.224	5.267	5.247	5.482	5.305
			(0.17)	(0.17)	(0.15)	(0.19)	(0.19)	(0.18)	(0.07)
		Other's Score	5.249	5.307	5.254	5.275	5.373	5.586	5.341
			(0.16)	(0.15)	(0.14)	(0.13)	(0.15)	(0.14)	(0.06)
	Interim	Own Score	5.882	6.325	5.684	5.824	5.946	5.922	5.930
			(0.32)	(0.32)	(0.36)	(0.35)	(0.35)	(0.35)	(0.14)
		Other's Score	6.007	6.284	5.964	6.276	6.063	6.312	6.151
			(0.24)	(0.22)	(0.24)	(0.23)	(0.25)	(0.25)	(0.10)
	Ex-Post	Actual Score	5.963	6.183	5.573	5.720	5.927	5.476	5.807
			(0.34)	(0.35)	(0.37)	(0.36)	(0.33)	(0.36)	(0.14)
		Other's Score	6.197	6.434	5.874	6.207	6.204	6.333	6.208
			(0.23)	(0.20)	(0.26)	(0.25)	(0.23)	(0.24)	(0.10)
Difficult	Ex-Ante	Own Score	6.415	5.699	5.514	5.364	5.368	5.419	5.630
			(0.19)	(0.18)	(0.15)	(0.17)	(0.19)	(0.20)	(0.08)
		Other's Score	6.467	5.645	5.386	5.250	5.365	5.377	5.582
			(0.17)	(0.13)	(0.13)	(0.12)	(0.14)	(0.11)	(0.06)
	Interim	Own Score	1.688	1.560	1.452	1.370	1.407	1.542	1.503
			(0.22)	(0.19)	(0.18)	(0.17)	(0.19)	(0.21)	(0.08)
		Other's Score	3.426	2.946	3.141	2.746	2.633	2.814	2.951
			(0.23)	(0.20)	(0.23)	(0.20)	(0.20)	(0.21)	(0.09)
	Ex-Post	Actual Score	0.463	0.732	0.451	0.488	0.634	0.646	0.569
			(0.10)	(0.13)	(0.09)	(0.10)	(0.12)	(0.11)	(0.04)
		Other's Score	2.834	2.542	2.441	2.049	2.049	2.205	2.353
			(0.24)	(0.20)	(0.23)	(0.20)	(0.18)	(0.20)	(0.09)

Table 2: Averages (and standard errors) of expected values of reported belief distributions.

Result	1		2	3		4
Dependant Variable	Score	E^1(Self)	E^0(Self) $-E^0$(Other)	E^1(Self) $-E^1$(Other)	Score $-E^2$(Other)	E^1(Self) $-$Score
Easy	**8.864**	**8.644**	0.008	**0.385**	**0.369**	**-0.219**
	(83.48)	(79.68)	(0.14)	(4.02)	(3.58)	(-3.98)
Medium	**5.925**	**5.930**	-0.036	**-0.221**	**-0.284**	0.006
	(55.80)	(54.67)	(-0.58)	(-2.30)	(-2.76)	(0.10)
Difficult	**0.693**	**1.503**	0.048	**-1.448**	**-1.660**	**0.810**
	(6.53)	(13.86)	(0.78)	(-15.12)	(-16.14)	(14.69)

Table 3: Dummy variable regressions demonstrating the four main results. Superscripts indicate ex-ante expectations (E^0), interim expectations (E^1), or ex-post expectations (E^2), and 'Score' refers to the subject's own score. Bold-faced entries are significant at the 5% level.

Phase	Quiz Difficulty	Expected Ranking	Actual Ranking Self<Other	Self=Other	Self>Other	Total	Correct Beliefs
Interim	Easy	E[Self]<E[Other]−1/2	15.2%	1.8%	0.6%	17.7%	86.2%
		E[Self]≈E[Other]	7.5%	17.5%	12.8%	37.8%	46.2%
		E[Self]>E[Other]+1/2	7.5%	21.1%	15.9%	44.5%	35.6%
	Medium	E[Self]<E[Other]−1/2	27.4%	2.6%	7.9%	38.0%	72.2%
		E[Self]≈E[Other]	8.9%	1.2%	13.0%	23.2%	5.3%
		E[Self]>E[Other]+1/2	9.3%	5.5%	24.0%	38.8%	61.8%
	Difficult	E[Self]<E[Other]−1/2	25.0%	27.4%	10.8%	63.2%	39.5%
		E[Self]≈E[Other]	8.7%	8.3%	7.1%	24.2%	34.5%
		E[Self]>E[Other]+1/2	2.0%	3.0%	7.5%	12.6%	59.7%
ExPost	Easy	E[Self]<E[Other]−1/2	17.1%	1.4%	1.6%	20.1%	84.8%
		E[Self]≈E[Other]	5.1%	18.5%	12.6%	36.2%	51.1%
		E[Self]>E[Other]+1/2	8.1%	20.5%	15.0%	43.7%	34.4%
	Medium	E[Self]<E[Other]−1/2	29.3%	2.4%	8.1%	39.8%	73.5%
		E[Self]≈E[Other]	7.3%	2.0%	9.6%	18.9%	10.8%
		E[Self]>E[Other]+1/2	9.1%	4.9%	27.2%	41.3%	66.0%
	Difficult	E[Self]<E[Other]−1/2	28.0%	28.5%	11.8%	68.3%	41.1%
		E[Self]≈E[Other]	6.7%	8.9%	6.5%	22.2%	40.4%
		E[Self]>E[Other]+1/2	1.0%	1.4%	7.1%	9.6%	74.5%

Table 4: Frequency of subjects exhibiting various rankings of own score vs. other's score, compared to actual score rankings.

Difficulty	Overplacement	Overestimation
Easy	-0.137	0.073
	(0.002)	(0.107)
Medium	-0.164	0.029
	(< 0.001)	(0.517)
Difficult	-0.177	0.021
	(< 0.001)	(0.645)

Table 5: Spearman correlation coefficients (and p-values) between ex-ante overprecision and interim overconfidence measures.