Bayesian Overconfidence*

Paul J. Healy† Don A. Moore‡

Abstract

Results from social psychology indicate that the overconfidence bias is not universal; underconfidence is often observed and correlated with task difficulty. We organize these various findings with a single experiment that tests three distinct forms of overconfidence: *overestimation* of one's actual performance, *overplacement* of one's performance relative to others, and *overprecision* in one's beliefs about private information. Task difficulty is negatively correlated with overestimation but positively correlated with overplacement. We construct a simple model of uncertainty about task difficulty that explains the data and unifies the three notions of overconfidence.

JEL Classification: C91; D01; D83

Keywords: Overconfidence; Bayes's rule; Hard-easy effect

---

*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.

†Corresponding author. Department of Economics, The Ohio State University, 1945 North High Street, Columbus, OH 43210. Phone: 1-614-247-8876. Fax: 1-614-292-3906. Email: healy.52@osu.edu.

‡Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.

1. Introduction

In recent decades social scientists have used overconfidence as an explanation for various phenomena relevant to economics, including costly delays in labor negotiations, excessive litigation, excessive market entry and subsequent entrepreneurial failure, excessive stock trading and subsequent market volatility, overinvestment by CEOs on internal projects, and even the initiation and prolonging of wars between countries (see, e.g., Neale and Bazerman 1985; March and Shapira 1987; Roll 1986; Camerer and Lovallo 1999; Odean 1998, 1999; Daniel, Hirshleifer, and Subrahmanyam 1998, 2001; Barber and Odean 2001; Statman, Thorley, and Vorkink 2006; Glaser and Weber 2007; Malmendier and Tate 2005; Howard 1983; and Johnson 2004).[1]

Despite claims that overconfidence is both prevalent and robust,[2] two pieces of evidence have emerged in the psychology literature that question the universality of the phenomenon. First, *under*confidence is observed in certain situations, and its occurrence is apparently linked to task difficulty. Second, the finding of overconfidence versus underconfidence depends critically on how one defines the concept. One possible form of overconfidence, which we call *overestimation*, occurs when a person's estimate of her own performance is greater than her actual performance. Researchers typically find overestimation on difficult tasks but find *under*estimation on easy tasks (see Lichtenstein and Fischhoff 1977 or Erev, Wallsten, and Budescu 1994, for example). According to Griffin and Tversky (1992, p. 427)—who also observe this result—"The difficulty effect is one of the most consistent findings in

[1] There is some debate, however, about whether behavioral biases can affect market outcomes in settings where some traders are rational; see Garcia, Sangiorgi, and Urosevic (2007), for example.

[2] De Bondt and Thaler (1995) claim that "perhaps the most robust finding in the psychology of judgement is that people are overconfident". According to Plous (1993, p. 217), "No problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence".

the calibration literature...". An alternative form of overconfidence, which we call *overplacement*, is the ranking of one's own performance above the performance of others.[3] When subjects are asked to rank themselves against others the link with task difficulty reverses; most studies use relatively easy tasks and find evidence of overplacement (Svenson, 1981, e.g.), but Kruger (1999); Moore and Small (2007); and Windschitl, Kruger, and Simms (2003) show that difficult tasks produce underplacement. Thus, easy tasks apparently lead to overplacement and underestimation while difficult tasks lead to underplacement and overestimation.[4]

The overconfidence literature has failed to provide a complete picture of these phenomena largely because authors either focus on only one type of overconfidence or muddle multiple definitions together. In this paper we piece together the puzzle of overconfidence by working within a single experimental and theoretical framework in which the various definitions of overconfidence can be considered simultaneously. We begin by formally defining three distinct and operable notions of overconfidence. We then measure each in an experiment with trivia quizzes and incentive-compatible belief elicitation. Finally, we show how the observed correlations among overconfidence measures and task difficulty can be explained by a simple Bayesian model where agents learn about a task's difficulty through experience. In addition to uniting and explaining the previous results on overestimation and overplacement, we discuss and explore the relation between these results and *overprecision* (perceiving more precision in private information than is warranted), which is commonly used to explain anomalies in the finance literature (Odean, 1999, e.g.).

[3]Note that under this definition even perfectly calibrated individuals can exhibit overplacement; a paradox occurs only when a large majority of individuals simultaneously hold such beliefs.

[4]Moore and Kim (2003) and Moore and Small (2007) have documented (but not explained) these connections between overestimation, overplacement, and task difficulty, and our work builds off of these results.

The operative assumptions in our model[5] are that tasks have unknown difficulties and that agents believe that their performance (and the performance of others) is determined by the overall difficulty of the task plus an individual-specific component that quantifies individual performance net of average overall performance. After performing a task, each agent is asked to estimate her own performance and the performance of a randomly-selected other participant. The result of Bayesian inference in this setting is that, after an unexpectedly easy task, a subject 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 (overplacement) and overconfidence about one's score relative to one's true score (overestimation); much of the previous literature confuses these two distinct concepts. Using this terminology, the model predicts overplacement and underestimation after unexpectedly easy tasks and underplacement and overestimation after unexpectedly difficult tasks. Furthermore, overprecision is predicted to be correlated with the other two forms of overconfidence, though the sign of the correlation depends critically on the unobservable source of the overprecision; this is detailed in Section 5.3.

The intuition behind 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 your competitors will also do well but that you have outperformed them by some degree.

As an example, suppose every manager in a particular emerging industry agrees

[5] The word "model" is perhaps an overstatement; our "model" is simply an application of Bayes's rule.

that the expected per-unit cost of a new product is $10 per unit. After production begins, however, each firm privately observes an actual per-unit 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 median) in exactly the same way that a large majority of drivers can believe that their driving ability is above the median (Svenson, 1981).[6] Had the firm's actual costs been higher than the prior estimate (ranging from $11 to $13, for example), the result would reverse and all managers would exhibit underplacement (believing its costs to be higher than the median). Thus, we can generally conclude that overplacement is more likely after unexpectedly easy tasks and underplacement is more likely after unexpectedly difficult tasks.

The logic for overestimation is similar; suppose now that firms build a prototype product before opening their production lines, and the cost of the prototype serves as an unbiased signal of the actual per-unit cost under full-scale production. If the range of prototype costs is lower than expected then 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 firm has realized a prototype cost of $8. But this firm's expectation about its true production cost would be higher—perhaps $9—because the firm's posterior expectation is a combination of the prior ($10) and its observed data ($8). An outsider who observes that true production costs ($8) are

[6]To elaborate, it may be that the typical driver finds driving to be easier than initially expected because crashes and citations are infrequent events. Since a driver observes her own driving record more completely than the record of anyone else she may use the same Bayesian logic to conclude that driving is somewhat easy for everyone, but also that she has done somewhat better than the average driver.

lower than the prior estimate ($10) will therefore observe that firms are, on average, overestimating their costs (at $9). By symmetry 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.

Note that from the perspective of our model overconfidence is a "bias" only in the sense that beliefs do not match the actual distribution of outcomes. Agents' beliefs are consistent with Bayes's rule and are therefore justifiable given agents' available evidence. Thus, overestimation and overplacement can exist as statistical biases—rather than behavioral biases—arising from incomplete information and uncertainty.

One notable difference between this and most other models of overconfidence is that overplacement and overestimation are results of the updating process and we therefore do not (necessarily) expect these phenomena before an agent experiences a task and updates her beliefs. In our experimental data we find a small degree of prior overplacement, though its direct is gender-specific: as in Niederle and Vesterlund (2007), men tend to exhibit overplacement while women tend to exhibit underplacement, and the two effects roughly cancel out in a mixed population. (We find no sign of prior overestimation either by gender or in the population as a whole.) This pattern of prior overplacement and underplacement can be incorporated easily in the model as an assumed prior bias, but its only affect would be to change the baseline level of overconfidence that the standard model assumes to be zero. Thus, the modified model with prior overplacement predicts increased overplacement after easy tasks and decreased overplacement (perhaps becoming underplacement) after difficult tasks; this is discussed in Section 5.4.2.

Existing models in the economics literature are designed to explain overplacement or overestimation, but do not capture the underplacement, underestimation, and correlation with task difficulty that we observe in our data. We review the models of overplacement by Van den Steen (2004) and Santos-Pinto and Sobel (2005) and the model of overestimation by Zabojnik (2004) in Section 6. We also consider the experimental evidence from the market entry game of Camerer and Lovallo (1999), which seems to indicate little to no prior overconfidence in the baseline treatment but excessive entry appears when subjects apparently fail to account for the fact that their competitors self-selected into the experiment knowing that payoffs would depend on trivia quiz performance. This suggests that other biases may exist that generate behavior that is observationally equivalent to overconfidence in some environments. Careful study of these biases is therefore necessary to disentangle their underlying causes.

We provide our formal definitions of the three notions of overconfidence in the next section. In Section 3 we discuss the design of our experiment. The results appear in Section 4. We then detail our theoretical model in Section 5 and compare its predictions to our experimental results. A more thorough review of the previous literature appears in Section 6, and the paper concludes with Section 7.

2. Three Definitions of Overconfidence

Because the term "overconfidence" has been used to explain a wide range of observed phenomena, we begin our study by formally defining three distinct notions of overconfidence: overplacement, overestimation, and overprecision.

Consider a pair of agents indexed by i and j that are independently engaging in some task in which their performance can be quantified unambiguously. Examples include examinations, athletic competitions, and product assembly procedures. Let X_i and X_j be random variables representing i and j's "score" on the task, with generic realizations x_i and x_j, respectively. Agent i receives private information at various points in time and updates her beliefs about X_i and X_j accordingly. For the present discussion let I_i represent any relevant information held by agent i, so that $E_i[X_i|I_i]$ and $E_i[X_j|I_i]$ represent i's expectation of her own score and j's score, respectively, given that her current information is I_i. Since we only consider the problem from i's perspective (because j's inference problem is identical), we henceforth drop the i subscript on the expectation operator. When more than two agents are involved, j represents a randomly-drawn individual from the group such that $j \neq i$.[7]

Definition 1. *Agent i exhibits* overestimation *(given information I_i) if* $E[X_i|I_i] > x_i$ *and* underestimation *(given I_i) if* $E[X_i|I_i] < x_i$. *If either is true, i exhibits* misestimation.

Definition 2. *Agent i exhibits* overplacement *(given I_i) if* $E[X_i|I_i] > E[X_j|I_i]$ *and* underplacement *(given I_i) if* $E[X_i|I_i] < E[X_j|I_i]$. *If either is true, i exhibits* misplacement.

Definition 3. *Agent i exhibits* overprecision *(given I_i) if the variance of the conditional random variable $X_i|I_i$ is strictly less than the variance of observed scores conditional on I_i,* underprecision *if the variance of $X_i|I_i$ is strictly greater than the variance of observed scores conditional on I_i, and* misprecision *if either is true.*

We reiterate that, under these definitions, overconfidence need not be irrational. In the case of misplacement it need not even be a statistical bias at the individual

[7]What matters for the specification of the comparison individual "j" is that (1) agent i knows neither more nor less about this individual (in expectation) than other individuals and (2) that j represent an actual individual rather than an order statistic (such as "the median score") since the distribution of the order statistic typically will be different from the distribution of any one individual.

level (consider the case where $I_i = \{x_i, x_j\}$ with $x_i \neq x_j$) but can appear "biased" at the aggregate level (for example, when all agents exhibit overplacement).[8] In this paper we seek to identify when these forms of overconfidence occur and ask whether a simple model of Bayesian updating might be able to account for them.

3. Experimental Design

To examine the various forms of overconfidence we abstract away from the details of competitive environments and study individuals' beliefs about their own performance and the performance of others in a series of trivia quizzes of varying difficulty.[9]

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. Each session 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

[8] Moore and Healy (2008) define overplacement as $E[X_i|I_i] - E[X_j|I_i] > x_i - x_j$ and find qualitatively similar results using the same data.

[9] It may be that certain forms of competition will generate other behavioral biases that would interact with overconfidence. Thus, 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 Camerer and Lovallo 1999) can then be used to paint a more complete picture of the overconfidence phenomenon.

[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 `http://cbdr.cmu.edu/roe` and log in using Participant ID 0000.

but do not know the scores of others.

In each of the three phases each subject is asked to submit a belief distribution about his or her 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 same 18 quizzes in some prior session.[11] Each probability distribution consisted of eleven probabilities, one for each of the possible scores (zero through ten). Subjects are shown eleven moveable horizontal bars—along with numerical values—to represent these eleven probabilities and can 'drag' each bar to the desired probability value.[12] Once a subject is satisfied with a particular distribution she clicks a button to submit the reported distribution.

Specifically, the timing of each round 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 (where the subjects' own scores are known) each subject reports a distribution for the RSPP's score.

Subjects earn money from two sources on each quiz in each period. First, if the subject's percentile rank on the quiz is $r \in [0,1]$ then she earns $\$25r$ for her perfor-

[11]Data from pilot sessions was used as the source for the RSPP in early sessions. Subjects were not explicitly informed about the pool of subjects from which the RSPP was drawn. Specifically, they were not told the number of subjects in the pool.

[12]Initial bar positions were randomly set each period. Moving one bar caused the ten other bars to adjust proportionally 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 instances of 'questionable' (misspelled or incomplete) answers being counted as correct. We did not remove these data from our analysis.

mance.[14] Second, each subject receives payments based on the accuracy of each of her five reported distributions. This is calculated using a quadratic scoring rule. Specifically, 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 payment for that 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). At the conclusion of the experiment five of the eighteen rounds were randomly selected as payoff rounds and subjects were paid the mean of their payoffs in these five rounds.

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. Since subjects earned an average of \$12.18 on the quiz and \$2.39 on the two reports about their own score, and since 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.[15]

The eighteen quizzes span six topics, each at three difficulty levels. The assignment of quizzes to the three difficulty levels (easy, medium, and difficult) was based

[14] 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.

[15] Exact conditions on the beliefs necessary for manipulation to be profitable are explored in a working paper version of this manuscript.

on previous experience with the questions in other studies.[16] The six topics were geography, movies, music, history, sports, and science. Each participant saw a different random order of the eighteen quizzes, subject to the constraint that each three-round block included one quiz at each difficulty level. The three difficulty levels were randomly ordered within each block, allowing for a relatively uniform distribution of quiz difficulty levels across the eighteen rounds while making it difficult for a subject to predict the difficulty or subject matter of an upcoming quiz.[17]

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 1. 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 incentivized subjects to reveal the data needed to compare the results to the predictions of the theory. For example, if subjects are manipulating their quiz performance to increase the accuracy of their predictions then stated beliefs will reflect expectations about manipulations—not true abilities—and the Bayesian model would not apply. Although small manipulations in performance would be difficult to detect, large manipulations are fairly obvious. Scores on easy

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

[17]One weakness of this design is 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, however, there will be no systematic effect on any one difficulty level. Alternative designs that do not group difficulties into blocks face the problem that a subject might encounter all six difficult questions early in the quiz and could then (rationally) expect no additional difficult questions in that quiz.

Difficulty	Phase	Distribution	Block 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 1: Averages (and standard errors) of expected values of reported belief distributions.

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.[18] Although these may represent true manipulations, they constitute only about 2 percent of the easy quiz 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, male subjects enter the experiment with a small degree of overplacement, female subjects enter with a small degree of underplacement, these effects roughly cancel out on aggregate, and subjects do not exhibit misestimation (in aggregate or by gender) prior to taking the quizzes. 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 provide a baseline environment for the model, while the last two results unify the previous findings from psychology and provide a coherent behavioral pattern for our theoretical model to rationalize.

These results are each demonstrated by regressions whose estimates and standard errors appear in Table 2. In each regression an appropriate dependant variable

[18]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.

Result #:	1		2	3		4
Quiz Difficulty	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 2: Coefficient estimates (and t-statistics) from 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.

is regressed against a full set of dummy variables indicating easy, medium, and difficult quizzes. Each quiz for each subject is treated as an independent observation in these regressions, for a total of 1,476 observations per regression. 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.[19] 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.

The first two regressions in Table 2 give the following result.

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.*

[19] 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.

This result is important in verifying that the three difficulty levels produce significantly different scores; if all quizzes produced similar scores then any correlation between overconfidence and actual task difficulty would become difficult to detect. It is clear from column (2) of Table 2 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.[20] 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 2 can be used to verify that subjects correctly perceive the differences in difficulty after taking the quiz. Subjects' expectations of their own scores 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.

Result 2. *Subjects exhibit no systematic misestimation before taking each quiz. Male subjects exhibit slight overplacement and female subjects exhibit slight underplacement before each quiz and these effects roughly cancel out when aggregated across gender.*

For misestimation in the first period, the median difference between reported expectations and actual scores is not significantly different from zero (sign test p-value of 0.581), indicating no first-period prior misestimation.[21] Of the 82 subjects,

[20] 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).

[21] The first-period median score is 6.00, which is insignificantly greater than the median reported expectation for subjects' own scores (5.35); the mean score (5.34), on the other hand, is insignificantly

38 report first-period expectations above their actual score. Looking within each subsequent period, prior misestimation does not achieve significance in any period except period 10 (sign test p-value of 0.035) and is not significant when all periods are aggregated (sign test p-value of 0.122).[22] Misestimation is also insignificant when controlling for gender.[23]

As for misplacement in the first period, the median difference between reported ex-ante expectations for self and expectations for the RSPP are 0.436 for men and -0.148 for women. The median for men is significantly positive (sign test p-value of 0.008) but the median for women is (marginally) insignificantly different from zero (p-value of 0.090). Of the 47 men, 33 report higher first-period expectations for their own score than for the RSPP, while the same is true of 12 out of 35 women. Aggregating across all periods, the medians are 0.126 and -0.058, respectively, both of which are significant (sign test p-values of < 0.001 and 0.006, respectively). Note, however, that the magnitudes of these effects are relatively small; each gender is misplacing by roughly one-tenth of a question (out of ten) on average.[24]

If the first-period misplacement data are aggregated across genders then significance disappears; the median difference between ex-ante expectations of subjects' own scores and expectations of the RSPP's score is 0.118 with a sign test p-value of 0.440. Significance returns when all periods are aggregated; the median difference is 0.020 with a sign test p-value of 0.022. Thus, the population as a whole exhibits

less than than the mean reported expectation (5.52).

[22]There does not appear to be anything peculiar about period 10; recall that when running eighteen tests we should expect roughly one significant difference at the 5% level under the null hypothesis.

[23]The mean and median of the differences between first-period expected scores and actual scores for men are 0.22 and -0.25, respectively, and 0.01 and -1.28, respectively, for women. Hence, men in this study exhibit greater first-period overestimation, but the gender difference is far from significant (Wilcoxon p-value of 0.69). Results reverse when aggregating across periods; men exhibit insignificantly *less* overestimation (p-value of 0.58).

[24]Significance is likely due to the large sample sizes of 846 observations for men and 630 observations for women.

a *very* slight—but statistically significant—tendency toward overplacement.[25] A regression of ex-ante overplacement on quiz difficulty aggregating across both genders and all periods (column (4) of Table 2) reveals no significant misplacement for any difficulty level.[26]

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 2 (columns (5)—(7)) test the predictions of Table 4. We examine two measures of overplacement: interim overplacement (when subjects are uncertain about their scores) and ex-post overplacement (when subjects know their own score). 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.

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 2, the *opposite* result obtains: sub-

[25] Again, the sample size for this test is quite large with 1,476 observations.

[26] Estimates are all insignificant if the same regression is run using lagged dummy variables. Thus, on aggregate, subjects do not exhibit significant overplacement before period t after taking an easy quiz in period $t-1$.

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%
Ex Post	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 3: Frequency of subjects exhibiting various rankings of own score vs. other's score, compared to actual score rankings.

jects 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 3). 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.[27]

Recall that the definition of overplacement used in this paper requires only that subjects *believe* they will score higher than the RSPP (in expectation); it does not require that this belief be incorrect. In Table 3 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 observed correlation from Result 3 (overplacement for easy quizzes, underplacement for difficult quizzes) is the modal ranking and in each of these four cases no more than 41.1 percent of the subjects had correct rankings of their expectations. Thus, the observed pattern of misplacement is the modal observation even though these beliefs are inaccurate the majority of the time.

[27] 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.

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

Table 4: The observed patterns of overconfidence.

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 2 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 insignificant.

4.2. Overprecision

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—imperfect signals of performance gained from taking the quiz—that cannot be observed by the experimenter. Since we cannot construct the appropriate empirical distribution that conditions on this information, we do not test for overprecision in estimates of subjects' own scores at the interim or ex-post stages.

The distribution of actual scores across all difficulty levels is highly bimodal, with

49.8% of all quiz scores equalling zero or ten.[28] 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.[29] 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.[30]

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.[31] 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.[32] Both are significantly greater than zero, indicating the presence of overprecision in interim beliefs about others' scores.

Finally, we explore the link between overprecision and the other forms of overconfidence. Table 5 reveals that the correlation between ex-anteoverprecision and interim overplacement is negative and significant while the correlation between ex-ante overprecision and interim overestimation is positive but insignificant.[33] Thus,

[28]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.

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

[30]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 ($p = 0.298$).

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

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

[33]Here, ex-ante overprecision is the variance of the distribution of all scores minus the variance of

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.

overprecision is related to—but distinct from—the other two measures of overconfidence. This relationship is explored in the theoretical model developed in the following section.

5. A Bayesian Explanation

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. Recall from Section 2 that we focus on agent i's beliefs and let X_i and x_i be the random variable representing i's score and a generic realization of i's score, respectively. X_j and x_j are defined analogously. We proceed by assuming that agent i believes that 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) and L_i is a mean-zero idiosyncratic component that determines the difference between i's

the subject's ex-ante distribution of her own score.

score and the overall average. We assume the mean of S exists and equals μ.[34] 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.[35] 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 believes $X_j = S + L_j$ (for some other agent $j \neq i$) and has well-defined priors on L_j, 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.

5.1. Overplacement and Underplacement

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 $E[X_j|x_i] = E[S|x_i]$.

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

[34] For simplicity, we assume throughout that all random variables have well-defined means.

[35] We discuss the case where $E[L_i]$ is non-zero in Subsection 5.4.2.

underplacement after unexpectedly difficult tasks; this matches Result 3 above.[36]

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.[37] The following examples highlight cases where $E[S|x_i]$ takes this particular form.

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 $\alpha = \sigma_L^2/(\sigma_S^2 + \sigma_L^2)$.[38] □

This example is a special case of a more general theorem due to Diaconis and 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 the prior on S is conjugate. Thus, we expect the predicted pattern of overplacement when, for example, X_i given S has a normal distribution with a normal prior on S, an exponential distribution with a gamma prior, a poisson distribution with a gamma prior, a geometric distribution

[36]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 of overconfidence about their *friend's* performance when the friend's performance is observable but others' performances are not. Our experiment does not test for this effect.

[37]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 $\alpha \in [0,1]$.

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

with a beta prior, or when X_i has a binomial distribution with a beta prior, as in the following example.

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).[39] 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.[40] □

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

[39]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.

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

previously expected.[41]

5.2. Overestimation and Underestimation

In modeling misestimation it becomes critical to specify the private information available to the agent at each phase in the experiment. The ex-ante and ex-post phases are of little interest since subjects either know nothing or know everything, so we focus on overestimation in the interim phase.

Formally, we assume that the act of taking a quiz provides each agent i with a noisy signal of her true score, denoted y_i. The signal y_i is believed (by i) 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 y_i makes inferences about S, L_i, and E_i, and forms the posterior expectation $E[X_i|y_i]$ that we compare against the agent's true score x_i.[42] 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.

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

[41]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.

[42]One interpretation of y_i is that it is the subject's initial 'gut feeling' about her performance, which she then integrates into her prior beliefs to generate the posterior expectation $E[X_i|y_i]$.

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; agents exhibit *under*estimation after easy tasks and *over*estimation after difficult tasks. This prediction matches the observations of Result 4.

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. □

As with overplacement, the prediction 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.

5.3. Overprecision and Underprecision

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 predictions on overplacement and overestimation; however, the *level* of precision in beliefs 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—than does 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 cannot 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.

5.4. Modifications and Extensions

The above theoretical model is, by design, simple and unsophisticated. Several embellishments could be added to make the model better fit 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 extensions to the model and argue that these changes do not (necessarily) invalidate the main conclusions of the simple model.

5.4.1. Multi-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 how 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 unobservable signal (conditional on the observable score x_i). 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 qualitatively alter the above analysis. Intuitively, we expect that r_i equals μ on average when prior beliefs are unbiased. In other words, we expect that this second signal only serves to strengthen the prior belief, pulling the posterior means for X_i and X_j toward μ on average. This 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

$$\left(\frac{\sigma_L^2 + \sigma_E^2 + \sigma_S^2}{\sigma_L^2 + \sigma_E^2 + 2\sigma_S^2}\right)\mu + \left(1 - \frac{\sigma_L^2 + \sigma_E^2 + \sigma_S^2}{\sigma_L^2 + \sigma_E^2 + 2\sigma_S^2}\right)x_i.$$

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.[43]

5.4.2. Ability and Prior Overconfidence

The simple model does not incorporate the possibility of prior overconfidence since it is not needed to explain the patterns of overconfidence observed in the experimental data; however, we do observe small degrees of gender-specific misplacement in prior beliefs. This may be due to misplacement gained from unobserved past experiences with trivia quizzes or it may be a true behavioral bias. Prior misplacement 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.

5.4.3. Non-Bayesian Updating

The above mathematical arguments make generous application of Bayes's rule, but the results may also hold for agents whose updating process is not perfectly

[43]If r_i is substantially lower than μ when $\mu < x_i$ then it is possible that the posterior expectation drops below μ. Similarly, if r_i is substantially greater than x_i then it is possible that the posterior expectation rises above x_i. Thus, we can observe some individuals whose posterior expectations do not lie between μ and x_i, but in expectation these opposing observations cancel out. In other words, the second signal adds noise to the data but does not change the expected conclusions.

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. Thus, the predictions can apply to Bayesians and non-Bayesians alike.

An important consequence of this observation is that our experiment is not a direct test of the details of the model; as with any experiment, we test only the *predictions* of the model. Since we do not impose specific assumptions about which distributions subjects use as their priors, we limit ourselves to examining directional and correlational predictions of the model rather than specific point estimates.[44] We therefore cannot reject any other model that generates qualitatively similar predictions. We are not aware, however, of any other model of overconfidence that predicts both *under*confidence and the observed correlations between overconfidence and task difficulty.

The fact that the same predictions can be generated by certain non-Bayesian behavior also means that our ability to generate the observed 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 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 or the exact nature of Bayes's rule.

[44]Although we can observe subjects' beliefs about X_i, beliefs about S, L_i, and L_j would be needed to generate point predictions. In principle one could elicit beliefs over S, L_i, and L_j, though this may lead subjects to dissect their predictions in ways that are not natural.

5.5. Further Tests and Implications

The previous discussion shows that the key predictions of the model conform to Results 3 and 4 from the experiment. Thus, the model rationalizes the data from a broad perspective. In this section we strengthen this claim by comparing additional model predictions to the experimental data.

One of the strongest assumptions of the simple model is that agents enter a task believing themselves to be no different than others. Although the model can be modified to include ability or prior overconfidence (see Section 5.4.2), our Result 2 indicates that in fact this assumption is reasonably accurate for our data; there is no evidence of systematic prior misestimation and only a small degree of misplacement that depends crucially on gender. This result matches previous findings; Niederle and Vesterlund (2007) observe the effect on gender and, as we discuss in Section 6, Camerer and Lovallo (1999) find little to no evidence of prior misplacement (on average) in their market entry game experiment. In fact, few studies in the psychology literature examine *ex ante* beliefs; most measure beliefs about some real-world event with which subjects typically have prior experience and find the above patterns of overconfidence as a result.[45]

In our model 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.[46] Recall from the first two columns of Table 2

[45]For example, Weinstein (1980) elicits beliefs about events such as suicide and having a happy marriage in the future. Although subjects presumably have not attempted suicide, they still have gained information about how easy it has been for them *not* to (want to) commit suicide or how difficult it has been to find an enjoyable date, and can therefore exhibit patterns of overconfidence as the result of Bayesian updating.

[46]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.

that the average score on easy quizzes is 8.864 while the average interim expected score after an easy quiz is 8.644, and the average score on difficult quizzes is 0.693 while the average interim expected score on these quizzes is 1.503. In both cases the average expected score is closer to the overall average of 5.36 than the average actual score, indicating a degree of regression in beliefs consistent with the betweenness condition at the aggregate level.

In fact, the betweenness condition is slightly 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 in 80.1 percent of quizzes in our data.[47]

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 of 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 1. Using the beta-binomial specification of Example 2, the resulting beta coefficients are $(\beta_1, \beta_2) = (3.39, 2.93)$, indicating a roughly symmetric prior over p with a mean of 0.5364 (since $\mu = 5.364$) and a skewness of only -0.095.[48]

As discussed in Section 5, if beliefs are highly bimodal then the betweenness condition may fail. Although subjects' actual quiz scores are highly bimodal, their prior *beliefs* are not (see Section 4.2), and so the difficulties with bimodality are avoided in our data. Had subjects' prior beliefs been better calibrated then the observed pat-

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

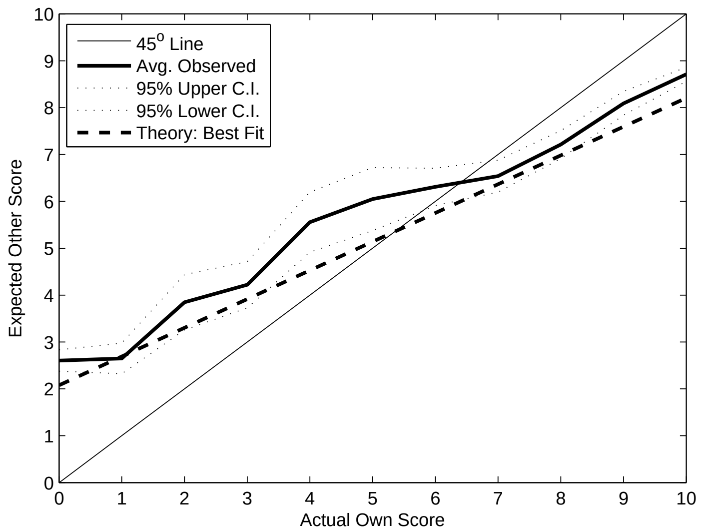

Figure 1: Average reported expectation of others' scores versus own score.

terns of overconfidence may have changed substantially. The question of when or why subjects' beliefs are poorly calibrated remains an open question.

One feature of any model of incomplete information is that experience and learning should diminish the effects of uncertainty. In the case of our experiment, however, the scope for learning is somewhat limited since each quiz is different, and so we expect that the overconfidence results will persist through the final periods.[49] This is consistent with our data; the same correlations between task difficulty and the various forms of overconfidence from Table 2 are observed if one restricts attention to the final block of three periods, though the magnitudes of the regression estimates are slightly smaller.[50]

[49]An additional conjecture is that subjects may eventually discover the three distinct difficulty categories (easy, medium, and hard), leading to a situation where a surprisingly easy quiz in the difficult category leads to overplacement and underestimation, for example. The data from later periods of our experiment do not indicate any such pattern.

[50]Some of the coefficients in the three-period regression are only marginally significant, though this is likely due to the fact that only one-sixth of the data is in use; details are provided in an earlier working paper.

The one measure in which there is scope for learning is in the accuracy of subjects' beliefs. A regression of the absolute difference between subjects' actual scores and reported prior expectations of their own score against the period number reveals a significant intercept (4.24, p-value < 0.001) and negative slope (-0.045, p-value < 0.001), indicating slow but highly significant improvements in calibration through time. A similar (though smaller) negative slope is found when using interim-stage reports (-0.011, p-value 0.017). Subjects' accuracy in predicting the RSPP's score is also improving in time, but the rate of improvement is marginally insignificant (slope from ex-ante reports: -0.015, p-value 0.050; slope from interim reports: -0.010, p-value 0.156).

Recall from Section 5.3 that overprecision should be correlated with the magnitudes of overplacement and overestimation, though the sign of this correlation cannot be predicted without knowing the underlying distributions for S and L_i. Table 5 in Section 4.2 revealed a strong negative correlation between ex-ante overprecision and interim overplacement and a weak positive correlation between ex-ante overprecision and interim overestimation. According to the logic of our model, 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.

6. Previous Literature

A variety of other papers explore overconfidence under differing assumptions.[51] Overconfidence has previously been modeled as stemming from other judgement biases Rabin and Schrag (1999) or from rational choice when beliefs are flexible and overconfidence affects other interactions and motivations (e.g., Benabou and Tirole 2002; Carillo and Mariotti 2000; or Compte and Postlewaite 2004). Other research (March and Shapira 1987; Odean 1998; Daniel et al. 2001, and Malmendier and 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 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.

Perhaps the most "rational" model of overconfidence in the economics literature is due to Zabojnik (2004), who assumes that agents can choose to participate in tests of their future productive ability but in doing so must sacrifice current-period production and its resulting consumption. If the payoff function is convex in ability then agents who believe their ability is higher face a larger expected sacrifice from participating in such tests. Thus, the optimal testing rule is asymmetric; agents whose tests indicate high ability will stop testing early while agents whose tests indicate low ability will continue testing longer. In the steady state we will observe more agents overestimating their ability than underestimating. Systematic *under*estimation would be predicted if the payoff function were concave in ability, and it may be that convexity and concavity are linked to our notions of an easy or difficult task; however, this

[51] Also note that Moore and Healy (2008) presents alternative analyses of the data presented in this paper.

model does not speak to the misplacement or misprecision phenomena that we also observe.

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^i_n. 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 has made an inferior choice. Thus, each will exhibit overplacement.

Santos-Pinto and 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.[52] 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.[53]

[52]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 $f_i(x) = \sum_{k=1}^{n} x_k p^i_k$.

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

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 assumes all agents are attempting to maximize the same objective function (total quiz score) but maximization is imperfect and is more difficult in some tasks than in others. It is the simultaneous inference about the task's difficulty and one's own performance that leads to the overplacement and underplacement observed in the data.

A second difference between VdS and SP&S and the current model is that the former necessarily imply prior overplacement but the latter does not. Our experiment reveals significant prior overplacement in men but significant prior *under*placement in women. Although the results of Camerer and Lovallo (1999) (henceforth C&L) are cited as evidence in favor of an overconfidence bias, they too find no strong evidence for prior overplacement in their baseline treatment. In their setting, subjects choose whether or not to enter a market in which entrants' profitability depends on their assigned rankings. In C&L's first treatment rankings are randomly chosen after the entry decisions are made. In the second and third treatments entrants are ranked based on their performance on 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 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.[54] Comparing the first two treatments reveals that subjects enter more frequently when rankings are quiz-

[54]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 5.4.2).

based, but the aggregate level of entry is at or below the risk-neutral equilibrium prediction in both cases.[55] 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 structure of incomplete information we assume has been used in other game theoretic models. In some cases, the results are indicative of the kind of 'rational overplacement' we describe. For example, 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.[56] In other 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).

[55]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.

[56]To see this from Shapiro's paper, simply compare the equilibrium with imperfect correlation ($\rho < 1$) to the equilibrium with perfect correlation ($\rho = 1$), which is the Nash equilibrium of the standard game with complete information.

7. Discussion

This paper accomplishes three goals. First, we clearly define various distinct notions of overconfidence that previous research has occasionally muddled. Second, we use a single experiment to paint a comprehensive picture of these three notions of overconfidence and how they are linked. Finally, we show how these patterns of overconfidence can be predicted in a simple Bayesian model without assuming any biases in judgement, as the testable predictions of the model are consistent with the experimental observations.

There have been a number of recent economic models that have attempted to explain how rational Bayesian agents could display overconfidence (see the paper cited in the previous section or Benabou and Tirole (2002); Bodner and Prelec (2003); or Rabin and Schrag (1999), for example). Although overconfidence has been widely observed, none of these models can parsimoniously account for the evidence from the present experiment because they predict neither the systematic *under*confidence nor the correlations between overplacement, overestimation, and task difficulty observed in these results.

We believe that the tendency for studies to focus on overconfidence (rather than underconfidence) may be attributable to methodology. For example, several studies examine individuals' beliefs about a single question, which confounds overestimation with overprecision since a more extreme probability estimate necessarily implies a lower variance (see, e.g., Alba and Hutchinson (2000) or Fischhoff, Slovic, and Lichtenstein (1977)), making it impossible to determine the degree to which each is responsible for the result. 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 (Juslin, Winman, and Olsson 2000; Malmendier and Tate 2005; or 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; Messick, Bloom, Boldizar, and Samuelson 1985; or Svenson 1981).

Although our model is Bayesian, there is ample reason to question whether people actually make judgments according to Bayes's rule. In some circumstances people appear to neglect priors (such as base rates), overweighting recent evidence (see, e.g., Grether 1980, 1990). In other circumstances people appear too conservative, overweighting priors and neglecting useful new evidence (e.g., Edwards 1968 or McKelvey and Page 1990). Which of these errors people commit depends on the order and form in which they acquire information (e.g. Hogarth and Einhorn 1992 or Wells 1992). What is important for our purposes here, however, is that although people are imperfect Bayesians, they rarely abandon Bayesian logic completely.[57] Overweighting the prior or overweighting the data still leads posterior means to lie somewhere between the prior mean and the observed data, generating the same patterns of overconfidence and underconfidence predicted by our model and observed in our data.

[57] See the discussion from Section 5.4.3.

References

Alba, J. W., Hutchinson, J. W., 2000. Knowledge calibration: What consumers know and what they think they know. Journal of Consumer Research 27, 123–156.

Barber, B. M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. The Quarterly Journal of Economics 116, 261–293.

Benabou, R., 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.

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

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

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

Casella, G., Berger, R. L., 2002. Statistical Inference, 2nd Edition. Duxbury Press, Pacific Grove, PA.

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

College Board, T., 1977. Student descriptive questionnaire. Educational Testing Service, Princeton, NJ.

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

Daniel, K. D., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53 (6), 1839–1885.

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

De Bondt, W., Thaler, R. H., 1995. Financial decision-making in markets and firms: A behavioral perspective. In: Jarrow, R. A., Maksimovic, V., Ziemba, W. T. (Eds.), Handbooks in Operations Research and Management Science: Finance. Vol. 9. North Holland, Amsterdam, pp. 385–410.

Diaconis, P., 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: Kleinmuntz, B. (Ed.), Formal representation of human judgment. Wiley, New York, pp. 17–52.

Erev, I., Wallsten, T. S., Budescu, D. V., 1994. Simultaneous over- and underconfidence the role of error in judgement processes. Psychological Review 101, 519–527.

Fischhoff, B., Slovic, P., Lichtenstein, S., 1977. Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance 3, 552–564.

Garcia, D., Sangiorgi, F., Urosevic, B., 2007. Overconfidence and market efficiency with heterogeneous agents. Economic Theory 30, 313–336.

Glaser, M., Weber, M., 2007. Overconfidence and trading volume. Geneva Risk and Insurance Review 32, 1–36.

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.

Griffin, D., Tversky, A., 1992. The weighing of evidence and the determinants of confidence. Cognitive Psychology 24, 411–435.

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

Howard, M. E., 1983. The causes of war and other essays. Harvard University Press, Cambridge, MA.

Johnson, D. D. P., 2004. Overconfidence and war: The havoc and glory of positive illusions. Harvard University Press, Cambridge, MA.

Juslin, P., Winman, A., 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.

Lichtenstein, S., Fischhoff, B., 1977. Do those who know more also know more about how much they know? Organizational Behavior and Human Performance 20, 159–183.

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

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

McKelvey, R. D., 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., Samuelson, C. D., 1985. Why we are fairer than others. Journal of Experimental Social Psychology 21, 480–500.

Moore, D. A., Healy, P. J., 2008. The trouble with overconfidence, forthcoming, *Psychological Review*.

Moore, D. A., 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., Small, D. A., 2007. Error and bias in comparative social judgment: On being both better and worse than we think we are. Journal of Personality and Social Psychology 92, 972–989.

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

Neale, M. A., Bazerman, M., 1985. The effects of framing and negotiator overconfidence on bargaining behaviors and outcomes. Academy of Management Journal 28.

Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W., 1996. Applied Linear Statistical Models, 4th Edition. McGraw-Hill, New York.

Niederle, M., Vesterlund, L., August 2007. Do women shy away from competition? do men compete too much? The Quarterly Journal of Economics 122, 1067–1101.

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.

Plous, S., 1993. The psychology of judgement and decision making. McGraw Hill, New York.

Rabin, M., 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., 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.

Statman, M., Thorley, S., Vorkink, K., 2006. Investor overconfidence and trading volume. Review of Financial Studies 19 (4), 1531–1565.

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.

Windschitl, P. D., Kruger, J., Simms, E., 2003. the influence of egocentrism and focalism on people's optimism in competitions: When what affects us equally affects us more. Journal of Personality and Social Psychology 85, 389–408.

Zabojnik, J., 2004. A model of rational bias in self-assessments. Economic Theory 23, 259–283.

APPENDICES*

A. Full Regression Results

The complete regression results (including difficulty, block, and interaction effects) are provided in Table 6. These regressions omit one dummy variable and code the remaining dummies as negative one for the omitted category. This procedure allows the inclusion of a constant term, giving an estimate for the overall average, and guarantees that the treatment effects sum to zero, as in an ANOVA procedure. The significance results are equivalent to a regression with a full set of dummies and no constant term, as in the manuscript (see, e.g., Neter, Kutner, Nachtsheim, and Wasserman, 1996).

B. The Quizzes

The eighteen trivia quizzes are shown in Table 7, with one quiz per page. The mean, median, and variance of scores on the quiz are shown for each quiz, along with the quiz ID number (1 through 18), the topic, and the difficulty level. Recall that quizzes were randomly placed into blocks with one difficulty level per block, and the order in which each subject encountered the six blocks was randomized. Therefore, the quiz ID numbers do not represent the order in which quizzes were shown to subjects.

*These appendices are for the reference of the reader and are not intended for publication.

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
Constant	**5.161**	**5.359**	0.007	**-0.428**	**-0.525**	**0.199**
	(84.06)	(85.34)	(0.19)	(-7.72)	(-8.86)	(6.26)
Easy	**3.703**	**3.285**	0.002	**0.813**	**0.894**	**-0.418**
	(42.65)	(36.99)	(0.03)	(10.38)	(10.66)	(-9.31)
Difficult	**-4.467**	**-3.856**	0.041	**-1.020**	**-1.135**	**0.611**
	(-51.46)	(-43.42)	(0.82)	(-13.03)	(-13.54)	(13.60)
Block 1	-0.043	-0.034	-0.061	-0.064	-0.096	0.009
	(-0.31)	(-0.24)	(-0.76)	(-0.51)	(-0.73)	(0.13)
Block 2	*0.234*	*0.243*	0.015	0.191	0.096	0.009
	(1.70)	(1.73)	(0.19)	(1.54)	(0.72)	(0.13)
Block 3	-0.055	-0.045	0.062	-0.059	-0.034	0.010
	(-0.40)	(-0.32)	(0.77)	(-0.48)	(-0.25)	(0.15)
Block 4	-0.055	-0.069	0.003	-0.098	0.021	-0.014
	(-0.40)	(-0.49)	(0.04)	(-0.79)	(0.16)	(-0.20)
Block 5	-0.055	-0.110	-0.060	0.039	0.041	-0.055
	(-0.40)	(-0.78)	(-0.75)	(0.31)	(0.31)	(-0.77)
B1*Easy	-0.016	-0.204	-0.005	0.068	*0.347*	*-0.187*
	(-0.08)	(-1.02)	(-0.04)	(0.39)	(1.85)	(-1.86)
B1*Diff.	-0.114	0.218	-0.039	-0.227	**-0.541**	**0.332**
	(-0.59)	(1.10)	(-0.35)	(-1.30)	(-2.89)	(3.31)
B2*Easy	0.024	0.034	-0.119	0.059	0.162	0.010
	(0.13)	(0.17)	(-1.05)	(0.34)	(0.87)	(0.10)
B2*Diff.	-0.183	-0.186	-0.009	-0.130	-0.233	-0.003
	(-0.94)	(-0.94)	(-0.08)	(-0.74)	(-1.25)	(-0.03)
B3*Easy	0.289	0.209	0.039	0.182	0.084	-0.080
	(1.49)	(1.05)	(0.34)	(1.04)	(0.45)	(-0.80)
B3*Diff.	-0.053	-0.007	0.018	-0.182	-0.162	0.046
	(-0.27)	(-0.03)	(0.16)	(-1.04)	(-0.86)	(0.46)
B4*Easy	-0.028	0.102	-0.088	-0.036	-0.182	0.130
	(-0.15)	(0.51)	(-0.78)	(-0.21)	(-0.97)	(1.30)
B4*Diff.	0.020	-0.064	0.063	0.170	0.249	-0.085
	(0.10)	(-0.32)	(0.56)	(0.97)	(1.33)	(-0.84)
B5*Easy	-0.309	-0.139	0.016	-0.248	**-0.425**	*0.170*
	(-1.59)	(-0.70)	(0.14)	(-1.41)	(-2.27)	(1.69)
B5*Diff.	0.081	0.014	0.015	0.183	0.289	-0.067
	(0.42)	(0.07)	(0.13)	(1.05)	(1.54)	(-0.67)

Table 6: Dummy variable regressions (with block and interaction effects) 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. Italicized entries are significant at the 10% level.

Quiz: 1	**Topic:** Geography	**Difficulty:** Easy
Mean: 8.817	**Median:**: 10	**Variance:** 3.435
Question 1	What continent lies directly south of Europe?	
	Answer: Africa	
Question 2	What African river (that empties into the Mediterranean sea through Egypt) is the longest river in the world?	
	Answer: Nile	
Question 3	In what North American country is the city of Toronto located?	
	Answer: Canada	
Question 4	The lowest temperature ever recorded on earth (-129.3žF) occurred on what southern continent?	
	Answer: Antarctica	
Question 5	In what western U.S. state is the ŞSilicon ValleyŤ?	
	Answer: California	
Question 6	Baghdad is the capital of what middle-eastern country?	
	Answer: Iraq	
Question 7	The Golden Gate Bridge is located in which Californian city?	
	Answer: San Francisco	
Question 8	What is the capital of the United States?	
	Answer: Washington D.C.	
Question 9	On what continent is France located?	
	Answer: Europe	
Question 10	What is the capital of and largest city in Japan?	
	Answer: Tokyo	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 2	**Topic:** Geography	**Difficulty:** Medium
Mean: 6.098	**Median:**: 7	**Variance:** 7.299
Question 1	In what U.S. city does the best-known celebration of ŞMardi GrasŤ take place? *Answer: New Orleans*	
Question 2	In what U.S. state is the ski-resort town of Aspen? *Answer: Colorado*	
Question 3	The "Ring of Fire" is located around what ocean? *Answer: Pacific*	
Question 4	In what U.S. state is Atlantic City located? *Answer: New Jersey*	
Question 5	The most famous "tea party" of the American Revolution took place in what city? *Answer: Boston*	
Question 6	The island of Honshu is part of what country? *Answer: Japan*	
Question 7	What is the name of the worldŠs largest coral reef, located off the coast of Australia? *Answer: Great Barrier Reef*	
Question 8	What is the highest mountain range in the world? *Answer: Himalayas*	
Question 9	What country is comprised of over 17,000 known islands? *Answer: Indonesia*	
Question 10	In what U.S. state is Disney World located? *Answer: Florida*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 3	**Topic:** Geography	**Difficulty:** Hard
Mean: 0.683	**Median:**: 0	**Variance:** 1.849
Question 1	What is South AmericaŠs highest peak? *Answer: Mt. Aconcagua*	
Question 2	What is the capital of Australia? *Answer: Canberra*	
Question 3	What two South American countries are land-locked? *Answer: Bolivia and Paraguay*	
Question 4	What geographical area was once referred to as “Seward’s Folly?” *Answer: Alaska*	
Question 5	What is the capital city of Uganda? *Answer: Kampala*	
Question 6	What Pacific island mountain claims to be the wettest spot on Earth? *Answer: Mt. Wai’ale’ale (on the Hawaiian island of Kauai)*	
Question 7	Sweden, Denmark, Poland, and Finland all border what sea? *Answer: Baltic Sea*	
Question 8	Bechuanaland was the colonial name of what country? *Answer: Botswana*	
Question 9	What is the capital of Turkey? *Answer: Ankara*	
Question 10	What two countries border Mexico to the south? *Answer: Belize and Guatemala*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 4	**Topic:** Movies	**Difficulty:** Easy
Mean: 8.939	**Median:**: 10	**Variance:** 5.12
Question 1	Leonardo DiCaprio and Kate Winslet starred in a 1997 film about the sinking of what famous ship after striking an iceberg on her maiden voyage? *Answer: Titanic*	
Question 2	Arnold Swartzenneger, current governor of California, is sometimes called the ŞGovernatorŤ Ű a nickname poking fun at his role in what 1984 film? *Answer: Terminator*	
Question 3	Toby Maguire starred in two movies as what web-slinging super hero with spider powers? *Answer: Spiderman*	
Question 4	Cinderella, The Little Mermaid, Aladdin, and The Lion King are all films produced by what famous entertainment company? *Answer: Disney*	
Question 5	Keanu Reeves starred as Neo, a computer hacker, in what film trilogy about machines taking over the earth? *Answer: The Matrix*	
Question 6	What was the title of George Lucas's original 1977 science fiction film about Luke Skywalker and Darth Vader, set "Long ago, in a galaxy far, far away"? *Answer: Star Wars*	
Question 7	What recent film trilogy was based on J.R.R. TolkeinŠs novels about the quest of a hobbit named Frodo to destroy a ring? *Answer: Lord of the Rings*	
Question 8	The quotes ŞLife is like a box of chocolatesĚŤ and ŞRun, Forrest, run!Ť are from what 1994 movie starring Tom Hanks? *Answer: Forrest Gump*	
Question 9	J.K. Rowling's books tell about a young wizard named Harry who goes to a school called Hogwarts. What is Harry's last name? *Answer: Potter*	
Question 10	What actor, formerly married to Nicole Kidman, starred in the films Rain Man, Minority Report, War of the Worlds, Mission Impossible, and Jerry Maguire? *Answer: Tom Cruise*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 5	**Topic:** Movies	**Difficulty:** Medium
Mean: 6.293	**Median:**: 7	**Variance:** 12.333
Question 1	Who starred as high school student Marty McFly in the hit science-fiction comedy “Back to the Future” (1985)? *Answer: Michael J. Fox*	
Question 2	Who wrote and directed ŞKill BillŤ Volumes 1 and 2? *Answer: Quentin Tarantino*	
Question 3	What actor plays an overweight college instructor & several other characters in The Nutty Professor? *Answer: Eddie Murphy*	
Question 4	Who was the first African American to win an Academy Award for best actress? *Answer: Halle Berry*	
Question 5	What actress co-starred with Tom Hanks in the movies Sleepless In Seattle, You’ve Got Mail, and Joe Versus the Volcano? *Answer: Meg Ryan*	
Question 6	Ben Affleck and Matt Damon won an Academy Award for writing the screenplay of what film, starring themselves and Robin Williams? *Answer: Good Will Hunting*	
Question 7	Mo, Larry, and Curly are a trio more commonly known by what name? *Answer: The Three Stooges*	
Question 8	What film about a female boxer, starring Hilary Swank and Clint Eastwood, won an Academy Award for best picture in 2005? *Answer: Million Dollar Baby*	
Question 9	Dorothy, Scarecrow, Tinman, and the Wicked Witch of the West are all characters from what movie? *Answer: The Wizard of Oz*	
Question 10	What was the full name of the cannibalistic main character in the film ŞSilence of the LambsŤ? *Answer: Hannibal Lecter*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 6	**Topic:** Movies	**Difficulty:** Hard
Mean: 0.561	**Median:**: 0	**Variance:** 1.039

Question 1	Who is the only actress to have been nominated for the “Best Actress” Academy Award 13 times? *Answer: Meryl Streep*
Question 2	Who played Mozart in the 1984 film “Amadeus””? *Answer: Tom Hulce*
Question 3	Who is the father of actress Gwyneth Paltrow? *Answer: Bruce Paltrow*
Question 4	In what film did actress Mae West say the line, “When I’m good, I’m very good, but when I’m bad I’m better”? *Answer: I’m No Angel*
Question 5	What actress performed the voice of Bo Peep in the film Toy Story? *Answer: Annie Potts*
Question 6	Who starred opposite Gene Wilder in the 1984 film “Woman in Red”? *Answer: Kelly LeBrock*
Question 7	What 1984 film was the big-screen debut of actress Sarah Jessica Parker? *Answer: Footloose*
Question 8	In Monty Python’s Holy Grail, the French soldier taunts King Arthur by telling him, “Your mother was a hamster and your father smelt of ________? *Answer: Elderberries*
Question 9	What actress played Ferris Bueller’s girlfriend in the 1986 film, “Ferris Bueller’s Day Off”? *Answer: Mia Sara*
Question 10	What actor holds the record for having been nominated most frequently for the “Best Actor” Academy Award (9 times)? *Answer: Spencer Tracy*

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 7	**Topic:** Music	**Difficulty:** Easy
Mean: 8.317	**Median:**: 10	**Variance:** 7.232

Question 1	John Lennon, Paul McCartney, George Harrison and Ringo Starr were the four members of what famous classic rock band? *Answer: The Beatles*
Question 2	Janet Jackson and Latoya Jackson are sisters of what famous and eccentric pop star? *Answer: Michael Jackson*
Question 3	What pop group, led by Bono, has created the following albums: The Joshua Tree, Rattle and Hum, Achtung Baby, Zooropa, and How to Dismantle an Atomic Bomb? *Answer: U2*
Question 4	The pop star known for such songs as Like a Virgin, Material Girl, and Like a Prayer, goes by what single name (rather than a first and last name)? *Answer: Madonna*
Question 5	What famous classical composer, who eventually went deaf, wrote 9 symphonies, the most famous of which is his 5th symphony? *Answer: Beethoven*
Question 6	Which famous ŞBritneyŤ of the pop music world recently married and had a baby with her back-up dancer? *Answer: Britney Spears*
Question 7	Ashlee Simpson is the younger sister of what pop singer and star of the reality show ŞNewlywedsŤ on MTV? *Answer: Jessica Simpson*
Question 8	Singer Celine Dion sang the hit song ŞMy heart will go onŤ for the soundtrack of what 1997 film about the sinking of a famous ship? *Answer: Titanic*
Question 9	Stevie Wonder and Ray Charles both wore sunglasses during performances and had what physical disability? *Answer: Blindness*
Question 10	Sporty Spice, Baby Spice, and Posh Spice were members of what musical group? *Answer: Spice Girls*

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 8	**Topic:** Music	**Difficulty:** Medium
Mean: 6.415	**Median:**: 8	**Variance:** 13.456
Question 1	What is the real name of the artist who once went by the pseudonyms “Puffy”, “Puff Daddy”, and “P. Diddy”? *Answer: Sean Combs*	
Question 2	What band was Justin Timberlake in? *Answer: N-Sync*	
Question 3	Who is considered ŞThe KingŤ of rock? *Answer: Elvis Presley*	
Question 4	Former pop star Paula Abdul is now a judge for what T.V. show? *Answer: American Idol*	
Question 5	Ozzy and Sharon Ozbourne took their music careers to television when they began a reality show about their family on what TV station? *Answer: MTV*	
Question 6	Jim Morrison (lead singer of The Doors), Elvis Presley, and Jimi Hendrix all died from what? *Answer: Drug overdoses*	
Question 7	Mick Jagger, at the age of 62, recently went on tour with what legendary rock band for which he sings lead vocals? *Answer: Rolling Stones*	
Question 8	What was the name of the famous 3-day music festival, held in 1969, which featured artists such as Bob Dylan, Janis Joplin, and Jimi Hendrix? *Answer: Woodstock*	
Question 9	Beyonce Knowles rejoined what musical group in 2003? *Answer: DestinyŠs Child*	
Question 10	Gwen Stefani launched a successful solo career, but is also the lead singer of what band? *Answer: No Doubt*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 9	**Topic:** Music	**Difficulty:** Hard
Mean: 0.451	**Median:**: 0	**Variance:** 1.806
Question 1	On which album does actor William Shatner sing “Lucy in the Sky with Diamonds”? *Answer: The Transformed Man*	
Question 2	What is the real name of U2’s guitarist “The Edge”? *Answer: David Evans*	
Question 3	What, according to Billboard, was the top-ranked movie soundtrack album for the year 2004, at #6? *Answer: O Brother, Where Art Thou?*	
Question 4	What album ranks #1 on the best-selling albums of all time, just ahead of Michael Jackson’s “Thriller”? *Answer: The Eagles: Their Greatest Hits*	
Question 5	The first video played on MTV was “Video killed the radio star” by the Buggles. Who was the Buggles’ lead singer? *Answer: Mike Scott*	
Question 6	Former Texas gubernatorial candidate Kinky Friedman sang the song “Get your biscuits in the oven and your buns in the bed” with which band? *Answer: Kinky Friedman and the Texas Jewboys*	
Question 7	Who wrote the theme song to the television series “Hill Street Blues”? *Answer: Mike Post*	
Question 8	What was the name of the person who shot and killed Marvin Gaye one day before his 45th birthday? *Answer: Marvin Gaye, Senior (his father)*	
Question 9	What three famous musicians died together in a plane crash on February 3rd, 1959? *Answer: Buddy Holly, Ritchie Valens, and J.P. Richardson (“The Big Bopper”)*	
Question 10	The band “New Kids on the Block” included Donnie Wahlberg, Daniel Wood, Jordan Knight, and Jonathan Knight. Who was the fifth member until 1985? *Answer: Jamie Kelley*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 10	**Topic:** History	**Difficulty:** Easy
Mean: 9.598	**Median:**: 10	**Variance:** 0.861
Question 1	What U.S. president faced an impeachment trial to investigate his alleged affair with White House intern Monica Lewinsky? *Answer: Bill Clinton*	
Question 2	What explorer is credited with discovering America in 1492? *Answer: Christopher Columbus*	
Question 3	In 1776, the United States declared independence from what country? *Answer: Great Britain or England*	
Question 4	On what country did the United States drop atomic bombs on during World War 2? *Answer: Japan*	
Question 5	Who was the leader of Germany's Nazi party during World War 2? *Answer: Adolph Hitler*	
Question 6	On what date in 2001 were airplanes crashed into New York's World Trade Centers, destroying them? *Answer: September 11th*	
Question 7	What structure was built over 2000 years ago in China and stretches 4,163 miles? *Answer: The Great Wall of China*	
Question 8	At the beginning of what year, known as "Y2K" were computers expected to crash due to the "Millennium bug"? *Answer: 2000*	
Question 9	Who was elected President of the United States in 2004? *Answer: George W. Bush*	
Question 10	How many World Wars have there been? *Answer: Two*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 11	**Topic:** History	**Difficulty:** Medium
Mean: 5.720	**Median:**: 7	**Variance:** 8.97
Question 1	Who painted the Sistine Chapel? *Answer: Michelangelo*	
Question 2	The ŞRed ScareŤ was a fear of what political system? *Answer: Communist*	
Question 3	The Korean War was fought in which decade? *Answer: 1950Šs*	
Question 4	The New Deal is most closely associated with what U.S. president? *Answer: Franklin Delano Roosevelt*	
Question 5	Who was the Native American woman who accompanied Lewis and Clark on part of their expedition in the Pacific Northwest? *Answer: Sacajaweeja*	
Question 6	What period in European cultural history took place from the 14th to the 17th century, after the Dark Ages? *Answer: Renaissance*	
Question 7	What famous believer in non-violent protest led India to freedom from the rule of Great Britain? *Answer: Mohandas “Mahatma” Gandhi*	
Question 8	The Spanish Conquistadors were soldiers who conquered land for Spain and spread what religion? *Answer: Christianity*	
Question 9	The Italian village of Pompeii was destroyed in 79 AD by what type of natural disaster? *Answer: Volcano*	
Question 10	Elliot Ness and the Untouchables are associated with the pursuit of what prohibition-era Chicago mafia figure? *Answer: Al Capone*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 12	**Topic:** History	**Difficulty:** Hard
Mean: 1.207	**Median:**: 1	**Variance:** 2.759
Question 1	In what year did Martin Luther post his 95 theses? *Answer: 1517*	
Question 2	In what year did Nigeria gain its independence from Great Britain? *Answer: 1960*	
Question 3	What former First Lady is credited with having coined the phrase "Absence makes the heart grow fonder"? *Answer: Eleanor Roosevelt*	
Question 4	In the story of the Trojan War, what son of King Priam dies in battle with the Greek warrior Achilles? *Answer: Hector*	
Question 5	In what year did Abraham Lincoln deliver the Gettysburg Address? *Answer: 1863*	
Question 6	In what year was Norwegian Viking Erik the Red lead the first European settlement of North America? *Answer: 986*	
Question 7	The stock market crash of 1929, remembered as the beginning of the Great Depression, occurred most dramatically on what day? *Answer: "Black" Monday (October 28, 1929)*	
Question 8	During the Second Punic War, what Carthaginian general led his army on a famous crossing of the Alps? *Answer: Hannibal*	
Question 9	What former U.S. vice president killed Alexander Hamilton in a duel? *Answer: Aaron Burr*	
Question 10	John Adams and Thomas Jefferson, the 2nd and 3rd Presidents of the United States, both died on what day (date and year)? *Answer: July 4, 1826*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 13	**Topic:** Sports	**Difficulty:** Easy
Mean: 8.537	**Median:**: 10	**Variance:** 7.289
Question 1	What does N.F.L. stand for ? *Answer: National Football League*	
Question 2	The Tour de France, which Lance Armstrong has won 7 times, is the premier competition in what sport? *Answer: Cycling*	
Question 3	What sport did Babe Ruth play? *Answer: Baseball*	
Question 4	What state do the Yankees play for? *Answer: New York*	
Question 5	What does WNBA stand for? *Answer: Women's National Basketball Association*	
Question 6	What Chicago Bulls guard wore number 23 and led his team to 6 championships in the 1990's? (Hint: his last name is Jordan) *Answer: Michael Jordan*	
Question 7	Ervin ŞMagicŤ Johnson plays what sport? *Answer: Basketball*	
Question 8	Mohammed Ali and Mike Tyson have each held the Heavyweight Championship Title in what sport? *Answer: Boxing*	
Question 9	What sport has been most famously played by Tiger Woods, Jack Nicklaus, and Arnold Plamer? *Answer: Golf*	
Question 10	In baseball, what is it called when a player hits a ball over the outfield wall and into the stands? *Answer: A home run*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 14	**Topic:** Sports	**Difficulty:** Medium
Mean: 4.598	**Median:**: 4	**Variance:** 9.182
Question 1	March Madness refers to what college sport's tournament? *Answer: Basketball*	
Question 2	What team did Larry Bird play for? *Answer: Celtics*	
Question 3	Which golf tournament's winner is traditionally awarded a green jacket? *Answer: Masters*	
Question 4	What sport does Sheryl Swoopes play? *Answer: Basketball*	
Question 5	What sport does Oscar De LaHoya participate in? *Answer: Boxing*	
Question 6	What city's NFL team is known as the Cowboys? *Answer: Dallas*	
Question 7	Who holds the NHL record for the most goals in a season? *Answer: Wayne Gretzky*	
Question 8	How many times was Michael Jordan MVP of the NBA Finals? *Answer: 6*	
Question 9	What is it called when a race horse wins the Kentucky Derby, the Preakness Stakes, and the Belmont Stakes? *Answer: Triple Crown*	
Question 10	What countryŠs team won the 2002 World Cup? *Answer: Brazil*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 15	**Topic:** Sports	**Difficulty:** Hard
Mean: 0.195	**Median:**: 0	**Variance:** 0.159
Question 1	Who is the only player to have scored 61 points in an NCAA Basketball Tournament game? *Answer: Austin Carr*	
Question 2	Who was the first MVP of the NBA? *Answer: Bob Pettit*	
Question 3	What college won the first Women's NCAA Basketball Tournament? *Answer: Louisiana Tech*	
Question 4	Who was the first NHL player to win the Conn Smythe Trophy? *Answer: Jean Beliveau*	
Question 5	Which College or University has won the most football bowl games? *Answer: University of Alabama*	
Question 6	Who won the first Super Bowl? *Answer: Green Bay Packers*	
Question 7	Who was the only unseeded man to win the Wimbledon singles title? *Answer: Boris Becker*	
Question 8	Who was the first jockey to ride 7,000 winners? *Answer: Willie Shoemaker*	
Question 9	Who won the Tour de France in 1997? *Answer: Jan Ullrich*	
Question 10	Who is the only hockey player to score at least 50 goals in 9 consecutive seasons? *Answer: Mike Bossy*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 16	**Topic:** Science	**Difficulty:** Easy
Mean: 8.976	**Median:**: 10	**Variance:** 3.629
Question 1	Where in the human body is the cerebellum located?	
	Answer: Brain (or Head)	
Question 2	The average length of a human pregnancy is how many months?	
	Answer: 9	
Question 3	Where in the human body is food digested?	
	Answer: the stomach (or intestines)	
Question 4	What is the smallest prime number greater than 3?	
	Answer: 5	
Question 5	What is the chemical symbol for water?	
	Answer: H_2O	
Question 6	The theory of evolution by means of natural selection is attributed to whom?	
	Answer: Darwin	
Question 7	The ____________ is the star at the center of our solar system.	
	Answer: Sun	
Question 8	An octopus has how many arms?	
	Answer: 8	
Question 9	What is the major pumping organ of the human circulatory system?	
	Answer: Heart	
Question 10	Carnivores are animals that eat meat, while herbivores are animals that eat what?	
	Answer: Plants (or Vegetation)	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 17	**Topic:** Science	**Difficulty:** Medium
Mean: 6.427	**Median:**: 7	**Variance:** 8.001
Question 1	What is the name for a group of lions? *Answer: pride*	
Question 2	What are the only mammals that truly fly? *Answer: bats*	
Question 3	Antibiotics kill what class of pathogen? *Answer: bacteria*	
Question 4	What African predator is the fastest land animal? *Answer: cheetah*	
Question 5	Deoxyribonucleic acid is better known as what? *Answer: DNA*	
Question 6	What is the largest species of whale? *Answer: blue whale*	
Question 7	What durable substance give human ears and noses their shapes? *Answer: cartilage*	
Question 8	What is normal human body temperature? *Answer: Roughly 98 F or 37 C*	
Question 9	Dobermans, schnausers, and alsations are all breeds of what kind of animal? *Answer: Dog*	
Question 10	How many meters are there in a kilometer? *Answer: 1000*	

Table 7: The 18 quizzes used in the experiment. (Continued on the next page.)

Quiz: 18	**Topic:** Science	**Difficulty:** Hard
Mean: 0.939	**Median:**: 1	**Variance:** 1.268

Question 1	Who is credited with inventing the wristwatch in 1904? *Answer: Louis Cartier*
Question 2	Laudanum is a form of what drug? *Answer: opium*
Question 3	The psychoactive ingredient in marijuana is THC. What does THC stand for? *Answer: delta-9-Tetrahydrocannabinol*
Question 4	What chemical element has the atomic number five? *Answer: Boron*
Question 5	The study of the structural and functional changes in cells, tissues and organs that underlie disease is called what? *Answer: Pathology*
Question 6	What does the suffix “itis” mean? *Answer: Inflammation*
Question 7	The bilby, bandicoot, and quokka are all representatives of what mammalian subclass? *Answer: Marsupials*
Question 8	Which one of the 50 United States is the only one never to have experienced an earthquake? *Answer: North Dakota*
Question 9	What evolutionary biologist wrote, “Creation science has not entered the curriculum for a reason so simple and so basic that we often forget to mention it: because it is false.”? *Answer: Stephen Jay Gould*
Question 10	What is the single most diverse phylum within the animal kingdom? *Answer: Arthropoda (arthropods, including crustaceans, insects, and spiders)*

Table 7: The 18 quizzes used in the experiment.