Experimental Laboratory Working Paper Series #2003-001
Dept. of Economics, Andrew Young School of Policy Studies, Georgia State University
Revised January 2005

KNOW THYSELF
INCOMPETENCE AND OVERCONFIDENCE

Paul J. Ferraro

Department of Economics

Andrew Young School of Policy Studies

P.O. Box 3992

Georgia State University

Atlanta, GA 30302-3992

Economic analyses of asymmetric information typically start with the assumption that individuals know more about their own characteristics than outside observers. This assumption implies that individuals can accurately assess their own competence in a given domain. However, individuals can only judge their competence if they are sufficiently competent. The relationship between competence and self-awareness explains a great deal of the overconfidence observed among economic agents. More specifically, overconfidence is inversely proportional to competence. Through a series of experiments and analyses of field data, the link between incompetence and overconfidence is confirmed and its implications for economic theory are explored.

JEL Codes: C72, C91, C93, D82

Key Words: overconfidence, competence, asymmetric information, gender, economic experiment

"He's stupid, but he knows he's stupid, and that almost makes him smart." Third Bass, *The Cactus Album*

I. Introduction

Classic economic analyses of asymmetric information typically start with the assumption that an individual knows more about his or her own characteristics (e.g., work ability, driving skill) than an outside observer (e.g., employer, insurance company). In particular, the widespread use of self-selection constraints in economic theory is based on the premise that individuals can correctly ascertain their own "types." An individual's type may not be observable to other market participants, but it is always assumed to be observable to the individual. Thus, in games of imperfect information when Nature chooses realizations of the random variable that determines each player's type, each player is assumed to observe the realization of his own random variable.

Psychologists, however, have long argued that people do <u>not</u> know their own type, but rather hold overly-favorable views of their abilities in many social and intellectual domains (e.g., 82% of drivers believe they are among the safest 30% of drivers; Svenson, 1981). Recent experimental research has been more precise about the nature of this overconfidence: overconfidence occurs because incompetent individuals lack the cognitive skills to recognize their own incompetence. Kruger and Dunning [1999] argue that the skills that *engender* competence in a domain are the same skills required to *evaluate* competence in that domain – one's own competence or anyone else's. In other words, if you are incompetent in a particular domain, you lack the ability to recognize your incompetence.[1] The Kruger-Dunning argument implies more than random error in self-assessment across the population. It implies that, in a

given domain, less competent individuals (e.g., poor performers) are unaware of their true types, while more competent individuals (e.g., high performers) know their types.

Some economists have incorporated overconfidence in their analysis. For example, see Golec and Tamarkin [1995], Kyle and Wang [1997], Daniel, Hirshleifer and Subrahmanyam [1998], Camerer and Lovallo [1999], Odean [1999], Barber and Odean [2001], Bernardo and Welch [2001], and Gervais and Odean [2001]. In contrast to psychologists, economists typically assume (sometimes implicitly) that all agents are overconfident or that overconfident personalities are randomly distributed through a population.[2] Moreover, these studies do not directly observe overconfidence: they either assume it exists or measure it indirectly (e.g., observations of excessive trading).

Thus psychologists have argued that incompetent agents are more likely to exhibit overconfidence whereas economists either (a) assume every agent knows his type or (b) do not specify the origin and distribution of overconfidence among decision makers. This difference between the perspectives of the two disciplines is important. If the psychologists are correct, many strands of the economic literature may need to be re-evaluated (e.g., models of screening contracts). Moreover, the perspective of psychologists implies far more structure on the errors that people make in their self-evaluations than economists have assumed to date. Thus, if the incompetence-overconfidence link is correct, economic models of individual decision-making may greatly benefit.

The results of psychologists, however, can be questioned on a number of grounds including the absence of salient incentives for accurate self-assessments and the absence of feedback to experimental subjects. On the other hand, the results of economists can be questioned based on implicit assumptions of the ability of all players to ascertain their type or the

lack of attention paid to the connection between competence and self-awareness. The purpose of this paper is to test hypotheses related to competence and self-awareness using a design that addresses the aforementioned weaknesses in the psychology and economics literature.

In Section II, the variables of interest are defined and seven hypotheses are presented. In sections III-V, field experiments and field data are used to test these hypotheses. The paper concludes by discussing the implications of the results for economic theory.

II. Hypotheses

In the psychology literature, "overconfidence" is used in two ways: (1) to characterize people who overestimate their ability and (2) to characterize people who see an outcome as more certain than it really is [Hvide 2002]. This paper focuses on overconfidence in the estimation of one's abilities.[3] In economic interactions, one's type may be determined by one's absolute ability (e.g., I can correctly complete 80% of the task) or by one's relative ability (e.g., I can complete the task better than 80% of the other players in the game). A subject's accuracy in assessing his or her absolute performance is defined as the absolute value of the difference between a subject's self-assessed performance and the subject's actual performance on a task (ABACCURACY). A subject's accuracy in assessing his or her relative performance is measured as the absolute value of the difference between a subject's predicted percentile ranking and the subject's actual percentile ranking (RELACCURACY). Overconfidence in absolute performance (ABCONFIDENCE) and in relative performance (RELCONFIDENCE) is measured by removing the absolute value operators: positive values indicate overconfidence.

Based on previous research (see *Introduction*), seven hypotheses are tested:

H1a: The less competent a subject is in a domain, the less accurate is the subject's evaluation of his or her absolute performance and the more overconfident the subject is (i.e., the larger and more positive is ABACCURACY and ABCONFIDENCE).

H1b: The less competent a subject is in a domain, the less accurate is the subject's evaluation of his or her relative performance and the more overconfident the subject is (i.e., the larger and more positive is RELACCURACY and RELCONFIDENCE).

H2a: Feedback on performance and the distribution of performances causes subjects to be more accurate and less overconfident in assessing their absolute performances over time. These improvements in self-assessment, however, are greatest among more competent subjects because they are more capable of correctly interpreting signals about their performance.

H2b: Feedback on performance and the distribution of performances causes subjects to be more accurate and less overconfident about their relative performances over time. Improvements in self-assessment, however, are greatest among more competent subjects because they are more capable of correctly interpreting signals about their performances.

H3a: Women are less overconfident about their absolute performances than men (i.e., ABCONFIDENCE smaller among women).

H3b: Women are less overconfident about their relative performances than men (i.e., RELCONFIDENCE smaller among men).

H4: Overconfidence in relative performances (i.e., how well a subject did compared to other subjects) is not driven by the subject's belief that other players are less competent than the subject, but rather by a subject's belief that she is more competent that she actually is. No one has attempted to differentiate between these two motivations for overconfidence in relative performances, but understanding such motivations is important in contexts such as

excess entry [Camerer and Lovallo]. The provision of information about the performance of others would decrease overconfidence in the former case, but not in the latter.

H5: In an insurance contract market, adverse selection will be attenuated by the link between competence and self-awareness.

H6: Traditional theory predicts that married women with a high probability of divorce will behave differently from women with a low probability of divorce in domains in which the probability of divorce affects payoffs. Theory that incorporates the link between competence and self-awareness predicts that high-risk married women are unaware of their higher risk for divorce and thus will not behave substantially different from low-risk married women.

H7: In a dynamic signaling game, decisions made by subjects will be consistent with a separating Nash equilibrium when subjects are told their player type by the experimenter, but substantially fewer decisions will be consistent with this equilibrium when subjects must infer their type based on their ability in a specific domain.

III. Experiment 1: Competence and Self-awareness (H1-H4)

We begin by conducting a framed field experiment [Harrison and List 2004] that confirms and elaborate the Kruger-Dunning results through two important experimental extensions: (1) subjects are given salient incentives to evaluate their absolute and relative performances on a task (payoffs in Kruger-Dunning were not conditional upon accuracy); and (2) subjects receive feedback on performance and repeat the task several times (subjects in Kruger-Dunning performed a task only once). The latter extension is motivated by the observation that, in naturally-occurring settings, individuals typically complete tasks more than

once and receive signals about their performance (e.g., an individual may think he is a great salesman, but after a month of making few sales, he may update his self-evaluation).

After a pre-test (Fall 2001), the experiment was run in three introductory microeconomics classes (Spring and Fall 2002). Students in these classes took three non-cumulative multiple-choice exams that made up the majority of their final grade in the class (data were not collected on the final exam, which contained new material and old material). Immediately after completing each exam, subjects were asked to estimate the number of questions that they answered correctly on the exam and their percentile ranking on the exam (percentile was defined in writing and orally to ensure subjects understood the concept). In Class 1, subjects who were closest to predicting their true number of correctly answered questions received $25 each and subjects who were closest to predicting their true percentile ranking received $25 each. In Classes 2 and 3, subjects received $5 if their prediction was accurate, $4 if it was within one question of the true number of correctly answered questions, and $1 if it was within two questions of the true number of correctly answered questions (for percentile: $5 if within one percentile point, $4 if within two percentile points and $1 within three percentile points).[4]

One hundred and five subjects took all three exams. Subjects who did not take all three exams were dropped from the analysis in order to focus on subjects who had received the same level of feedback and had the same level of experience taking exams in their class.[5] After each exam, subjects were told the number of questions they answered correctly and descriptive statistics about the grade distribution (e.g., mean, median, letter grade frequencies). Subjects were not told their exact percentile ranking. Based on the descriptive statistics, however, they could infer their approximate position relative to the 50^{th} percentile. Subjects were not told their percentile ranking because we wanted to maintain the classroom environment as natural as

possible. Students are typically not told their exact percentile ranking after an exam. In most naturally-occurring contexts, one may be able to assess own absolute performance, but one typically has only a general idea of the overall distribution of performances.

The environment of a real-world economics class rather than a laboratory was used for four reasons: (1) a field experiment looking at overconfidence and economic behavior had not yet been conducted, (2) the class environment ensured that the domain in which subject's self-assessments were being elicited was not an artificial one in which subjects had little experience or interest in performing well,[6] (3) the class provided naturally-occurring feedback with its concomitant time delays and potential for inter-player communication, rather than the typical "stationary replication" of laboratory experiments in which subjects perform a task repeatedly in rapid succession; and (4) a subject's ability to reason on economics exams provides a close parallel to the subject's ability to reason in naturally-occurring economic decisions.

To test hypotheses *H1*-*H3*, the subjects' absolute performances and relative performances were contrasted with the subjects' self-assessments. A subject's exam score reflects the subject's competence. The qualitative results are unchanged by the use of a subject's percentile ranking or final grade as the measure of competence.

A. Summary Statistics

Prior to performing regression analyses on the panel data from the three exams, we present summary statistics. On the first exam, the mean score was 69% and the mean predicted score was 77%, which are significantly different (one-sided, paired t-test: $p < 0.001$). On the first exam, 80% of the subjects believed they were above the 50^{th} percentile (including almost 60% of the bottom quartile). Thus, on average, subjects were overconfident.

Average behavior, however, masks important differences across competence levels. Table I presents statistics on subjects' accuracy and overconfidence in their absolute and relative performances. Subjects are assigned to three arbitrarily chosen categories: low competence (less than 65% correct), middle competence (65% to 80% correct) and high competence (greater than 80% correct). The first row reports the mean overconfidence in absolute and relative performances. Positive values indicate overconfidence; negative values indicate underconfidence. The second row reports mean accuracy in absolute and relative performances. The last row reports p-values for a t-test that tests whether the observed overconfidence (or underconfidence) is significantly different from zero in each competence category.

The results suggest that, whether one is measuring performance in absolute terms or in relative terms, overconfidence decreases and accuracy increases as competence increases. For example, on the first exam, low-competence subjects predicted, on average, that they scored substantially higher than they actually did. Their mean score was 56.5%, but they believed, on average, that they scored 70.1% (i.e., they performed in the bottom quintile, but believed they scored slightly above average). On the same exam, low-competence subjects also overestimated their relative performance: their mean percentile ranking was the 18th percentile, yet they believed they performed, on average, at the 54th percentile. Poor-performing subjects did not necessarily believe they had a top-quartile performance (although 25% did). They did, however, fail to realize the magnitude of their shortcomings: two-thirds believed they were above the 50th percentile. In contrast, the high-performing subjects were, on average, the least overconfident of the subject pool and the most accurate. Thus the summary statistics provide some support for *H1*: the more competent the subject, the more accurate and less overconfident the subject was.

[Table I about here]

The summary statistics also provide support for *H2a*, which claims that overconfidence in one's absolute performance will decrease over time and that this decrease will be greatest among more competent subjects because they are more capable of interpreting signals about their performance. The subject pool can be sorted by their total scores on all 3 exams. The average subject in the bottom 50% of scorers was overconfident by 10.4% (ABCONFIDENCE) on the first exam and 9.7% on the third exam, implying that the average level of overconfidence did not change over time (paired t-test p-value = 0.239). In contrast, the average subject in the top 50% was overconfident by 6.8% on first exam, but by the third exam was slightly underconfident by -0.6% ($p<0.0001$). The 32 subjects (31% of the pool) who scored less than 70% on all 3 exams provide further support that less competent subjects did not reduce their overconfidence over time despite receiving repeated signals that their self-assessments were grossly overconfident. These subjects overestimated their score by an average of 12% on the first exam, 14% on the second exam and 12% on the third exam and they became <u>less</u> accurate over time: moving from a mean error (ABACCURACY) of 13% to one of 16% ($p=0.07$).

There is no support, however, for *H2b*, which predicts declining overconfidence in assessments of relative performance. By the <u>third</u> exam, 83% of all subjects still believed they performed above the 50^{th} percentile. The average subject in the bottom 50% of scorers went from being about 30 percentile points overconfident on the first exam to about 25 percentile points on the third exam ($p=0.130$). The average subject in the top 50% went from being overconfident by 11 points on the first exam to 7 points on the third exam ($p=0.234$; the variance of the top-scorers' self-assessments did, however, decline significantly). The average subject who scored less than 70% on all three exams consistently overestimated his percentile ranking by about 30 percentile points.

Support for *H3*, which suggests that women are less overconfident than males, can be found by observing that on the first exam, 82% of males (n=44) and 72% of females (n=61) were overconfident, yet there was no significant difference in the average scores of the two sexes. Males, on average, overestimated their absolute performance by 11% compared to 6% for females (one-side t-test: p = 0.032).

B. Regression Results

In order to control for subject differences and differences in each subject's competence across exams, a random effects model using the GLS estimator was employed.[7] There are four dependent variables (*Y*): (1) ABCONFIDENCE (Predicted Percentage of Correct Answers – Actual Percentage of Correct Answers), (2) ABACCURACY (Absolute Value of ABCONFIDENCE), (3) RELCONFIDENCE (Predicted Percentile Ranking – Actual Percentile Ranking on Exam (in decimals)), and (4) RELACCURACY (Absolute Value of RELCONFIDENCE). Each dependent variable is regressed on the same variables:

$$Y_{it} = \beta_0 + \beta_1 Score_{it} + \beta_2 Ex2 * Score_{it} + \beta_3 Ex3 * Score_{it} + \beta_4 Female_i + \beta_5 Econ_i + \beta_6 Class2_i + \beta_7 Class3_i + \upsilon_i + \varepsilon_{it}.$$

Y_{it} represents subject's *i*'s dependent variable at time *t* (*i* =1,…,105; *t* = 1, 2, 3), υ_i is the subject-specific residual, ε_{it} is the standard residual associated with the panel, *Score* represents the subject's score on the exam, *Ex2* and *Ex3* are dummy variables for the second and third exam, *Female* is a dummy variable for female subjects, *Econ* is a dummy variable for subjects who had taken both a college and a high school economics course before the current class, and *Class2* and *Class3* are dummy variables for two out of the three classes from which data were obtained.

We begin by examining subjects' overconfidence and accuracy in their absolute performances. If incompetence drives inaccurate self-assessments and overconfidence when subjects evaluate their absolute performance on an exam (*H1a*), one would expect the coefficient on the exam score (*Score*) to be negative, significant and substantial in both regressions. The results in columns (1) and (2) of Table II confirm this expectation. Alternate specifications or estimators do not change this conclusion.

[Table II about here]

H2a states that subjects will become more accurate and less overconfident after receiving feedback on their performance, and these improvements are conditional on subject competence. The hypothesis implies that the coefficients on the interaction terms *Ex2*Score* and *Ex3*Score* should be negative, significant and substantial in both regressions and the coefficient on *Ex3*Score* should be significantly larger than the coefficient on *Ex2*Score* (by the third exam, subjects have had more opportunities to receive feedback on their economic reasoning skills).

In the regression for accuracy, there is no evidence that subjects become more accurate in their self-assessments on the second and third exams. In the regression for overconfidence, the coefficient on *Ex2*Score* is not significantly different from zero. The coefficient on *Ex3*Score* is negative and significantly different from zero (and from the coefficient on *Ex2*Score*), but its magnitude is not substantial, particularly for the poor performers. The coefficient on *Ex3*Score* suggests that there are small reductions in overconfidence over time and the reductions are conditional on competence: although the less competent subjects have the greatest gap to close between actual and perceived performance, they are least able to close it. For example, the model predicts that the average poor-performing male subject from Class 1 with no economics

experience who consistently scored 65% on the exams would overestimate his performance by 15% on the first exam and almost 11% on the third exam.

About 40% of the subjects had already taken both a college-level and a high school-level economics class (including 50% of the subjects who scored below 65% on the first exam). These subjects had thus already received substantial feedback on their performance in the discipline of economics. The insubstantial and insignificant coefficient on *Econ*, however, suggests that substantial previous exposure to economics classes had no effect on subject self-assessments. Adding interaction terms for the exam numbers with *Econ* yields the same insignificant result.

The lack of strong evidence for improved accuracy and substantially lower overconfidence with feedback is particularly surprisingly in light of the clarity of the signals that subjects receive. There are few opportunities in life to receive explicit numerical measures of one's performance. Moreover, subjects had a lot of time to ruminate over the signals (there were periods of weeks between exams) and had an opportunity to discuss with other "players" what had transpired in previous rounds of the game. These results are, however, consistent with other studies that suggest overconfidence persists over time [e.g., Camerer and Lovallo].

The coefficients on *Female* are negative and significantly different from zero, thus providing support for *H3a*, which states females are more accurate and less overconfident than males on average. Although the difference between males and females is not insubstantial, the average difference in overconfidence among poor performers and high performers is much larger than that between males and females.

The analysis of subjects' overconfidence and accuracy in their relative performances yields similar results. Columns (3) and (4) in Table II support *H1b*. The coefficient on SCORE is

negative, substantial and statistically different from zero in both regressions: the more competent the subject, the more accurate and less overconfident the subject was.

There is no evidence that subjects become more accurate in their self-assessments of their relative performances over time (*H2b*). There is weak evidence that subjects become less overconfident by the third exam, but the coefficient on *Ex3*Score* is not substantial and the magnitude of the reduction is conditional on competence. Furthermore, previous exposure to economics had no discernible effect on a subject's self-assessment of relative performance. The data do, however, support *H3b*: females were significantly more accurate and less overconfident than males in assessing their relative performances.

We now turn to *H4*. Overconfidence in relative performance can stem from two sources: (1) an overconfident assessment of one's own absolute performance; and (2) an inaccurate assessment of the absolute performance of others on the same task. The data support (1) rather than (2). Had subjects correctly answered the number of questions they believed they answered correctly, their predicted percentile ranking would have been quite accurate: on average, subjects were off by less than one percentile point. In other words, the mean difference between subjects' predicted percentiles and the percentiles they would have achieved had they answered as many questions correctly as they thought they had answered correctly was not significantly different from zero. For example, consider a subject in Class 1 who answered 60% of the questions correctly on the first exam and was at the 20^{th} percentile. The same subject believed he answered 80% of the questions correctly and was at the 70^{th} percentile. Had the subject actually answered 80% of the questions correctly, he would have been at the 72^{nd} percentile. Thus, on average, subjects' overconfidence in their relative performances does not seem to derive from beliefs that

they are more competent relative to the other subjects in the classroom, but rather from beliefs that they are more competent on an absolute scale than they actually are.[8]

Despite the stark results from the experiment, one might argue that the results derive from the experimental design rather than the hypothesized link between competence and self-awareness. When overconfidence is measured by the differences between the predicted and actual scores (or percentiles), the maximum possible overconfidence of high-scoring subjects is lower than that of low-scoring subjects (i.e., the results might represent a regression to the mean). Such an argument, however, is not consistent with improved self-awareness by top performers over time and the persistent absence of self-awareness among the poor performers.

Moreover, Kruger and Dunning conducted an experiment in which the top quartile and bottom quartile performers from a previous experiment were invited to return to the lab after several weeks. They gave each group the tests of five of their peers to evaluate and then asked them to re-evaluate their own tests. The five tests were selected to have the same mean and standard deviation as observed in the subjects' sessions and subjects were told this. The bottom quartile subjects not only were unable to judge the quality of the five tests, but there was evidence that they, on average, revised their own test evaluations upwards after reviewing the performances of their peers. In contrast, top quartile subjects became more accurate in their self-evaluations after viewing the performances of their peers. Thus, less competent individuals failed to gain insight into their own incompetence by observing the behavior of others.

There are two other plausible explanations of the Experiment 1 results: (1) utility may be increasing in perceived ability and the payoffs in the experiment were not high enough to induce low performers to abandon their high-performing self-image; and (2) causality runs in the reverse direction whereby overconfident subjects do not invest in improving their competence.

To evaluate these other conjectures, we conducted two other experiments (not reported here; see referee's Appendix). In the first, subjects evaluate the quality of two exams completed by someone else (thus breaking the connection between self-image and the evaluation). Less competent subjects were more likely to overestimate the quality of the poor exams in their possession and underestimate the quality of the good exams (reversing the quality categories in most cases). In the second experiment, we induce incompetence in competent subjects (by increasing the complexity of the task in a given domain) and demonstrate that subjects who formerly were not overconfident become overconfident.

IV. Experiment 2: An Insurance Market (H5)

In an attempt to continue to maintain some control over the data generation process, but to allow more realistic responses to economic environments, we return to another Introduction to Microeconomics class (Fall 2003) and create a real market for grade insurance.[9] Sixty-four students were registered for a class in which 3 regular exams and a final cumulative exam were offered. After each of the first 3 exams, students were asked to estimate their absolute and relative performances using the instructions from Experiment 1. After each exam, students discovered their scores and the mean, median, standard deviation, and distribution of letter grades. The information was announced in class and posted on the class website.

Thus, prior to the final exam, students had not only received three signals pertaining to their and others' performances in answering microeconomics questions, but they also had three opportunities to evaluate their absolute and relative performances and to observe how accurate their self-evaluations were (the results of these evaluations follow Experiment 1's pattern; e.g., on the third exam, only 5% of subjects believed they scored below the 50^{th} percentile).

On their final exam, students were offered insurance contracts that could be purchased with points deducted from their final exam. In the event of a predefined "accident" (poor grade), the contract would pay the purchaser in points on their final exam. The contracts, which are detailed below, were designed so that if subjects were aware of their likelihood of an accident, they could select a contract that would offer them a positive expected gain. Traditional economic theory predicts that adverse selection will be a large problem in such a market. Theory that incorporates the link between competence and self-awareness predicts much weaker adverse selection effects because many high-risk individuals are unaware that they are high-risk.

To focus on the degree to which adverse selection affects outcomes in this market, we (1) introduced the insurance market in a way that reduces the likelihood of moral hazard and (2) offered insurance contracts that, if subjects were to correctly ascertain their type, would cause the insurer (the professor) to lose a substantial number of points.

To reduce the incentive for students to reduce their studying effort because of their access to insurance (moral hazard), we did not inform students of the availability of grade insurance until the official starting time of the final exam. However, we were concerned that students might be confused if confronted by a novel grade insurance market immediately before the final exam. Thus, during a review session for the third exam, the professor related a story about a colleague at a different university who offered grade insurance to his economic students where students could spend points from their exam to insure against a bad grade (a true story). Not surprisingly, several students in the class stated aloud that their professor should offer such insurance on the third exam. The same anecdote was repeated after the third exam results were announced in class. Attendance was taken in both of these classes; all but one of 59 students who participated in the market were present when the grade insurance market was described.

At the beginning of the final exam period, but before any student had seen the final exam, the professor announced that grade insurance would be offered. Students had a sheet of paper at their desks that described the insurance program (see referee's Appendix). Each student could choose to purchase Contract A, Contract B, or no contract (a choice had to be made). Contract A required payment of a 10-point premium and paid out 20 points in the event that the student fell below the 50th percentile on the 100-point final exam. Contract B required payment of a 2-point premium and paid out 4 points in the event that the student fell below the 75th percentile and at or above the 50th percentile. Students were orally advised that should they purchase a contract, points would indeed be taken from their exam to cover the premium and points would be added to the exam if and only if their performance matched the "accident" described in their policy.

Sixty-one students took the final exam, but 2 students showed up late and were not given the opportunity to purchase insurance. Two students took only 1 of the 3 previous exams and are excluded from the analysis.[10] The analysis thus uses a sample of 57 subjects; 46 took 3 exams and 11 took 2 exams prior to the final exam.

Traditional economic theory predicts that many students would purchase Contract A. About 50 percent of the students are guaranteed to fall below the 50th percentile (students were aware that slightly more than 50 percent fell below the 50th percentile on the three previous exams). Moreover 19 of the 57 subjects scored below the 50th percentile on <u>every</u> previous exam (mean = 18th; max = 44th), while 9 more of the subjects scored below the 50th percentile on two out of three of their previous exams. Twenty-eight students scored consistently below the 60th percentile on every previous exam (mean = 26th).

Three attributes of the market lead one to conclude that fewer students would purchase Contract B: (1) by definition, fewer students could perform in the percentile range deemed an

"accident," (2) the net payoff was small, and (3) few students consistently performed in this range in previous exams, thus making predictions more difficult (only 3 scored in this range on each previous exam and 7 more scored in this range on 2 out of 3 exams).[11]

Economic theory that incorporates the link between self-awareness and competence, however, predicts that (1) Contract B would be as popular, or more popular, than Contract A, and (2) the majority of individuals would not purchase any insurance. The expected popularity of Contract B and no insurance derives from two sources: (1) most of the poor performers would be unaware of their likelihood of performing poorly on the final exam and believe instead that they would score above average on the exam (thus making Contract B or no contract the best choice) and (2) many of the average or slightly above average performers believe they would score in the top quartile (thus making no contract the best choice).

Only 19 of 57 subjects purchased an insurance contract. Thirty-one subjects performed below the 50th percentile, but only 6 of these subjects purchased Contract A. An additional subject who scored above the 50th percentile (as she had on 2 out of the 3 previous exams) also purchased Contract A. In contrast, 12 subjects purchased Contract B, of whom 10 scored below the 50th percentile (mean percentile = 22nd); only two scored (barely) above the 50th percentile.

If subjects had perfect information about their final exam outcomes and purchased the appropriate insurance contracts, they would have purchased 31 Contract As and 16 Contract Bs. In this case, the insurer would incur a net lost of 342 points. Actual purchase decisions generated a net loss of only 30 points.

Clearly, however, not every student who fell below the 50th percentile on the final exam could be considered, *ex ante*, a "high risk" for performing poorly and thus should have purchased Contract A. In Experiment 1, we classified players according to their competence on a given

exam, which was observable to the experimenter *ex post*. In Experiment 2, the insurance decision has to be made prior to beginning the exam just as purchasing car insurance happens prior to going out on the highway. Thus each player's "type" is unobservable to the experimenter. Just as a driver's road accident does not prove the driver is high-risk, a subject's final exam performance below the 50th percentile does not imply the subject was high-risk and thus should have purchased insurance. However, just as insurance companies use driving histories as proxies of competence, we can use subjects' past exam performances.

Of the 19 people who scored below the 50th percentile on every previous exam, only 4 purchased Contract A, while 6 purchased Contract B and the others chose not to purchase a contract.[12] Of the 9 additional students who scored below the 50th percentile on 2 of their 3 previous exams, 7 scored below the 50th percentile on the final exam: none of these students purchased Contract A and 3 purchased Contract B. Adopting a less restrictive definition of who was at risk, only 5 of the 28 students who scored consistently below the 60th percentile on every previous exam purchased Contract A; 11 chose Contract B. Twenty of them scored below the 50th percentile on the final exam (mean of 28 students = 32nd). If all had purchased Contract A, the insurer would have lost 120 points.

As with any empirical analysis, there are alternative explanations of the results. Students who consistently perform poorly on exams may be risk-preferring and thus forgo insurance. This explanation is plausible but unlikely. Most experimental and field data suggest that few students are risk-preferring. As in the previous experiment, the expected payoff from the insurance contracts may simply not been have been high enough for a subject to abandon his or her positive self-image. This alternative explanation was addressed in Section III.

In order to maintain some control over the data generating process and allow for easier comparability with Experiment 1, the insurance market in Experiment 2 was implemented in the domain of "answering microeconomics question." However, the decision to purchase an insurance contract was a real economic decision with clear and salient incentives for cognitive investment. In the next section, further control over the data generating process is ceded to examine behavior in another domain in which one might expect the link between competence and self-awareness to be important.

V. Field Data: Retirement Investment Behavior of Married Women (H6)

Baker and Emery [1993] observed that individuals have accurate perceptions of the likelihood of divorce and its financial and personal effects in the larger population. The same individuals, however, assign extremely low probabilities to their own likelihood of divorce and are overly optimistic about the likely effects on their lives if their marriages fail.

Like success in any domain, success in marriage requires competence in a variety of skills (e.g., mate selection, bargaining, cooperation, stress management). An individual's expectation of how long a marriage will last affects important decisions in the individual's life [Becker et al. 1977]. If, however, those individuals who are most likely to divorce are substantially overconfident about their ability to avoid a costly divorce, they will not behave rationally when addressing decisions such as the number of children to have, the necessity of pre-nuptial agreements, or the desirability of starting or ending careers during marriage.

Data on the "marital skills" of large numbers of individuals do not exist. Insurance companies face a similar problem in the car insurance market: they do not have observations on a driver's skill. They do, however, have observations on an individual's driving record in which

the number of accidents and traffic violations in a driver's life are a strong indicator of the quality of a driver's skill. In a similar fashion, the number of divorces in an individual's life may be a strong indicator of the quality of his or her marital skills. As noted by Becket et al. [p.1157], people dissolving their marriages are not selected at random, but are selected by characteristics that increase their probability of dissolution.

The MINT I model was developed to project retirement income for current and future Social Security beneficiaries [Butrica *et al.* 2001]. As an intermediate step, the model uses data from the U.S. Census Bureau's Survey of Income and Program Participation (SIPP) 1990-93 panel to estimate a divorce hazard model (the likelihood of transitioning out of marriage). The number of times an individual has previously divorced has a significant and substantial effect on increasing the probability of divorce. A previous divorce doubles the hazard rate of divorce in the current marriage and two or more previous divorces lead to a four-fold increase in the hazard rate. Similar results have been observed in other studies [e.g., Bramlett and Mosher 2001; Becker et al. 1977; Monahan 1958, 1959].

Becker et al.'s "rational divorce model" predicts (p.1144, 1179) that, after receiving a (noisy) signal about their marriage skills, self-aware divorced people should become aware of their high probability of divorce (and its costs) and be less likely to marry in the future. Butrica *et al.*'s estimated hazard model for marriage, however, suggests that the aphorism "once bitten, twice shy" does not apply to the average individual in the marriage market: coefficients on the dummy variables for previous marriages are positive and substantial. Despite their high probability of a future divorce, previously divorced individuals are more likely to re-enter a marriage contract, and their hazard rate increases with the number of previous divorces.

The same empirical pattern, however, has an alternative explanation: previously divorced individuals may be self-aware of their higher probability of divorce, but simply have strong preferences for the act of marriage. If they were self-aware, however, their decisions while married should reflect their higher probability of divorce.

To test this conjecture, we examine the decisions of married women to participate in individual retirement accounts. Such accounts are not held jointly by the couple, but rather in the name of an individual. Married men, on average, are more likely to have retirement accounts than married women [e.g., Even and Macpherson 2000]. Although married women may view the funds in retirement accounts as joint income for retirement, divorce destroys the joint nature of these funds. A woman can claim a share of her husband's accounts, but doing so requires substantial effort, an understanding of the value of these accounts, and legal expertise and a willingness to force the judicial system to consider the division of these funds (a divorce decree that includes a QDRO and transfer incident language related to retirement assets).[13] The average divorcing woman does not receive half of the household retirement account value at the time of divorce and most investment advice texts for women encourage married women to avoid depending on their husband for retirement and to have their own accounts.[14]

A formal model of the investment decision is presented in the Appendix. From the model, one derives the intuitive conclusion that women who are much more likely to divorce should have a greater incentive to direct household funds to their own retirement accounts. Including loss aversion in the model only strengthens this prediction, as does including the learning about the costs of divorce that takes place after a divorce.[15] However, if these women were substantially overconfident about the likelihood of a joint retirement, one would see little or no

difference in investment patterns across married women. One might even see women at greater risk of divorce being less likely to invest in their own accounts.

The only data set to have both the marital history and investment behavior of married women ages 18-64 is the Survey of Income and Program Participation (SIPP; the same survey used by Butrica *et al.*). To explore the connection between competence and self-awareness in the retirement investment behavior of married women, we construct a data set from the 1996-1999 SIPP panel. The relationship between the number of times a married woman previously divorced and her retirement investment decisions sheds light on whether or not individuals correctly infer their probability of divorce and make appropriate decisions conditional on that probability.

We construct two models: (1) a Probit model that estimates the likelihood of a married woman having her own Individual Retirement Account (IRA) in 1997 (n=5837) and (2) a Probit model that estimates the likelihood of her having her own Employer-Sponsored Pension Plan (ESP) in 1997 (n=4059). Note that before 1998, a nonworking spouse could make tax-deductible contributions up to $2000 to an IRA and Roth IRAs were not yet available. We also estimate a recursive bivariate Probit model that portrays the decision to participate in an IRA and an ESP as a simultaneous decision, but we cannot reject the null that there was no correlation of the errors across the models (chi2 = 0.957). Thus we present the results from the two univariate Probit models and note the conclusions are not altered in the context of a bivariate Probit model.

In a model in which all agents are self-aware of their probability of divorce, one would predict that previously-divorced women (*Divorcee* = 1), on average, would be more likely to have their own retirement accounts. Ideally, one wants also to observe actual contributions in addition to ownership, but the SIPP has many missing observations on actual contributions. Based on econometric analyses of retirement investment behavior [e.g., Even and MacPherson

1994; Clark et al. 2000; van Derhei and Copeland 2001; Hrung 2004], we control for other relevant covariates: high school dropout (*Dropout*), college educated (*College*), age (*Age*), household income (*Income*), race (*Black, Other Race*), and the number of children under 18 in the household (*Children*). In the model of ESP participation, we also control for the size of the company (*Large Firm, Medium Firm*), the number of hours worked per week (*Hours Worked*), and job tenure (*Job Tenure*). In both models, we also control for the length of the current marriage (*Months*). Nine states are community-property law states, in which a 50-50 split of household assets is the starting point for divorce proceedings (although not necessarily the ending point), and thus we also control for an individual's residence in such states (*Community*). Based on previous research, we add squared terms for age, income and job tenure. A dummy variable for participation in an ESP is included in the equation for IRA participation for two reasons: because IRA eligibility is affected by ESP participation and because evidence suggests that individuals may have a target level of retirement savings in mind [e.g., Clark et al. 2000] and thus if they participate in an ESP, they may have fewer incentives to own an IRA.[16]

Results from the models are presented in Table III. The signs of the marginal effects predicted by theory or previous empirical work are in parentheses adjacent to the variable label. With the exception of *Divorcee*, every sign is as predicted and most of the marginal effects are significantly different from zero in both models.[17] In the IRA model, the marginal effect of *Divorcee* is negative and significantly different from zero (removing seven influential observations increases the marginal effect by over 35% and reduces the standard error, yielding p=0.007). Creating two dummy variables, one for women only divorced once and one for women divorced more than once only strengthens the results (the effect of multiple divorces more than doubles the marginal effect; p=0.004).

[Table III about here]

In the ESP model, the marginal effect of *Divorcee* is also negative, but not significantly different from zero. Of course, failure to reject the null hypothesis does not imply acceptance of the null. We can, however, perform an equivalency test [Parkhurst 2001] by choosing a value that represents the threshold above which one might agree that previous divorces have a substantial positive effect on investment behavior. We choose a value equal to one-half the absolute magnitude of the dummy variable with the smallest, significant Probit coefficient (i..e, *College*). Using a Chi-squared statistic (4.61), we reject the null at the $p = 0.0317$ level; the null hypothesis that the effect of *Divorcee* is positive and of the same absolute magnitude as *College* can be rejected at the $p = 0.0007$ level.

As a "stress test," the sample was restricted to previously-divorced women married for less than 10 years and never-divorced women married for more than 10 years. The ten-year threshold is important for two reasons: (1) most divorces happen during the first 10 years of marriage and (2) a woman married for less than 10 years has no claim on her husband's social security checks (divorcees are entitled to spousal benefits only if the marriage lasted 10 years and the claimant does not remarry). Thus these two sub-groups represent women with the highest and lowest average probabilities of divorce and corresponding highest and lowest incentives for participating in their own retirement account. In the re-estimated models, *Divorcee* now represents the joint dichotomous attributes of previously-divorced and married for less than 10 years. In both models, the marginal effect of *Divorcee* remains negative and more than doubles ($p=0.004$ in IRA model and $p=0.183$ in ESP model).

Thus the data fail to support the predictions of conventional theory and are consistent with a theory that incorporates the link between competence and self-awareness. As with any

econometric analysis, the inferences above may be incorrect due to misspecification or unobserved heterogeneity among individuals. Different specifications of the model did not generate a positive coefficient on *DIVORCEE* significantly different from zero. With regard to unobserved heterogeneity, if previously-divorced women are self-aware of their higher risk of divorce but also have higher discount rates or are more risk-loving than other women, one might observe the same results (see model in Appendix). We know of no data that supports or rejects these conjectures. Although unobserved heterogeneity is plausible, the data are consistent with the experimental results from the previous sections.

VI. Further Implications for Economic Theory

In the previous sections, we used experimental and field data to demonstrate a link between competence and self-awareness in a given domain. Just as importantly, we demonstrated how this link implies a specific structure on decision errors that can be exploited for making economic predictions. Note that the assumption of rationality is maintained - beliefs are simply erroneous. Unlike past analyses that include erroneous beliefs, however, the nature of the error in this case has a specific structure that leads to more precise predictions.

For example, consider again the context of an insurance market. An economist observes a high-risk person pretending to be a low-risk person and believes the motivation is a combination of inappropriate incentives (the high-risk person is trying to mingle with low-risk people to secure a lower rate) and the inability of low-risk individuals to credibly signal their type. As a solution, the economist instructs insurance companies to design contracts that force consumers to reveal their true types and stay with their own kind; i.e., an application of incentive compatibility constraints. An analyst aware of the incompetence-overconfidence link observes the same

behavior and sees the incentive compatibility constraints as useless because most of the high-risk people genuinely believe they are low-risk.

More specifically, consider screening models, such as those first analyzed by Rothschild and Stiglitz [1976] in the market for insurance. These models generally do not allow for pooling equilibria; only separating equilibria and cycling are possible. However, if a large proportion of one type in the population (e.g., high-risk drivers) believes it is another type (e.g., low-risk drivers), separating equilibria are unlikely. In their analysis of the U.S. term life insurance market, Cawley and Philipson [1999] indeed found that the pricing schedule was incompatible with risk sorting across contracts in the separating equilibrium predicted by theory (they also found a negative covariance between risk rates and the quantity of insurance purchased).

Given the inability of less competent agents to recognize their true type, we would expect firms and consumers to depend on costly-to-fake signals from potential suppliers or partners rather than on screening contracts that force players to reveal their types. For example, insurance companies would expect high-risk (incompetent) drivers to be unaware of their type and thus would depend heavily on driving histories as a costly-to-fake, albeit noisy, signal of driving skill. In an attempt to explain their anomalous data, Cawley and Philipson hypothesize that insurance companies may have better information than consumers about consumers' risk types (e.g., by observing systematic patterns in claims over time).[18]

In empirical analyses of risk taking, a behavior that an economist might describe as indicating a "taste for risk" may in fact derive simply from the inability of less competent agents to determine their own type. Consider, for example, the decision to engage in or forgo risky sexual activity. When researchers asked gay men to rate the riskiness of their own sexual conduct for contracting HIV, most respondents who engaged in high-risk activity did not rate their own

risk as high [Bauman and Siegel 1987]. Their erroneous self-assessments were largely based on their beliefs in the risk-reducing power of practices that had no effect on reducing risk (e.g., inspecting their sexual partners for lesions, showering after sex). Such considerations also have implications for calculations of the Value of a Statistical Life (VSL) from labor market data. If those who are least able to reduce their likelihood of experiencing an accident (and thus should demand a higher wage premium) believe they are in fact highly capable of reducing their exposure to risk, the VSL will be underestimated.

Recent theoretical work on the effect of overconfidence on decisionmaking may benefit from considering the connection between incompetence and overconfidence. For example, Gervais, Heaton and Odean [2002] present a theoretical model in which overconfident managers can increase the value of a company more than risk-averse managers because overconfident managers are more willing to undertake projects and exercise options. However, a key implicit assumption in their analysis is that overconfident managers and risk-averse managers are equally competent and simply differ in the direction they move from the risk neutral strategic choice.

Recent empirical work may likewise benefit from re-examination in the light of the overconfidence-incompetence link. The results of Barber and Odean [2001] suggest that younger investors trade more than older investors (a proxy for overconfidence) and earn lower returns. Gervais and Odean [2001] note that these data suggest that overconfidence diminishes with experience through signals of past performance. These same data, however, are consistent with the idea that overconfidence is driven largely by incompetence. Experienced traders are less overconfident because (1) they are more competent than they were when they were young and (2) they were more competent than their peers at early stages of their career and thus remain in the business, while less competent peers have exited (a selection effect).

The effects of the overconfidence-incompetence link will also likely be felt in such disparate fields as the economics of unemployment and the economics of health. Researchers have found that, on average, people underestimate their own chance of experiencing unemployment or disease [see references in Hanson and Kysor 1999]. Fundamental ideas such as comparative advantage and the ability of individuals to self-select in career choice (e.g., entry into small business) may need to be reconsidered in the light that not everyone can recognize their own competence. For example, the excess entry and expenditures observed in winner take-all-markets, tournaments, rent-seeking opportunities, and crowded industries may derive less from uncertainty and random error and more from incompetence and overconfident self-assessments (i.e., those with little chance of succeeding in these domains will be spending substantial resources to gain entry rather than reducing their efforts as predicted by theory). In research on the growth of social capital, analysts would have to consider that the least literate parents are unable to judge the academic potential of their children and evaluate the returns to investment in children, thus potentially exacerbating unequal distributions of social capital over time. There are clearly many more examples of economic issues that may be fruitfully examined in light of the link between competence and self-awareness.

VII. A Dynamic Signaling Game with Imperfect Information (H7)

In this section, the gains from and the obstacles to incorporating the link formally into standard game theory are highlighted. Consider Goeree and Holt's Dynamic Signaling Game with Incomplete Information [2001: 1414-1416]. Senders are of Type A or Type B. The Sender chooses a signal, *L* or *R*. This signal determines whether the payoffs on the right or left side of Table IV are used. This signal is communicated to a Responder that is anonymously matched

with the Sender. The Responder sees the Sender's signal, *L* or *R*, but not the Sender's type, and then chooses a response: *C, D,* or *E*. Payoffs are then determined by Table IV, in which the Sender's payoff is to the left in each cell.[19]

[Table IV about here]

In Goeree and Holt, every player knows the ex ante probability of Type A Sender is one-half and Senders are self-aware of their type before making their decisions (and Responders know this). A choice of *L* if Type B and *R* if Type A, followed by a response of *C* to an *L* or *R* is a separating Nash equilibrium. Neither type of Sender would benefit from sending the other signal and the Responder cannot do any better by choosing *D* or *E*. In this equilibrium, the signal reveals the Sender's type. The only other equilibrium in this game is a pooling equilibrium in which responders respond to *R* with *D* and *C* to *L*, and both Sender types send *L* to avoid receiving *D*. Such beliefs, however, do not satisfy the intuitive criterion [Cho and Kreps 1987].

In Goeree and Holt's experimental implementation of the game, all Type B Senders sent *L*, 7 out of 10 Type A Senders sent *R*, and all Responders chose *C*. Thus their results are largely consistent with the separating equilibrium prediction. We replicated their experiment with twenty-six subjects from an introductory economics class (Spring 2003). Payoffs were awarded in the form of extra credit points on the final grade in the class (a payoff of 500 was equivalent to 1.67 points out of the 100 points of a student's final grade). As in Goeree and Holt, the experimenter informed Senders if they were Type A or Type B. Each subject first made a decision as a Sender and then made a decision as a Responder in response to a randomly assigned signal from one of the other subjects. All 13 of the Type B senders sent *L* and 12 of the 13 Type A senders sent *R*. All responses to *L* were *C* and 9 of the 12 responses to *R* were *C*

(there were also one *D* and two *E*).[20] Thus, as in the Goeree and Holt experiment, the majority of the pairs (~85%) made choices that were consistent with the separating Nash equilibrium.

Consider a new version of the game with one change: Sender type reflects ability in some domain and each Sender must infer his or her own type. Let high-ability individuals be Type A players and low-ability individuals be Type B players. Standard economic theory would predict no change in outcomes between the new version and the original version. The self-awareness/competence link, however, implies that Type A players are aware that they are Type A players, while Type B players believe they are also Type A. Making predictions about play of this game using standard game theory syntax and equilibrium notions is not straightforward for reasons that will become clear below. We explore how one might begin to make such predictions, but note that an in-depth, generalized analysis of such games is beyond the scope of this paper.

We can consider two possible formulations of the game: one in which Responders are aware of overconfident Senders and one in which they are not aware of such Senders. In the first formulation, we assume that any given Sender believes he is Type A. The Responder knows that the Sender believes he is Type A, the Sender knows that the Responder knows this, etc. We then assume the Responder has, as part of her private information, a belief that the Sender sending her a signal is Type A with probability λ and Type B with probability 1 - λ.

We define an equilibrium in this game as a strategy profile in which neither player has an incentive to change their strategy given their beliefs (i.e., a best response to the other players' strategies given given their beliefs). Given Sender beliefs, which are common knowledge, given Responder beliefs, which are private information, and given a utility function over the payoffs, three equilibria are possible. Assuming risk-neutrality, if $\lambda = 0$, then (*R*, *D*) is an equilibrium. If

$\lambda = 1$, then (*R*, *C*) is an equilibrium. If we assume risk-aversion and $\lambda \in (0.50,1)$, then (*R*, *E*) can be supported as an equilbrium (e.g., u = $\sqrt{payoff}$; $\lambda = 0.6$). The same three equilibria can result from other combinations of Responder beliefs and utility functions. A Sender sending *L* is not part of any equilibrium.

The second formulation of the game is intuitively easier to comprehend, but harder to model formally: all Senders believe they are Type A and both Senders and Responders are unaware of this – in other words, Senders and Responders are not aware that play of the game will be any different from the original Goeree and Holt version. We also assume that every player recognizes that sending *R* as a Type A and responding to *R* with *C* is part of a separating equilibrium <u>in the game in which there are no overconfident Senders</u>. The difficulty in incorporating an unaware agent into a standard state-space game is well known [Dekel et al. 1998]. Standard game theory syntax and equilibrium notions do not allow an agent to be unaware of an event. Despite the difficulty in formally modeling this situation, however, it seems reasonable to predict that if Senders believe they are Type A, and Senders and Responders are unaware that the game has changed from its original version, all Senders will send *R*, and Responders will respond with *C* to an *L* or an *R*. Note this outcome does not correspond to any equilibrium notion.

Several weeks after the control session in which Senders were told their type, 26 subjects from the <u>same</u> introductory economics class participated in a treatment session (all but one had participated in the control session). <u>Prior</u> to this treatment session and immediately after the students' first exam, subjects were asked to assess their absolute and relative performances on the exam as in Section III. As in the previous experiments, the average subject demonstrated

substantial overconfidence (e.g., 90% believed they performed above the 50th percentile) and again, the degree of overconfidence was strongly related to competence.

Subjects were also informed of the aggregate pattern of results from the control session (i.e., what Type As and Type Bs sent, and how Responders responded to each signal). They were not informed of their specific payoffs from the control session. Subjects were informed of the previous session's aggregate results to (1) ensure all subjects understood the game and (2) allay fears that some players may not have understood the game or recognized the separating Nash equilibrium. Thus in the treatment session, the players could focus on how the treatment condition changed the incentives, rather than on confirming their understanding of the original version or determining if everyone else had the same understanding.

In the class period immediately after their exam, *before any feedback on their exam performances had been given*, the subjects played the game in Table IV again. In contrast to the control session in which each subject was told his or her type, subjects in the treatment session were told that they were Type A if they had scored at the 50th percentile or higher on the exam, and Type B otherwise. Thus, approximately half of the class was Type A and the other half Type B (not exactly half due to lumpy distribution of grades).

The traditional theory in which subjects are self-aware would not predict any substantial change in outcomes between control and treatment conditions (feedback from the control session results should actually ensure that all outcomes would be consistent with the separating equilibrium). In contrast, a theory that incorporates the link between competence and self-awareness predicts that (1) most Senders will send *R* and (2) some Responders, aware that there are overconfident players in their midst, will choose to respond to *R* with *D* or *E*, while other Responders, some unaware of the change in the game, will continue to respond to *R* with *C*.

The data reflect these predictions. Thirteen of the 14 Type B Senders sent *R*, as well as 11 of the 12 Type A Senders. Moreover, only 14 of the 24 responders who received an *R* signal chose to respond with *C*. Five chose to respond with *E* and 3 chose to respond with *D*. Thus when subjects were not explicitly told their type by the experimenter, only a little over 20% of the outcomes were consistent with the separating equilibrium, while 92% were consistent with predictions from a theory that incorporates the link between competence and self-awareness. The dramatic outcome in this simple game suggests substantial gains to devising formal models to incorporate the self-awareness/competence relationship.

VII. Conclusion

Recognition that self-awareness is difficult is not new. The great philosophies and religions, and many of the great works of literature, describe the dangers of overconfidence, the honor of humility, and the necessity of self-introspection that leads to knowing oneself. Given the link between competence and self-awareness, however, it is not surprising that more than a millennium of efforts to develop social norms to constrain overconfidence has done little to solve the problem of overconfidence. In a way, the problem is unsolvable: the less one knows, the less one is able to know that he does not know. In order for overconfident individuals to see that they are grossly overconfident, they must become competent – but if they are competent, they are unlikely to be grossly overconfident.

Modern social norms in Western, industrialized nations seem, in fact, to encourage overconfidence. A corporate employee in the United States will have the daily opportunity to read motivational posters with such aphorisms as “Experience tells you what to do; confidence allows you to do it,” and “Have confidence that if you have done a little thing well, you can do a

bigger thing well too." Perusing the self-help books at Amazon.com, a consumer will run across titles such as *Unstoppable Confidence***;** *Selling Yourself: Be the competent, confident person you really are!***;** *Getting What You Deserve***,** *Earn What You Deserve: How to stop underearning and start thriving*; and *Getting What You Deserve From Rotten Bosses, Demanding Spouses, Phony Friends, Prying Parents, Annoying Neighbors and Other Irritating People*. However, a consumer interested in finding a title along the lines of *How to Recognize When You're Not as Competent as You Think You Are* will walk away unsatisfied.

We are all incompetent in some domains. Thus, as Kruger and Dunning have pointed out, we are all doubly cursed in these domains: we are not only incompetent, but we are also unable to recognize the extent of our incompetence and thus are destined to be overconfident. We are therefore surprised when faced with signals that hint at our overconfidence, but we are unable to understand fully the meaning of the signal, for that would require competence in the domain. Thus a student who has performed poorly on an exam may blame the teacher who allegedly posed vague questions; a professor whose article has been rejected for publication may blame the reviewers who could not see the alleged brilliance of the arguments presented.

Adam Smith, in the *The Wealth of Nations* [p.209], noted that "[t]he overweening conceit which the greater part of men have in their own abilities is an ancient evil remarked by the philosophers and moralists of all ages." Economists, in their attempt to reconcile theory and empirical observations of human behavior, may do well to revisit Smith's observation and consider the role that competence plays in the ability of economic agents to know themselves and understand their place in the games in which they play.

References

Baker, Lynn A. and Robert E. Emery. 1993 When Every Relationship is Above Average: perceptions and expectations of divorce at the time of marriage. Law and Human Behavior 17: 439-450.

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

Bauman, Laurie J. and Karolynn Siegel. 1987. Misperception Among Gay Men of the Risk for AIDS Associated with Their Sexual Behavior. Journal of Applied Social Psychology 17: 329-45.

Becker, G.S., E.M. Landes, and R.T. Michael. 1977. An Economic Analysis of Marital Instability. Journal of Political Economy 85(6): 1141-1187.

Bernardo, Antonio and Ivo Welch. 2001. On the Evolution of Overconfidence and Entrepreneurs. Journal of Economics and Management Strategy 10(3): 301-330.

Bramlett M.D. and W.D. Mosher. 2001. Marriage dissolution, divorce, and remarriage: United States. Advance data from vital and health statistics; no. 323. Hyattsville, Maryland: National Center for Health Statistics.

Brenner, Lyle A., Derek J. Koehler, Varda Liberman, and Amos Tversky. 1996. Overconfidence in Probability and Frequency Judgments: A Critical Examination. Organizational Behavior and Human Decision Processes 65(3): 212–219.

Brinig, Margaret and Douglas W. Allen. 2000. These Boots are Made for Walking: why most divorce filers are women. American Law and Economics Review 2(1): 126-169.

Butrica, Barbara A., Howard M. Iams, James H. Moore, and Mikki D. Ward. 2001. Methods in Modeling Income in the Near Term (MINT I). ORES Working Paper no. 91. Office of Research, Evaluation and Statistics, Office of Policy, Social Security Administration.

Camerer, Colin and Dan Lovallo. 1999. Overconfidence and Excess Entry: an experimental approach. American Economic Review 89(1): 306-318

Cawley, John and Tomas Philipson. 1999. An Empirical Examination of Information Barriers to Trade in Insurance. American Economic Review 89(4): 827-846.

Cho, In-Koo and David M. Kreps. 1987. Signaling Games and Stable Equilibria. Quarterly Journal of Economics 102(2): 179-221.

Clark, Robert L., Gordon P. Goodfellow, Sylvester Schieber and Drew Warwick. 2000. Making the Most of 401(k) Plans: who's choosing what and why? In Mitchell, Olivia, P. Brett Hammond, and Anna M. Rappaport (eds.), Forecasting Retirement Needs and Retirement Wealth. Philadelphia, PA: University of Pennsylvania Press, pp. 95-138.

Daniel, Kent D., David Hirshleifer and Avanidhar Subrahmanyam. 1998. Investor Psychology and Security Market Under- and Over-reactions. Journal of Finance 53(5): 1839-1886.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini. 1998. Standard State-Space Models Preclude Unawareness 66(1): 159-173.

Dobbs, Allen R. 1997. Evaluating the Driving Competence of Dementia Patients. Journal of Alzheimer's Disease and Associated Disorders, 11(suppl.1): 8-12.

Even, William E. and David A. Macpherson. 1994. Employer Size and Compensation: the role of worker characteristics. Applied Economics 26(9): 897-907.

Even, William E. and David A. Macpherson. 2000. The Changing Distribution of Pension Coverage. Industrial Relations 39(2): 199-227.

Gervais, Simon and Terrance Odean. 2001. Learning to be Overconfident. Review of Financial Studies 14 (1): 1-27.

Gervais, Simon, J.B. Heaton, and Terrance Odean. 2002. The Positive Role of Overconfidence and Optimism in Investment Policy. Wharton School Working Paper. University of Pennsylvania. Philadelphia, PA.

Goeree, Jacob and Charles Holt. 2001. Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. American Economic Review 91(5): 1402-1422.

Golec, Joseph H. and Maurry Tamarkin. 1995. Do Bettors Prefer Longshots Because They are Risk-Lovers or are They Just Overconfident? Journal of Risk and Uncertainty 11: 51-64.

Griffin, Dale and Amos Tversky. 1992. The Weighing of Evidence and the Determinants of Confidence. Cognitive Psychology 24: 411-429.

Hanson, Jon D. and Douglas A. Kysar. 1999. Taking Behavioralism Seriously: the problem of market manipulation. New York University Law Review 74: 630-749.

Harrison, Glenn and John List. 2004. Field Experiments. Journal of Economic Literature 42(4):

Holt, Charles A. and Susan K. Laury. 2002. Risk Aversion and Incentive Effects. American Economic Review 92(5): 1644-1655.

Hrung, Warren B. 2004. Information, the Introduction of Roths, and IRA Participation. Contributions to Economic Analysis and Policy 3(1), Article 6: 1-17.

Hvide, Hans K. 2002. Pragmatic Beliefs and Overconfidence. Journal of Economic Behavior and Organization 48: 15-28.

Keren, Gideon. 1987. Facing Uncertainty in the Game of Bridge: a calibration study. Organizational Behavior and Human Decision Processes 39: 98-113.

Kruger, Justin and David Dunning. 1999. Unskilled and Unaware of It: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology 77(6): 1121-1134.

Kyle, Albert S., and F. Albert Wang. 1997. Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test? Journal of Finance 52(5): 2073-2090.

Lichtenstein, Sarah, Baruch Fischoff, and Lawrence D. Philips.1982. Calibration of Probabilities. State of the Art to 1980. In Daniel Kahneman, Paul Slovic and Amos Tversky (eds.), Judgement under Uncertainty: heuristics and biases. New York: Cambridge University Press.

Monahan, T.P. 1958. The Changing Nature and Instability of Remarriages. Eugenics Quarterly 5(June):73-78

Monahan, T.P. 1959. The Duration of Marriage to Divorce: Second Marriages and Migratory Types. Marriage and Family Living 21(May): 134-38.

Odean, Terrance. 1999. Do Investors Trade Too Much? American Economic Review 89: 1279-1298.

Parkhurst, David F. 2001. Statistical Significance Tests: Equivalence and Reverse Tests shoud reduce misinterpretation. Bioscience 51(12): 1051-1057.

Rothschild, Michael D. and Joseph E. Stiglitz. 1976. Equilibrium in Competitive Insurance Markets: an essay in the economics of imperfect information. Quarterly Journal of Economics 90(4): 629-649.

Smith, Adam. 1776 (1982). The Wealth of Nations: Books I-III. Andrew Skinner, editor. Penguin Classics, New York, New York.

Svenson, Ola. 1981. Are We All Less Risky and More Skillful Than Our Fellow Drivers? Acta Psychologica 47: 143-146.

van Derhei, Jack and Craig Copeland. 2001. A Behavioral Model for Predicting Employee Contributions to 401(k) Plans: preliminary results. North American Actuarial Journal 5(1): 80-94.

Appendix: Two-period Investment Model with Divorce Hazard

A married woman has an initial endowment M_1. She may have earned M_1 or simply have influence over its allocation. In period 1, she consumes C_1 and invests the remainder $(M_1 - C_1)$ in retirement accounts, which can be consumed in Period 2. Let S_w denote the amount she invests in her retirement account and S_h denote the amount she invests in her husband's account. There is a probability *p* that she will divorce before period 2. Denote the returns to her account and her husband's account by r_w and r_h respectively, where $r_h > r_w > 1$. The latter assumption can represent a husband's better investment information or the transaction costs associated with deviating from traditionally defined cultural roles in which the male makes household investment decisions (these costs can be costs of information acquisition or costs of marital discord involved with a wife's insistence on directing household funds towards her account).

If she remains married, her consumption in period 2 will be $C_2 = r_w S_w + r_h S_h$. If she divorces, she will loses some fraction of the value of her husband's account and therefore her consumption will be $C_2 = r_w S_w + r_h a S_h$, where $0 < a < 1$. Assume if a divorce takes place, the woman would prefer, *ex post*, to have invested in her own account: $r_w > a r_h$.

The woman's utility function is assumed additive across both time and states of nature and thus has the form: $U(C_1, C_2) = u(C_1) + \delta E u(C_2)$, where $0 < \delta < 1$ is a rate of time preference. Then, the woman's investment decision problem can be expressed as:

$$\underset{S_w, S_h}{\text{Max}}\ u(M_1 - S_w - S_h) + \delta(1-p)u(r_w S_w + r_h S_h) + \delta p u(r_w S_w + a r_h S_h).$$

For interior solutions, $\frac{\partial S_w}{\partial p} > 0$ and $\frac{\partial S_h}{\partial p} < 0$. It is well known that people are particularly averse to losses, which can make high-variance gambles less attractive if there is a safe option that avoids the chance of a loss. Incorporating such loss aversion would only strengthen the magnitude of the woman's response, not change the direction. For a corner solution in which the woman invests only in her own account, the comparative statics do not change. For a corner solution in which the woman invests only in her husband's account, it is possible that an increase in *p* will not generate an increase in a woman's investment into her own account.[21] Such an outcome is possible if the return to the husband's account is substantially higher than that of the woman's and thus a small increase in *p* might not change the desirability of investing in her husband's account. Finally, increases in the discount factor (lower discount rates) and decreases in the fraction of the husband's account obtained in the event of a divorce make the woman more likely to invest in her own account.

---

[1] Kruger and Dunning's results were foreshadowed by: many authors (e.g., Lichtenstein, Fischoff and Philips [1982], Griffin and Tversky [1992]) who find subjects exhibit more pronounced overconfidence when presented with more difficult judgment tasks; Karen [1987], who finds that amateur bridge players exhibit significant overconfidence in predicting game outcomes; and Dobbs [1997], who interviewed drivers with clinically significant cognitive impairments and found 98% considered themselves to be as good as or better than other drivers their age (evaluations by driving experts overwhelmingly contradicted the self-assessments).

[2] An exception is Barber and Odean. Based on results in the psychology literature, they assume males are more overconfident than females. Their analysis, however, did not actually measure overconfidence nor did it control for competence; i.e., women may be less overconfident because they are more competent in the tasks.

[3] The second stream of overconfidence studies include "calibration studies," in which subjects answer a factual question and are asked to estimate the probability that their answers are correct (actual and predicted rates of correct answers are then compared). There is still a substantial debate in the psychological literature focused on whether overconfident calibration is real or an artifact of experimental designs and statistical methods, and, if it exists, whether it is task dependent [Brenner et al. 1996].

[4] Depending on the exam, 8% – 13% of the subjects in Class 1 received the $25 payoff. Depending on the exam, 30-50% of the subjects in the Classes 2 and 3 received a positive payoff. We observed at least one (high-performing) student deliberately leaving a question on his exam blank in order to increase his estimation accuracy, which suggests that the payoff structure was salient.

[5] For each class, the average performance of the subjects who took each of three exams is not significantly different from that of subjects who took only one or two of the exams.

[6] Kruger and Dunning tested subjects on grammar, logic and humor competence (subject competence in humor was judged using a reference set of evaluations made by a panelist of professional comedians).

[7] All regressions were performed in Stata v.8. A Breusch-Pagan test for random effects returned significant results ($p < 0.001$) on all models. A Hausman specification test, under the assumption of correct specification, tests the appropriateness of the random-effects estimator applied to the data. For all regressions, the test yielded an insignificant result at conventional levels ($p>0.15$), which provides evidence that the random effects and the regressors are not correlated and thus the random effects model is appropriate.

[8] Although not the focus of this study, one can also use the panel to test for the so-called "self-serving" (or "self-attribution") bias, which would imply that the more successful the subject had been on earlier exams, the more overconfident he would be, *ceteris paribus*. Adding a one-period lagged score variable to a regression on observations of ABCONFIDENCE from the second and third exams supports a weakly significant and positive relationship between past performance and current overconfidence (LagScore = 0.10; t-stat = 1.75). Adding a one-period lagged variable to the regression on RELCONFIDENCE supports a significant and positive relationship (LagPercentile = 0.17; t-stat = 2.10). The other coefficients move in directions that either do not affect or strengthen the conclusions in the text.

[9] As with all the previous experiments, human subjects approval for establishing this market was obtained.

[10] One who scored in the bottom quartile on both the first and final exam chose contract B. One who scored in the 3rd quartile on Exam 1 and the bottom quartile the final exam chose contract B. Inclusion of these subjects only strengthens the conclusions drawn in this section.

[11] After three colleagues (economists) observed these contracts, each predicted that Contract A would be the most popular and that the professor would have to award many points in payouts.

[12] One subject who scored consistently below the 50th percentile on all exams, but did not purchase insurance on the final exam, was in the author's Introduction to Microeconomics class the following semester (she needed to retake the course to meet the grade requirements for her major). When asked why she did not purchase insurance on the final exam, she stated "I had studied really hard for that exam and I really thought I was going to get a top score."

[13] As noted in many investment books and websites for divorcing women, QDROs (Qualified Domestic Relations Order) take substantial expertise to execute in a manner that will ensure the wife receives an equitable share of the husband's retirement accounts. Note that an "equitable" share as determined by the courts is not necessarily an "equal" share. Furthermore, if not done correctly the transfer of retirement assets can be subject to taxes and early distribution penalties.

[14] Becker et al.'s model predicts that the initiator of the divorce has more to gain from the divorce and thus has fewer incentives to demand an equal share of household assets as a condition of the divorce: more than two-thirds of divorces are initiated by the woman [Brinig and Allen 2000, who also cite evidence to that women tend to be happier after divorce despite being worse off financially].

[15] Similarly, Becker et al.'s model of marital dissolution implies that increases in the probability of marital dissolution will decrease a woman's incentive to invest in "marital-related capital" that will be less valuable to the woman after a divorce.

[16] Although is plausible that having an IRA can affect ESP participation, the motivation for including ESP in the IRA equation only is that not everyone has access to ESPs whereas everyone has access to an IRA. Including an IRA dummy variable in the ESP model does not change the coefficient on *Divorcee* and it generates an *IRA* coefficient that is negative but not significantly different from zero ($p=0.19$). Excluding *ESP* from the IRA model changes the coefficient on *DIVORCEE* at the fourth decimal place only.

[17] Conclusions are the same whether one computes marginal effects at the mean, computes average marginal effects (the average of partial and discrete changes over the observations; i.e., the delta method) or if one simply examines the Probit model coefficients.

[18] Of course, for a costly-to-fake signal to work, the person on the receiving end of the signal must be competent. The long-running popularity of Dr. Laurence J. Peter's "Peter Principle" suggests that this may not always be the case.

[19] The only difference between Goeree and Holt's payoffs (Table 7 in their article) and those in Table IV is the Responder payoff when a Type A Sender chooses *R* and the Responder chooses *C*. In Goeree and Holt's version the payoff is 900, whereas in Table III it is 500. This change does not change the equilibria in the game.

[20] The subject who responded with *D* to an *R* signal also sent *L* as a Type A. Although this behavior is consistent with the pooling equilibrium, the post-experiment explanation form filled out by the subject suggests the subject was confused.

[21] A decrease in investment to the husband's account as a result of an increase in p cannot be ruled out for all functional forms, but it can be ruled out under standard utility functions such as negative exponential, quadratic, logarithmic, power, and hybrids such the power-expo function [Holt and Laury 2002].

Table I – Self-assessments of absolute performance by subject competence (Exam 1)

	Absolute Performance[1]				Relative Performance		
	<65%	**65%-80%**	**>80%**		**<65%**	**67%-80%**	**>80%**
Mean Over-confidence (ABCONFIDENCE) [2]	+13.6%	+6.8%	+1.7%	***Mean Over-confidence (RELCONFIDENCE)*** [4]	+36 pts	+16 pts	-8 pts
Mean Error (ABACCURACY) [3]	15.4%	10.0%	4.3%	***Mean Error (RELACCURACY)*** [5]	37 pts	21 pts	13 pts
t- test – H0: No Overconfidence (ABCONFIDENCE=0) vs. H1: ABCONFIDENCE>0 (or <0)							
p-values	< 0.0001	< 0.0001	0.060		< 0.0001	< 0.0001	0.038
N	44	38	23		44	38	23

[1] *Percentage categories refer to the percentage of questions answered correctly by each subject in the category.*

[2] *Predicted Percentage of Correctly Answered Questions – Actual Percentage of Correctly Answered Questions*

[3] *Absolute Value of ABCONFIDENCE.*

[4] *Predicted Percentile Ranking – Actual Percentile Ranking*

[5] *Absolute Value of RELCONFIDENCE.*

Table II –Self-evaluations of absolute (AB) and relative (REL) performances

	Absolute Performance		**Relative Performance**	
	(1) Overconfidence (ABCONFIDENCE)	**(2) Accuracy (ABACCURACY)**	**(3) Overconfidence (RELCONFIDENCE)**	**(4) Accuracy (RELACCURACY)**
Independent Variables	*Coefficient (standard error)*	*Coefficient (standard error)*	*Coefficient (standard error)*	*Coefficient (standard error)*
Constant	50.746 ***	41.951***	1.072 ***	0.833 ***
	(3.694)	(2.790)	(0.071)	(0.054)
Score	-0.546 ***	-0.406 ***	-0.011 ***	-0.007 ***
	(0.044)	(0.035)	(0.0009)	(0.0007)
Ex2∗Score	-0.021	-0.008	-0.0004	-0.0003
	(0.016)	(0.013)	(0.0003)	(0.0003)
Ex3∗Score	-0.067 ***	-0.008	-0.0006 *	-0.0004
	(0.016)	(0.013)	(0.0003)	(0.0003)
Female	-4.032 **	-1.903 *	-0.082 ***	-0.065 ***
	(1.662)	(1.132)	(0.029)	(0.020)
Econ	-1.592	-1.524	-0.009	-0.015
	(1.628)	(1.110)	(0.029)	(0.020)
Class2	-2.525	-1.563	-0.118 ***	-0.077 ***
	(1.878)	(1.287)	(0.033)	(0.024)
Class3	-4.402 *	-4.162 ***	-0.093 **	-0.081 ***
	(2.272)	(1.557)	(0.040)	(0.053)
	Overall R^2 = 0.37	Overall R^2 = 0.38	Overall R^2 = 0.41	Overall R^2 = 0.34
	Wald Chi^2 = 202.9 ***	Wald Chi^2= 158.5***	Wald Chi^2 = 199.8 ***	Wald Chi^2 =142.3***

, **, and * indicate t-statistic p-values less than 0.10, 0.05 and 0.01, respectively.*

Table III –Married Women's Retirement Investment Behavior (Probit)

	(1) Individual Retirement Account (IRA) Participation	(2) Employer-sponsored Pension Plan (ESP) Participation
Covariates (Predicted Sign)	*Marginal Effect* *(standard error)*	*Marginal Effect* *(standard error)*
Divorcee (+)	-0.031** (0.015)	-0.024 (0.026)
Community (-)	-0.014 (0.011)	-0.001 (0.019)
Age (+)	0.012*** (0.001)	0.012* (0.007)
Age^2 (-)	-0.000 (0.000)	-0.000* (0.000)
Months (?)	0.000 (0.000)	0.000 (0.000)
Children (-)	-0.023*** (0.005)	-0.011 (0.008)
Black (-)	-0.122*** (0.015)	-0.068** (0.028)
Other Race (?)	-0.019 (0.025)	-0.020 (0.041)
Dropout (-)	-0.139*** (0.013)	-0.093*** (0.031)
College (+)	0.121*** (0.014)	0.070*** (0.023)
Income (+)	0.006*** (0.001)	0.006*** (0.002)
Income^2 (-)	-0.000*** (0.000)	-0.000 (0.000)
ESP (-)	-0.029*** (0.011)	
Large Firm (+)		0.304*** (0.028)
Medium Firm (+)		0.208*** (0.063)
Job Tenure (+)		0.004*** (0.000)
Job Tenure^2 (-)		-0.000*** (0.000)
Hours Worked (+)		0.011*** (0.001)
	McKelvey-Zavoina's R^2 = 0.33 Wald Chi^2 = 766***	McKelvey-Zavoina's R^2 = 0.77 Wald Chi^2 = 1063***

, **, and * indicate p-values less than 0.1, 0.05 and 0.01, respectively. "Other Race" includes all other non-white individuals. Marginal effect for dummy variables is for discrete change from 0 to 1.*

Table IV – Signaling game payoff matrix (Sender's payoff, Responder's payoff)

	Response to *Left* signal				Response to *Right* signal		
	C	**D**	**E**		**C**	**D**	**E**
Type A Sends "**L**"	300, 300	0, 0	500, 300	Type A Sends "**R**"	450, 500*	150, 150	1000, 300
Type B Sends "**L**"	500, 500*	300, 450	300, 0	Type B Sends "**R**"	450, 0	0, 300	0, 150

* *Asterisks mark the separating equilibrium.*