Overconfidence is a Social Signaling Bias[*]

Stephen V. Burks[1], Jeffrey P. Carpenter[2],
Lorenz Goette[3], Aldo Rustichini[4]

October 12, 2010

Abstract

Evidence from psychology and economics indicates that many individuals overestimate their ability, both absolutely and relatively. We test three different theories about observed relative overconfidence: Bayesian updating, self-image concerns and social signal, using data on 1,016 individuals' relative ability judgments about two cognitive tests. We reject the first two theories, and provide evidence that personality traits strongly affect relative ability judgments in a pattern that is consistent with the third, social signal, theory. Our results together suggest that overconfidence in statements is most likely to be induced by social concerns, and significantly affected by personality traits.

[1] Division of Social Sciences, University of Minnesota, Morris, MN 56267-2134
[2] Department of Economics, Middlebury College, Middlebury, VT 05753
[3] Department of Economics, University of Lausanne, Lausanne, Switzerland
[4] Department of Economics, University of Minnesota, Minneapolis, MN 55455-0462

1 Introduction

Well-calibrated judgments about one's abilities are important in many economic decisions. However, evidence from psychology and economics suggests that individuals may have excessive confidence in their abilities. This excessive confidence may be absolute (subjects predict that they will perform better than they really do) or relative (subjects state their performance ranks higher, compared to that of others, than it really does). In this paper we will use the term overconfidence to describe the relative version of excessive confidence. In principle, individuals might err in either the direction of over or under confidence, but overconfidence seems the dominant behavior, although individuals may be under-confident in a few cases (Hoelzl and Rustichini, 2005). For example in a typical study few individuals rate themselves in the bottom 40 percent of a distribution, largely independent of the skill in question (See Alicke et al., 1995; Dunning, 1989; Svenson, 1981). Studies also link measures of overconfidence to behavior, and show that more confident judgments are associated with more daring behaviors. Malmendier and Tate (2008) show that more confident CEOs make more daring merger decisions (see also Malmendier and Tate (2005)). Barber and Odean (2001) show that men engage in more frequent trading in common stock, consistent with the evidence from psychology that men are more overconfident than women. This trading reduces their returns substantially relative to women. Thus, if overconfidence is truly a judgment bias, these studies should raise concern, as they raise the possibility that individuals systematically make suboptimal decisions because they choose based on biased beliefs.

The first question we therefore address is whether overconfidence should be viewed as a bias in judging one's ability, or whether there is some natural way in which rational agents could appear overconfident. If individuals have perfect knowledge of their abilities, results showing that, e.g., 50 percent of the individuals rate themselves in the top 25 percent of an ability distribution necessarily imply a judgment bias.[1] However, assuming perfect knowledge of one's ability may not be realistic. Rather, individuals may only vaguely know their abilities, and update their beliefs as new information arrives. A recent paper by Benoît and Dubra (2007) shows that, if individuals have imperfect knowledge of their own ability, even perfectly rational Bayesian updaters may report overconfident beliefs in a typical study.

Benoît and Dubra point out that, in most studies, individuals are asked to indicate

---

[1]Merkle and Weber (2007) do show overconfidence leads to bias in beliefs. Their test is based on eliciting the c.d.f. of beliefs over abilities for which it is difficult to pin down the true distribution. This allows them to reject Bayesian updating without even knowing what the true distribution of ability is.

their most likely place in the ability distribution. They provide a general characterization of the information structure leading to results such as, for example, 50 percent of the individuals rationally putting themselves in the top 25 percent of the ability distribution. Intuitively, this can arise if the signals individuals receive become noisier the better the signal values are, akin to taking an easy test. Everyone who fails the test can be sure that his ability is low; however, low-ability types sometimes also pass the test by sheer luck. But still, passing the test rationally leads individuals to believe it is more likely that they have high ability, therefore creating 'overconfidence' by this measure.

Several papers (Kőszegi, 2006; Weinberg, 2006) also provide plausible psychological underpinnings for this idea, showing that such types of information structures can arise endogenously. They argue that self image concerns lead individuals with high beliefs to refrain from seeking more information, leading to an information structure that is conducive to creating overconfidence. Yet, while overconfidence in that sense may prevail in the population, the beliefs are still unbiased, as the individuals who think they are in the upper half of the distribution recognize that this is not sure and, in their implicit internal model, attach the correct probability to this state. As a consequence, individuals may take actions based on these beliefs, which may not be optimal compared to the case in which the individual knew his true type. However, conditional on the information that the individual received, the judgments are consistent and rational, and their actions are optimal. We call here the hypothesis that judgments on relative positions of individuals are produced by Bayesian updating (from a common prior) and truthful revelation of the posterior, or some statistics of it, the Bayesian hypothesis.

However, we show in this paper that if judgments are the result of Bayesian information processing from a common prior, then testable restrictions are placed on the beliefs as a function of the individual's true ability. In particular it must be true that of all individuals placing themselves in ability quantile k, the largest (modal) share of them must actually fit in quantile k. Therefore, we can base a test of the Bayesian hypothesis on whether this is the case. We test the model with data from a large sample of subjects who judge their ability for each of two cognitive tests that we administer to them. We clearly reject the restriction: in general, individuals from an ability quantile $j < k$ are more likely to think they are in quantile k. Our test is general in the sense that it rejects *any* model that relies on Bayesian updating in forming overconfident relative ability judgments, independently of the motives behind the formation of the judgments. Our test thus rejects the joint hypothesis of Bayesian updating, the common prior assumption, and truth-telling, leaving unanswered these questions: which part of the joint hypothesis has failed, and which theory can explain

our data?

We then turn to the examination of factors behind overconfident relative ability judgments and test some of their implications independently of the auxiliary assumptions of common priors and Bayesian updating. Theoretical literature on self-confidence assumes that individuals have reason to hold correct beliefs, since this knowledge helps in the choice of better actions, but other factors may move preferences over beliefs in the optimistic direction. This literature proposes three broad reasons for the existence of optimistic rather than realistic self-assessment (Bénabou and Tirole, 2002): consumption value (individuals like to have positive self-image as a good in itself), motivation value (optimistic assessments may induce higher second-stage effort, and hence better outcomes, than correct ones), and signaling value (positive self-confidence makes a positive external representation of oneself easier). The first two reasons motivate an individual in isolation, independently of his social relations.

The first is little more than the assumption that self-confidence is sought after–while it may be true, it is *ad hoc* in the absence of further analysis. However the second explanation provides a potential functional role for optimistic beliefs: a better belief in one's skill may counter time inconsistency in choices (induced for example by quasi-hyperbolic preferences), and provide incentives for the *ex ante* right amount of effort. A way in which optimistic beliefs are produced is described in models of image concerns. In these models, individuals like to believe that they have high ability, but their beliefs are constrained by Bayesian updating (Kőszegi, 2006; Weinberg, 2006). The core idea is that once individuals are sufficiently certain that they are of high ability, they stop seeking information, as this only offers the risk of revising their beliefs downward. By contrast, individuals with a low relative ability self-assessment seek information as long as there is a chance for improvement.

The social signaling interpretation given in Bénabou and Tirole (2002) focuses on a different aspect: the idea that the easiest way to lie is to lie to yourself first. This moves back in the direction of a bias in judgment, but a bias with social roots, not individual ones, and could offer a functional explanation for the existence of a preference for high self-confidence. Of course, social signaling may also have a more direct and strategic interpretation: subjects called upon to provide a self evaluation may consider this as a social act, with possible social consequences, and may consciously choose to report a higher estimate of their own abilities than they actually hold. We report a more precise definition of social signaling, and discuss its implications and predictions, later (see section 6). Our data allow a sharp separation between the first two theories (which appeal to "bare" consumption value and self-motivation value) and the third explanation provided by social signaling, and we focus on this contrast here.

We tested the information acquisition prediction of the self-image management theory by offering our subjects the opportunity to find out exactly how well they did in the tests relative to the other participants. Our data strongly reject the prediction of self-image models: We find that individuals with high beliefs are *more likely* to demand information about their ability. Thus, beliefs do play an important role in demanding information, but not in ways that are consistent with preserving self-image.

Our results are more in line with a model in which individuals enjoy acquiring evidence confirming a positive belief, and enjoy sending public signals based on such evidence, rather than preserving a fortuitous positive self-image. We further corroborate this interpretation by examining how individual personality differences affect relative ability judgments. Consistent with our interpretation, we find that more socially dominant individuals (high on the Social Potency scale) make more confident judgments, holding constant their actual ability. This effect is also quantitatively large: Of those individuals with a below-median score in social dominance, only 33 percent think they are in the top 20 percent of the IQ distribution. Of the individuals with an above-median score in social dominance, 55 percent think they are in the top 20 percent, when, in fact, 20 percent of both groups are in the top 20 percent. Further, the broad trait of Neuroticism, and more specifically the scale of Stress Reaction, predicts instead reducing the ranking of one's own performance.

In summary, our results show that overconfidence cannot arise from Bayesian updating on signals about one's ability. Our results also lend little support to the view that overconfidence is the result of indirect self-deception through the management of information acquisition, as we find that individuals with optimistic beliefs about themselves seek more information, in contradiction to those models. Instead our findings suggest that overconfidence is likely to arise in the process of communicating judgments about one's relative performance to others.

The remainder of this paper is structured as follows: Section 2 describes our empirical setup. Section 3 presents the basic findings on overconfident relative ability judgments. Section 4 introduces a framework of incomplete information about one's own ability, derives restrictions that this places on relative ability judgments, and tests them. Section 5 discusses image preservation as a source of overconfidence, and provides an empirical test. Section 6 presents evidence on how personality traits are related to overconfidence. Section 7 concludes.

2 Design of the Study

The data was collected from 1,063 trainee truck drivers at a driver training school in the upper Midwest of the United States, on Saturdays that fell in the middle of a two-week basic training course the subjects were undertaking in order to earn a commercial driver's license. The two tests were part of a larger data collection process for the Truckers and Turnover Project (Burks et al., 2009), which was administered to participants in groups of twenty to thirty from December, 2005 to August, 2006. At the beginning of each session, subjects were guided through a consent form that explained all the conditions for participation in the study. A central point in the informed-consent process was to explain to the participants that their employer would see none of the individual data collected in the project (see Burks et al., 2008, for more details).

The subjects participated in two tests of cognitive ability, in which appropriate incentives were provided. The first was part of a standard non-verbal IQ test, Ravens Progressive Matrices (Raven, 2000), which involves identifying visual patterns; this was administered on notebook computers. The second was a section of the Adult Test of Quantitative Literacy (here after Numeracy), from the Educational Testing Service (ETS), which involves reading text samples and solving arithmetic problems that are based on the text; this was administered using paper and pencil. The IQ test was administered before the numeracy test. In total, 1,063 subjects participated in this study; however, because we tried out a different IQ measure before switching to the Raven's, 1,016 subjects have non-verbal IQ data.

The sequence of events was the same in both tests. First, using the standard instructions that came with each test, the nature of the test was explained, directions about how to complete questions were given, and a sample question was provided and the correct solution presented. After the instructions, we recorded the first self-assessment of the subjects' abilities: the subjects were asked how well they thought they would do in this test relative to the rest of the session's participants by identifying the quintile of group performance in which their score would fall. After the test was completed, the subjects were asked to self-assess a second time, by again picking the quintile of group performance in which their own score would fall.

We paid subjects for their attendance and their performance (Borghans et al., 2008). Each subject took part in two sessions, each two-hours long; both cognitive skills tests were in the second session. We paid an initial amount of $20 for participation at the beginning of each session. In addition, for each cognitive skill test, we randomly selected two subjects from each group after the test and paid each of these persons one

dollar for every correct answer in the IQ test (maximum possible payout of $48), and two dollars for every correct answer in the Numeracy test (maximum possible payout of $24). We also paid each subject $2 each time the subject correctly identified the quintile into which his or her own score actually fell (maximum possible payout of $4 per test). Payments depending on performance were explained before each test, as part of the test's instructions.

Because the payout calculations for the Numeracy task were manual, and because subjects were enrolled in a course that continued for another week, we paid out all the earnings from participation beyond the show-up fees at the beginning of the work week following the Saturday test administration. This provided us with the opportunity to also ask subjects, immediately after their second self-assessment response on each test, whether they would like to learn on payout day both their exact score, and what their actual relative performance was, i.e. which quintile they were actually in. Those who answered "no" only received their payout, and not this extra information. Thus, this answer is our measure of each subject's demand for information about their relative performance: "yes" signaled the desire to know. We added this question after data collection began, so there are 839 subjects that indicated their demand for information on the IQ test, and 889 that did so on the Numeracy test.

In addition to providing a clear measure of the demand for information about one's relative performance, this design provides incentives to truthfully report one's self-assessment of relative performance, and to make that estimate as accurate as possible. A strength of the design is that we asked subjects about their performance relative to a specific group of people, whom they had known for more than a week by the time of the experiment. Therefore, unlike the most common studies of overconfidence in the psychology literature, our design rules out that subjects were comparing themselves to groups outside the lab. Finally, it avoids the ambiguities of earlier studies that asked individuals whether they were above or below the mean.[2]

During the entire experiment we collected a variety of additional demographic and socio-economic information. Subjects also filled out the Multidimensional Personality Questionnaire (MPQ). The MPQ is a standard personality profile test (Patrick et al., 2002; Tellegen, 1988; Tellegen and Waller, 1994). It consists of questions concerning 11 different scales that represent primary trait dimensions: wellbeing, social potency, achievement, social closeness, stress reaction, alienation, aggression, control, harm avoidance, traditionalism, and absorption. In our study we used the short version

---

[2]If, e.g., the median of abilities is significantly above the mean, a fraction significantly larger than half could correctly answer that they are better than average, which makes the interpretation of these studies difficult.

(Patrick et al., 2002), which has 154 multiple choice questions.

3   Evidence of Overconfidence

Table 1 presents some basic descriptive statistics. The first panel in the table shows the number of correct answers in the two cognitive tests. In Burks et al. (2008), it is shown that the distribution of the score in the Raven's task in our sample is close to that of representative samples, although slightly lower: for example, the median score in our sample is 47.5, in the representative sample (reported in Raven, 2000, the median is 52). Turning to the demographics of our sample, we see that the most frequent education level in our sample is a high school degree, though some have also degrees from technical schools, and a significant fraction has at least some college education. The table shows that our sample is predominantly Caucasian, male, and were on average in their late thirties. See Burks et al. (2008, 2009) for a more extensive discussion.

Overconfidence in Relative Ability Judgments

In this subsection, we present the basic evidence on overconfidence in our study. This serves two purposes: to show that our results are comparable to overconfident judgments found in other studies, and, to motivate the theoretical model we discuss later.

Figure 23 displays the distribution of relative ability judgments across all individuals. It shows a typical pattern found in a large number of studies: Very few individuals rate their ability in the bottom 40 percent of the ability distribution. By contrast, well above 60 percent think they are in the top 40 percent. The figure shows a very similar pattern for the relative ability judgments in the two tests.

Figure 2 displays relative ability judgments as a function of the true ability in the IQ test, reporting under- and overconfident judgment relative to the true ability of the individual. Shadings indicate the extent of overconfidence: Light shading indicates that the individual is just one quintile off, darker shading indicates that the individual is more than one quintile off. Panel A displays relative ability judgments before the IQ test. The figure shows that overconfident judgments are pervasive across the ability spectrum, except where impossible by definition in the top ability quintile. The figure also shows that the relative ability judgments are strongly asymmetric: underconfidence is much rarer than overconfidence. Panel B in Figure 2 displays relative ability judgments after the IQ test and shows essentially the same pattern: taking the test

does not qualitatively change the distribution of beliefs compared to those reported earlier, after just the instructions and practice question. The relative ability judgments for the Numeracy test are presented in Figure 3: The results are very similar to the case of IQ test.

4 The Bayesian Model

In this section we establish the benchmark model of the behavior for a population that is forming beliefs about their own ranking using belief updating based on the information they have available. Our aim is to show that such a model can produce some features of the relative ability judgments that we showed in the previous section, but to also derive testable restrictions imposed by the Bayesian theory.

In the model we consider a large population of individuals, each one endowed with a type t, which is the value of a specific characteristic. For example the type of an individual might be his height, something easily determined and observed. Another more interesting example is his ability to score high on an intelligence test, a quality that we briefly described as the individual's *IQ*. We are interested in types that are ordinal quantities. In what follows, we will restrict attention to judgments about the individual's position in the distribution of outcomes. As in our empirical study we elicit judgments about the quintiles, so we also restrict our notation in the model to quintiles.

The type of each individual is determined independently, according to a known probability measure on the set of types. Thus, the population has a common prior on the distribution of types, which (since types are percentiles) is the uniform distribution. Individuals do not know their type, but during their life they gather information by observing private signals about it. On the basis of this information they update in a Bayesian fashion their belief about their type, which initially was the common prior, and therefore also they update the belief they have about their own relative position in the population with respect to the characteristic we are considering. For example, through their school performance, job performance, as well as occasional exchanges with other people they form an opinion about their IQ, and hence of their relative standing with respect to this characteristic within the population. Formally, we assume that individuals observe an outcome $x_i \in X, i = 1, ...n$ from some signal space X, where we assume n is larger than 5, the number of quintiles. [3] A subject participating

---

[3]Notice that we restrict attention to one draw from a signal structure, rather than, e.g., a dynamic acquisition of signals. Dynamic acquisition signals can be redefined as a single draw from a single signal structure.

in an experiment like ours comes to the laboratory with this posterior belief about his ability. Denote the probability that an individual receives signal x_i given that he is of ability t_k by $p_k(x_i)$. Then the individual's posterior beliefs about his ability is given by

$$\Pr(t_k|x_i) = \frac{p_k(x_i) \cdot \frac{1}{5}}{\sum_j p_j(x_i) \cdot \frac{1}{5}} = \frac{p_k(x_i)}{\sum_j p_j(x_i)} \tag{1}$$

The signal structure $p = (p_k(x_i))_{k=1,\ldots,5,i=1,\ldots,n}$ is the true information structure. We have very little hope of determining this object empirically in a direct way. So suppose we ask the individual to predict the quintile in which his IQ score will fall, and promise him a payment if his prediction is correct. Let us assume that our incentives are sufficient motivation for him to state the truth, and that he believes that our test is unbiased. Then an individual who observes the signal x_i will pick the most likely quintile given x_i, i.e. the individual will indicate that he is *most likely* in ability quintile $s(i)$, where:

$$s(i) = \arg\max_j Pr(t_j|x_i) = \arg\max_j p_j(x_i) \tag{2}$$

We call the theory that subjects follow this procedure of deriving posterior beliefs with Bayesian updating from a common prior and then truthfully report to us the most likely quintile the *Bayesian model*.

A large fraction of subjects thinking that they are in the top two quintiles is consistent with this model. To illustrate, consider an example with only two types, good and bad. The top two quintiles (40 percent) are good types, and the remaining three quintiles are bad types. This is the distribution of types and the common prior. The only source of information for individuals is a test that everybody takes. Good types pass the test for sure, bad types only pass it with probability 50 percent. The posterior probability that an individual is a good type if he passes the test is:

$$\Pr(\text{good type}|pass) = \frac{1 \cdot 0.4}{1 \cdot 0.4 + 0.5 \cdot 0.6} = \frac{4}{7} \tag{3}$$

so individuals who pass the test and answering truthfully state that their most likely type is the good type. A fraction of 70 percent of the population passes the test (all the good types, plus half the bad types): Thus in this population, 70 percent truthfully report that they belong to the top 40 percent, much as we observe in the data presented in the previous section. Beliefs are on average correct: 70 percent of the population

believe that they are good with probability 4/7, and 30 percent believe that they are good with probability 0. Overconfidence in beliefs arises because the test was easy (all good types and half of the bad types pass the test). If the test were hard (for example, all bad types and half of the good fail), underconfidence would arise, and only 20 percent would state that they are good types.

4.1 Testable Restrictions on Beliefs

Incomplete information about one's abilities, and a particular feature of the signal structure (an easy test) may lead to overconfident beliefs. However, the Bayesian model imposes testable implications on how the distribution of relative ability judgments should be related to true abilities. These are testable because the experimenter also observes the true score of the individual in the test, so he has at the end of the experiment for each subject a pair of observations, (*true score, stated quintile*). The true score is not a precise measure of the IQ of an individual, of course, but it is good enough so that we can ignore sampling error with respect to the quintiles.

Since individuals have an incentive to choose the most likely quintile, the Bayesian model requires them to use (1) to form their posterior and (2) to select their statement. Denote the expected fraction of individuals in true ability quintile k assigning themselves to quintile j based on the signal structure provided in (1) by $q_k(s_j)$. We call the function $(q_k(s_j)_{k,j=1,...,5})$ allocating each type k to five quintiles in specific proportions, the theoretical allocation function. It defines a 5-by-5 matrix of relative ability judgments. Note that for every true ability quintile k, $\sum_j q_k(s_j) = 1$. The items in the diagonal denote the fraction that hold the correct beliefs about their abilities. Entries $q_k(s_j)$ with $k < j$ indicate individuals who hold overconfident beliefs, while entries with $k > j$ indicate the fraction of individuals holding underconfident beliefs. What restrictions does Bayesian updating place on this matrix? Because individuals pick the most likely quintile given the signal x_i that they received, the mode of individuals thinking they are in quintile k must actually have true ability quintile k. That is, Bayesian updating imposes that:

$$q_k(s_k) = \max_l q_l(s_k) \tag{4}$$

In the appendix, we characterize this property more fully. The theoretical allocation function allows us to sidestep a problem that has no easy direct solution: what is the true information structure p? If the behavior we want to describe only depends on the posterior distribution over quintiles given the signal, then we may assume that the

10

true information structure takes values in the simple signal space given by the set of quintiles. To see this, consider an information structure where individuals observe some signal x in some arbitrary signal space X, compute the posterior on their type, and state the most likely quintile. This information structure, in our environment in which the only task of the individuals is to state the most likely quintile, is equivalent to a simple information structure where individuals are directly communicated the quintile they should state (so the signal space is the set of quintiles), and they do so (because the diagonal condition (4) insures this behavior is incentive compatible). The theoretical allocation function derived from equations (1) and (2) can be considered a canonical information structure. The harder problem: "Is there an information structure that can generate the data?" has been replaced by the easier problem: "Is there a canonical information structure that can generate the data?" This problem has an answer, that we present in the next section. [4]

4.2   Rejection of the Bayesian Model

We have seen that Bayesian updating implies condition (4), which we may call diagonal condition, because if the theoretical allocation function is read as a matrix, then the entries with the largest values are on the diagonal. Table 3 presents an illustration of an allocation function satisfying this condition. But how can restrictions imposed by (4) be tested against the *empirical allocation function* $\hat{q}_k(s_j)$, i.e., the empirical distribution of relative performance judgments as a function of the individuals' true ability? Intuitively, strong evidence that the main diagonal condition is violated rejects the Bayesian model.

Table 2 displays the empirical allocation function for the numeracy and IQ test. The table shows that in both cases, the empirical frequencies violate the diagonal condition. For example, in the numeracy test, only 18 percent of the individuals from the third quintile put themselves into the third quintile. By contrast, 40 percent from the first quintile and 27 percent from the second quintile put themselves in the third quintile, in violation of the diagonal condition (4). But is the violation significant? Since we don't know the underlying signal structure, how likely is it that a signal structure satisfying (4) generated the data in Table 2? We propose a test that gives the Bayesian model the best chance not to be rejected.

We estimate the parameters of the theoretical signal structure by maximum likelihood

---

[4]Notice that we have so far assumed that all individuals draw signals from a common signal structure. This, however, is not a crucial assumption. If different individuals drew signals from different signal structures, this can be modeled as a meta signal-structure, in which individuals first observe from which sub-structure they will draw signals.

subject to the constraint imposed by (4). That is, we compute the $q = (q_k(s_j))_{k,j}$ that solves:

$$\max_q \sum_{j,k} n_{kj} \log(q_k(s_j)) \tag{5}$$

subject to for every $k, j, q_k(s_j) \geq 0, \sum q_k(s_j) = 1$ and

$$\text{for every } k, q_k(s_k) = \max_l q_l(s_k)$$

where n_{kj} is the number of individuals of ability quintile k saying that they are in quintile j. This is a concave problem and maximization is straightforward with numerical methods. Denote the solution to (5) by $q_k^{ML}(s_j)$. Notice that this gives the best chance to the null hypothesis of Bayesian updating, since we pick q^{ML} as the one satisfying (4) that best fits the observed data. The constrained maximum likelihood estimator for Numeracy and IQ test are reported in table 3.

We then calculate the fit of q^{ML} to $\hat{q}$ as the mean square root error from each cell:

$$\hat{d} = \frac{1}{25} \sqrt{\sum_{j,k} (\hat{q}_k(s_j) - q_k^{ML}(s_j))^2} \tag{6}$$

The distance measure is $\hat{d}^{IQ} = 0.026$ for the IQ test, and $\hat{d}^{Num} = 0.033$ for the numeracy test. That is, the average deviation from the ML estimate of q is 2.6 percentage points in the IQ test and 3.3 percentage points in the Numeracy test. In order to assess whether the fit $\hat{d}$ is improbably bad, we generate 100,000 simulations of the same sample size as our data using q^{ML} as the data generating mechanism and calculate the distances d_n for each trial n. This provides us with an empirical distribution function for the distance measure d to calculate the probability that a draw from q^{ML} has a worse fit than the empirical allocation function $\hat{q}$. The p-values are $p = 0.005$ for the IQ test, and $p = 0.001$ for the numeracy test. Therefore, we clearly reject the hypothesis that our data is generated by the joint hypothesis of imperfect information about ability, Bayesian updating from a common prior using this information, and truthful revelation of the belief thus formed.

5 Do Self-Image Concerns create Overconfidence?

The previous section tests and rejects a wide class of models that rely on Bayesian updating from a common prior after exogenous arrival of information. Other models have been developed to explain overconfidence arising endogenously as a function of individuals' choices.

Two recent papers (Kőszegi, 2006; Weinberg, 2006) have argued that a concern for self-image can lead to overconfidence. If individuals' utility depends on their belief about their ability, this can lead to an endogenous mechanism that produces results as if they were drawing signals from "easy test" signal structure in Benoît and Dubra (2007). This requires that utility is sufficiently "kinked" in the belief. Kőszegi (2006) provides an example in which an individual's utility discretely increases by some fixed amount v if the individual believes that the chance that his ability t is below some threshold $\hat{t}$ is small. Formally, utility is given by

$$U(c, \hat{t}) = u(c) + v \cdot I(F(\hat{t}) \leq x) \tag{7}$$

where F is the c.d.f. of the individual's current belief over his ability. To see how this can lead to overconfidence, assume that the individual's belief currently is that $F(\hat{t}) < x$ and that he is offered more information about his ability. Suppose that the only change in utility he can have from further information is from the possible change in self-image. Then he will never seek more information, because more information only harbors the risk of revising his belief downward. Conversely, if $F(\hat{t}) > x$, the individual will seek more information. If his belief is further revised downward, this leaves utility unchanged. If the individual receives a positive signal, he will gain utility v if $F'(\hat{t}) < x$ where $F'()$ is the c.d.f of beliefs incorporating the new information. Thus, this model can generate a pattern in which individuals with low beliefs will seek all the information they can find, while individuals with high beliefs will have less accurate information: of all the individuals with initially low beliefs, all those with high ability will revise their views upward. By contrast, some of the individuals who initially had high beliefs will have received good signals by chance, but will not discover their mistake. The result is that too many individuals will believe they have high abilities.

5.1 Demand for Information

We test the prediction of this model by testing the implication that individuals with high beliefs should be less likely to seek more information about their ability. Recall that after each test, we offered the subjects the opportunity to find out exactly how well they did relative to the others. We thus gave the individuals the chance to obtain more information, exactly as required in the model. This test also has the feature that it does not rely on the assumption of common prior. Rather, it measures the demand for information directly as a function of the individuals' beliefs.

Figure 4 displays the fraction of individuals demanding information about their performance as a function of how well they did in the test and their stated belief about their performance. Because of the small number of observations, we exclude individuals with

13

beliefs in the bottom two quintiles. Panel A in Figure 4 displays the results for the IQ test, while the results for the Numeracy test are displayed in Panel B. Both Panels show a strong impact of beliefs on the demand for information. However, in contrast to what is predicted by models in which the belief about ability enters the utility function in the manner specified above, among our subjects individuals with a higher belief are more likely to ask for the performance information. The figure also controls in a rudimentary way for differences in true abilities by splitting the sample into the top and bottom half of the performers. Thus, by comparing individuals with identical beliefs in the top and bottom half of the true abilities, we can gauge the impact of true ability on the demand for information. There is, essentially, no relationship between ability and the demand for information.

To formally test the model, we estimate the following probit equation

$$\Pr(seek_i = 1 | q_i, x_i) = \Phi(\gamma q_i + \beta' x_i) \tag{8}$$

where $seek$ is an indicator variable equal to 1 if the individuals seeks information about his performance in the test, and zero otherwise. Φ is the cumulative normal distribution. We estimate the equation separately for the IQ and numeracy tests. Our variable of interest is stated belief of the individual $q \in \{1, 2, ..., 5\}$ regarding the most likely quintile. The vector of control variables x includes controls for test performance. We estimate a five-part linear spline in test performance, with the splines defined over quintiles in order to control for test performance in a flexible way. We also include personality characteristics as measured by the Multidimensional Personality Questionnaire (Patrick et al., 2002). Our estimates also include a large set of controls for socio-demographic differences across subjects: 5 dummy variables for education levels, 5 categories for ethnicity, a gender dummy, age and age squared, and household income.

The results are displayed in Tables 4 and 5 for the demand for information about one's performance in the IQ and numeracy test, respectively. The table displays marginal effects on the probability of seeking information, rather than the bare coefficient estimates. Both tables are structured the same way. In the first column, we test whether, as indicated by the figure, a higher belief increases the likelihood of demanding information. Column (1) in Table 4 controls for test performance using a flexible functional form. It shows that conditional on actual performance, the subject's belief about their performance predicts whether or not they seek information. More optimistic beliefs increase the likelihood of seeking information: a one-quintile increase in beliefs is associated with a 3 percentage point higher probability of demanding information about

14

the test.

The results are even stronger for the numeracy test, where a one-quintile increase in the belief leads to almost a 6 percentage point increase in the likelihood of seeking information. In both cases, the effects are statistically highly significant. Column (2) adds personality traits as controls, obtained from the MPQ. The only significant trait is Harm Avoidance, a measure of the relative preference of individuals for less risky situations. The effect is negative and small, and lends itself to a plausible interpretation that individuals who are less risk averse are more likely to seek information, preferring the extreme values to their expected value. In column (3), we add the socio-economic control variables. However, they have no effect on the coefficient of interest. As a robustness check in column (4), we use the belief before the test as the independent variable to explain the demand for information. In both tests, the belief before the test is significant as well. As a final step, we build on this last specification to examine whether it is the current belief the subject holds that determines the demand for information, or just some general notion of confidence that may be reflected in all of the beliefs. Therefore, we also add the beliefs about the ability before the test as well as the beliefs about the ability in the other test as explanatory variables. Some individuals do change their evaluation over the course of the test (correlation between pre and post test beliefs: $\rho = 0.64$ for IQ and $\rho = 0.74$ for numeracy). Similarly, while beliefs are correlated across tests, they are not perfectly correlated ($\rho = 0.54$ for beliefs after the test). This allows us to examine the specificity of the link between beliefs and the demand for information. Our results show that the link is highly specific. In Table 4, we see that only the most recent belief is significantly correlated with the demand for information. Confidence in the numeracy test is uncorrelated with the demand for information about IQ, and so is confidence before the test, *ceteris paribus*. Our results are slightly weaker for numeracy, where we find a weak effect of confidence in IQ on the demand for information about relative performance.

Overall these results clearly reject the driving force for overconfidence postulated by models of self-image concerns (Kőszegi, 2006; Weinberg, 2006). In fact, we find the opposite of what these models predict: More confident individuals are more likely to seek information. This is consistent with a model in which individuals value the signals they send about their ability, not the resulting belief. However, this mechanism also tends to undermine overconfidence, as individuals with high relative ability judgments are more likely to seek information, thus throwing into sharper relief the question how overconfidence comes about in our subjects. One possibility is that individuals do not process information in a Bayesian manner. This interpretation is consistent with our evidence from the previous section, rejecting specifically that relative performance

judgments are formed in a way consistent with Bayesian updating from a common prior and truthfully reported. In particular, this explanation suggests that personality characteristics may be related to the misinterpretation of information. We explore this explanation in the next section.

6 Personality Traits affect Overconfidence

Individuals differ not just in their level of cognitive skills, but in other dimensions of their preferences and personality that can help predict how their relative ability judgments match their actual ability. Therefore additional insights into the causes of overconfidence may be provided by information on their personality traits. These traits can affect relative ability judgments in two different ways. First, they can affect the information that an individual collects during his life. This is true even if we consider choices about information gathering as part of a single player problem. In this case an individual in general should want to be as well informed as he can. However, different personality traits may influence the choice of signal structure he uses (for example the information that he is gathering, or he is paying attention to) among several incomparable ones. Notice, however, that differences in information acquisition due to differences in personality traits alone cannot explain overconfident judgments as we found them in section 4, as individuals should properly discount the fact that different individuals seek different information in forming their beliefs.[5] Second, personality traits may affect the way in which individuals process the same information, and signal their opinion to the outside world: they affect either what people think of themselves (giving rise to what might be called a socially-rooted bias), or what people strategically choose to state about themselves (giving rise to strategic lying), or both. We have already seen that Bayesian updating alone cannot explain the pattern of confidence we observe in our data, so we focus on how personality traits can affect social signaling.

One important factor is the importance individuals attach to the opinion of others about their own skills. Consider, to analyze these effects, a population that is heterogeneous in these two dimensions: cognitive skill and the utility derived from the opinion of others about the level of their cognitive skills. The information gathered in life has provided each individual with some knowledge of his or her cognitive skill.

---

[5]In our model, differences in personality can be modeled as a meta-signal structure, in which individuals first receive information about their personality type, and then receive a signal about their ability. In this context, a version of the model by Santos-Pinto and Sobel (2005) may explain how differences in personality traits lead to overconfidence: Loosely speaking, in their model overconfidence can arise as individuals neglect the fact that different individuals may seek different information in forming their judgments.

Suppose now that the individuals in the population are asked to provide some signal about their skill based on the private information they have about it. The public will observe this statement as well as a noisy direct signal about the skill, and will be able to compare the two. Individuals will pay a cost that is increasing in the difference between what their statement said and what the signal shows. For example, in our experiment the cost is that the payment to subjects is decreasing in the distance between the stated quintile and the true quintile; but in social life there are several other ways in which this cost of being "found out" may occur.

Subjects who do not derive utility from the high opinion of others will focus on the direct cost, and state to the best of their knowledge a correct evaluation of their skill. Subjects who instead care that others have a good opinion of them will bias their statement upward, even at the cost of some utility loss that might follow from the disagreement between what they say and what they are. In a world of rational agents, people who receive this information will take into account this bias. Since there are individuals who do not bias their statement, the signal is still informative, and members of the public will still upgrade their beliefs after confident signals, and the people who derive utility from the opinion of others will still bias their statements upward.

The main hypothesis we are proposing here is that subjects differ in the strength of preference for a positive view that others have about them. Subjects who have stronger preference are more inclined to send a signal about their skills which is more positive than their information would grant, even at some cost. In our experiment the cost is reducing the probability of obtaining the monetary prize: but there are of course many social costs that are attached to such discrepancy. Perhaps the underlying psychology is that these subjects process the information they have received in a biased manner for this social reason, and thus misrepresent their real skill to themselves. Or, it may be that they strategically lie, misleading others, but not themselves. We do not suggest one of these two possibilities is exclusively correct. Probably a little of both is true in the population, and perhaps to some degree also in many individuals. As Bénabou and Tirole (2002) suggest, a very good way to lie to others is to lie to yourself first. What is crucial to our hypothesis (and this is the reason for describing it as social signaling) is that the main motivation for a misrepresentation is that it affects the individual's social standing.

We test this theory by focusing on dimensions that can readily be measured using personality scales, such as the individual's desire to be in a dominant position relative to others. One of the MPQ traits, Social Potency, provides a good measure of the strength of this preference. We predict that individuals who score high on Social Potency will state higher relative ability judgments than warranted given their characteristics. An

important concern in this context is separating the desire to dominate others from other social motivations or from an absolute desire to achive. The MPQ also allows us to distinguish this from a more general desire to be connected to others, which is measured in the Social Closeness scale. The MPQ also allows us to distinguish the desire to dominate from general drive to achieve, by including the Achievement scale as a control.

The second important dimension is how individuals respond to negative social feedback. We hypothesize that if individuals are worry-prone and feel vulnerable, this may moderate their stated beliefs about themselves to make it less likely to experience these negative social emotions. The MPQ also allows us to control for other aspects of risk preferences, such as a more general tendency towards prudence, as measured by the Harm Avoidance scale, and general pessimism captured by the Alienation scale.

6.1 Descriptive Evidence

Figure 5 provides a first summary of the evidence. It shows relative ability judgments and actual abilities for individuals who have different scores in personality traits. Each panel reflects a different personality trait. In each case, we cut the sample by the median trait score. For example, in Panel A, the first graph shows that about 30 percent of the individuals scoring below the median in social potency think they belong to the top 20 percent in the IQ distribution. By contrast, 55 percent of the individuals scoring above the median in social potency think they are in the top 20 percent. Each graph also contains the actual fraction of individuals scoring in the top 20 percent for each subsample.

The graph shows virtually no difference between high- and low- social potency individuals in terms of actual ability. The results for relative ability judgments in the numeracy test are very similar. Thus, social dominance appears to pick up quantitatively important differences in the overconfidence of judgments, while being unrelated to differences in actual abilities. Turning to the graph that cuts the sample by social closeness, we see no differences in relative ability judgments. Thus, it appears that individuals who care more about sociability are not more confident in general; the relationship is limited to the aspect of dominance relative to others. The third graph cuts the sample by the median of the stress reaction score. Individuals who are highly sensitive to social stress have substantially more timid judgments about their ability, as can be seen in the graph, while this is again not related to differences in actual abilities. Again, a very similar pattern emerges when we examine relative ability judgments regarding the numeracy test in Panel B.

6.2 Econometric Results

In order to examine these hypotheses using a formal statistical test, we proceed in two steps to make transparent the role of the econometric structure imposed in the estimation. In a first step, we estimate an ordered probit model of *confidence* judgments as a function of personality traits. The individual believes his most likely quintile is $b_i = k$ if

$$\alpha_k \leq \gamma' MPQ_i + \beta' x_i + \epsilon_i < \alpha_{k+1} \tag{9}$$

where $1 \leq b_i \leq 5$ is the individual's belief about his most likely quintile. MPQ is the vector of 11 personality traits, and x is a set of control variables, α_k are the judgment cutoffs, and ϵ_i is a standard normal residual. This gives rise to an orderd probit model that can me estimated by maximising the likelihood

$$\Pr(b_i = k | MPQ_i, x_i) = \Phi(\alpha_{k+1} - \gamma' MPQ_i - \beta' x_i) - \Phi(\alpha_k - \gamma' MPQ - \beta' x) \tag{10}$$

The function $\Phi()$ is the cdf of the standard normal distribution. This specification allows us to test whether, conditional on a broad set of controls, personality characteristics affect confidence judgments in the predicted way.

In a second step, we then impose slightly more structure and estimate an ordered probit model of *overconfidence.* Our theory predicts relationships between personality traits and overconfidence, and it is desirable to test these directly. We model over- and underconfidence in an ordered model where the difference between individual i's confidence judgment b_i and his actual ability q_i is $b_i - q_i = k$ if

$$\alpha_k \leq \gamma' MPQ_i + \beta' x_i + \epsilon_i < \alpha_{k+1} \tag{11}$$

This implies for the probability of $b_i - q_i = k$

$$\Pr(b_i - q_i = k | MPQ_i, x_i) = \Phi(\alpha_{k+1} - \gamma' MPQ - \beta' x) - \Phi(\alpha_k - \gamma' MPQ - \beta' x) \tag{12}$$

which we estimate by maximum likelihood. However, we also need to take into account the truncation of $b_i - q_i$ induced by the actual ability q_i. In this specification of the model, individuals of the top quintile cannot overestimate their ability, thus if $b_i - q_i = 0$ for $q_i = 5$, we only know that $\gamma' MPQ_i + \beta' x_i + \epsilon_i \geq \alpha_0$ (and not $\alpha_1 > \gamma' MPQ_i + \beta' x_i + \epsilon_i \geq \alpha_0$). Similarly, individuals with $q_i = 4$ can only overestimate their ability by one quintile, and we take this into account analogously. See the appendix for details.

Table 6 presents the results from the model of confidence judgments, as specified in equations (9) and (10), displaying directly the marginal effects on believing that one is

in the top 20 percent, for confidence judgements in the IQ test in columns (1) to (3), and the numeracy test in columns (4) to (6). The first column in each group presents the partial correlations with the only the personality traits included. While social potency and stress reaction are both significant with the predicted sign, other personality traits are significant, as well in contrast to what we expected (in particular, absorption, traditionalism and social closeness). However, as we progressively include more stringent controls, these variables are no longer significant. In our strictest specification (in columns (3) and (6), respectively), social potency and stress reactions still have their predicted sign (together with harm avoidance in the specification for IQ, which is also consistent with our model). Differences in the personality traits of social potency and stress reaction have quantitatively large effects on confidence, conditional on all of our controls. An increase of 8 index points for social potency (social potency's interquartile range) increases the probability that the individual thinks he is in the top 20 percent of the distribution by 8.8 percentage points, while an increase of 9 index points for stress reaction (stress reaction's interquartile range) reduces the probability that the individual thinks he is in the top 20 percent of the distribution by 7.2 percentage points

We now turn to the estimates of the impact of personality traits on our direct measure of overconfidence, as specified in equations (11) and (12). The results are presented in Table 7, calculating directly the marginal effects for being overconfident. While the identification of the personality traits on overconfidence in equation (9) was achieved by conditioning on actual performance, the results here are identified directly by the difference between the belief b_i and the actual ability q_i and thus impose somewhat more structure. The results in all three columns and for both measures of overconfidence (IQ and numeracy test) closely parallel those in table 6, but display the specificity of social potency and stress reaction for overconfidence even better.

Again, the estimates conform well to our theory. Even though we include the entire set of personality characteristics and some aspects are highly correlated, we find highly specific effects exactly as predicted by the theory. In contrast to the estimates based on equation (9), specifying the model directly in terms of overconfidence makes the estimates more stable and less dependent on the conditioning variables.

It is also worth pointing out that the results show a very strong specificity of personality traits for overconfidence. For example, the results show that it is the desire to dominate others that is predictive of overconfidence, not the desire to socialize with others (social closeness) or the desire to perform well (achievement). Even though social potency and social closeness are strongly correlated ($\rho = 0.39$) as well as achievement ($\rho = 0.23$), it is only social potency that is predictive of being overconfident. Similarly, stress reaction and general pessimism (alienation) are highly correlated ($\rho = 0.58$), yet it is only stress

reaction that is predictive of less overconfidence. A final way to assess the specificity of the effects is to ask if stress reaction and social potency were also significant if one took the agnostic null of no relationship between personality characteristics and overconfidence. With 11 variables in a regression, by the definition of the 5 percent significance level, there is a much higher chance that some are significant than 5 percent. First, notice that an F-test that the coefficients γ of the personality traits are zero is overwhelmingly rejected ($p < 0.001$ in all cases). Second, we can also apply the Holm (1979) correction to adjust the critical values to 11 hypothesis tests. Even when we apply this correction, stress reaction and social potency are still significant at the 1 percent level for numeracy, while stress reaction is only significant at the 6 percent level with the Holm correction.

The results show a quantitatively large effect of of social potency and stress reaction on overconfidence. Based on even the most conservative estimates in column (3) and (6) respectively, raising social potency (stress reaction) by it's interquartile range, increases the probability that an individual makes an overconfident judgment in the IQ test by 6.3 (6.4) percentage points for social potency (stress reaction). The quantitative implications are very similar for overconfidence in the numeracy test.

7   Conclusions

We have examined in an experimental setup evidence for overconfidence of individuals about their intelligence and its possible motivation. We reported three main findings. First, we rejected the Bayesian model, that is, the hypothesis that overconfidence results from incomplete information about one's own ability, Bayesian updating from a common prior, and truthful revelation. The test we use is general, and may be used to probe the same hypothesis in similar studies. In our data, the level of overconfidence in our subjects' statements is beyond what can occur in a world of truthful Bayesians.

Second, we rejected the hypothesis that optimistic beliefs about one's abilities lead individuals to avoid new information about their absolute or relative performance. As an implication of this finding, we reject a central prediction of models of self-image management. These models assume that individuals derive utility directly from better beliefs about their own skills, and predict that when individuals optimally manage information acquisition those with better beliefs will be more reluctant to search and observe further information about their abilities. In our data the opposite is true: we find a positive and highly significant association between optimism of beliefs and demand for information about one's relative performance. This relationship is, as we

have shown, specific to the belief about one's relative performance in the test at hand. Further, it is the belief after the test, not the belief about one's ability before the test, that predicts the demand for information. Individuals are more likely to demand feedback on performance when they have just received a positive impression of their performance, and this is precisely when self-image management concerns should lead to choosing ignorance.

Third, we show that specific measures of personality traits affect significantly the stated level of confidence (that is, the quintile of test performances in which the subject locates himself). The personality traits that affect the statement, and the direction of the effect, are consistent with the idea that the explanation of confidence is the social signal that positive confidence produces. Specifically, social potency, an indicator of personal inclination to a dominant role, strongly increases the probability that a subject states a higher level of relative performance, holding actual performance constant. Stress reaction and traditionalism have the opposite effect, reducing the level of confidence. Personality traits do not significantly affect the demand for information. This is consistent with the additional finding that personality traits which should affect the accuracy of self-evaluation (Control) do not appear to affect either demand or overconfidence, whereas traits that measure motivation for ranking (Social Potency) significantly affect statements but not demand. In Bénabou and Tirole (2002)'s classification, optimistic self-assessment seems motivated by its signaling value, that is, by its potential effect on the opinion of others. As we mentioned earlier, the individuals who give optimistic self-assessment may believe what they say, or may try to deceive others: we do not advance either explanation to the exclusion of the other, and our data cannot really provide a way to separate them.

These experimental findings are consistent with the current evaluation of the importance of self-esteem as a predictor of individual performance and success. In recent years, a re-examination of the correlation between self-esteem and outcomes of interest has consistently found a weak relationship to school performance (Davies and Brember, 1999; Kugle et al., 1983), and IQ (Gabriel et al., 1994). In addition, the causal direction is likely to go from performance to self-esteem as much as it is going in the opposite direction. The survey in Baumeister et al. (2003) is a thorough discussion of the evidence in favor of a positive effect of self-esteem on a range of performance measures, including happiness and healthy lifestyle, and the overall conclusion is that the evidence of a causal relation is weak at best. Similar results are reported in other surveys (Mecca et al., eds, 1989; Leary, 1999). If the utility from positive self-image has no individual functional basis and a positive self-image offers no improvement in any significant performance index, then it is natural to consider the possibility that

the roots of overconfidence lie in the value of over-confidence as a social signal (Leary and Downs, 1995; Leary et al., 1995). These findings also point to the importance of personality traits in predicting economic and strategic behavior (Rustichini, 2009).

References

Alicke, M. D., D. L. Klotz, T. J. Breitenbecher, J. Yurak, and D. S. Vredenburg, "Personal contact, individuation, and the better-than-average effect," *Journal of Personality and Social Psychology*, 1995, *68* (5), 804 – 825.

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

Baumeister, R., J. D. Campbell, J. I. Krueger, and K. D. Vohs, "Does High self-esteem cause better performance, interpersonal success, happiness or healthier lifestyles?," *Psychological Science in the Public Interest*, 2003, *4* (1), 1 – 44.

Bénabou, Roland and Jean Tirole, "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 2002, *117* (3), 261 – 292.

Benoît, Jean-Pierre and Juan Dubra, "Overconfidence?," MPRA Paper 6017, University Library of Munich, Germany October 2007.

Borghans, L., H. Meijers, and B. ter Weel, "The Role of Cognitive Skills in Explaining Cogntive Test Scores," *Economic Inquiry*, 2008, *46* (1), 2 – 12.

Burks, S.V., J.P. Carpenter, L. Goette, and A. Rustichini, "Cognitive skills affect economic preferences, strategic behavior, and job attachment," *Proceedings of the National Academy of Sciences*, 2009, *106* (19), 7745.

_ , _ , _ , K. A. Monaco, K. Porter, and A. Rustichini, "Using Behavioral Economic Field Experiments at a Large Motor Carrier: The Context and Design of the Truckers and Turnover Project," in S. Bender, J. Lane, K. Shaw, F. Anderson, and T. von Wachter, eds., *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches.*, NBER and University of Chicago Press, 2008.

Davies, J. and I. Brember, "Reading and mathematics attainments and self-esteem in years 2 and 6 - and eight-year cross-sectional study," *Educational Studies*, 1999, *25*, 145 – 157.

Dunning, D., "Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessments of Ability," *Journal of Personality and Social Psychology*, 1989, *57* (6), 1082 –1090.

Gabriel, M. T., J. W. Critelli, and J. S. Ee, "Narcisisstic Illusions in Self-Evaluations of Intelligence and Attractiveness," *Journal of Personality*, 1994, *62*, 143 – 155.

Hoelzl, E. and A. Rustichini, "Overconfident: do you put your money on it?," *Economic Journal*, 2005, *115*, 305 – 318.

Holm, Sven, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 1979, *6*, 65 – 70.

Kőszegi, Botond, "Ego utility, overconfidence, and task choice," *Journal of the European Economic Association*, 2006, *4* (4), 673–707.

Kugle, C. L., R. O. Clements, and P. M. Powell, "Level and stability of self-esteem in relation to academic behavior of second graders," *Journal of Personality and Social Psychology*, 1983, *44*, 201 – 207.

Leary, M. R., "The social and psychological importance of self-esteem.," in R. M. Kowalski and M. R. Leary, eds., *The social psychology of emotional and behavioral problems: interfaces of social and clinical psychology*, American Psychological Association, 1999, pp. 197 – 222.

_ **and D. L. Downs**, "Interpersonal functions of the self-esteem motive: The self-esteem system as a sociameter," in M. H. Kernis, ed., *Efficacy, agency, and self-esteem*, Plenum Press, 1995.

_ **, E. S. Tambor, S. J. Terdal, and D. L. Downs**, "Self-esteem as an interpersonal monitor," *Journal of Personality and Social Psychology*, 1995, *68*, 518 – 530.

Malmendier, U. and G. Tate, "CEO overconfidence and corporate investment," *Journal of Finance*, 2005, pp. 2661–2700.

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

Mecca, A. M., N. J. Smelser, and Vasoncellos J., eds, *The social importance of self-esteem*, Berkeley, CA: University of California Press, 1989.

Merkle, C. and M. Weber, "True Overconfidence," Working Paper, University of Mannheim 2007.

Patrick, C., J. Curtin, and A. Tellegen, "development and validation of a brief form of the multidimensional personality questionnaire," *Psychological Assessment*, 2002, *14* (2), 150 – 163.

Raven, J. C., *Raven's Standard Progressive Matrices (SPM) and Raven's Standard Progressive Matrices Plus (SPM Plus)*, Pearson, 2000.

Rustichini, A., "Neuroeconomics: what have we found, and what should we search for," *Current Opinion in Neurobiology*, December 2009, *19* (6), 672–677.

Santos-Pinto, L. and J. Sobel, "A model of positive self-image in subjective assessments," *American Economic Review*, 2005, *95* (5), 1386–1402.

Svenson, O., "Are we all less risky and more skillful than our fellow drivers?," *Acta Psychologica*, 1981, *94*, 143 – 148.

Tellegen, A., "The analysis of consistency in personality assessment," *Journal of Personality*, 1988, *56*, 621 – 663.

_ **and N.G. Waller**, "Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire," *Personality measures: Development and evaluation*, 1994, *1*, 133–161.

Weinberg, B., "A model of overconfidence," Working Paper, Ohio State University 2006.

A Figures and Tables

Figure 1: Distribution of beliefs about ability in IQ test and Numeracy test.
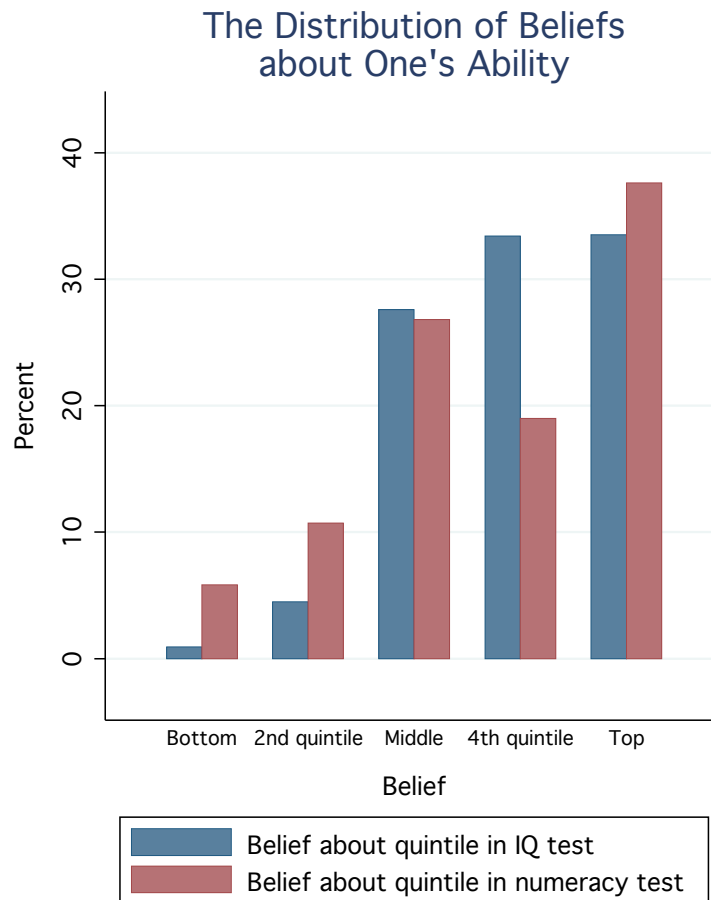
Figure 2: Relative performance judgments as a function of actual ability: IQ.
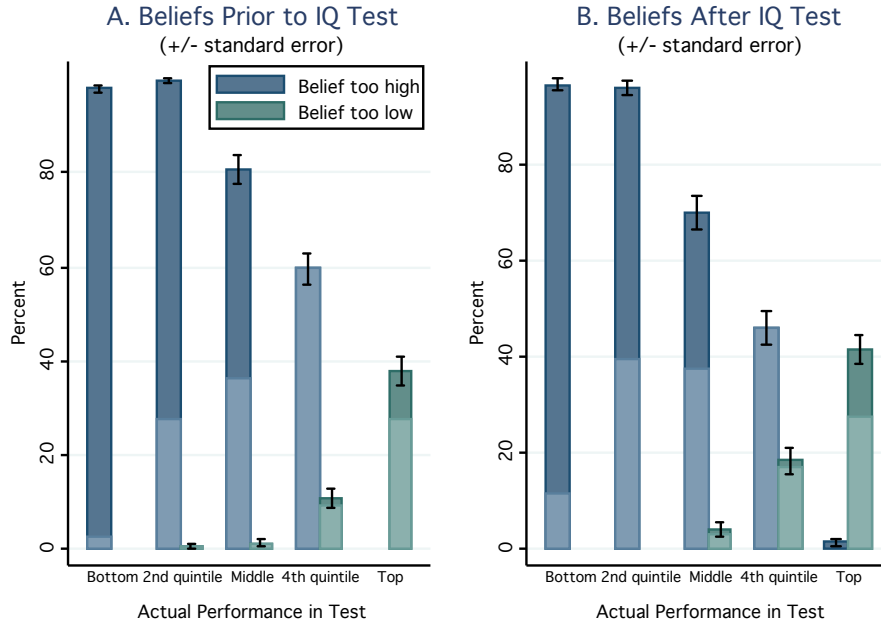
Figure 3: Relative performance judgments as a function of actual ability: Numeracy.
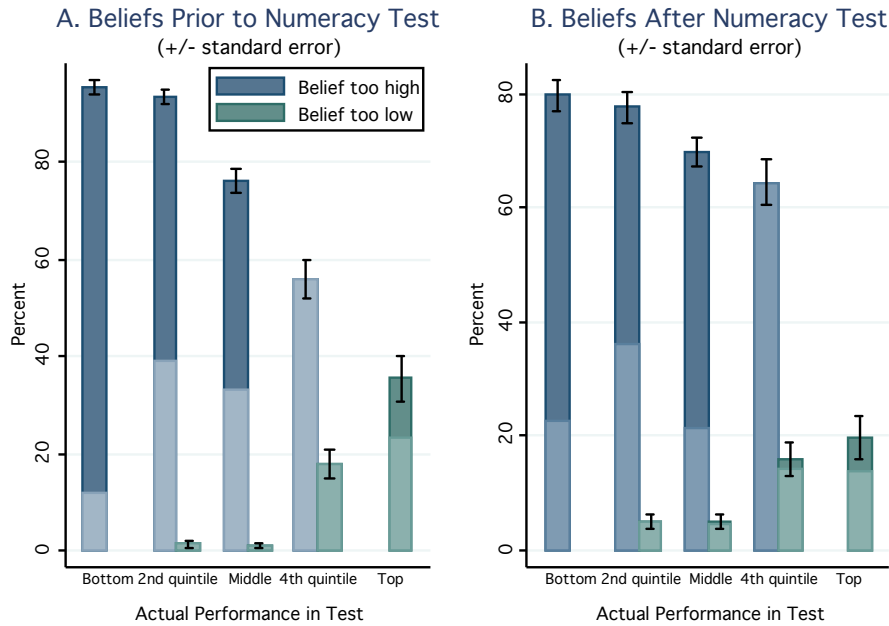
Figure 4: The demand for information
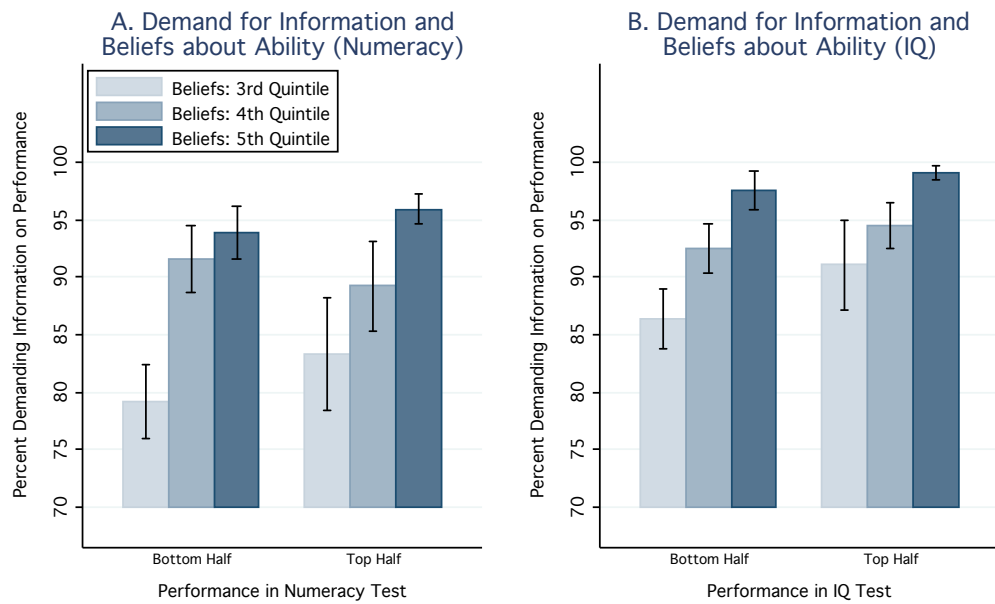
Notes: Caps indicate standard error of the mean.

Figure 5: Personality characteristics and relative performance judgments

Panel A: IQ

Confidence judgments in IQ test as a function of
the strength of trait relative to the median

Below · Above
Social Potency

Below · Above
Social Closeness

Below · Above
Stress Reaction

% believing to
be in top 20%

actual %
in top 20%

Below · Above
Harm Avoidance

Panel B: Numeracy

Confidence judgments in numeracy test as a function of
the strength of trait relative to the median

Below · Above
Social Potency

Below · Above
Social Closeness

Below · Above
Stress Reaction

Below · Above
Harm Avoidance

Notes: Caps indicate standard error of the respective mean.

Table 1: Descriptive Statistics

Test Scores: Number of correct answers.

	Mean	Standard Deviation	Min	Max
Numeracy Test	8.54	2.55	1	12
IQ Test	45.33	8.15	1	60

Education: Highest level attained

Middle School	3.9%
High School	39.3%
Technical School	14.9%
Some College	33.2%
College	6.5%
Graduate School	2.3%

Ethnic Categories:

Caucasian	82.7%
African-American	14.1%
Indian	2.8%
Asian	0.7%
Latino	1.8%
Other	1.0%

Other Demographics:

	Mean	Standard Deviation	Min	Max
Age	37.43	10.90	21	69
Male	88.7%			
Household income (in thousands of US Dollars)	52.66	27.07	10	150

Notes: $N = 888$ individuals.

Table 2: The Empirical Allocation functions $\hat{q}_k(s_j)$

	Numeracy Test				
	s_1	s_2	s_3	s_4	s_5
t_5	0.0	0.0	0.1	0.27	**0.62**
t_4	0.004	0.009	0.091	**0.298**	0.59
t_3	0.0	0.0125	**0.181**	0.362	0.443
t_2	0.004	**0.0**	0.272	0.377	0.345
t_1	**0.02**	0.02	0.401	0.376	0.175

	IQ Test				
	s_1	s_2	s_3	s_4	s_5
t_5	0.004	0.016	0.121	0.271	**0.579**
t_4	0.0	0.014	0.168	**0.355**	0.461
t_3	0.006	0.031	**0.262**	0.375	0.325
t_2	0.0	**0.04**	0.39	0.363	0.204
t_1	**0.033**	0.11	0.42	0.322	0.104

Notes: The empirical allocation function indicates for each ability quintile k, what fraction of individual put themselves in ability quintile j.

Table 3: Constrained Maximum Likelihood estimators of the allocation function $q_k^{ML}(s_j)$.

Numeracy Test

	s_1	s_2	s_3	s_4	s_5
t_5	0	0	0.121	0.232	**0.646**
t_4	0	0	0.159	**0.335**	0.504
t_3	0	0.007	**0.364**	0.275	0.352
t_2	0.012	**0.106**	0.364	0.335	0.180
t_1	**0.071**	0.106	0.364	0.335	0.122

IQ Test

	s_1	s_2	s_3	s_4	s_5
t_5	0	0	0.101	0.277	**0.621**
t_4	0.004	0.008	0.080	**0.378**	0.528
t_3	0	0.01	**0.343**	0.290	0.355
t_2	0.004	**0.015**	0.269	0.373	0.337
t_1	**0.012**	0.015	0.343	0.378	0.251

Notes: The Maximum likelihood estimator is the solution of the problem described by equation (5). It indicates for each ability quintile k, what fraction of individual receives a signal that would induce him to choose the quintile j as most likely.

Table 4: The Demand for Information: IQ Test.

Dependent Variable: Demand Information (=1)
Marginal Effects from Probit Estimates

	(1)	(2)	(3)	(4)	(5)
q_i^{IQ} after test	0.031***	0.029***	0.029***		0.029***
	(0.009)	(0.009)	(0.008)		(0.010)
q_i^{IQ} before test				0.018**	$-$ 0.004
				(0.009)	(0.011)
q_i^{NT} after test					0.005
					(0.007)
Piece-wise linear profile in test score					
first quintile	0.000	0.000	0.000	0.001	0.000
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
second quintile	0.001	0.000	0.000	0.000	0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
third quintile	0.017	0.016	0.015	0.018	0.014
	(0.012)	(0.012)	(0.011)	(0.012)	(0.011)
fourth quintile	$-$ 0.008	$-$ 0.006	$-$ 0.007	$-$ 0.007	$-$ 0.007
	(0.011)	(0.010)	(0.009)	(0.010)	(0.009)
fifth quintile	0.006	0.006	0.006	0.010	0.006
	(0.011)	(0.010)	(0.009)	(0.010)	(0.009)
Harm Avoidance		$-$ 0.003**	$-$ 0.003**	$-$ 0.004**	$-$ 0.003**
		(0.002)	(0.001)	(0.002)	(0.001)
Social Closeness		0.002*	0.002	0.002	0.002
		(0.001)	(0.001)	(0.001)	(0.001)
Social Potency		$-$ 0.001	$-$ 0.001	$-$ 0.000	$-$ 0.001
		(0.002)	(0.002)	(0.002)	(0.002)
Stress Reaction		0.000	0.000	$-$ 0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)
Demographic controls?	No	No	Yes	Yes	Yes
p	0.000	0.001	0.003	0.003	0.005
N	838	838	826	825	825

Notes: ***, **, * indicate significance at the 1, 5, 10 percent level, respectively. The model estimated here is described in section 5, see in particular equation 8.

Table 5: The Demand for Information: Numeracy Test.

Dependent Variable: Demand Information (=1)
Marginal Effects from Probit Estimates

	(1)	(2)	(3)	(4)	(5)
q_i^{NT} after test	0.060***	0.057***	0.059***		0.040***
	(0.010)	(0.011)	(0.011)		(0.013)
q_i^{NT} before test				0.063***	0.018
				(0.013)	(0.017)
q_i^{IQ} after test					0.028**
					(0.014)
Piece-wise linear profile in test score					
first quintile	0.022	0.022*	0.022	0.029**	0.022*
	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)
second quintile	$-$ 0.006	0.001	0.002	0.011	0.003
	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)
third quintile	0.011	0.009	0.010	0.017	0.008
	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)
fourth quintile	0.003	0.008	0.016	0.021	0.017
	(0.041)	(0.040)	(0.039)	(0.039)	(0.039)
fifth quintile	$-$ 0.001	0.009	0.013	0.019	0.008
	(0.048)	(0.045)	(0.044)	(0.043)	(0.045)
Harm Avoidance		$-$ 0.005**	$-$ 0.005*	$-$ 0.005**	$-$ 0.004*
		(0.002)	(0.002)	(0.002)	(0.002)
Social Closeness		0.003	0.002	0.002	0.002
		(0.002)	(0.002)	(0.002)	(0.002)
Social Potency		$-$ 0.000	0.001	0.001	0.001
		(0.002)	(0.002)	(0.002)	(0.002)
Stress Reaction		0.002	0.002	0.001	0.002
		(0.002)	(0.002)	(0.002)	(0.002)
Demographic controls?	No	No	Yes	Yes	Yes
p	0.000	0.001	0.003	0.005	0.005
N	888	886	873	873	873

Notes: The model estimated here is described in section 5, see in particular equation 8. ***, **, * indicate significance at the 1, 5, 10 percent level, respectively.

Table 6: Personality characteristics and confidence judgments.

Marginal effects on the probability of in the top 20 percenet from ordered probit model

| | Confidence in IQ test | | | Confidence in numeracy test | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Absorption	0.007**	0.005	0.005*	0.004*	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Achievement	0.004	-0.001	-0.001	0.004	-0.003	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Aggression	0.000	-0.001	0.000	-0.002	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Alienation	-0.004	0.000	0.001	-0.007**	0.004	0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Control	-0.001	0.001	0.001	-0.001	0.003	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Harm Avoidance	-0.009***	-0.008***	-0.007**	-0.007***	-0.006**	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Social Closeness	-0.007***	-0.004	-0.003	-0.007***	-0.004	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Social Potency	0.014***	0.013***	0.011***	0.013***	0.013***	0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Stress Reaction	-0.007***	-0.009***	-0.009***	-0.006***	-0.007***	-0.007***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Traditionalism	-0.007**	0.000	-0.001	-0.004	-0.001	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Wellbeing	-0.003	-0.002	-0.001	-0.009	-0.001	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Skill controls?	No	Yes	Yes	No	Yes	Yes
Demographics?	No	No	Yes	No	No	Yes
F-test for joint significance of personality traits	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
pseudo-R^2	0.01	0.27	0.28	0.01	0.12	0.13
N	1060	1060	1060	1060	1060	1060

Notes: ***, **, * indicate significance at the 1, 5, 10 percent level, respectively. The model estimated here is described in section 5, see in particular equation 8.

Table 7: Personality characteristics and overconfident judgments

Marginal effects on the probability of being overconfident from modified ordered probit model

| | Overconfidence in IQ test | | | Overconfidence in numeracy test | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Absorption	0.000	0.003	0.004	0.001	0.002	0.002
	(0.003)	(0.012)	(0.002)	(0.003)	(0.003)	(0.003)
Achievement	-0.004	0.000	0.000	-0.005	-0.004	-0.003
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Aggression	-0.003	-0.002	-0.001	0.000	-0.001	-0.001
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Alienation	0.005*	0.000	0.002	0.005*	0.003	0.004
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Control	0.000	0.001	0.000	0.002	0.002	0.000
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Harm Avoidance	-0.004	-0.005***	-0.005**	-0.004	-0.005*	-0.003
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Social Closeness	-0.002	-0.004*	-0.003	-0.003	-0.004	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Social Potency	0.009***	0.010***	0.008***	0.012***	0.013***	0.010***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Stress Reaction	-0.005**	-0.007***	-0.007***	-0.006***	-0.006***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Traditionalism	0.003	0.000	0.000	-0.001	-0.002	-0.003
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Wellbeing	0.003	-0.002	-0.001	0.002	0.000	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Skill controls?	No	Yes	Yes	No	Yes	Yes
Demographics?	No	No	Yes	No	No	Yes
F-test for joint significance of personality traits	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
pseudo-R^2	0.01	0.24	0.25	0.02	0.05	0.06
Log likelihood	-1463.912	-1129.03	-1117.221	-1317.37	-1282.838	-1265.65
N	1060	1060	1060	1060	1060	1060

Notes: ***, **, * indicate significance at the 1, 5, 10 percent level, respectively. The model estimated here is described in section 5, see in particular equation 8.

B   Restrictions imposed by the Bayesian Model

We provide here the conceptual structure to set up the empirical test of the Bayesian hypothesis, that that statements of individuals about their most likely percentile are produced by truthful reporting of Bayesian updating on the basis of private information.

B.1   Private Information

Prior to the experimental session, each individual has observed in his lifetime possibly complex signal on his intellectual abilities. These signals may include all sorts of different personal experiences: their success in school, on the job, in day to day comparison with others, including their speed in solving Sudoku games. All these signals are summarized in our model by a single observation. This signal is his private information, and is produced by an experiment (in the sense of statistical theory), which is a function from the set of types to distribution on signals. We take as set of signals the real line, X, endowed with the Borel σ-algebra $\mathcal{B}(X)$.

So the private experiment is:

$$(X, \mathcal{B}(X), (P_\theta)_{\theta \in \Theta}) \tag{13}$$

where for every θ, $P_\theta \in \Delta(X, \mathcal{B}(X))$, the set of probability measures on X.

We do not know or observe the experiment P, so we are trying to estimate the most likely experiment given our data; and to test whether the overall hypothesis that the data are produced by Bayesian updating is supported or rejected by the data.

In the Bayesian model, a subject with a type θ observes a signal x with probability induced by P_θ, and then computes the posterior given the signal, which we denote

$$m(\cdot|x) \in \Delta(\Theta, \mathcal{B}(\Theta)) \tag{14}$$

Over and under confidence

Let $S \equiv \{s^i : i = 1, \ldots, 5\}$ be the set of statements that the subject can make, where s^i is interpreted as *"I am in the ith quintile"*. Given the signal x he has observed, the subject determines which of the 5 quintiles has the largest probability according to his posterior, that is, he solves:

$$\max_{i=1,\ldots,5} m(R^i|x) \tag{15}$$

and then states s^k if k is the solution of the problem (15).

Definition 1. *A subject in the quintile R^i stating s^j is overconfident if $j > i$ and underconfident if $i > j$.*

The model implicitly describes a function giving for every θ a probability over the set of quintiles. Note that only we, the experimenters, observe θ, although with some noise due to the imprecision of the task.

Allocation functions

An allocation function is a function $q : \Theta \to \Delta(S)$. An allocation function is induced by an experiment P with the distribution m over the type space Θ if it can be obtained from Bayesian updating according to P. Formally:

Definition 2. *An allocation function q is induced by an experiment P with m the prior distribution over the type space Θ if there exists a choice function $C : X \to \Delta(S)$ such that*

$$\text{if } C(x, s^j) > 0 \text{ then } m(R^j|x) = \max_k m(R^k|x) \tag{16}$$

and such that for every θ and s^j,

$$q_\theta(s^j) = \int_X P_\theta\,(dx)\,C(x, s^j) \tag{17}$$

We denote by $A(P)$ the set of allocation functions induced by an experiment P.

The allocation function of an experiment is not unique because the choice function C is not unique. Note that $(S, \mathcal{P}(S), (q_\theta)_{\theta \in \Theta})$, where $\mathcal{P}(S)$ is the set of all the subsets of S, is an experiment on Θ, dominated by P in the Blackwell order, since it is obtained from P though the Markov kernel C. The function q depends on the experiment P (and is a coarsening of P): we may use the notation q^P when we want to emphasize this dependence.

We denote $X^i \equiv \{x : \ \text{argmax}_j m(R^j|x) = i\}$. We can also define the average theoretical allocation function

$$A_q(R^i, s^j) = \int_{R^i} q_\theta(s^j)dm(\theta). \tag{18}$$

An allocation function displays overconfidence (respectively underconfidence) at $\theta \in R^i$ if $q_\theta(s^j) > 0$ for $j > i$ (respectively $j < i$).

Finite types

For our intended application, providing a test of the Bayesian model in our experimental data, a finite type space is enough. We consider a type space where a quintile

coincides with a type. An individual has type θ^i if his IQ score in the Raven's matrices task is in the ith quintile. So formally we have:

$$\Theta \equiv \{\theta^i : i = 1, \dots, 5\} \tag{19}$$

From the point of view of our more general model with a continuum of types, this simplification ignores the problem of aggregation of the different types within a quintile and simply assumes that all the individuals in a quintile are identical. We lose some information (for example, it seem natural that people with higher IQ score have more optimistic beliefs that those with lower score in the same quintile), but we gain in simplicity in the analysis of the data.

Experiments and allocation functions

To make the search for the experiment P more systematic we may proceed as follows. First we pose the problem: in our simple environment (with finite types, signals and states), when can an observed empirical allocation function possibly be produced as the allocation function of some experiment, when the prior is uniform over the types? The answer turns out to be simple: if and only if each quintile considers itself more likely than any other quintile does. Formally:

Theorem 3. *Let q be an allocation function. The following conditions are equivalent:*

1. *There exists an experiment $(X, \mathcal{X}, (P_\theta)_{\theta \in \Theta})$ over some signal space X such that q is one of its allocation functions;*

2. *For every i*
$$q_{\theta^i}(s^i) = \max_k q_{\theta^k}(s^i) \tag{20}$$

Proof

Let $(X, \mathcal{X}, (P_\theta)_{(\theta \in \Theta})$ be the experiment and C the choice function inducing q. Then for every i,

$$q_{\theta^i}(s^i) = \int_X P_{\theta^i}(dx)C(x, s^i)$$

By the definition of choice function, if $C(x, s^i) > 0$ then

$$m(R^i|x) = \max_k m(R^k|x). \tag{21}$$

But in the present case $R^k = \{\theta^k\}$, and the m is uniform, so 21 is equivalent to

$$P_{\theta^i}(x) = \max_k P_{\theta^k}(x) \tag{22}$$

and therefore for every k:

$$
\begin{aligned}
q_{\theta^i}(s^i) &= \int_{\{x:C(x,s^i)>0\}} P_{\theta^i}(dx)C(x,s^i) \\
&\geq \int_{\{x:C(x,s^i)>0\}} P_{\theta^k}(dx)C(x,s^i) \\
&\equiv q_{\theta^k}(s^i)
\end{aligned}
$$

Conversely, let q be an allocation function that satisfies (20). We construct an experiment inducing q as its allocations function. Let $X = S$, and for every i and j let $P_{\theta^i}(s^j) = q_{\theta^i}(s^j)$. This is an experiment: we only need to construct a choice function for this experiment that induces q. Let $C(s, s^j) = \delta_s(s^j)$ (that is, $= 1$ if and only if $s = s^j$ and $=0$ otherwise). The condition (16) on the choice function follows from the assumption (20), and the induced allocation is

$$
\sum_s q_{\theta^i}(s)\delta_s(s^j) = q_{\theta^i}(s^j).
$$

QED

C   The modified ordered probit model

The dependent variable in our model is the difference between the individual's belief b_i about his most likely quintile and his ability quintile q_i. That is, $b_i - q_i = 0$ implies that the individual is correct in her assessment, while $b_i - q_i > 0$ implies various degrees of overconfidence, the greater $b_i - q_i$. We model this as an ordered variable

$$
b_i - q_i = k \Rightarrow \alpha_k < x_i'\beta + \epsilon_i \leq \alpha_{k+1} \tag{23}
$$

and $b_i - q_i = -4 \Rightarrow x_i'\beta + \epsilon_i \leq \alpha_{-3}$ and $b_i - q_i = 4 \Rightarrow x_i'\beta + \epsilon_i > \alpha_4$.

C.1   An ordered model of over- and underconfidence

However, (23) the standard ordered probit model needs to be amended because of the following problem: The dependent variable we examine is the difference between the belief over the performance quintile b_i and the actual performance q_i. Suppose an individual's ability quintile is $q = 1$, i.e., it's impossible that this individual can underestimate his quintile, because, well, he's not the brightest crayon in the box, and there is no way to underestimate his quintile. Thus, suppose that this individual answers $b - q = 0$. The only thing we can learn for this individual (with characteristics x) is that $x\beta + e < \alpha_1$. Now, suppose that an individual has ability level $q = 2$ and answers $b = 1$. What do we know about this individual? Again, we are not sure about

the degree of underconfidence, because in $q = 2$, you can underestimate your ability by at most one quartile. Thus, we can only conclude that $x\beta + \epsilon < \alpha_0$ for this individual. At the same time, this individual can only overestimate his probability by 3 quartiles. Thus, if he overestimate his ability by 3 quartiles, all we know is that $x\beta + \epsilon > \alpha_3$.

Thus, in general, for an individual of performance quintile $q = 1$, the model is

$$
\begin{aligned}
b_i - q_i = 0 & \Rightarrow & x_i'\beta + \epsilon_i \le \alpha_1 \\
b_i - q_i = 1 & \Rightarrow & \alpha_1 < x_i'\beta + \epsilon_i \le \alpha_2 \\
b_i - q_i = 2 & \Rightarrow & \alpha_2 < x_i'\beta + \epsilon_i \le \alpha_3 \\
b_i - q_i = 3 & \Rightarrow & \alpha_3 < x_i'\beta + \epsilon_i \le \alpha_4 \\
b_i - q_i = 4 & \Rightarrow & \alpha_4 < x_i'\beta + \epsilon_i
\end{aligned}
\tag{24}
$$

Assuming that e_i has a standard normal distribution, the likelihood of an observation having $b_i - q_i = 1$ is given by

$$
\Pr(\alpha_1 < x_i'\beta + \epsilon_i \le \alpha_2) = \Phi(\alpha_2 - x\beta) - \Phi(\alpha_1 - x\beta)
\tag{25}
$$

The probability of $b_i - q_i = 4$ is given by

$$
\Pr(\alpha_4 < x_i'\beta + \epsilon_i) = 1 - \Phi(\alpha_4 - x\beta)
\tag{26}
$$

and it is obvious how to derive the other probabilities.

Continuing this logic, for an individual of performance quintile $q = 2$, the model is

$$
\begin{aligned}
b_i - q_i = -1 & \Rightarrow & x_i'\beta + \epsilon_i \le \alpha_0 \\
b_i - q_i = 0 & \Rightarrow & \alpha_0 < x_i'\beta + \epsilon_i \le \alpha_1 \\
b_i - q_i = 1 & \Rightarrow & \alpha_1 < x_i'\beta + \epsilon_i \le \alpha_2 \\
b_i - q_i = 2 & \Rightarrow & \alpha_2 < x_i'\beta + \epsilon_i \le \alpha_3 \\
b_i - q_i = 3 & \Rightarrow & \alpha_3 < x_i'\beta + \epsilon_i
\end{aligned}
\tag{27}
$$

For an individual of performance quintile $q = 3$, the model is

$$
\begin{aligned}
b_i - q_i = -2 & \Rightarrow & x_i'\beta + \epsilon_i \le \alpha_{-1} \\
b_i - q_i = -1 & \Rightarrow & \alpha_{-1} < x_i'\beta + \epsilon_i \le \alpha_0 \\
b_i - q_i = 0 & \Rightarrow & \alpha_0 < x_i'\beta + \epsilon_i \le \alpha_1 \\
b_i - q_i = 1 & \Rightarrow & \alpha_1 < x_i'\beta + \epsilon_i \le \alpha_2 \\
b_i - q_i = 2 & \Rightarrow & \alpha_2 < x_i'\beta + \epsilon_i
\end{aligned}
\tag{28}
$$

For an individual of performance quintile $q = 4$, the model is

$$
\begin{aligned}
b_i - q_i = -3 & \Rightarrow & x_i'\beta + \epsilon_i \le \alpha_{-2} \\
b_i - q_i = -2 & \Rightarrow & \alpha_{-2} < x_i'\beta + \epsilon_i \le \alpha_{-1} \\
b_i - q_i = -1 & \Rightarrow & \alpha_{-1} < x_i'\beta + \epsilon_i \le \alpha_0 \\
b_i - q_i = 0 & \Rightarrow & \alpha_0 < x_i'\beta + \epsilon_i \le \alpha_1 \\
b_i - q_i = 1 & \Rightarrow & \alpha_1 < x_i'\beta + \epsilon_i
\end{aligned}
\tag{29}
$$

Finally, for a smartass with $q = 5$, the model is

$$
\begin{aligned}
b_i - q_i = -4 &\Rightarrow x\beta + \epsilon_i \leq \alpha_{-3} \\
b_i - q_i = -3 &\Rightarrow \alpha_{-3} < x_i'\beta + \epsilon_i \leq \alpha_{-2} \\
b_i - q_i = -2 &\Rightarrow \alpha_{-2} < x_i'\beta + \epsilon_i \leq \alpha_{-1} \\
b_i - q_i = -1 &\Rightarrow \alpha_{-1} < x_i'\beta + \epsilon_i \leq \alpha_0 \\
b_i - q_i = 0 &\Rightarrow \alpha_0 < x_i'\beta + \epsilon_i
\end{aligned}
\tag{30}
$$

Thus, for each performance, we can specify an ordered probit model that we can then piece together to one overall model. This model fixes the problem that there is a mechanical relationship between performance quintiles and overconfidence and takes this explicitly into account. To see this, consider how a "normal" ordered probit model would consider an individual with characteristics x who is the the top quintile, and thinks he is in the fifth quintile. The model would assume that $\alpha_0 < x\beta + e \leq \alpha_1$. One of the x's is going to be the actual performance. Thus, the model would force a coefficient on the performance to ensure that the above inequality holds. By contrast, the modified ordered probit model recognizes that all we can know is $\alpha_0 < x\beta + e$, relaxing the weight it will put on actual performance depending on how much probability mass remains to to the right of $\alpha_0 - x\beta$. The more probability mass remains to the right of that cutoff, the less weight it will put on the variables to estimate their magnitude.

C.2   Marginal Effects

We calculate the marginal effects and its standard errors for an average individual. Since the marginal effects are non-linear in the data x, we need to choose how to calculate them: We calculate the marginal effects at sample means $\bar{x}$.[6] The marginal effects of interests are the the effect of a one-unit increase in x_{ik} on the probability of be'ing overconfident. The statistical model implies

$$
\begin{aligned}
\Pr(i \text{ is overconfident}) &= \Pr(\bar{x}'\beta + \epsilon_i > \alpha_1) = \Pr(\epsilon_i > \alpha_1 - \bar{x}'\beta) \\
&= 1 - \Phi(\alpha_1 - \bar{x}'\beta) = \Phi(\bar{x}'\beta - \alpha_1)
\end{aligned}
$$

where the last equality follows from the symmetry of the normal distribution.[7] Define the marginal effect as

$$
p_k(\theta) \equiv \frac{\partial \Pr(i \text{ is overconfident})}{\partial x_k} = \phi(\bar{x}'\beta - \alpha_1)\beta_k
$$

---

[6]The marginal effects are a function $f()$ of the data x. Hence, if $x_i \xrightarrow{p} \mu_x$, then $f(x_i) \xrightarrow{p} f(\mu_x)$ by Slutzky's theorem. Also by Slutzky's theorem, $f(\bar{x}) \xrightarrow{p} f(\mu_x)$. Thus, the marginal effects can be consistently estimated if evaluated at sample means or if averaged across individuals.

[7]Remember that $1 - \Phi(z) = \Phi(-z)$ because of the symmetry.

where $\theta = (\alpha_1, \beta')$. The standard errors of the marginal effect can be calculated using the Delta method, which is really highly painful in this particular application. Here it goes: the variance of $p_k(\theta)$ is given by

$$V(p_k(\theta)) = \frac{\partial p_k(\theta)}{\partial \theta}' \Sigma_\theta \frac{\partial p_k(\theta)}{\partial \theta}$$

where Σ_θ is the covariance matrix of θ. Tedious calculations show that

$$\frac{\partial p_k(\theta)}{\partial \alpha_1} = -\phi'(\bar{x}'\beta - \alpha_1)\beta_k = (\bar{x}'\beta - \alpha_1)\phi(\bar{x}'\beta - \alpha_1)\beta_k \qquad (31)$$

where the last equality follows from the fact that $\phi'(z) = -z\phi(z)$. It follows for the elements in β that

$$\begin{aligned}
\frac{\partial p_k(\theta)}{\partial \beta_k} &= \phi(\bar{x}'\beta - \alpha_1) + \phi'(\bar{x}'\beta - \alpha_1)\beta_k \bar{x}_k \\
&= \phi(\bar{x}'\beta - \alpha_1) - (\bar{x}'\beta - \alpha_1)\phi(\bar{x}'\beta - \alpha_1)\beta_k \bar{x}_k \\
&= \phi(\bar{x}'\beta - \alpha_1)(1 - (\bar{x}'\beta - \alpha_1)\beta_k \bar{x}_k) \qquad (32)
\end{aligned}$$

and for β_j with $j \neq k$:

$$\begin{aligned}
\frac{\partial p_k(\theta)}{\partial \beta_j} &= \phi'(\bar{x}'\beta - \alpha_1)\beta_k \bar{x}_j \\
&= -(\bar{x}'\beta - \alpha_1)\phi(\bar{x}'\beta - \alpha_1)\beta_k \bar{x}_j \qquad (33)
\end{aligned}$$

In the program calculating the standard errors, we make use of the following three recursions:

$$\frac{\partial p_k(\theta)}{\partial \alpha_1} = p_k(\theta)(\bar{x}'\beta - \alpha_1)\beta_k \qquad (34)$$

$$\frac{\partial p_k(\theta)}{\partial \beta_k} = \phi(\bar{x}'\beta - \alpha_1) - \frac{\partial p_k(\theta)}{\partial \alpha_1}\bar{x}_k \qquad (35)$$

$$\frac{\partial p_k(\theta)}{\partial \beta_j} = -\frac{\partial p_k(\theta)}{\partial \alpha_1}\bar{x}_j \qquad (36)$$