Does the Better-Than-Average Effect Show That People Are Overconfident?: An Experiment.

Jean-Pierre Benoît and Juan Dubra and Don Moore

LBS, Universidad de Montevideo, Tepper School of Business, Carnegie Mellon University

2. February 2009

Does the Better-Than-Average Effect Show That People Are Overconfident?: An Experiment.*

Jean-Pierre Benoît
London Business School

Juan Dubra†
Universidad de Montevideo

Don Moore
Tepper School of Business, Carnegie Mellon University.

Abstract

We conduct a proper test of the claim that people are overconfident, in the sense that they believe that they are better than others. The results of the experiment we present do not allow us to reject the hypotheses that the data has been generated by perfectly rational, unbiased, and appropriately confident agents.

Keywords: Overconfidence; Better than Average; Experimental Economics; Irrationality; Signalling Models.

Journal of Economic Literature Classification Numbers: D11, D12, D82, D83

1 Introduction

A large body of literature across several disciplines, including psychology, finance, and economics, purports to find that people are generally overconfident.[1] For economists, the issue

*This paper was previously circulated as "A Proper Test of Overconfidence". We thank Uriel Haran for help with data collection as well as the staff and facilities of the Center for Behavioral Decision Research at Carnegie Mellon University.

†email: dubraj@um.edu.uy

[1] Papers on overconfidence in economics include Camerer and Lovallo (1999) analyzing entry in an industry, Fang and Moscarini (2005) analyzing the effect of overconfidence on optimal wage setting, Garcia, Sangiorgi and Urosevic (2007) analyzing the efficiency consequences of overconfidence in information acquisition in financial markets, Kőszegi (2006) who studies how overconfidence affects how people choose tasks or careers, and Menkhoff et al. (2006) who analyze the effect of overconfidence on herding by fund managers. In finance, papers include Barber and Odean (2001), Bernardo and Welch (2001), Chuang and Lee (2006), Daniel, Hirshleifer and Subrahmanyam (2001), Kyle and Wang (1997), Malmendier and Tate (2005), Peng and Xiong (2006), and Wang (2001). See Benoît and Dubra (2008) for a discussion of some of the literature.

of overconfidence is of paramount importance as it affects the equilibrium outcomes in almost every market. Although the term "overconfidence" has been used rather broadly, Moore and Healy (2008) point out that, in fact, three distinct varieties of overconfidence have been examined in the literature: (1) a person *overestimating* his or her performance or abilities, (2) a person *overplacing* himself relative to others, and (3) a person having excessive confidence in the accuracy of his beliefs, or *overprecision*. In this paper, we focus on the second type of overconfidence, overplacement.

For the most part, researchers have not directly observed overplacement but, rather, inferred a tendency for individuals to rank themselves too highly from a tendency for a majority of people to claim to be superior to the median person – the so-called *better-than-average effect*. The better-than-average-effect has been noted for a wide range of simple skills, from driving, to spoken expression, to the ability to get along with others, to test taking on easy tests.[2] While this effect is well established, Benoît and Dubra (2008) (henceforth B&D) have recently questioned its significance. They show that better-than-average data in and of itself merely gives the *appearance* that (some) people must be overplacing themselves, but does not indicate *true* overplacement, which carries with it the implication that people have made some kind of error in their self-placements.[3] Because of this reason, almost none of the existing experimental literature on the better-than-average effect can actually claim to have found overplacement.[4] Moreover, most of the experiments by their very design do not even have the potential of showing overplacement. In this paper, we report on an experiment designed to provide a proper test of overplacement. The following example, taken directly from B&D, illustrates the basic flaw in previous tests.

Consider a large population with three types of drivers, low skilled, medium skilled, and high skilled, and suppose that the probabilities of any one of them causing an accident in any single period are $p_L = \frac{4}{5}, p_M = \frac{2}{5}$, and $p_H = 0$. In period 0, nature chooses a skill level for each person with equal probability. Initially, no driver knows his or her own skill level, and so each person (rationally) evaluates himself as no better or worse than average. In period 1, everyone drives and learns something about his skill, based upon whether or not he has caused an accident. Each person is then asked how his driving skill compares to the rest of the population. How does a driver who has not caused an accident reply?

[2] While early research pointed towards a universal better-than-average effect, more recent work indicates that the effect is primarily for easy tasks and may be reversed for difficult tasks.

[3] Other papers which also question this stance include Zábojník (2004) and Brocas and Carillo (2007).

[4] We note, however, that many papers on the better-than-average effect have a different goal than simply demonstrating overplacement. For instance, Kruger (1999) tests for the relationship between the better-than-average effect and egocentrism.

Using Bayes' rule, he evaluates his own skill level as follows:

$$\begin{aligned} p\left(\text{Low skill} \mid \text{No accident}\right) &= \frac{\frac{1}{3}\frac{1}{5}}{\frac{1}{3}+\frac{1}{3}\frac{3}{5}+\frac{1}{3}\frac{1}{5}} = \frac{1}{9} \\ p\left(\text{Medium skill} \mid \text{No accident}\right) &= \frac{\frac{1}{3}\frac{3}{5}}{\frac{1}{3}+\frac{1}{3}\frac{3}{5}+\frac{1}{3}\frac{1}{5}} = \frac{1}{3} \\ p\left(\text{High skill} \mid \text{No accident}\right) &= \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{3}\frac{3}{5}+\frac{1}{3}\frac{1}{5}} = \frac{5}{9} \end{aligned}$$

Such a driver thinks there is over a $\frac{1}{2}$ chance (in fact, $\frac{5}{9}$) that his skill level is in the top third of all drivers. His mean probability of an accident is $\frac{5}{9}0+\frac{1}{3}\frac{2}{5}+\frac{1}{9}\frac{4}{5}=\frac{2}{9}$, which is better than for $\frac{2}{3}$ of the drivers, and better than the population mean. Furthermore, his beliefs about himself strictly first order stochastically dominate the population distribution. Any way he looks at it, a driver who has not had an accident should evaluate himself as better than average. Since $\frac{3}{5}$ of drivers have not had an accident, $\frac{3}{5}$ rationally rank themselves as better than average.

As this example shows, the fact that 60% of drivers rank themselves above the median does not indicate erroneous self-evaluations. In fact, Theorem 1 below shows that any fraction of people could rank themselves as being in the top half of the population without any overplacement being implied. Therefore, any experiment designed just to show that more than half the population rank themselves as better than average cannot possibly show overplacement. Experiments with more detailed information on how subjects place themselves in percentiles have the potential to show overplacement, but even these must be carefully interpreted.

We conduct a test that has the potential to reveal that people are not making rational assessments of their abilities. The experiment is based upon the theory developed in B&D, which we briefly review in Section 2. Although the subjects in our experiment also give the superficial appearance of being overconfident by overplacing themselves, we do not find any evidence that they are in fact overconfident. While this finding by itself hardly proves that people do not overplace themselves, it does not stand alone. Two other experiments which conduct careful, proper tests of overplacement are Clark and Friesen (2008) and Moore and Healy (2008), and they also do not find such a bias. Furthermore, as is argued in B&D, the well-known experiment by Camerer and Lovallo (1999) which is usually interpreted as showing overplacement is better interpreted as showing no overplacement.[5]

[5]In the experiment, N subjects ("firms") must decide whether to play In or Out. After the entry decisions are made, the subjects who have played In are ranked. The payoff to playing In is greater than the payoff to playing Out if and only if an entrant is ranked in the top $k < N$. There are two treatments, one in which subjects are ranked randomly and one in which they are ranked according to their performance on a test. More subjects enter under the test treatment than the random treatment, and Camerer and Lovallo conclude

The most common type of experiment in this field asks subjects how they rank compared to others. For instance, Weinstein (1980) asks students to compare themselves to the average student on a variety of attributes, including their chances of getting a good job offer before graduation and their chances of developing a drinking problem. Similarly, Svenson (1981) asks subjects in a room to estimate how their driving compares to the other subjects, and to make estimates of the form "I drive better than x% of the people in this room".

There are at least four criticisms that can be made of this type of experiment, though not every criticism applies to every experiment:

1. Participants have no material incentive to answer the question accurately and internal motivations to answer accurately are likely to compete with other motivations, such as appearing competent, self-confident, or modest.

2. It may be unclear to the subjects what is meant by an "average" student. In particular, should the average be interpreted as the mean or median (or something else still)?

3. Subjects may be uncertain of their *own* skill levels, making the meaning of their answers unclear.

4. The research design does not allow subjects to demonstrate their degree of confidence in their self-placement.

The first two criticisms are quite familiar, so let us turn to the last two. Consider a subject who is asked to rank himself on IQ, given that the median IQ is 100. If he has not actually taken an IQ test then he must guess at his IQ. Suppose, for the sake of argument, that he believes that his IQ is 80 with probability 0.45, 110 with probability 0.45, and 115 with probability 0.1. How should he rank himself? He could reasonably respond that be believes himself to be of above average intelligence, given that there is over a 50% chance that his IQ is above average. On the other hand, he could just as reasonably respond that he is of below average intelligence, given that his mean IQ is only 97. Thus, the subject's answer to the question gives no clear indication of its meaning. By the same token, we have no way of knowing his degree of confidence when he utters a statement like "I believe I have a higher IQ than the average person". As we will discuss, both these ambiguities have important implications. Note, however, that if, as a matter of fact, subjects have very tight estimates of their types then both these issues become moot – the various meanings they could have for their answers converge and subjects will be almost 100% confident in their

that this indicates that the subjects are overconfident. However, as we show in B&D, increased entry does not show overconfidence. Overconfidence would be indicated if subjects earned negative expected profits or utility. However, these quantities are both positive (even for large degrees of risk aversion) so that, in our view, this experiment is better interpreted as also not finding overconfidence.

self-placements. Therefore, in addition to testing for overconfidence we test the hypothesis that subjects do not have very tight estimates of their types.[6] Note that in the previous driving example, $\frac{3}{5}$ of the drivers' believe that their mean abilities and median abilities are better than average, justifying their overconfident seeming answers. At the same time, each of these drivers thinks there is a $\frac{4}{9}$ chance that he is not above average, and even a $\frac{1}{9}$ chance that he is below average.

Even if we grant that subjects with no material incentive respond to questionnaires as accurately as possible, so that point 1) above is not an issue, an experiment that fails to pay attention to *any* one of the remaining points may fail as a test of overplacement, as we show in the following section.

2 Background

When should we say that a person is *overconfident*? An immediate proposal is that an overconfident person is not as "skillful" as she thinks she is. However, making such a determination may be problematic, as many skills are not easily measured. For instance, consider a person who asserts "I am a very good driver". Even supposing that we can make the notion of "very good" precise and that we can agree on what constitutes a very good driver, how are we to determine if the statement is true? Giving the person a driving test may not be practical. Moreover, the skills measured in such a test may not match up very well with the day-to-day skills reflected in the driver's self-assessment.

Researchers have circumvented these problems by considering entire populations at once and asking subjects how their skills compare to each other. Beyond circumvention, there are at least two reasons to be interested in this overplacement. Firstly, in many domains people may well have a better idea of their relative placements than their absolute placements. Thus, we might expect students to have a better idea of their math abilities relative to their classmates, than of their absolute abilities. Secondly, in many areas of interest, relative ability is of primary importance. For instance, in many jobs success depends primarily on a person's abilities relative to his or her peers.

The basic idea behind the relative population approach is that, since not more than 50% can be in the top 50% in skill level, if more than half the people in a population claim to be in the top half – or make choices which reveal such a belief – they "must" be making an error. However, as the example in the introduction shows, this idea is flawed. Obviously, it

[6] Within the behavioral economics literature, a number of papers, including Bénabou and Tirole (2002) and Kőszegi (2006), start from the premise that people are continually learning about their types. Several strands of the psychology literature also stress that people are uncertain of their types, including Festinger's (1954) influential social comparison theory, Bem's (1967) self-perception theory, and Amabile (1983).

is important to have a proper theoretical framework for discussing overconfidence.

Clearly, the implication in terming a population overconfident is that the members of the population have made some errors or have some inconsistencies in their self-evaluations.[7] Thus, B&D proposes that data be called *overconfident* only if it cannot be obtained from a population which derives its beliefs in a fully rational and consistent manner. A fairly standard model for a population deriving its beliefs in such a manner is as follows:

Definition 1 *A* ***signalling structure*** *is a triplet* $\sigma = (S, \Theta, f)$*, where* S *is a set of signals,* $\Theta \subset \mathbf{R}$ *is a type space, and* $f = \{f_\theta\}_{\theta \in \Theta}$ *is a collection of probability distributions over* S*.*

Definition 2 *A* ***signalling model*** *consists of a population of individuals and a signalling structure* $\sigma = (S, \Theta, f)$ *such that:*
i) In period 0, nature picks a type $\theta \in \Theta$ *for each individual, resulting in some distribution* p*; initially, each person's belief about her own type is given by this distribution.*
ii) In period 1, an individual of type θ *receives a signal* $s \in S$ *according to the probability distribution* f_θ*; each person updates her initial belief using Bayes' rule.*

- Throughout this paper we assume that higher types are more skillful.

We say that a person of type t is in the top x of a population if the fraction of people whose type is greater than or equal to t is at most x. Thus, in a population of 100 people *at most* 25 can be in the top $\frac{1}{4}$.

Definition 3 *Suppose that a fraction* y *of a population of* N *people believe that there is a probability* q *that their type is in the top* x *of the population. These beliefs can be* ***rationalized*** *if there is a signalling model with* N *individuals in which the expected fraction of people who will have these beliefs after updating is* y*.*

Notice that by asking that y be the *expected* fraction of people who will hold the particular beliefs, the definition is demanding: Data cannot be rationalized simply because it is possible that it could arise in a stochastic environment. If the data from an experiment can be rationalized, there is no prima facie reason to call it overconfident.

The following Theorem, taken from B&D, provides the basis for our tests of overconfidence.

Theorem 1 *Consider a population of* N *people and two integers* $0 \leq m \leq N$ *and* $1 \leq r \leq N$*. Suppose a fraction* $y = \frac{m}{N}$ *of the population believe that there is a probability at least* q *that their types are in the top* $x = \frac{r}{N}$ *of the population. These beliefs can be rationalized if and only if* $qy \leq x$*.*

[7] These errors can be expected to lead to further errors, such as too many people attempting to become professional athletes.

The following example illustrates the Theorem. Consider ten people who are to take a math test. First suppose that 7 of them believe that there is at least a $\frac{1}{2}$ probability that their type is in the top $\frac{3}{10}$ (so that $qy > x$). If this belief were rational, then on average at least $\frac{1}{2} \times \frac{7}{10} = \frac{7}{20}$ of the population would be in the top $\frac{3}{10}$, a clear absurdity. On the other hand, suppose instead that $\frac{3}{5}$ of the people believe that there is at least a $\frac{1}{2}$ probability that their type is in the top $\frac{3}{10}$. How could these beliefs rationally arise? One simple way is as follows. Before the test, a brief conversation reveals that six of them have an advanced degree in mathematics, whereas the remaining four have only high school mathematics. With no further information, the six can rationally believe they will place in the top six, with the precise order being uniformly random. Hence each of the six believes there is a $\frac{1}{2}$ chance he or she will place in the top 30%.

Armed with Theorem 1, we are in a position to better appreciate the four criticisms of prior experiments made in the introduction.

Consider a person who is given the choice between a 50% chance at a prize, and the prize if she places in the top half of a subject pool on a test. The person has been incentivized and the meaning of "average" is irrelevant, so Criticism 1 and 2 do not apply. Suppose the person strictly prefers the prize based on her test placement. The meaning of her preference is clear –she believes that there is more than a 50% chance that she places in the top half – so that Criticism 3 does not apply either. However, the strength of this belief – exactly how much more than 50% – is unclear, so that Criticism 4 does apply. Theorem 1 tells us that almost everybody could rationally prefer the placement alternative, rendering the experiment useless as a test for overconfidence.[8]

Svenson (1981) finds that 82.5% of (American) subjects in his experiment claim to be in the top 30% of subjects in their driving skill level. His subjects are not incentivized. More importantly, even granting the veracity of their answers, the meaning of these claims is unclear (Criticism 3). If the subjects, who presumably are uncertain of exactly how skillful they are as drivers, are answering based upon their self-beliefs about their median type, then Theorem 1 shows that the subjects are displaying overconfidence. However, if the subjects are answering based upon their self-beliefs about their mean type, then Theorem 4 in B&D shows that their answers are consistent with purely rational self-assessments.[9]

[8] Even everybody (as opposed to almost everybody) strictly prefering the placement bet is consistent with rationality, given an inevitable "sampling error" arising from the finite population.

[9] See B&D for a detailed discussion of what happens when people base their answers on their mean beliefs.

3 The experiment

On the positive side, Theorem 1 paves the way for experiments (in which people place themselves) that provide the potential of detecting overplacement. We can infer overconfidence if a sufficient fraction of people (variable y in the theorem) believe sufficiently strongly (variable q) that they rank sufficiently high (variable x). We conduct two tests of overplacement. We test if more than 60% of the subjects believe that there is at least a 50% chance that their type is in the top 30%. Recall that Svenson found that over 80% of his American subjects placed themselves in the top 30%, but it was unclear what they meant by this placement. We also test if more than 83.3 % of the subjects feel that there is more than a 60% chance that they are better than the median. We choose 60% because we are independently interested in whether a relatively small increase in the chance of receiving a prize randomly – from 50% in a benchmark test to 60% here – makes many people change their choice behavior. While these are the explicit tests we conduct, as discussed below implicitly there are more tests.

We were interested in the extent to which previous findings of apparent overplacement could be shown to be actual overplacement. Prior experimental work and the theory in B&D demonstrate that populations exhibit the better-than-average effect more markedly on easy tasks than difficult ones.[10] Accordingly, we gave our subjects an easy test.

Subjects were 134 individuals recruited through the web site of the Center for Behavioral Decision Research at Carnegie Mellon University <http://cbdr.cmu.edu/experiments/>. We report the data for the 129 subjects who gave complete responses to the three choices with which they were presented; the results are unchanged when we analyze, for each question, all the answers we have for that question.

The experiment was advertised under the name "Test yourself" along with the following description: "Participants in this study will take a test with logic and math puzzles. How much money people make depends on their performance and on how they choose to bet on that performance." This wording of the recruitment instructions was chosen to be conductive to more "overconfident looking data" (Camerer and Lovallo (1999) find that excess entry into their game (their measure of overconfidence) is much larger when subjects volunteer to participate in the experiment knowing that payoffs will depend on skill).

Subjects had a mean age of 25 years (SD = 6.4) and 42 percent of them were male. All subjects took a 20-item quiz of math and logic puzzles. They made a series of three choices between (1) bets on their test performance (skill) and (2) chance gambles of known probability. Subjects had to choose one of the two for each of the three pairs of bets. The three pairs of bets are listed below.

[10] The theory in Moore and Healy (2008) predicts that a test that is *easier than expected* should yield more overconfident looking data.

Skill Option	Chance Option
1. You will receive $10 if your test score puts you in the top half of previous test-takers. In other words, if your score is better than at least 50% of other test-takers, you will get $10. . .	1. There is a 50% chance you will receive $10. We have a bag with 5 blue poker chips and 5 red poker chips. You will reach in to the bag without looking and randomly select one of the poker chips. If the poker chip is blue, then you will get $10. If it is red, you will get nothing
2. You will receive $10 if your test score puts you in the top 30% of previous test-takers. In other words, if your score is better than at least 70% of other test takers, you will get $10. . .	2. There is a 50% chance you will receive $10. We have a bag with 5 blue poker chips and 5 red poker chips. You will reach in to the bag without looking and randomly select one of the poker chips. If the poker chip is blue, then you will get $10. If it is red, you will get nothing.
3. You will receive $10 if your test score puts you in the top half of previous test-takers. In other words, if your score is better than at least 50% of other test takers, you will get $10 . .	3. There is a 60% chance you will receive $10. We have a bag with 6 blue poker chips and 4 red poker chips. You will reach in to the bag without looking and randomly select one of the poker chips. If the poker chip is blue, then you will get $10. If it is red, you will get nothing.

Subjects were randomly assigned to experimental conditions that crossed two treatment variables: motivation and feedback.

The motivation manipulation varied what subjects were told about the test they were about to take. By introducing a manipulation of motivation we hoped to observe the effect of inducing a motive to be overconfident. Many theories of overconfidence assume that the belief that one is better than others is driven by the desire to actually be better than others (Benabou & Tirole, 2002; Köszegi, 2006; Kunda, 1990). Therefore, people's propensity to overplace their performances relative to those of others ought to be greatest under those circumstances when they are most motivated to achieve (see Krizan & Windschitl, 2007). Those in the high motivation condition read:

"In this experiment, you will be taking an intelligence test. Intelligence, as you know, is an important dimension on which people differ. There are many positive things associated with higher intelligence, including the fact that more intelligent people are more likely to get better grades and advance farther in their schooling. It may not be surprising to you that more intelligent people also tend to earn more money professionally. Indeed, according to research by Beaton (1975) ten IQ points are worth about four thousand dollars in annual salary. Children's intelligence is a good predictor of their future economic success according to Herrnstein and Murray (1994). Of course, this is partly because, as documented in research by Lord, DeVader, and Alliger (1986) intelligent people are perceived to have greater leadership potential and are given greater professional opportunities. But what may be surprising to you is that intelligent people also tend to have significantly better health and longer life expectancies (see research by Gottfredson & Deary, 2004)."

Those in the low motivation condition read: "In this experiment, you will be taking a test of math and logic puzzles."

Then subjects saw a set of sample test items. In order to constitute this set of sample items, we began with a larger set of 40 test items. One half of this set was randomly chosen for Test Set S. The other half belonged to Test Set M. Those participants who were to take Test S saw sample items from Set M, and vice versa.

Half of the subjects (those in the feedback condition) received a histogram showing how others had scored on the test they were about to take.

Next, subjects chose between skill and chance options for each of the three bets. The order in which the three bets appeared was varied randomly, as was whether the chance or the skill option appeared first for each bet. Participants were told that they would make the three choices again after taking the test, and that one of these six choices would be randomly selected at the end of the experiment to count for actual payoffs.

Then subjects took the twenty-item test under a ten-minute time limit. The two test sets appear in Appendix A. Subjects earned $.25 for each test question they answered correctly. Then subjects chose between the skill and chance options for each of the three bets again. Subjects then answered a series of questions regarding what they thought their score would be, how they felt during the experiment, etc.

Finally, if a subject chose to bet on chance (rather than their test performance) for the one bet that counted, an experimenter had the subject draw from the relevant bag of poker chips to determine whether he or she won the $10 prize.

4 The data

Before taking the test, each subject was presented with the three previously listed groups of choices. The order in which subjects were presented with these choices was randomized among subjects. The choices can be summarized as:

1. **Benchmark Choice**: A 50% chance of a prize (as determined by a random draw), or to be awarded the prize if your score on the test places you in the top 50% of previous test takers.

2. **High Placement Choice**: A 50% chance of a prize (as determined by a random draw), or to be awarded the prize if your score on the test places you in the top 30% of previous test takers.

3. **Strength Choice**: A 60% chance of a prize (as determined by a random draw), or to be awarded the prize if your score on the test places you in the top 50% of previous test takers.

There are 5 variables, none of which had any effect on the choice behavior of subjects (or their scores – except for the High Motivation treatment, which decreased scores, see below).

First, as expected, neither of the following three randomizations had any effect:

- The order of the presentation of the bets (123, 132, 213, etc).
- Whether the skill or random bet was presented first in each pair.
- Whether subjects saw sample M and took test S, or saw S and took M.

Second, we didn't have a prior belief of how the feedback manipulation would affect scores or choices between bets; it had no effect. Finally, and surprisingly to us, the Motivation manipulation had no effect either. Hence, we discuss only aggregate data, without discriminating by treatments.

Of paramount importance to a subject is her score on the test. Thus, it is most convenient to model a subject's "type" as just being this score.[11] This means that at the time she makes her decision, the subject does not yet have a type. Rather, her type is a random variable to be determined later. Formally, this poses no difficulties. Based on her life experiences and the sample test she sees, the subject has a distribution over her possible types, i.e., test scores. In the Benchmark Choice, a subject (presumably) prefers to be rewarded based on her placement if there is more than a 50% chance her type is in the top 50%. In the High Placement Choice, a subject prefers to be rewarded based on her placement if there is more than a 50% chance her type is in the top 30%. In the Strength Choice, a subject prefers to be rewarded based on her placement if there is more than a 60% chance that her type is in the top 50%.

As expected, in the Benchmark Choice, the population displays apparent overplacement: 74% choose to be rewarded based upon their placement. Barring too many equally skilled subjects (and ignoring the possibility of errors), such a result is usually interpreted as 74% place themselves in the top half of test takers. However, this statement is imprecise, if not misleading. A more precise interpretation is that 74% believe that there is at least a

[11] Other ways to model the subjects type are possible, however.

50% chance that they are in the top half (or more than 50% chance if we interpret their preferences as being strict).

Note that these two interpretations are different and have different implications for rationality. In the first interpretation, if we assume "place themselves" indicates (near) certainty, then the population displays overconfidence, not just apparent overconfidence. But the more precise interpretation, the second interpretation, shows that the choice behavior of the subjects is consistent with rationality, as indicated by Theorem 1. Overplacement can be inferred only if the subjects' belief that they are in the top half is sufficiently more than 50% or if they believe they place sufficiently high within the top half.

Before turning to the question of overplacement, we consider the question of how certain a subject is of her type. Of the 74% who opt for placing in the top half over a 50% random draw, 22% switch and choose a 60% random draw over placing in the top half.[12] Thus, a significant fraction of the subjects do not show much confidence in their belief that they are better than average. This fact supports the underlying premise of B&D (2008), and of Moore and Healy (2008), that people are uncertain of their types.[13] In particular, it suggests that prior work on overconfidence cannot be justified by an untested presumption that people are certain, or nearly certain, of their types. Presumably, if we had asked people to vote for their placement versus a 70% or higher random draw we would have found even more people defecting from the placement option.

We turn now to the question of overconfidence. As noted, Theorem 1 proves that the Benchmark Choice cannot show overconfidence, since every subject could prefer the placement option even in a rational population. However, the Strength Choice and High Placement Choice do have the potential to show overconfidence.

From Theorem 1, the population exhibits overconfidence if more than 60% vote for the skill bet in the High Placement pair of bets (i.e. place in the top 30% vs \$10 with 50% chance), or if more than 83.3% vote for the skill bet in the Strength pair of bets (i.e. place in the top half vs. get \$10 with 60% chance). In fact, only 51.9% and not 60% vote for the skill bet in the High Placement pair of bets (51.9% is different from 60% at the 3% significance level). Also, only 64.3% and not 83.3% choose the skill bet in the Strength pair (64.3% is different from 83.3% at significance levels lower than 1%).

Observe that Theorem 1 shows that 64.3% of the population could rationally prefer to be paid based on placing in the top half to receiving the prize randomly with a probability of up to 77.8%. Thus, assuming that the number of people who bet on their placement

[12] We note that 6% of the subjects favor a 50% draw over their placement, but their placement over a 60% draw. We have no explanation for this inconsistent behaviour.

[13] However, our experiement does not provide a definitive test of the subjects' uncertainty about their types as they may also have been concerned about randomness in the test itself (although concern about this randomdess should be mitigated because subjects were shown a quite representative sample test).

would not increase as the probability of receiving the prize increases (in the random bet), we cannot reject the no-overconfidence hypothesis for a range of prize probabilities beyond the 60% we test for directly. We note that, on the one hand, the figure 77.8% overstates the range as it accepts the 64.3% of the population as a precise count without conducting a significance test, while on the other hand it understates the range as, surely, far less than 64% of the population would have voted for the test had the alternative been the prize with a probability of 78% or greater.

Although it is not the focus of our study, we mention one intriguing finding. While the high/low motivation treatment does not affect the betting behaviour of our subjects, the subjects have significantly lower scores under the high motivation treatment. Those in the high motivation condition answered 16.6 questions correctly, whereas those in the low motivation condition answered an average of 18 questions correctly, and an independent samples t-test reveals this difference to be significant at significance levels below 1%. Thus, our subjects appear to "choke" under pressure, as has been documented by other studies, including Ariely, Gneezy, Loewenstein, and Mazar (2005), Beilock and Carr (2001), Dohmen (2005), and Markman and Maddox (2006). In the present context, this finding is interesting in that it speaks to the potential adaptiveness (or lack thereof) of motivations to be confident.

4.1 A Single Model

Theorem 1 indicates that the results from our three questions can all be generated in a rational fashion. More precisely, the theorem tells us that the data from these three choices can be rationalized by three different rational models (three populations, three signalling structures, etc). However, our data comes from a single subject pool in a single experiment. We now show by construction that the aggregate data can also rationally be generated by a single experiment in which all the participants are fully rational.

There are twenty-one possible scores in our experiment, and so we build a model with twenty-one types. Subjects receive signals of their types. Given the nature of the experiment, the simplest model to generate the data is one in which the population divides into three "equivalence classes". Types in the lowest equivalence class, θ_l, score in the bottom 50% of subjects; types in the middle equivalence class, θ_m, score in-between the bottom 50% and the top 30% of subjects; types in the highest class score in the top 30%. Each type in a given equivalence class receives one of fours signals s_1, s_2, s_3, s_4 according to the same probability distribution. The joint probability distribution of types and signals is

	θ_l	θ_m	θ_h	Marginal
s_1	.2599081	.000087	0000049	26%
s_2	.0499	.0393	.0108	10%
s_3	.051987	.043823	.03419	13%
s_4	.1382049	.11679	.2550051	51%
Marginal	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{3}{10}$	

The numbers in the above chart are not particularly "nice" as they must be chosen to fit the data. Importantly, however, the signalling structure itself is nice in that it satisfies the monotone likelihood ratio property (for type θ' larger than type θ, we have that $\Pr_{\theta'}(s_i) / \Pr_{\theta}(s_i)$ is increasing in s_i).

The following table shows the posterior beliefs over types given each signal s_j, $\Pr(\theta_i | s_j)$,

	θ_l	θ_m	θ_h
s_1	0.99965	.00033462	.000018846
s_2	0.499	0.393	0.108
s_3	0.399 9	0.337 1	0.263
s_4	0.270 99	0.229	0.50001

Thus, a person who sees the signal, say, s_4, believes she has just above a 27% chance of placing in the bottom 50%, just below a 23% chance of placing higher than the bottom 50% but lower than the top 30%, and just above a 50% of placing in the top 50%. Such a person will always vote for the placement option rather than one of the random choices, since there is a 73% chance that she places in the top 50% and over a 50% chance that she places in the top 30%. The following table indicates how people who receive the different signals should vote:

	Pr	top half vs 50%	top 30% vs 50%	top half vs 60%
s_1	26%	Random	Random	Random
s_2	10%	Placement	Random	Random
s_3	13%	Placement	Random	Placement
s_4	51%	Placement	Placement	Placement
	Placement Total	74%	51%	64%

As the bottom row of the table shows, this signalling model generates the data found in our experiment.

5 Conclusion

As in much previous experimental work, we find a better-than-average effect among our subjects. Since the task we assigned the subjects was an easy one, the theory in B&D

led us to expect this finding. In contrast to previous work, we inquire further to see if the subjects exhibit behaviour that cannot be explained rationally but, rather, is unambiguously indicative of biased and erroneous beliefs. We do not find such evidence, even though we pushed in the direction of overconfidence by recruiting subjects through instructions that would lead to self selection and by motivating them. We also test whether subjects are uncertain of their types and find evidence that they are quite unsure. This is important because previous work on the better-than-average effect could be interpreted as showing overconfidence if subjects are (almost) certain of their types. In contrast to previous work (Moore and Healy (2008), Clark and Friesen, 2008) that has also failed to find overplacement, while conducting a proper test, our experiment is based on subjects' estimations just of their relative rankings.[14]

Our experiment can be viewed as a test of the null hypothesis that people are behaving rationally (at least in so far as they are not overconfident). We cannot reject that hypothesis. Of course, this is not to say that we can rule out the hypothesis that people are overconfident, either. One reason is that we did not (and could not) carry out all the tests implied by Theorem 1. Therefore, for instance, we do not know how many people would have been willing to bet that they would place in the top 20% of test takers. Another reason is that, by their very design, these types of experiments are ill-suited to rule out overconfident, or underconfident, behaviour. To understand this claim, suppose that ten subjects are to be given a Japanese vocabulary test and that nine of them have absolutely no knowledge of Japanese, while the tenth is Japanese. The nine subjects, who will answer questions randomly, each have about a $\frac{4}{9}$ chance of finishing in the top half while the Japanese subject will almost certainly finish in the top half. If the subjects are behaving rationally, only 10% of the people should prefer betting that they place in the top half rather than accepting a 50% chance at a prize. Therefore, if 30% vote for the placement option, the subjects, as a whole, are overconfident even though they naively appear to be underconfident. We used our experimental design, despite its inability to rule out overconfidence, because there is a vast literature with experiments of this type purporting to show overconfidence, and we wanted to see if, in fact, overconfidence could be found here. Naturally, a priori it seemed quite possible that we would find overconfidence.

The difficulty with making strong conclusions about whether people are overconfident given their limited information and imperfect signals is that we cannot observe all their information or their private signals, which include events from their lifetime experience.

[14] The literature on better-than-average experiments can be divided into two types: Ranking Experiments, such at the one in this paper (and myriad others), where subjects indicate their beliefs just about their relative placement, and Scale Experiments where subjects indicate exactly where they place on a scale. The experiments by Clark and Friesen (2008) and Moore and Healy (2008) are implicitly Scale Experiments. B&D contains a detailed discussion.

Thus, we do not know what rational Bayesian agents would believe had they known what our subjects knew after they saw the practice test. We make no claims, however, that our subjects are rational Bayesians. Many researchers have argued that Bayes' Rule is not a good description of intuitive human judgment (Kahneman & Tversky,(1972), Grether, 1980). At the same time, there is evidence that their judgments nevertheless roughly follow the logic underlying Bayes' Rule: their posterior beliefs lie somewhere between their priors and the signal they receive (Grether (1990), McKelvey & Page (1990).

We do not purport to show that there are no circumstances under which people believe irrationally in their own superiority. No study can do that. Our study questions the generality of the conclusion that people have biased beliefs that they are better than others. It may be argued that the experimental design we have chosen, in which performance can be measured unambiguously, and our subjects expected that their claims about performance would therefore be subject to verification, may undermine subjects' willingness to indulge their motivation to believe that they are better than others (Dunning, Meyerowitz, & Holzberg, 1989). We can only respond that such objective measurement is necessary in order to assess the accuracy of subjects' beliefs.

Some have noted the potential adaptive benefits of beliefs in one's own superiority (Armor, Massey, & Sackett, in press; Benabou & Tirole, 2002). If self-confidence increases the probability of success, then even a belief that may be demonstrably inaccurate could nevertheless be rational in the larger sense. Nevertheless, any such claim must contend with evidence suggesting that belief in one's own superiority can undermine subsequent performance, such as when a student's assurance that he will perform well on a test leads him to not study (Stone, 1994), when motivation to perform leads to choking under pressure (Ariely, Gneezy, Loewenstein, & Mazar, 2005; Beilock & Carr, 2001; Dohmen, 2005; Markman & Maddox, 2006), or when inflated belief in one's social status reduces one's popularity (Anderson, Srivastava, Beer, Spataro, & Chatman, 2006). Moreover, there is the disappointment that is likely to follow inflated expectations of performance (McGraw, Mellers, & Ritov, 2004).

It is not fair to assume that people are omniscient or perfectly prescient. Theories of human behavior must take into account the limited, imperfect, and biased information people have at their disposal when they make important assessments, such as their abilities relative to others. Our psychological and economic theories will be better to the extent that they do so.

6 Appendix A: Test items from the two tests

1S) Susie has a cake that she splits into six pieces to share with all her friends. If each person with a piece of cake then splits their piece in half to give to another friend, how many pieces of cake are there in the end? 12

1M) The Maroons are first in the league and the Browns are fifth while the Blues are between them. If the Grays have more points than the Violets and the Violets are exactly below the Blues then who is second? The Grays

2S) A bridge consists of 10 sections; each section is 2.5 meters long. How far is it from the edge of the bridge to the center? 12.5 m

2M) Five friends share three oranges equally. Each orange contains ten wedges. How many wedges does each friend receive? 6

3S) There are four equally spaced beads on a circle. How many straight lines are needed to connect each bead with every other bead? 6

3M) Fall is to Summer as Monday is to _____? Sunday

4S) HAND is to Glove as HEAD is to _____? Hat

4M) What is the minimum number of toothpicks necessary to spell the word "HAT". (You are not allowed to break or bend any toothpicks, or use one toothpick as a part of more than one letter.) 8

5S) John needs 13 bottles of water from the store. John can only carry 3 at a time. What's the minimum number of trips John needs to make to the store? 5

5M) Milk is to glass as soup is to _____? bowl

6S) LIVED is to DEVIL as 6323 is to _____? 3236

6M) Which number should be next in the sequence: 2, 4, 8, 16, 32, ? 64

7S) If the day before yesterday is two days after Monday then what day is it today? Friday

7M) A rancher is building an open-ended (straight) fence by stringing wire between posts 25 meters apart. If the fence is 100 meters long how many posts should the rancher use? 5

8S) Which number should come next in the series: 3, 9, 6, 12, 9, 15, 12, 18, ? 15

8M) "Meow" is to a cat as "Moo" is to _____? Cow

9S) Which letter logically follows in this sequence: T, Q, N, K, H, ? E

9M) Which word does not belong in the group with the other words? Brown, Black, Broom, Orange, Bread Orange

10S) If two typists can type two pages in five minutes, how many typists will it take to type twenty pages in ten minutes? 10

10M) If a woman is 21 and is half the age of her mom, how old will the mom be when the woman is 42? 63

11S) Tiger is to stripes as leopard is to _____? Spots

11M) Which number should come next: 514, 64, 8, 1, 1/8, ? 1/64

12S) Brother is to sister as nephew is to _____? Niece

12M) Which number should come next in this series: 1 - 1 - 2 - 3 - 5 - 8 - 13 - ? 21

13S) Desert is to oasis as ocean is to _____? Island

13M) If 10 missionaries have 3 children each, but only two thirds of the children survive, how many children survive? 20

14S) Kara has $100. She decides to put 20% in savings, donate 20% to a charity, spend 40% on bills, and use 20% for a shopping spree. How much money does she have left over afterwards? $0

14M) Kimberly makes $20 per hour and works for 20 hours each week. How much does she make in a week? 400

15S) How many straight lines are needed to divide a regular hexagon into 6 identical triangles? 3

15M) Which number should come next in this series: 1,4,9,16,25,? 36

16S) What is the average of 12, 6 and 9? 9

16M) DIDIIDID is to 49499494 as DIIDIIDD is to _____? 49949944

17S) There are three 600 ml water bottles. Two are full, the third is 2/3rds full. How much water is there total? 1600ml

17M) If a wood pile contains 30 kilos of wood and 15.5 kilos are burned, how many kilos are left? 14.5

18S) Which letter does not belong in the following series: D - F - H - J - K - N - P - R K

18M) Joe was both 5th highest and 5th lowest in a race. How many people participated? 9

19S) If a certain type of bug lives for only 20 days, how old is the bug when it has lived half of its lifespan? 10 days

19M) PEACH is to HCAEP as 46251 is to _____? 15264

20S) Begin is to began as fight is to _____? Fought

20M) Nurse is to hospital as teacher is to _____? school

References

Amabile, T. (1983) "The social psychology of creativity: A componential conceptualization," *Journal of Personality and Social Psychology*, **45(2)**, 357-76.

Anderson, C., S. Srivastava, J.S. Beer, S.E. Spataro and J.A. Chatman (2006), "Knowing your place: Self-perceptions of status in face-to-face groups," *Journal of Personality and Social Psychology*, **91(6)**, 1094-1110.

Ariely, D., U. Gneezy, G. Loewenstein and N. Mazar, (2005), "Large stakes and big mistakes," available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=774986

Armor, D. A., C. Massey, and A.M. Sackett (in press), "Prescribed optimism: Is it right to be wrong about the future?" *Psychological Science*.

Barber, B. and T. Odean (2001), "Boys Will Be Boys: Gender, Overconfidence, And Common Stock Investment," *Quarterly Journal of Economics*, 116(1), 261-92.

Beilock, S. L. and T.H. Carr (2001), "On the fragility of skilled performance: What governs choking under pressure?" *Journal of Experimental Psychology: General*, **130(4)**, 701-25.

Bem, D.J. (1967), "Self-perception theory: An alternative interpretation of cognitive dissonance phenomena," *Psychological Review*, **74(3)**, 183-200.

Bénabou, R. and J. Tirole (2002), "Self Confidence and Personal Motivation," *Quarterly Journal of Economics*, **117(3)**, 871-915.

Benoît, J-P. and J. Dubra (2008), "Overconfidence?" mimeo available at ssrn.

Bernardo, A. and I. Welch (2001), "On the Evolution of Overconfidence and Entrepreneurs," *Journal of Economics & Management Strategy*, **10(3)**, 301-330.

Brocas, I. and J. Carrillo (2007), "Systematic errors in decision-making," mimeo.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: an experimental approach', *American Economic Review*, **89(1)**, pp. 306–18.

Clark, J. and L. Friesen (2008), "Rational Expectations of Own Performance: An Experimental Study," forthcoming *Economic Journal*.

Chuang, W. and B. Lee, (2006), "An empirical evaluation of the overconfidence hypothesis," Journal of Banking & Finance, 30(9), 2489-515.

Daniel, K., D. Hirshleifer and A. Subrahmanyam (2001), "Overconfidence, Arbitrage, and Equilibrium Asset Pricing," *Journal of Finance*, **56(3)**, 921-65.

Dohmen, T. J. (2005), "Do professionals choke under pressure?" Unpublished manuscript.

Dunning, D., J.A. Meyerowitz and A.D. Holzberg (1989), "Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability," *Journal of Personality and Social Psychology*, **57(6)**, 1082-90.

Fang, H. and G. Moscarini, (2005) "Morale Hazard," *Journal of Monetary Economics*, **52(4)**, 749-777.

Festinger, L. (1954) "A Theory of Social Comparison Processes," *Human Relations*, **7(2)**, 117-140.

Garcia, D., F. Sangiorgi and B. Urosevic, (2007), "Overconfidence and Market Efficiency with Heterogeneous Agents," *Journal Economic Theory*, **30(2)**, 313-36.

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

Grether, D. M. (1990), "Testing Bayes rule and the representativeness heuristic: Some ex-

perimental evidence," *Journal of Economic Behavior and Organization*, **17**, 31-57.
Kahneman, D. and A. Tversky (1972), "Subjective probability: A judgment of representativeness," *Cognitive Psychology*, **3(3)**, 430-454.
Kőszegi, B., (2006), "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, **4(4)**, 673-707.
Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. Psychological Bulletin, 133(1), 95-121.
Kruger, J. (1999), "Lake Wobegon Be Gone! The "Below-Average Effect" and the Egocentric Nature of Comparative Ability Judgements", *Journal of Personality and Social Psychology*, **77(2)**, 221-232.
Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin, 108(3), 480-498.
Kyle, A. and F.A. Wang, (1997), "Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?" *Journal of Finance*, **52(5)**, 2073-90.
Malmendier, U. and G. Tate (2005), "CEO Overconfidence and Corporate Investment," *Journal of Finance*, **60(6)**, 2661-700.
Markman, A. B., W.T. Maddox, (2006), "Choking and excelling under pressure," *Psychological Science*, **17(11)**, 944-48.
McGraw, A. P., B.A. Mellers and I. Ritov (2004), "The affective costs of overconfidence," *Journal of Behavioral Decision Making*, **17(4)**, 281-295.
McKelvey, R. D. and T. Page (1990), "Public and private information: An experimental study of information pooling," *Econometrica*, **58**, 1321-39.
Menkhoff, L., U. Schmidt and T. Brozynski, (2006) "The impact of experience on risk taking, overconfidence, and herding of fund managers: Complementary survey evidence," *European Economic Review*, **50(7)**, 1753-66
Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517.
Peng, L. and W. Xiong, (2006), "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, **80(3)**, 563-602.
Stone, D. N. (1994),"Overconfidence in initial self-efficacy judgments: Effects on decision processes and performance," *Organizational Behavior and Human Decision Processes*, **59(3)**, 452-74.
Svenson, O., (1981), "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica*, **94**, pp 143-148.
Wang, A. (2001), "Overconfidence, Investor Sentiment, and Evolution," *Journal of Financial Intermediation*, **10(2)**, 138-70.
Weinstein, N. (1980), "Unrealistic Optimism about Future Life Events," *Journal of Personality and Social Psychology*, **39(5)**, 806-20.

Zábojník, J. (2004), "A Model of Rational Bias in Self-Assessments," *Economic Theory,* **23(2)**, 259–82.