Overconfidence?

Jean-Pierre Benoît and Juan Dubra

London Business School, Carnegie Mellon University and
Universidad de Montevideo

2008

Overconfidence?

Jean-Pierre Benoît
London Business School

Juan Dubra*
Carnegie Mellon University
and Universidad de Montevideo

Abstract

Many studies have shown that people display an apparent overconfidence. In particular, it is common for a majority of people to describe themselves as better than average. The literature takes for granted that this better-than-average effect is problematic. We argue, however, that, even accepting these studies completely on their own terms, there is nothing at all wrong with a strict majority of people rating themselves above the median.

When it comes to overconfidence, there is a consensus on a consensus: writers agree that researchers have found overconfidence to be common. Typical comments include "Dozens of studies show that people...are generally overconfident about their relative skills" (Camerer, 1997), "Perhaps the most robust finding in the psychology of judgment is that people are overconfident" (DeBondt and Thaler, 1995), and "The tendency to evaluate oneself more favorably than others is a staple finding in social psychology" (Alicke et al. 1995). While the study of overconfidence originated in the psychology literature, the phenomenon has migrated into the economics and finance literature, taking its place in the growing list of "irrational" aspects of human attitudes and behaviour that were once on the fringes but are now on the forefront of mainstream thinking.[1]

*We thank Stefano Sacchetto for his research assistance. We also thank Ariel Rubinstein, Rafael Di Tella, Federico Echenique, Emilio Espino, PJ Healy, Richard Lowery, Henry Moon, Don Moore, Nigel Nicholson, Luís Santos-Pinto, and Madan Pilutlla for their comments.

[1]Papers on overconfidence in economics include Camerer and Lovallo (1999), Garcia, Sangiorgi and Urosevic (2007), Hoelzl and Rustichini (2005), Koszegi (2006), Menkhoff et al. (2006), Noth and Weber (2003), Van den Steen (2004), Zabojnik (2004). In finance, recent (published) papers include Barber and Odean (2001), Biais et al. (2005), 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), Wang (2001).

If people are indeed overconfident there are important implications for our understanding of the economy and for public policy. For instance, a basic principle of competition is that firms will enter an industry only up to the point that they earn zero expected profits. But if, to take one industry, restaurateurs overestimate their abilities, we can expect too many restaurants to open only to close shortly thereafter, and the restaurant business to lose money on average. At the same time, if people overestimate their driving ability, then merely informing them of general risks will not induce them to take sufficient care.

Overconfidence has been reported in peoples' beliefs in the precision of their estimates, in their beliefs about the likelihood their answers to questions are correct, and in their appraisal of their relative skills and virtues. In this paper, we are concerned with the last form of overconfidence.[2] As Myers $(1999, p.57)$ writes, "on nearly any dimension that is both *subjective* and *socially desirable*, most will see themselves as better than average." As evidence, he cites research showing that most people perceive themselves as more intelligent than their average peer, most business managers rate their performance as better than their average fellow manager, and most high school students rate themselves as more original than the average high-schooler. In an oft-quoted study, Svenson found that 77% of Swedish subjects felt they were safer drivers than the median, and 69% felt they were more skillful. These findings, and others like them, are typically cited as evidence of overconfidence, at least in peoples' estimation of their relative skills, without any explanation as to why such data is indicative of mistaken self-appraisals. For instance, Alicke et al. (1995) simply assert that "the better-than-average effect provides compelling evidence that people maintain *unrealistically* positive images of themselves relative to others" [italics added]. Presumably, the reason for the lack of explanation is that, since "it is logically impossible for most people to be better than average" (Taylor and Brown (1988)), it seems obvious that some people must be making errors in their self-evaluations. But the simple truism that most people cannot be better than average – more precisely, the median – does not imply that most people cannot rationally rate themselves above average. Indeed, we will show that *none* of the evidence cited above is evidence of unrealistically positive images at all. This is true even if we accept the evidence described on its own terms (and do not, for example, argue that subjects misunderstood the questions or disagreed in their interpretations). Failure to recognize this fact comes from a failure to frame the issue of overconfidence precisely.

To illustrate the main point of this paper, 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 $p_H = 0$,

[2] Some authors, such as Burson et al. (2005) reserve the term "overconfidence" for people who overestimate their absolute abilities, but we see no reason for this restriction. In any case, the literarture uses the term in all the ways we have listed, and more.

respectively. 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?

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

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

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, $\frac{3}{5}$ rank themselves better than average. Thus, the population of drivers seems overconfident on the whole. However, rather than being *overconfident*, which implies some error in their judgements, they are simply using all the information available to them in the best possible manner.

We emphasize that in this paper, we do not provide an explanation for overconfidence. Quite the contrary, we show that, as in the above driving example, much of the supposed evidence for overconfidence does not indicate overconfidence at all; the apparent overconfidence may be an illusion. At the same time, we do not argue that people are, in fact, not overconfident. Rather, we argue that better-than-average data does not provide evidence one way or the other. Thus, for example, a finding that 80% of the people in a population rate themselves as above median intelligence does not work against the null hypothesis that no one suffers from overconfidence.

Missing from the discussion in the literature has been the recognition that when people rank themselves, their rankings are just summary statistics which provide only limited information about the entire distributions of their beliefs. In the above driving example, although the statement "$\frac{3}{5}$ of the people rank themselves above average" appears to be problematic, an examination of the complete beliefs of the population shows that there is no anomaly. Indeed, since the $\frac{2}{5}$ of the population that do cause an accident have beliefs "I am low skill with probability $\frac{2}{3}$, medium with probability $\frac{1}{3}$, and high with probability 0," the beliefs of

the population average out to the actual population distribution, as they should.

We will also show that not only can a majority of people rationally rank themselves above the median, but, depending upon the definitions one adopts, even 99% of the population can rank itself in the top 1% without any cognitive error being implied. Moreover, the failure to properly frame the issue has led authors to make plausible sounding statements that are simply wrong, such as Camerer's (1997, p. 173) claim that two firms cannot both think they are each more likely to have the most skill (assuming a common prior). We will return to Camerer's claim in Section 1.3.

In some experiments, subjects are asked to take actions, rather than answer questions. Overconfidence is then inferred from their actions. But actions, too, provide only a summary statistic of beliefs, and the same errors that have been made in interpreting answers to questionnaires have been made in interpreting the actions that subjects take.

The remainder of this paper is organized as follows. In Section 1 we provide a careful framework for analyzing overconfidence. We distinguish between *apparent* overconfidence, which gives a possibly misleading impression of overconfidence (as in the above driving example), and (true) overconfidence. We show that much of the evidence in the literature purporting to show overconfidence does nothing of the sort. At the same time, we identify two types of evidence that can show overconfidence (Theorems 3 and 6). In this section, we analyze experiments in which subjects are asked to evaluate themselves. In Section 2 we look at two papers in which subjects are asked to take actions, rather than make statements. Using the framework developed in Section 1, we show that the experiments in these papers do not provide proper tests of overconfidence. In Section 3 we suggest two experiments that do provide a proper test of overconfidence.

Recent work has challenged the universality of the finding that most people rate themselves as above average. In particular, there is evidence that this effect is attenuated when the attribute under consideration is objectively measurable, and even reversed when the skill under consideration is a difficult one to master. In Section 4 we provide an explanation for these findings. In Section 5 we discuss why it is important to distinguish between apparent overconfidence and overconfidence. In Section 6 we present some evindence on our approach. In Section 7 we review some of the literature. Section 8 is the conclusion.

1 Questionnaires

Consider a person who asserts "I am very intelligent." How are we to tell whether or not this confidence is merited? It may well be impossible, given the vagueness of the term "very". Suppose that instead the person asserts "I am more intelligent than most people." The concept "more than most" is clearer than "very," but the statement remains difficult to

assess, since it is unclear how to measure intelligence, IQ tests notwithstanding.

Researchers have attempted to surmount these difficulties by considering entire populations at once. The idea is that, while it may be difficult to judge whether or not a specific individual is overconfident, it may be possible to determine that a population is overconfident on the whole. For instance, if everyone in a room asserts "I am definitely the most intelligent person in this room", it could be concluded that all but one of them is overconfident, at least in their evaluations of their relative abilities[3] (and assuming that they agree on what constitutes intelligence). Note that for many economic problems, these types of relative rankings are the relevant ones. For instance, the wisdom of attempting a career as a professional football player depends on a person's ability relative to other would-be footballers.

This research can be divided into two types: work that proceeds by means of questionnaires and work that asks subjects to take actions. We consider the questionnaire work first, and subdivide this work into two categories, one ordinal and one cardinal:

1. Ranking questionnaires: People are asked to rank themselves relative to others on some attribute ("I am more intelligent than 80% of the people in this room").

2. Scale questionnaires: People are asked to compare themselves to the population on a scale ("On a scale of -5 (much less intelligent than average) to 5 (much more intelligent than average), I am a 3."

We consider the ranking literature first.

1.1 Ranking Questions

Svenson's (1981) work is a prototypical example of a ranking questionnaire. Svenson gathered subjects into a room and presented them with the following instructions (among others):

> We would like to know about what you think about how safely you drive an automobile. All drivers are not equally safe drivers. We want you to compare your own skill to the skills of the other people in this experiment. By definition, there is a least safe and a most safe driver in this room. We want you to indicate your own estimated position in this experimental group. Of course, this is a difficult question because you do not know all the people gathered here today, much less how safely they drive. But please make the most accurate estimate you can.

[3]Each individual may or may not also have an overflated opinion of his absolute level of intelligence. In fact, individuals may not even have a clear notion of what this absolute level is.

Each subject was then asked to place himself or herself into a safety decile. In one respect, Svenson was very careful. Realizing that the subjects had little information about the other drivers in the room, he explicitly stated: *Of course, this is a difficult question because you do not know all the people gathered here today, much less how safely they drive.* But there is another aspect he left unaddressed: Does a driver know how safely he himself drives? Of course, a driver has more information about himself than about a stranger (for instance, he knows the number of accidents he has had), but there is no reason to presume that he knows precisely how safe his driving is[4] (even assuming that he knows exactly what it means to drive "safely"[5]). This raises the question of what exactly a respondent means when he ranks himself as being, say, in the 7^{th} decile of drivers when it comes to safety.

To isolate the nature of the problem, let us consider a more carefully delineated problem. Suppose a subject is asked to rank her "reaction time". She is told that reaction time is measured to the nearest tenth of a second, and that it varies in the general population uniformly from 1 second to 0.1 seconds, so that, for instance, a time of .3 seconds places a person in the 8^{th} decile (smaller reaction times are better). She is asked to estimate her position in the population. Suppose that her beliefs about her own reaction time are given by Chart I below (she estimates that with probability 0.16 her reaction time is .5 seconds, with probability 0.2 her reaction time is .4 seconds, etc...). The chart also lists population deciles.

Reaction Time	1	.9	.8	.7	.6	.5	.4	.3	.2	.1
Prob. own time	0	0	0	0.16	0.16	0.16	0.20	0.30	0.02	0
Decile	1	2	3	4	5	6	7	8	9	10

Chart I

In what decile will she place herself? Three reasonable answers immediately come to mind, corresponding to three common notions of "average".

(α) She can reasonably place herself in the 7^{th} decile. After all, there is over a 50% chance that her reaction time will be .4 or better, and this is the fastest time for which she can make such a claim. Another way of saying this is that .4 is the median of her distribution, and this places her in the 7^{th} decile .

(β) She can reasonably place herself in the 6^{th} decile. After, all .462 is her mean reaction time, which rounds up to .5 which is in the 6^{th} decile.

[4]Many authors explicitly acknowledge that people are not likely to be certain of their skill levels. Thus, Alicke et al. (1995) ask subjects to make "point estimates" of their skill, while Hoelzl and Rustichini (2005) note that a subject in their experiment has only "some idea of his skills in it". However, these and other authors do not fully explore the implications of this uncertainty.

[5]Dunning et. al (1989) argue that people may have different notions of what it means to drive safely, so that the data is not what it appears to be. Here, we give the best case for the data and assume that all subjects agree on the meaning of a safe driver.

(μ) She can reasonably place herself in the 8^{th} decile. After all, her modal time .3 places her in the 8^{th} decile.

Certainly, this list is not exhaustive. Thus, when a person places herself in a certain decile, we, as researchers, have no real way of knowing the significance of her answer. Is there a "correct" or "rational" answer? No; she is being asked to summarize her beliefs with a single parameter, but no single reply supplies the best information for all circumstances. For instance, if she is risk-neutral, and we are interested in knowing whether she would place an even money bet that her reaction time is better than that of $x\%$ of the population, then her median belief provides the requisite information. On the other hand, if we would like to know whether she would place a bet where she receives a payment based on how much faster or slower she is than other people, her mean answer may be more informative. If she is not risk-neutral, then no single parameter may be of much use.

Returning to Svenson's driving question, a person may consider herself to be quite a safe and skillful driver since she has never had an accident and always manoeuvres well in traffic, but at the same time realize that her limited experience restricts her ability to make a precise self-appraisal. In ranking herself, she must estimate her own ability as well as that of the others in the room. As a result, it is unclear what to make of her answer to Svenson's question and, hence, of Svenson's data. Rather than ascribe a particular meaning, we will consider several possibilities.

1.1.1 Population Median Data

Consider a population where each person is asked to rank his or her skill level relative to the other people in the population. (We use the word "skill" loosely here to denote the attribute under consideration.) The literature is not always very careful in defining when this population displays overconfidence (or, equivalently for this paper, when the population displays the so-called better-than-average effect), but the general idea is that there is overconfidence if, as Myers $(1999, p.57)$ writes, most people "see themselves as better than average."

An immediate difficulty with this formulation is the ambiguity in the notion of "average" – does this refer to the mean or the median? It is easy to see that the mean cannot possibly provide the right definition, at least when the underlying trait distribution may be skewed. For example, in a population of ten people, one who has scored 0 on a test and nine who have scored 50, the nine have, in fact, performed better than the mean, so there is certainly nothing wrong in them believing that they have. Thus, Definition 1 below, which uses the median, is what the literature has in mind, whether explicitly or implicitly.[6]

[6] For instance, Hoelzl and Rustichini say that a population exhibits overconfidence if "a majority of people estimates their skills or abilities to be better than the median".

Given a set of individuals who are asked to rank themselves, let *population median data* x be the fraction of people who rank themselves strictly above the median.

Definition 1 *Population median data x is **apparently overconfident** if $x > \frac{1}{2}$.*

The reader will have noticed that we have used the word "apparently" in the above definition. To understand the reason, we must ask why a population that ranks itself highly on average should be called *over*confident, rather than simply confident. Clearly, the idea is that in an overconfident population there is something incorrect, or at the very least inconsistent, in people's self-evaluations. To determine if a population that is apparently overconfident is truly overconfident, we need a notion of what it means for peoples' self-evaluations to be correct and consistent. Fortunately, we have such a notion readily available, given by the Harsanyi common prior approach in which nature picks a skill level, or type, for each person, and over time each person receives information about her type and updates her beliefs about herself using Bayes' rule. We formalize this below.

Definition 2 *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.*

If, for example, S is finite, $f_\theta(s)$ is the probability that a person of type θ receives the signal s.

- Throughout this paper we interpret higher types as more skillful.

Definition 3 *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.

This definition of a signalling model reflects a standard approach to a situation of incomplete information.[7] Although it is not necessary for us, we posit that each individual knows the distribution of types in the population. On the one hand, this is certainly a plausible condition. For instance, an individual might know the distribution of the number of accidents a person can expect to have in a lifetime, or the distribution of IQs in the population,

[7] In the literature, it is often assumed that nature chooses a type for each person independently. Imposing this restriction would not modify our results.

without knowing either figure for herself.[8] More importantly, this condition makes our task more difficult: we will show that apparently overconfident data can rationally arise, even if everyone has a perfect understanding of the level of skills in the population. Note that, although it is not explicit in the definition of a signalling model, a dynamic time structure is allowed. In particular, the signal that an individual receives in "period 1" may consist of various pieces of information obtained over time (for instance, from her driving experience).

This is quite a rational model; indeed for many it is the definition of full rationality. As such, it provides a proper foundation for judging the rationality of a population. Note that since agents in a signalling model use Bayes' rule, for a large population their beliefs average out to the (true) population distribution.

The following proposition indicates that apparent overconfidence is incompatible with rationality *when people know their skill levels exactly.*

Proposition 1 *Consider a signaling model. If everybody's updated beliefs after receiving their signals are degenerate, the population median data cannot be apparently overconfident.*

Proof. All proofs not in the text are in the appendix ■

If people are certain of their types, then a strict majority of them cannot rationally believe they are strictly better than the median.[9] However, this certainty is a rather implausible condition – in most, if not all, situations each person will have only an imperfect indication of his own skill.[10] How imperfect can a rational persons' self-knowledge be? A priori, it seems difficult to require more of a rational population than that its members derive their beliefs in a rational and consistent manner; that is, that they derive them from a signalling model.[11]

In ranking questionnaires, people do not report their full beliefs, but only a ranking. Rational and consistent individuals report rankings that come from beliefs that are derived from a signalling model; however, as discussed in the previous section, there are many legitimate ways to report a ranking derived from a particular belief. Therefore, in interpreting ranking data we need to consider several possibilities.

[8] In experiments, the comparison population is sometimes the group of subjects in the room. Even if a subject knows the skill distribution in the general population (or relevant subpopulation, for instance, students), he may not know the distribution of skills in the room. However, the best case for the validity of an experiment is when the subject group is large enough to be representative of the larger population.

[9] Propostion 1 remains true if we modify the definition of a signalling model to allow people to be uncertain of the population distribution. For instance, nature could pick one of several population distributions with probabilities that are common knowledge.

[10] As Benabou and Tirole (2002) write, "The psychology literature generally views introspection as quite inaccurate (Nisbett and Wilson (1977)), and stresses that learning about oneself is an ongoing process."

[11] We might further ask that this signalling model be, in some sense, reasonable. We explore this issue in Section 1.1.3.

Suppose that people use their median beliefs about themselves in their self-evaluations (as in α of Section 1.1). The next definition says that a fraction x can rationally and consistently report themselves as being better than the population median, if, starting from a common prior, and using Bayes' rule, a fraction x can come to believe their median type is better than the population median. To avoid ambiguities, we require that the distribution of types has a unique median.

Definition 4 *Population median data x can be α-**rationalized** if there exists a signalling model in which the distribution of types has a unique median and x is the expected fraction of people who will believe that their median type is strictly greater than the population median, after receiving their signals and updating.*

More formally, x can be α-rationalized if there exists a signalling model as follows: Let m be the median of the prior p. Let $S_{med} \subset S$ be the set of signals such that an individual who receives a signal $s \in S_{med}$ has a median belief about himself that is strictly greater than the (unique) population median. Thus, $s \in S_{med}$ if and only if $P(\theta \leq m \mid s) < \frac{1}{2}$. Let F denote the probability distribution of the signals in S, often called the "marginal". That is, for each (measurable) $T \subset S$ let

$$F(T) = \int_{\Theta} \int_{T} df_{\theta}(s)\, dp(\theta)$$

Then, x can be α-**rationalized** if $x = F(S_{med})$.

In a stochastic environment it is possible to "explain" a wide range of experimental data as the outcome of a random, although possibly unlikely, process. Definition 4 avoids this "cheat" by insisting that the data x be the *expected* fraction of people who believe themselves to be above average. This can also be interpreted as restricting ourselves to data that comes from large populations.[12] Thus, Definition 4 is a demanding notion of rationalizing. When people self-evaluate using their median types and the data can be α-rationalized, there is no prima facie case for calling it "overconfident".

The following definition is for a population in which people use their mean beliefs about themselves (as in β of Section 1.1) for their self-evaluations.

Definition 5 *Population median data x can be β-**rationalized** if there exists a signalling model in which the distribution of types has a unique median and x is the expected fraction of people who will believe that their mean type is strictly greater than the population median, after receiving their signals and updating.*

[12] In Section 1.3 we briefly discuss small populations, where more extreme data can be rationalized.

Although reporting a modal belief strikes us as a plausible way to answer a questionnaire, it also strikes us as less compelling than reporting either a median or a mean belief. Therefore, in the interest of space, from now on we no longer consider mode reports. (Considering mode reports would not modify our results in any essential way.) Instead, we turn to a slightly different approach than the one we have adopted so far.

It is reasonable for a person with no information about herself, other than that she is a random member of the population, to rate herself as average.[13] Suppose the person now receives a signal that causes her beliefs about her own type to strictly first order stochastically dominate the population distribution. It is natural for this person to now rank herself above the median person.[14] This leads to the following definition.

Definition 6 *Population median data x can be γ-**rationalized** if there exists a signalling model in which the distribution of types has a unique median and x is the expected fraction of people who, after receiving their signals and updating, will have beliefs about their own type that strictly first order stochastically dominate the population distribution.*

The existing literature assumes that apparent overconfidence implies cognitive errors, or inconsistencies, on the part of (some) respondents without considering the meaning of their replies. In our terms, the literature assumes that apparent overconfidence implies that the data cannot be rationalized without specifying which sense of rationalizing. In some cases, there may be a reason to focus on a particular sense (as in Section 2.2). Absent such a reason, a stringent definition of overconfidence requires that population median data be called overconfident only when it cannot be rationalized using any of the above concepts, for only then can we be sure that there is a "problem" with the data[15]; a lax definition requires only that data be called overconfident when it cannot be rationalized using at least one of the concepts.[16]

[13] A person who has no private information about herself and who self-evaluates using the median of her type, ranks herself as equal to the population median; if she uses her mean type, she ranks herself as equal to the population median if the prior is symmetric, but not necessarily otherwise (which may be an argument against the reasonableness of the mean).

[14] Note, however, that if the population distribution is not symmetric, the fact that a person's beliefs about herself strictly first order stochastically dominate the population distribution does not imply that either her median or mean type is strictly better than the population median type.

[15] In fact, even then we could not be sure as there could be still other reasonable ways for people to evaluate themselves. Morevover, although in the interest of space, we have assumed that the entire population self-evaluates in the same way, nothing precludes different people using different ways. For instance, $\frac{1}{3}$ of the population could self-evaluate with their mean type and $\frac{2}{3}$ with their median type.

[16] It is the *data* (i.e., the evidence at hand) which we are defining as overconfident, or not, rather than the population. Note, for instance, that even a group of people that is apparently underconfident could, in fact, be overconfident if they are ranking themselves more highly than a rational appraisal of their life experiences would justify (although it might be difficult, or impossible, for an analyst to determine this).

Definition 7 *Population median data x is **strongly overconfident** if it is apparently overconfident and it cannot be α-rationalized, and cannot be β-rationalized, and cannot be γ-rationalized.*

Definition 8 *Population median data x is **weakly overconfident** if it is apparently overconfident and it cannot be α-rationalized, or it cannot be β-rationalized, or it cannot be γ-rationalized.*

In a *symmetric* signalling model, the population distribution is symmetric.[17] In what follows, we note when the data can (also) be rationalized by a symmetric signalling model (and hence the rationalizing does not depend upon a discrepancy between the mean and median).

The following theorem shows that population median data cannot prove even the weak version of overconfidence.

Theorem 1 *Apparent overconfidence of population median data x implies weak overconfidence only if $x = 1$. Put differently, it is possible to α-rationalize, and β-rationalize, and γ-rationalize any fraction $x \in [\frac{1}{2}, 1)$ of the population rating themselves above the median. Moreover, these rationalizations can be done with symmetric signalling models.*

Theorem 1 shows that when people have imperfect information about their skills, and receive information about these skills over time, there is nothing wrong with a strict majority of them ranking themselves above the median. Thus, apparent overconfidence should not be used as an indication of overconfidence. For instance, Svenson's (1981) finding that "a majority of subjects regarded themselves as more skillful and less risky than the average driver" is unproblematic. Note that Theorem 1 restricts x to be greater than $\frac{1}{2}$ only because we are concentrating on overconfidence. The theorem remains true for all $x \in (0, 1)$, so that apparent *underconfidence* is also not problematic. Although the three notions of rationalizing used in Theorem 1 are independent of each other, the drivers example in the introduction illustrates the theorem for all three.

The above theorem concerns a population of individuals who place themselves relative to the median person. The next theorem is even more dramatic: almost everyone can rationally believe that their *mean* skill level is strictly higher than the skill level of any fraction of the population (even if the population distribution is symmetric).

Theorem 2 *Data in which any fraction $x \in (0, 1)$ of the population ranks itself strictly higher than any fraction $q \in (0, 1)$ of the population can be β-rationalized by a signalling model. In particular, 99% of the people can rationally believe that their mean skill level is*

[17] If P is the distribution, and h is the midpoint of the support , $P(\theta \leq h - y) = P(\theta \geq h + y)$ for all y.

strictly higher than the skill level of 99% *of the people. Moreover, this rationalizing can be done with a symmetric signalling model.*

1.1.2 Population Ranking Data

Although the results of ranking experiments are typically summarized by the number of people who rank themselves above the median, most of these experiments collect more detailed data, such as the deciles into which subjects place themselves. While the previous section shows that population median data is essentially useless for determining whether or not people are overconfident, this more complete data is potentially helpful.

Suppose that each person is asked to place himself into a "k-cile", where to be in the j^{th} k-cile means that the person ranks himself strictly above the fraction $\frac{j-1}{k}$ of the population, but not strictly above the fraction $\frac{j}{k}$.[18] *Population ranking data* is a vector $x \in \mathbf{R}^k$, $\sum_1^k x_i = 1$, where x_i, $i = 1, ..., k$ is the fraction of people who rank themselves in the ith k-cile. We have the following:

Definition 9 *The population ranking data x is* ***apparently overconfident*** *if x strictly first order stochastically dominates $\left(\frac{1}{k}, ..., \frac{1}{k}\right)$.*

If neither one of x and $\left(\frac{1}{k}, ..., \frac{1}{k}\right)$ first order stochastically dominates the other, then the data has neither an unambiguously overconfident nor underconfident appearance. For instance, the data $\left(\frac{1}{5}, \frac{1}{5}, \frac{3}{10}, 0, \frac{3}{10}\right)$ contains a disproportionately large number of people who consider themselves to be in the top fifth, but a disproportionately small number who place themselves in the top two fifths.

Definition 10 *The population ranking data x can be α-****rationalized*** *if there exists a signalling model in which nature assigns a fraction $\frac{1}{k}$ of the population to each k-cile and the expected number of people whose updated beliefs will place their median type in the j^{th} k-cile is x_j, $j = 1, ..., k$.*[19]

The following theorem says that when people report the median of their beliefs, a rational population can be "twice as confident" as reality would suggest, but no more. For instance, suppose that people place themselves into deciles ($k = 10$). Then apparently overconfident data in which $\frac{2}{10}$ of the people rank themselves in the top decile, $\frac{4}{10}$ rank themselves in the

[18] By definition, each person ranks himself striclty above the empty set, so that everyone is in a (unique) k-cile.

[19] The definition assumes that nature places a fraction $\frac{1}{k}$ of the population in each k-cile in order to avoid trivialities. For instance, if the entire population were assigned a single type then, even without receiving any signals, 100% of the population would place themselves in the 1st decile. Most of the experimental work seems to carry a presumption that the population divides evenly into the k-ciles.

top two deciles, and $\frac{2i}{10}$ rank themselves in the top i deciles for $i = 3, 4, 5$ can be rationalized. However, data in which $\frac{3}{10}$ of the population place themselves in the top decile can not be explained as rational. (Although we have been emphasizing overconfident looking data, there is a similar constraint put on underconfident looking data, and that is captured by the second inequality in the theorem.) Let $\lceil n \rceil$ denote the least integer weakly greater than $n \in \mathbf{R}$.

Theorem 3 *The population ranking data x can be α-rationalized if and only if*

$$\begin{aligned} \sum_{i}^{k} x_j &\leq \frac{2}{k}(1+k-i),\ i = \left\lceil \frac{k+1}{2} \right\rceil, ..., k \qquad \textit{and} \\ \sum_{1}^{i} x_j &\leq \frac{2}{k} i,\ i = 1, ..., \left\lceil \frac{k-1}{2} \right\rceil \end{aligned}$$

Moreover, the rationalizing can be done with a symmetric signalling model.

Corollary 1 *Population ranking data in which the median declared placement is as high as the 75^{th} percentile, but no higher, can be α-rationalized.*

While almost everyone can rationally think they are better than the median, only half can rationally think they are better than the 75^{th} percentile.

Theorem 3 provides hope for detecting overconfidence by the use of ranking questionnaires. It is worth looking at Svenson's data in greater detail than that provided by his population median data. Svenson questioned students in Sweden and the United States, asking them both about their driving safety and driving skill. Swedish drivers placed themselves into deciles in the following proportions when asked about their safety:

Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Reports (%)	0.0	5.7	0.0	14.3	2.9	11.4	14.3	28.6	17.1	5.7

We first note that although the population ranking data has an overconfident feel to it, the data is not, strictly speaking, apparently overconfident, since fewer than 10% of the population ranks itself in the top 10%. More importantly, Theorem 3 implies that the data can be α-rationalized, so that it is not indicative of cognitive biases.[20] Note, for instance, that while 65.7% of the people rank themselves among the top 40%, Theorem 3 would allow up to 80% to rationally do so. Furthermore, the median ranking is between the seventh and eighth decile, which Corollary 1 permits of a rational population. In fact, out of Svenson's four questions, only half yield answers that cannot be α-rationalized – namely, the American answers. The median safety placement for American students is between 81% and 90%,

[20] Theorem 4 below shows that the data can also be $\beta-$rationalized.

which violates Corollary 1. While the median skill placement is in the acceptable $61-70\%$ range, 46% of the population places itself in the top 20% of skill level, which is too many to α-rationalize. Thus, Svenson does find some evidence of overconfidence (*if* his subjects based their answers on their median types), but it is not as strong as commonly believed. Note that when 46% of the population place themselves in the top 20% this is only 6% too many, not 26%.

Theorem 3 is our first positive result: if people self-evaluate using their median rankings, then questionnaires have the potential to detect overconfidence. Unfortunately, our next result indicates that if people use their mean rankings, then even the more complete population ranking data is useless.[21]

Definition 11 *The population ranking data x can be β-**rationalized** if there exists a signalling model in which nature assigns a fraction $\frac{1}{k}$ of the population to each k-cile and the expected number of people whose updated beliefs will place their mean type in the j^{th} k-cile is x_j, $j = 1, ..., k$.*

Theorem 4 *Any population ranking data can be β-rationalized.*

1.1.3 Reasonableness

The previous discussion has been in the rather abstract language of signalling models. Some readers may wonder if the results depend upon signalling models that are somehow "bizarre". In this section, we address this concern.

One obvious reason for a person to consider herself a safe driver is that she has not had any accidents. Thinking of driving as a "test", and not having an accident as passing the test, motivates the next definition.

Definition 12 *A **testing model** consists of a population of individuals and a type space $[0, 1]$ such that:*
i) Nature chooses a type $\theta \in [0, 1]$ for each individual independently, resulting in some distribution p; initially, each person's beliefs about his own type are described by p.
ii) A person of type θ receives a signal "pass" with probability θ and "fail" with probability $1 - \theta$.
iii) Each person updates his beliefs about himself using Bayes' rule.

[21] Theorem 4 is the first result that relies on a signalling model that is not symmetric. This could be considered to be a defect of the theorem, if there is a reason to believe that the trait under consideration is symmetrically distributed in the population. As far as we know, no one in the literature has argued that their data is significant because the trait distriubtion is symmetric.

A testing model is a natural and simple signalling model. In a *symmetric* testing model, the prior distribution is symmetric.

Theorem 5 *Population median data x, for any $x \in (0,1)$, can be α-rationalized, β-rationalized, and γ-rationalized by a symmetric testing model. Furthermore, data in which any fraction $x \in (0,1)$ of the population rank themselves strictly above any fraction $q \in (0,1)$ of the population can be β-rationalized by a symmetric testing model.*

Theorem 5 implies Theorems 1 and 2 of Section 1.1.1. As it depends only upon a symmetric testing model, which is quite simple and straightforward, it shows that those results do not depend upon a strained signalling model.

While the simplicity of a testing model is a virtue, it has a cost: since it involves only two signals, pass and fail, it can only generate data in which the population's updated beliefs divide into at most two sets. Thus, while population median data can be rationalized by a testing model, population ranking data in which the population places itself into more than two k-ciles ($k > 2$) cannot be. At the same time, a driver, for example, self-evaluates using not only the number of accidents she has had, but also the number of near-accidents, her beliefs about her reflexes and eyesight, and myriad other factors, which may be better captured by the abstractness of a signalling model than by a more concretely specified model.

What makes for a reasonable signalling model? A standard restriction found in the literature is that a signalling structure (S, Θ, f) should satisfy the monotone likelihood ratio property (mlrp): for all $\theta' > \theta$, $\frac{f_{\theta'}(s)}{f_\theta(s)}$ is increasing in s. The following proposition shows that the mlrp has implications for population ranking data.

Proposition 2 *For $\varepsilon < \frac{1}{14}$, the population ranking data $\left(\varepsilon, \varepsilon, \frac{1}{2} - \varepsilon, \frac{1}{2} - \varepsilon\right)$ cannot be $\alpha-$rationalized by a signalling model with a signalling structure that satistfies mlrp.*

Proposition 2 stands in contrast to Theorem 3 of Section 1.1.2. This proposition leaves open the possibility that population ranking data might be more useful than is implied by that theorem. For instance, one might hope to argue that Svenson's Swedish data is, in fact, indicative of some overconfidence, as it could not be rationalized by a "reasonable" signalling model. While this possibility is intriguing, the following example shows that imposing mlrp still leaves plenty of room for overconfidence.

Example 1 *The data $x = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{3}, \frac{1}{3}\right)$ can be $\alpha-$rationalized by a signalling model with a signalling structure that satisfies mlrp. In particular, let the type space be $\Theta = \{1,2,3,4\}$, let the set of signals be $S = \{1,2,3,4\}$, and let the probability with which type θ receives signal*

s be given by, for $\varepsilon < \frac{7}{180}$, $f_\theta(s)$*:*

$$f_1(s) = \begin{cases} \frac{1}{3} + 4\varepsilon & s = 1 \\ \frac{1}{3} - 2\varepsilon & s = 2 \\ \frac{1}{3} - 3.5\varepsilon & s = 3 \\ 1.5\varepsilon & s = 4 \end{cases}, \quad f_2(s) = \begin{cases} \frac{1}{3} - 5\varepsilon & s = 1 \\ \frac{1}{3} & s = 2 \\ \frac{1}{3} + 3\varepsilon & s = 3 \\ 2\varepsilon & s = 4 \end{cases}$$

$$f_3(s) = \begin{cases} \frac{5}{6}\varepsilon & s = 1 \\ \frac{8}{5}\varepsilon & s = 2 \\ \frac{1}{2} + \frac{1}{2}\varepsilon & s = 3 \\ \frac{1}{2} - \frac{44}{15}\varepsilon & s = 4 \end{cases}, \quad f_4(s) = \begin{cases} \frac{1}{6}\varepsilon & s = 1 \\ \frac{2}{5}\varepsilon & s = 2 \\ \frac{1}{6} & s = 3 \\ \frac{5}{6} - \frac{17}{30}\varepsilon & s = 4 \end{cases}$$

Finally, let the types be uniformly distributed. It is easily verified that the signalling structure satisfies mlrp. Moreover the fraction of people who see the signal $s = 1$ *is* $\frac{1}{6}$ *and the median of their posteriors is* $\theta = 1$*; the fraction of people who see the signal* $s = 2$ *is* $\frac{1}{6}$ *and the median of their posteriors is* $\theta = 2$*; the fraction of people who see the signal* $s = 3$ *is* $\frac{1}{3}$ *and the median of their posteriors is* $\theta = 3$*; the fraction of people who see the signal* $s = 4$ *is* $\frac{1}{3}$ *and the median of their posteriors is* $\theta = 4$*.*

As to $\beta-$rationalizing, Theorem 5 shows that even a simple testing model permits any fraction of the population to place their mean type in the top 1% of the population, so that imposing mlrp has no hope of eliminating extremely overconfident looking data.

1.2 Scale Questions

In scale questionnaires, participants are asked to make evaluations using a scale. There are variations, but in the version we consider people are asked to compare themselves to the average person on a designated scale.[22] For instance, Alicke et al. (1995) present subjects with a personality trait, such as intelligence or dependability, and ask them "to rate the extent to which the trait describe(s) themselves relative to the average college student of the same sex, on a single 9-point scale (*0=much less than the average college student;4=about the same as the average college student;8=much more than the average college student).*"

More generally, in a scale survey (of this type), a comparison scale T and "average" m are specified, and each person i is asked to compare himself to this average by choosing an $x_i \in T$. **Population scale data** is a quadruple $(T, m, n, \bar{x})$, where $T \subset \mathbf{R}$ is the scale used in the questionnaire, $m \in T$ is the specified scale average, n is the number of individuals,

[22] In a common variant each person places him or herself on a scale from 1 to T and indicates the fraction of the population that falls into each scale position. This is formally equivalent for defining apparent overconfidence. However, the notions of rationalizing becomes relatively involved, as the different population distributions must also be explained.

and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, where each $x_i \in T$. The literature uses the notion that the population scale data is **apparently overconfident** if $\bar{x} > m$.

Before proceeding, we must note that the fundamental methodology in many scale questionnaires seems a bit dubious. The basic idea, apparently, is that in a rational population, the answers given should average out. But given the subjective nature of many of these scales, it is unclear why this should be so, even when there is no uncertainty at all and everyone is in perfect agreement. Imagine the following question posed to two drummers, A and B, in a room:

> On an integer scale from -5 (much worse), 0 (the same) to 5 (much better), please rate yourself as a drummer compared to the other drummer.

Suppose that, as it happens, both A and B are of the opinion that the only skill that matters in a drummer is accurate time keeping. Furthermore, they both agree that A's time keeping is 80% accurate, while B's is 30% accurate. There seems to be little scope for "true" overconfidence (or underconfidence). There also seems to be nothing wrong with A rating himself with a 4, and B rating herself with a -2. After all, the scale markings are subjective. Nonetheless, the standard measure indicates overconfidence since the average rating is 1, not 0.[23]

Despite our reservations, we will proceed as if the scales are interpreted in a consistent manner by all concerned. Alternatively, we will only be considering scale questionnaires in which there is an objective scale. As an example, Weinstein (1980) asks students how their chances of obtaining a good job offer before graduation compare to those of other students at their college, with choices ranging from 100% less than average to 5 times the average. Here there is no ambiguity in the meaning of the scale, but two ambiguities remain; namely, what is meant by an average student,[24] and what a subject means by a point estimate of his or her own type?

As we saw, in ranking questionnaires the median is the only notion of an average student that it is reasonable for us to consider.[25] It turns out that in scale questionnaires, only the mean provides us with a useful notion of an average student, although it is reasonable for respondents to use either the mean of the median.

[23] There are still other potential problems with scale questions. For instance, Schwarz et al. (1991) ask subjects how successful they have been in life, and find that answers differ significantly depending upon whether subjects are presented with a scale from -5 to 5, or 0 to 10.

[24] Weinstein asks subjects to compare themselves to "other Cook students" using terms such as "50% less than average". Alicke et al. ask their subjects to compare themselves to the average student.

[25] More precisely, if the trait described is not symmetrically distributed, there is no reason for 50% of the people to be ranked above the mean. If the trait is symmetrically distributed, then the mean equals the median anyway.

To illustrate, suppose for the sake of discussion that all of Weinstein's subjects agree that there are two types of students at their college, low and high, with job offer probabilities $p_L = 0.3$ and $p_H = 1$, and that 80% of the population are low type. A reasonable interpretation of an average student is one whose chance of obtaining an offer is 0.3. Consider a respondent who thinks that there is a 50% chance that she is a low type. Her probability of obtaining a good job offer is $(.5 \times 0.3) + (.5 \times 1) = 0.65$. A perfectly reasonable response to Weinstein's question is that her chances are 35% above average. Note that we are claiming that a reasonable, perhaps the most reasonable, way to answer uses the median (or mode) in determining the population average, but the mean (of her own beliefs) for self-evaluating. It is not necessary that the reader accept this as the *most* reasonable way of answering, but merely that he or she accepts this as a plausible way. Of course, this is not to deny that it is also reasonable for a respondent to use the population mean in defining the average student and her own mean for self-evaluating. Moreover, for other questions, in particular those not involving probabilities, it may be reasonable for subjects to use their median type, rather than mean type, when self-evaluating. Thus, just considering medians and means, there are four ways to interpret answers to scale questions.

We will spare the reader the formalization of all four treatments. It is fairly obvious that in the two cases where people self-evaluate using the median of their beliefs, apparent overconfidence will not imply overconfidence, since there is no particular reason for the weighted average of medians to equal the population median or mean. Example 2 below shows that apparent overconfidence also does not imply overconfidence when people self-evaluate using their mean belief, and the population "average" is taken to be the median. Theorem 6 below covers the fourth case.

Example 2 *Consider an experiment in which 150 subjects are asked to compare themselves to others on a scale* $T = \{0, 1, ..., 9, 10\}$*, where* 5 *is "average",* 0 *indicates "much below average", and* 10 *indicates "much above average." Suppose that the number of people who place themselves at* 5 *is* 50*, and that for each of 6, 7, 8, 9, and 10 the number is 20. Clearly, the data is apparently overconfident, since nobody places himself below the average, and 2/3 of the people place themselves strictly above. We now show how to rationally explain this data when the average* 5 *represents the median type, and respondents self-evaluate using their mean type.*

Suppose that, in fact, of the 150 *subjects,* 90 *are* 5*'s and* 60 *are* 10*'s. Note that the median type is* 5*. This general information is common knowledge. Beyond this, each person receives a signal giving him further information about his own type. The set of signals is* $S = \{0, 1, ..., 9, 10\}$ *and, for the two types in the population, the probability of receiving a signal*

are given by:

$$f_5(s) = \begin{cases} 0 & s < 5 \\ \frac{5}{9} & s = 5 \\ \frac{20-2s}{45} & s > 5 \end{cases} \text{ and } f_{10}(s) = \begin{cases} 0 & s \leq 5 \\ \frac{s-5}{15} & s > 5 \end{cases}$$

Some simple calculations show that the expected number of people who will receive the signal 5 *is* 50*, and that any one receiving this signal knows that his type is* 5*. For each* $i = 6, ...10$*, the expected number who will receive the signal is* 20*, and, using Bayes' rule, a person receiving the signal* i *has a mean type of* i*. Therefore, we expect* 50 *people to rationally rate themselves as* 5*s, and* 20 *people to rationally rate themselves each of* 6*, 7, 8, 9, and 10, as reported in the experiment.*

The next theorem contrasts markedly with this example. It shows that if "average student" is interpreted as the mean of the population, and people self-evaluate using the mean of their beliefs, then apparent overconfidence implies overconfidence. More precisely, consider a (large) population whose mean type is m, and whose members know the overall distribution of types and learn about themselves over time. Then, if individuals report their mean type, at any point in time, the expected average report must be m.

Theorem 6 *Consider a population where individual* $i = 1, ..., n$ *is of type* $t_i \in T \subset R$*, and* m *is the mean type. Suppose that each person knows the distribution of types in the population and receives a signal about his own type. Then* $E\left(\frac{1}{n}\sum_{i=1}^{n} \bar{t}_i\right) = m$*, where* $\bar{t}_i$ *is the mean of person* i*'s updated beliefs.*

Theorem 6 provides the first case where the standard interpretation of the data found in the literature has merit. Weinstein finds that his subjects display apparent overconfidence. If we assume that his subjects consider the average subject to be represented by the mean of the population, and that they self-evaluate using their mean type, then his subjects also display overconfidence. Since Weinstein asks his subjects for probability information, it does seem most reasonable to interpret their responses as reflecting their mean self-evaluations. On the other hand, as we argued earlier, it may be more reasonable for subjects to consider the average of the population to be the median individual, rather than the mean.

1.3 Small Samples

As we discussed in Section 1.1.1, our definitions of rationalizing, which rely on "expected data", can be understood as implicitly insisting that the data be generatable by infinitely large populations. Naturally, an experimenter prefers to have as large a subject pool as possible, so that her results cannot be dismissed as a statistical aberration (or experimental error). Thus, our implicit assumption that the data comes from an infinitely large population

provides the best case for the data. Nonetheless, actual experimental populations are, of course, finite, and many experiments involve quite small numbers. It is worth noting that finite samples permit data that is even more (seemingly) anomalous.

Consider the following quote from Camerer (1997):

> The now-standard approach to games of imperfect information pioneered by John Harsanyi presumes that players begin with a "common prior" probability distribution over any chance outcomes. As an example, consider two firms A and B, who are debating whether to enter a new industry like Internet software. Suppose it is common knowledge that only one firm will survive-the firm with more skilled managers, say- so firms judge the chance that their managers are the more skilled. The common prior assumption insists both firms cannot think they are each more likely to have the most skill.

Despite the apparent plausibility of this statement, we now show that, using the "standard approach," each of two firms can, in fact, concurrently hold the belief that it is more likely to be the more skillful. Furthermore, this state of affairs can arise with probability arbitrarily close to one. The following is a simple and straightforward model yielding this result.

1. Nature moves first. With probability $\frac{1}{2}$, Firm 1 has high skill, Firm 2 has low skill; with probability $\frac{1}{2}$ Firm 1 has low skill, Firm 2 has high skill. Both firms know this, but neither is told Nature's choice.

2. Each firm takes a test. The high skill firm passes with probability p_h, the low skill firm passes with probability p_l, where $p_h > p_l$.

3. Each firm uses Bayes' rule to determine the probability that it is the more skillful.

As an illustration, if $p_h = 0.99$ and $p_l = 0.98$, then, from Bayes' rule, a firm that passes the test believes there is a 0.502 chance that is the more skillful. Moreover, with probability 0.97 both firms will pass the test. Thus, with probability 0.97 each firm will believe that it is more likely to be the more skillful. If $p_h = 1$ and $p_l = 0.5$, then with probability $\frac{1}{2}$ both firms will pass the test, and each will then simultaneously believe it has a $\frac{2}{3}$ chance of being the more skillful. The following proposition generalizes these possibilities.

Proposition 3 *For any $p \in (0,1)$, the parameters in the above model can be chosen so that with probability p each firm will believe there is strictly more than a $\frac{1}{2}$ chance that it is the more skillful. In particular, the parameters can be chosen so that with a 50% chance each firm will believe there is a $\frac{2}{3}$ chance that it is the more skillful; in order for the chance to be more than 50%, the belief must be smaller than $\frac{2}{3}$, in order for the belief to be greater than $\frac{2}{3}$, the chance must be less than 50%.*

Proof of Proposition 3. The probability that a firm that passes the test is high skill is:

$$p\left(\theta = H \mid pass\right) = \frac{\frac{1}{2}p_h}{\frac{1}{2}p_h + \frac{1}{2}p_l} > \frac{1}{2}$$

The probability that both firms past the test is $p_h p_l$. Fixing $p \leq p_l < \sqrt{p}$ (to ensure that $1 \geq p_h > p_l$), and setting $p_h = \frac{p}{p_l}$ establishes the first part.

To establish the second part, note that the solution to the problem

$$\max_{p_l, p_h} \frac{\frac{1}{2}p_h}{\frac{1}{2}p_h + \frac{1}{2}p_l}$$

subject to $p_h p_l \geq \frac{1}{2}$ and $p_h > p_l$, is $\frac{2}{3}$ at $p_h = 1, p_l = \frac{1}{2}$. ■

We note that the first part of the proposition has nothing to do with there being two firms – an arbitrary finite number of firms can all believe they are the best firm with any probability $p \in (0, 1)$.

2 Actions

We have emphasized the ambiguity inherent in the interpretation of replies to questionnaires. A different approach to the study of overconfidence circumvents this ambiguity by asking subjects to take actions. The subjects' beliefs are then inferred from their actions. In this section we look at two such studies.

2.1 Entry

In an oft-cited paper, Camerer and Lovallo (1999) test for overconfidence using an experiment meant to model firms' entry decisions. 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 $k < N$ (hence, all subjects who play In do better than all subjects who play Out if fewer than k choose In). There are two treatments.

1. Treatment 1. The subjects who play In are ranked randomly.

2. Treatment 2. The subjects who play In are ranked according to their results on a skill or trivia test. (The test is administered after the entry decisions, but subjects are given sample questions beforehand).

Since the number of subjects that can profitably play In is independent of the treatment, the authors test for overconfidence on the part of subjects by testing if the number of

entrants is greater under Treatment 2 than under Treatment 1.[26] They find that, indeed, more subjects enter under the second treatment than the first. But is this an indication of overconfidence, with its implication of irrationality, or apparent overconfidence, with no such implication?[27]

We now show that if more subjects enter under Treatment 2 than Treatment 1, this only shows apparent overconfidence.

We proceed with a slightly simpler setup than the one used by Camerer and Lovallo; the basic methodology and conclusions remain valid for their more intricate setup. Suppose there are two subjects ($N = 2$), and that only one of them can profitably play In ($k = 1$). Specifically, if both subjects enter, the higher ranked one earns 1, the lower ranked one loses 3. A subject who enters alone, again earns 1. A subject who does not enter earns 0. We can write "the subjective expected payoff matrix" for the game as follows, where p_i is subject i's belief that he or she will be the higher ranked:

	In	Out
In	$p_1 - 3(1-p_1), p_2 - 3(1-p_2)$	$1, 0$
Out	$0, 1$	$0, 0$

Since there are only two participants, the two treatments in this case are:

1. Under Treatment 1, if both subjects enter, the higher ranked subject is randomly chosen with probability $\frac{1}{2}$.

2. Under Treatment 2, if both subjects enter, their ranking is determined by a trivia test. The test is administered only after entry decisions have been made, but each subject is shown a sample question before.

To explain Camerer and Lovallo's finding of more entry under Treatment 2 than Treatment 1 as the result of apparent overconfidence, rather than overconfidence, we must specify a signalling model (preferably, a "reasonable" one); Section 1.3 provides one. Suppose that, before seeing the sample question, each subject believes there is a $\frac{1}{2}$ chance that he will do better on the test, and (correctly) believes that if he is to do better (resp., worse) on the test, he will know the answer to the sample question with probability p_h (resp, p_l, $p_l < p_h$).

[26] As they write "The *difference* in the number of entrants in the random and skill conditions is the primary measure of interest."

[27] In fact, as a test of any kind of overconfidence, the results of the paper are muddied by several complicating issues, including the risk attitudes of the participants and their ability to play to an equilibrium. We consider the best case scenario (for their paper) in which participants are risk-neutral and play to an equilibrium.

Suppose also that $p_h p_l > \frac{1}{2}$, so that there is a greater than 50% chance that both subjects will correctly answer the test (and each will think he is likely to be the more skillful).

We find symmetric equilibria of the games induced by the two treatments:

1. Under the first treatment, each subject enters with probability $\frac{1}{2}$. The expected number of subjects that enters is 1.

2. Under the second treatment, each subject adopts the following strategy: If I answer the the sample question correctly, I enter with probability $\frac{p_h+p_l}{4p_h p_l} > \frac{1}{2}$, if I answer incorrectly, I do not enter. The expected number of firms that enter is $\frac{1}{4p_h p_l}(p_h + p_l)^2$.

Note that $\frac{1}{4p_h p_l}(p_h + p_l)^2 > 1$. Therefore, an experimenter should expect to find that more subjects enter under Treatment 2 than under Treatment 1, but this is only an indication of apparent overconfidence, not overconfidence.[28] Interestingly, this increased entry occurs even though a subject that answers the sample question correctly realizes that there is a good chance that the other subject will also consider himself to probably be the more skilled.

2.2 A Vocabulary Test

Hoelzl and Rustichini (2005) divide subjects into groups and present them with two options:

1. Option 1. You will be given a monetary prize M with probability $\frac{1}{2}$, as determined by the toss of a die.

2. Option 2. Everyone in your group will be administered a vocabulary test. You will be given M if your score places you in the top half of your group.

Hoelzl and Rustichini write, "Since only half of the subjects will win if the test decides the winner, any excess over a half of the subjects voting for the test indicates an erroneous evaluation of their own relative skills". This statement, which forms the basis of their analysis, is incorrect.

To see why the statement is incorrect, first note that a subject will prefer the test condition if she thinks that there is more than a 50% chance that her performance will be in the top half. If we assume, as Hoelzl and Rustichini do, that each person believes that test taking ability can be summarized by a single parameter, or type, then a person will

[28] A different test is on the expected profits that subjects make, rather than the number of firms. If subjects are (truly) overconfident under Treatment 2, rather than ostensibly overconfident, they will make negative expected profits (see Section 5). However, Camerer and Lovallo do not find that subjects make negative profits (in the setting that does not suffer from a self-selection problem, which is a separate issue that the authors identify)

prefer the test if the median of her beliefs about herself is better than the group median. If more than 50% of the people (strictly) prefer the test, then the population displays apparent overconfidence; this apparent overconfidence "indicates an erroneous evaluation" only if the data cannot be α-rationalized (given that people are using their median types). However, as Theorem 1 indicates, any amount of apparent overconfidence can be α-rationalized.[29] Thus, there is no error in judgement implied when more than half the subjects vote for the test.

In fact, even without the formalization in this paper, it is rather trivial to see that there is nothing wrong with more than half the subjects voting for the test. Imagine, for the sake of argument, 10 subjects who reach the following conclusion in the waiting room: Nine of them are native English speakers, with a perfectly ordinary command of the language, while one is a Haitian with a more recent knowledge of English. There is certainly nothing wrong with the nine native speakers voting for an English vocabulary test on the grounds that this gives each of them a $\frac{5}{9} > \frac{1}{2}$ chance of winning the prize. Thus, the experimental design of Hoelzl and Rustichini cannot possibly prove what they set out to prove.

The authors actually run two treatments of their experiment, one in which the vocabulary test is easy, one in which it is difficult. In both cases, subjects vote on the options before the test is administered, but are shown sample questions (either easy or difficult) before the vote. Hoelzl and Rustichini find that more subjects vote for the test option when the test is easy, than when it is difficult (64% vote for the test when it is easy, 39% vote for the test when it is difficult). They write "Two interpretations of this result are possible. The first interpretation is that subjects confuse 'being good' with 'being better'... A second interpretation is possible in terms of ambiguity aversion". In fact, a third interpretation is possible within a completely rational framework. In Section 4.2 we show that one should expect apparent overconfidence when the test is easy and apparent *underconfidence* when the test is difficult.[30] To understand this result in the present context, imagine that the difficult (on average) test is a Haitian Creole vocabulary test. Now we should expect nine of the subjects to vote against the test, on the grounds that it would give each of them a $\frac{4}{9} < \frac{1}{2}$ chance of winning.

[29] More precisely, Theorem 1 says that any degree of apparent overconfidence short of 100% can be α-rationalized. Given that any experiment must allow a margin for error, one could argue that even 100% apparent overconfidence is not problematic. Moreover, α-rationalization is (implicitly) for an infinite population; for a finite rational population 100% ostensible overconfidence can be obtained without experimental error.

[30] In fact, Propostion 3 can also be turned on its head to the same effect.

3 Testing for (true) Overconfidence

While the experiment run by Hoelzl and Rustichini does not provide a proper test of overconfidence, we now provide two modifications that do provide a proper test. The validity of both experiments derives from Theorem 7 below, which generalizes the necessary conditions of Theorem 3.

First Experiment

Suppose subjects are given the following two options:

1. Option 1. You will be given a monetary prize M with probability $\frac{1}{2}$, determined by the toss of a die.

2. Option 2. Everyone in your group will be administered a vocabulary test. You will be given M if your score places you in the top $x\%$ of your group.

A subject prefers the test condition if she thinks that there is more than a 50% chance that she will perform in the top $x\%$ of her group. That is, a person will prefer the test if her median type is in the top $x\%$. From Theorem 7, at most $2x\%$ of the people can rationally hold such a belief (at least if the population is large), so that a choice of x smaller than 50 by the experimenter provides a viable test. For instance if $x = 30$ so that a subject wins if he or she places in the top 30%, then at most 60% can rationally vote for the test.[31]

Second Experiment

Suppose that 10 subjects are given the option between winning M with probability $\frac{6}{10}$, and winning M if they place among the top 5 in a vocabulary test. As the following theorem indicates, at most 8 of the subjects can rationally prefer the test.

Theorem 7 *A fraction y or greater of a population can rationally believe that there is at least a probability q that their types are strictly better than the worst type in the top $x\%$ of the population if and only if $qy \leq x$.*

4 Apparent Overconfidence?

Recent work has questioned the universality of overconfidence. In particular, there is evidence that when the skill under consideration is objectively measurable, populations do not display much apparent overconfidence, and that they even display apparent underconfidence when the skill is a difficult one to master. In this section, we explain these facts within our rational framework.

[31] We gratuitously (since we have not run an experiment) remark that we doubt that more than 60% would vote for the test.

4.1 Objectivity

Moore (2007) writes "Attributes that are specific, public, and objectively measurable tend to show the weakest BTA [better-than-average] effects, whereas vague, private and subjective attributes tend to show the strongest BTA effects". It is reasonable to presume that people have more information and, hence, tighter estimates of their own abilities for attributes that are more objectively measurable. The following proposition indicates that when people are quite certain of their types, there cannot be much apparent overconfidence (or underconfidence).

For simplicity, suppose the type space Θ is finite. For any two probability distributions p and q on Θ, let the distance between them be

$$d(p,q) = \max_j |p_j - q_j| .$$

We say that a distribution q is $\delta-$**close to a degenerate** if there exists a $j \leq |\Theta|$ such that for e^j (the j^{th} canonical vector) $d(q, e^j) \leq \delta$.

Proposition 4 *Consider a signalling model in which the proportion of individuals strictly above the median of the population distribution is π. For all ε there is a δ and a t such that, if the expected fraction of the population who have beliefs (after receiving their signals) that are δ-close to a degenerate is at least t, then $|x_m - \pi| < \varepsilon$ and $|x_\mu - \pi| < \varepsilon$, where x_m (x_μ) is the expected population median data when people self-evaluate using their median type (mean type).*

This proposition is a "continuous" analogue of Proposition 1, which shows that population median data cannot be apparently overconfident when people's beliefs about themselves are degenerate.

4.2 Confidence and Difficulty

Kruger (1999) finds a "*below-average effect* in domains in which absolute skills tend to be low". Moore (2007), summarizing similar work, writes "When the task is difficult or success is rare, people believe that they are below average". Standard explanations for the confidence discrepancy between easy and difficult skills focus on cognitive errors, such as a tendency for people to "focus egocentrically on their own skills and insufficiently take into account the skills of the comparison group" (Kruger (1999)). In this section, we offer a rational explanation (of course).

Imagine a large pool of subjects who are informed that the level of "g-ability" in the population varies uniformly from 0 to 1. Since none of them has ever heard of g-ability, each

one (rationally) considers that his or her g-ability is equally likely to be anywhere from 0 to 1. Now suppose that they are given a test which measures g-ability. They are told that the test is an easy one: a person with ability t will pass with probability $0.7 + 0.2t$.

What will a person who passes the test conclude about his ability? He certainly will not be surprised that he passed, since he expected to pass the test regardless of his ability. Nonetheless, his confidence in his g-ability will increase, if only slightly. More precisely, using Bayes' rule, he will ascribe probability $\frac{0.7+0.2t}{.8} = 0.25t + 0.875$ to being of type t. His median type will be .53 and his mean type will be .52,[32] both better than the median (or mean) of the population. Thus, whether he ranks himself by his median type or mean type, he will consider himself to be better than average. Furthermore, his beliefs about himself will strictly first order stochastically dominate the prior. But, (about) 80% of population will pass the test, and so the population will exhibit apparent overconfidence. While the population is apparently overconfident, it is not overconfident.

Now let us change just one thing: The test is a difficult one, and everyone is so informed.[33] Specifically, the probability of passing the test is now $0.1+0.2t$, so that 80% of the people are expected to fail. Using Bayes' rule, those who fail will have a median type of 0.47, and mean type of 0.48, and the population distribution will strictly first order stochastically dominate their beliefs. Thus a population facing a difficult test will exhibit *apparent underconfidence*.

Theorem 8 below formalizes the above argument using a testing model (defined in Section 1.1.3). Let p be the prior that nature uses in the testing model and P be the associated cumulative distribution function. Then $E(\theta) = \int \theta dP$ is the average number of people who pass the test, and this number is naturally interpreted as the difficulty of the test. We say that a test is **easy** if $E(\theta) < \frac{1}{2}$ and **difficult** if $E(\theta) > \frac{1}{2}$. Let m be the median type.

Theorem 8 *Suppose that beliefs are generated by a symmetric testing model with a non-degenerate prior distribution p. Following the test, on average a fraction $E(\theta)$ will consider that their mean type is strictly better than the population median and will have beliefs about themselves that strictly f.o.s.d the population distribution; a fraction $1 - E(\theta)$ will consider that their mean type is strictly worse than the population median and will have beliefs about themselves that are strictly f.o.s.d by the population distribution. Furthermore, the fraction $E(\theta)$ will consider that their median type is weakly better than the population median. If p has a density, or $p(m)$ is small enough, the fraction $E(\theta)$ $(1 - E(\theta))$ will consider that their median type is strictly better (worse) than the population median. Thus, if the test is easy the population will exhibit apparent overconfidence, while if it is difficult, the population will*

[32] The median is the solution to $\int_0^t (0.25t + 0.875)\, dt = \frac{1}{2}$, while the mean is given by $\int_0^1 (0.25t + 0.875)\, tdt$.

[33] The crucial difference between our discussion here and the reasoning of Healy and Moore (2007), is that here everybody knows precisely the difficulty of the test they are taking. See Section 7 for a discusion of Healy and Moore (2007).

exhibit apparent underconfidence.

Thus, when a "task is difficult or success is rare" we can expect apparent underconfidence, and when a task is easy, we can expect apparent overconfidence.

5 Does it Matter?

In Section 2.1, we saw that apparently overconfident entrepreneurs might enter an industry at a greater rate than entrepreneurs with neutral beliefs about their abilities. The same, of course, is true about (truly) overconfident entrepreneurs. This invites the question: does it really matter if a population is apparently overconfident but not overconfident, or is the distinction merely semantic minutia? In this section, we argue that the distinction is important.

Recall that in the two player game analyzed in Section 2.1, under the random ranking treatment the equilibrium expected number of firms is 1 while under the test treatment – which leads to apparent overconfidence – the equilibrium expected number of firms is greater than 1. Crucially, however, the expected profits of each firm is zero *under both treatments* (as is easily verified). In contrast, consider what happens when firms are truly overconfident. For instance, suppose each firm irrationally believes there is a $\frac{3}{4}$ chance that it is the more skillful. Given these beliefs, entering is a dominant strategy for each firm, and in equilibrium each firm earns -1. Thus, while the presence of overconfidence requires us to rethink the basic economic tenet that firms will not enter an industry if there are negative profits to be made, the presence of apparent overconfidence does not.

Some authors use overconfidence as a springboard for assuming that agents have an irrational bias. For instance, in an influential finance paper, Malmendier and Tate (2005) write "Our overconfidence story builds upon a prominent stylized fact from the social psychology literature, the "better-than-average" effect. When individuals assess their relative skill, they tend to overstate their acumen relative to the average". Armed with this supposed evidence of irrationality, Malmendier and Tate go on to assume that the CEOs in their model overestimate the returns to their projects. However, if, as we have argued, the evidence is only that people are apparently overconfident, then the evidence is consistent with agents using all the information available to them in the best possible way, in which case it supplies no justification for assuming that agents make biased estimates. In particular, apparently overconfident CEOs who are not overconfident will, on average, have a correct estimate of the returns to their projects.

Should authorities regulate the behaviour of drivers by imposing speed limits, mandating seat belt use, etc..., rather than simply informing them of the risks? One argument in favour of such regulation would be that drivers have too much confidence in their abilities.

As Svenson writes, "Why should we pay much attention to information directed towards drivers in general if [we believe] we are safer and more skillful than they are?" But if drivers are only *apparently* overconfident, they may well pay attention. Note that in the driving example from the introduction, the $\frac{3}{5}$ of the population who rate themselves above average still believe that with a probability of $\frac{4}{9}$ they are no more skilled than average, so there is certainly no reason for them to simply ignore advice pitched at the general population.

6 Evidence on Our Approach

The basic principle driving our results is that people who receive good signals will tend to be confident, while those who receive bad signals will tend to be unconfident, regardless of the underlying probabilities of receiving the signals. Thus, the previously noted finding that people tend to be apparently underconfident on difficult tasks (where they fail relatively often) and apparently overconfident on easy tasks lends support to our approach. In a similar vein, recall that Hoelzl and Rustichini show people difficult sample vocablulary questions (which are likely to send a person the bad signal that he could not answer the question) and easy sample questions (which are likely to send a good signal). A sample question is only one signal among a lifetime of signals, so that it is unclear how a person should rank himself given this piece of information. Nonetheless, ceteris paribus, a positive signal should induce a higher ranking than a negative signal, and Hoelzl and Rustichini do find that people rank themselves higher following an easy sample question than a difficult one.

An accident is, presumably, a negative signal about a driver's ability. Preston and Harris (1965) ask drivers who have recently had accidents to rate themselves on a scale from 0 to 9. The mean of their self-ratings turns out to be "almost identical" to the mean self-rating of a control group that has not had any accidents (the actual data is not reported). This finding tends to go against our reasoning. However, it should be noted that the accident drivers were interviewed while still in the hospital, and many faced severe legal and financial consequences as a result of the accidents, so there is considerable reason to view their answers with skepticism. Indeed, only 15 drivers admitted responsibility for their accidents, while police reports blamed 34; given the circumstances, these 19 denials can hardly be deemed irrational or surprising. While Preston and Harris' evidence may be dubious, Marotolli and Richardson (1998) find similar results in a more neutral setting. They interview drivers (not selected for accident histories) in their homes, and still find no difference in the confidence levels of drivers who have had adverse driving events (according to their own reports) and the confidence levels of those who have not. Specifically, 68% of both groups rate themselves as being better than the average driver, and none rate themselves as being below average. On the other hand, in our favour Groenger and Grande (1996) find that drivers' self assessments

are positively correlated to the average number of accident-free miles they have driven. Importantly, they also find that these self-assessments are uncorrelated to the number of accidents the drivers have had. The number of accident-free miles would seem to be the more relevant statistic, so Groenger and Grande's finding may help explain Marotolli and Richardson's negative result.[34]

We interpreted Proposition 4 of Section 4.1 as showing that populations should display less apparent overconfidence or underconfidence with reference to objectively defined traits, as researchers have found to be the case. Technically, the proposition says that populations in which people have tight estimates of their own skill levels should not display much overconfidence or underconfidence. Bayesian updaters have tighter estimates of their own skill as they gather more information. This suggests that we should expect drivers with more experience to display less apparent overconfidence. In line with this prediction, Walton (1999) interviews professional truck drivers, who have considerable driving experience (each one drives approximately $100,000$ kilometers per year), and finds no bias in their self-assessments of their relative skills. Walton does find a majority of the truckers claiming to be safer than average.[35] However, it is quite possible that most of the truck drivers had not had an accident, and had only had safe drving experiences, so that a majority should rank themselves as safer than average.[36] Mathews and Moran (1986) question young drivers (18-25) and older drivers (35-50). They find that while the young drivers rate themselves as being better overall drivers than their peers, the older drivers rank themselves comparably to drivers in their age group. Similarly, Holland (1993) finds no self-bias for drivers in their fifties, or for drivers in their seventies either.[37] On the other hand, Marotolli and Richardson (1998) find a pronounced better than average effect in a sample of drivers 72 years and older, as does Cooper (1990) in a sample aged at least 55 years.

Thus, we find existing evidence in favour of our approach, although not all the evidence is in our favour.[38] We note that it would be wrong-headed to attempt a direct test of our model by, say, trying to determine whether or not people actually know the distribution of types in the population. That is at once only a simplifying assumption and one that makes it harder

[34] To push the point further, one might expect better drivers to drive more, raising their number of accidents, so that the number of accidents itself would be a poor indicator of ability.

[35] Walton (1999) asks the drivers many questions, but skill and safety are the only ranking questions.

[36] That is, the notion that experience should dissipate apparent overconfidence is too coarse. In the "short run", this need not be the case.

[37] Holland does find that drivers in their seventies consider themselves less prone to non-driving mishaps (e.g., losing an important document) than average. However, Holland speculates that in this regard the drivers may be comparing themselves to *all* seventy year olds, not just the driving subpopulation, so that there is no apparent problem.

[38] In line with the rest of our paper, we accept the evidence without examining the quality of the work (with the exception of the Preston and Harris paper).

to rationalize the data – we could easily relax the definition of a signalling model to one in which nature first picks one of several distributions, and people then receive signals about their own types and the population distribution, without altering our basic conclusions.

7 Literature

There is a vast literature on overconfidence, both testing for it and providing explanations for it. We have already mentioned some of the evidentiary literature. On the explanatory side, most of the literature accepts that there is something wrong when a majority of people believe they are above average, and either explains the phenomenon as resulting from a psychological "error", or finds a rational way around the data.

The first category of explanations include *egocentrism* (Kruger (1999)), *incompetence* (Kruger and Dunning (1999)) and *self-serving biases* (Greenwald (1980)). Bénabou and Tirole (2002) introduce a behavioral bias that causes people to become overconfident.

In the second category of explanations, Dunning et al. (1989) find that people may have varying notions of what, say, constitutes a good driver. If people are interpreting the question differently, there is not even an apparent contradiction when most people report themselves to be better than average (although there may be a self-seving bias in their interpretations). Van den Steen (2004) and Santos-Pinto and Sobel (2005) push this further and propose that as a result of these variances, people invest in skills in different ways. In the model of Santos-Pinto and Sobel, "Without the ability to add to skills... precisely p percent of the population would claim to be better than $1-p$ percent of the others." However, we have shown that there is no need for there to be this perfect calibration, even if everyone agrees on the evaluative criteria of skill and has no opportunity to invest in skills, provided only that people do not know their own skills exactly.

In Healy and Moore (2007), people take a test that may be either difficult or easy.[39] Each person is uncertain about his own ability, and about the difficulty of the test he is taking. A person who takes the easy test and does well, is uncertain if this is because he is of high ability or because the test is easy; hence he rationally assigns weight to both possibilities and considers himself to be above average. Since most people who take the easy test do well, those who take the easy test appear to be overconfident. By the same token, those who take the difficult test appear to be underconfident. More precisely, following the logic of Healy and Moore, averaging over the entire population of test takers (those who take the easy and the difficult test), one should find no apparent overconfidence or underconfidence,[40] but the

[39] Healy and Moore (2007) have several purposes to their paper. To compare their work to ours, we focus on the extent to which their theory can provide a completely rational explanation of the overconfidence data. However, that is not their primary focus. We encourage the reader to consult their paper.

[40] One could find apparent overconfidence if the subjects underestimate the fraction of tests that are

analyst makes a mistake by focussing on the groups separately. Again following their logic, if subjects understand the difficulty of their tasks, then their respective populations, even viewed in isolation, should not rate themselves above or below average. However, we have shown that there may be apparent overconfidence and underconfidence even in subjects who understand their environments perfectly. As noted in Section 2.2, Hoelzl and Rustichini find apparent overconfidence when their subjects are given an easy test and underconfidence when they are given a difficult test. In their experiment subjects are told which test they are taking (and it is obvious), so that Healy and Moore's analysis does not apply.[41] Similarly, Kruger's (1999) subjects consider easy tasks (e.g., riding a bicycle) and difficult tasks (e.g., juggling) that are clearly delineated. Nonetheless, there are doubtless areas where it is hard to judge difficulty and, even though that additional assumption is not necessary to generate apparent overconfidence, it may yield greater overconfidence.

While the above literature tries to explain how anomalous data can arise, we show that the data is, in fact, not anomalous at all. In this regard, Zábojnik (2004) is more in the spirit of our paper, although his approach is very different. In his model, agents can choose to forego consumption in order to learn about their abilities, which may be either high or low. Given certain technical assumptions – in particular $U(a_t)$, an individual's utility as a function of his ability, is strictly convex, and $EU(a_t)/U(a_t)$ is decreasing – the optimal learning rule of agents leads them to halt their learning in a biased fashion. As a result, a disproportionate number end up ranking themselves as high in ability, despite the fact that they are all rational Bayesians. Hence, like us, Zábojnik finds that the mere fact that a disproportionate number of people rank themselves in the top $x\%$ does not indicate a problem. However, as Zábojnik readily admits, he is not really concerned with explaining much of the experimental evidence, where his experimentation story is not very compelling. Furthermore, he requires technical assumptions which play no role in our work. For instance, we show that apparent overconfidence can rationally arise whether or not an agent's utility is convex. Beyond this, Zábojnik does not consider the possibility that people use their median types to self-evaluate, and has a model with only two types of agents and two signals.

In a tangentially related paper, Ledyard (1986) argues that Bayesian rationality imposes almost no restrictions on an agent's behaviour. Specifically, only the use of a dominated strategy can be ruled out. Our paper differs in many respects. First of all, at a purely technical level, the two sets of results are unrelated. Second of all, we do obtain important

easy. However, as previously noted, we are interested in the extent to which Healy and Moore provide an explanation that does not rely on errors.

[41] More precisely, the analysis of Healy and Moore does not apply if subjects know which test is easy and have a good idea of just how easy and difficult the tests are. There is evidence that the subjects did understand the degree of difficulty: Hoezl and Rustichini ask them to guess the mean scores of the tests and, for the tests they actually take, their estimates are remarkably close to the actual results.

restrictions, as shown by Theorems 3, 6, and 7, so that the underlying messages of Ledyard's paper and ours are quite different. Finally, as Section 1.1.3 makes clear, our results do not depend upon pushing the Bayesian assumption to its "limits".

8 Conclusion

Consider two populations. In population A, about 50% of the people declare themselves to be better than the median. In population B, about 80% of the people make the same declaration. On the face of it, population A has good self knowledge, while population B is somewhat delusional. Nevertheless, we have shown that unless agents have a very good idea of their own skills, the declaration of population B is no more aberrational than the declaration of population A. Put differently, there is no more reason to try and "justify" B's data than A's; both sets of data can naturally arise as the result of agents receiving information about their skills.

Is there a tendency for people to be overconfident? We do not know, and have no stake in the answer one way or the other. Psychologists do have theories independent of the "better-than-average" evidence we have cited for believing in overconfidence. However, it is imperative to have a clear statement of the problem (which we hope to have provided), and to conduct proper empirical tests (such as the ones suggested in Section 3). As it stands, much of the supposed evidence for overconfidence shows nothing of the sort, even if the evidence is taken completely on its own terms.

At the most abstract level, our reasoning makes it no more likely that a population will rate itself above average than below average. Thus, if populations consistently rate themselves as being better than average, a puzzle remains. However, recent evidence has shown that the supposed universality of overconfidence has been overstated. In particular, populations appear to be underconfident with respect to difficult skills, and there is little apparent overconfidence, or underconfidence, when the skill is objectively defined. Our theory predicts these differences. Moreover, even if it were to turn out that the relative number of studies pointing to overconfidence rather than underconfidence cannot be fully accounted for by an easy/difficult skill distinction, it is important to emphasize that the statement, "there are many studies, which taken together *as a whole* indicate that people are overconfident" is a very different, and weaker, statement than "there are many studies, *each of which individually* indicate that people are overconfident" .

We emphasize that our goal is not so much to provide a concrete explanation for the better-than-average data as much as it is to show that the data has been misinterpreted. Thus, we have argued that *even* making the extreme, and possibly unrealistic, assumptions that everybody knows the distribution of types in the population exactly, that they use

Bayes' rule perfectly, and that the subject pool is arbitrarily large, apparently overconfident data can be generated. There is no need for recourse to errors of any kind. At the same time, in keeping with our goal, we have accepted the experimental as given, although criticisms could be made.[42]

Some psychologists will reject our approach out of hand, on the prior grounds that individuals do not use Bayes' rule and, for that matter, may not even understand simple probability. Even for these researchers, however, the basic challenge of this paper remains: To indicate why a finding that a majority of people rank themselves above the median is indicative of *overconfidence*. If such a finding does not show overconfidence in a Bayes' rational population, there can be no presumption that it indicates overconfidence in a less rational population.

9 Appendix

Proof of Proposition 1. We present the proof for the case in which the type space and the signal space are finite; the proof when either is infinite is technically more involved, but the essential reasoning is the same.

Let p describe the distribution of types in the population. Suppose that person i is of type $\hat{\theta}^i$, $p\left(\hat{\theta}^i\right) > 0$, and that i has received the signal s_i. Since the posterior of i is degenerate, there is a $\widetilde{\theta}^i$ such that

$$\begin{aligned} \Pr\left(\widetilde{\theta}^i \mid s_i\right) &= 1 \Leftrightarrow \frac{\Pr\left(s_i \mid \widetilde{\theta}^i\right)\Pr\left(\widetilde{\theta}^i\right)}{\sum_{i=1}^{n}\Pr\left(s_i \mid \theta^i\right)\Pr\left(\theta^i\right)} = 1 \\ &\Leftrightarrow \forall \theta \neq \widetilde{\theta}^i \, s.t. \; p(\theta) > 0, \; \Pr\left(s_i \mid \theta^i\right) = 0. \end{aligned}$$

Since only type $\widetilde{\theta}^i$ receives the signal s_i with strictly positive probability, we have $\hat{\theta}^i = \widetilde{\theta}^i$. Thus, if beliefs are generated by a signalling model and these beliefs are degenerate, they are correct, and it cannot be that more than half the population believes it is better than the median. ■

Proof of Theorem 1. Fix the population median data that is to be rationalized, $1 > x > 1/2$ and fix $y < \frac{1}{2} - \frac{x}{1+x}$. We now define the elements of a symmetric signalling model that will α, β, and γ rationalize the population data x. The signalling structure consists of the signal space $S = \{0, 1\}$, the type space $\Theta = \left\{\frac{3x-1}{2}, x, \frac{1+x}{2}\right\}$, and the probability

[42] For instance, most researchers fail to determine whether subjects are comparing themselves to the population median or mean. Also, the ecological validity of many experiments in unclear. In a slightly different context Gigerenzer et al. (1997) write "there is apparently not a single study on confidence in knowledge where a reference class has been defined and a representative (or random) sample of general-knowledge questions has been drawn from this population."

distributions $f_\theta(1) = \theta$. The types are distributed in the population according to the prior p, which assigns probability $2y$ to x, and $\frac{1}{2} - y$ to each of $\frac{3x-1}{2}$ and $\frac{1+x}{2}$. The expected fraction of people who observe $s = 1$ is x. A person that observes 1 has a posterior belief given by

$$p(\theta \mid 1) = \begin{cases} \frac{1+x}{2x}\left(\frac{1}{2} - y\right) & \theta = \frac{1+x}{2} \\ 2y & \theta = x \\ \frac{3x-1}{2x}\left(\frac{1}{2} - y\right) & \theta = \frac{3x-1}{2} \end{cases}.$$

We now check that this posterior distribution (strictly) first order stochastically dominates the prior:

$$\begin{aligned} p\left(\frac{1+x}{2} \mid 1\right) &= \frac{1+x}{2x}\left(\frac{1}{2} - y\right) > \frac{1}{2} - y = p\left(\frac{1+x}{2}\right) \\ p(x \mid 1) &= 2y = p(x) \\ p\left(\frac{3x-1}{2} \mid 1\right) &= \frac{3x-1}{2x}\left(\frac{1}{2} - y\right) < \frac{1}{2} - y = p\left(\frac{3x-1}{2}\right) \end{aligned}$$

Thus, the population ranking data x can be γ-rationalized.

To see that the population ranking data x can be β-rationalized, notice that the posterior mean of a person who observes 1 is strictly larger than the prior mean, x, which is also the prior median.

To see that the population ranking data x can be α-rationalized, notice that the posterior median of a person that observes 1 is $\frac{(1+x)}{2}$, since

$$p\left(\frac{1+x}{2} \mid 1\right) > \frac{1}{2} \Leftrightarrow \frac{1+x}{2x}\left(\frac{1}{2} - y\right) > \frac{1}{2} \Leftrightarrow \frac{1}{2} - \frac{x}{1+x} > y$$

by construction. Furthermore, $\frac{(1+x)}{2} > x$. ■

Proof of Theorem 2. Pick any $1 > x, q > 1/2$. Let the signalling structure be given by $S = \{0, 1\}$, $\Theta = \left\{\frac{3x-1}{2}, x, \frac{1+x}{2}\right\}$, and $f_\theta(1) = \theta$. On Θ, define a probability distribution p that assigns probability $2q - 1$ to x, and $1 - q$ to each of $\frac{3x-1}{2}$ and $\frac{1+x}{2}$. Notice that the fraction of the population whose type is at most x is q.

An observation of 1 occurs with probability x. A person that observes 1 has a posterior belief given by

$$p(\theta \mid 1) = \begin{cases} \frac{1+x}{2x}(1 - q) & \theta = \frac{1+x}{2} \\ 2q - 1 & \theta = x \\ \frac{3x-1}{2x}(1 - q) & \theta = \frac{3x-1}{2} \end{cases}$$

and an expected type that is strictly larger than x. Therefore, the expected fraction of people who will have a mean type strictly greater than the fraction q of the population is x.

The case $0 < x, q < \frac{1}{2}$ is symmetrical, while the remaining cases are easy. ■

We prove necessity in Theorem 3 with a lemma that will be used elsewhere.

Lemma 1 *A fraction y or greater of a population can rationally believe that there is at least a probability q that their types are: strictly better than the worse type in the top $x\%$ of the population only if $qy \leq x$; weakly better than a type θ^* such that $p(\theta \geq \theta^*) = x$ only if $qy \leq x$; weakly worse than a type θ_* such that $p(\theta \leq \theta_*) = x$ only if $qy \leq x$*

Proof. For the first claim, note that given p, the worse type in the top $x\%$ is $\theta^* = \min\{z : p(\theta \leq z) \geq 1 - x\}$, which implies $p(\theta > \theta^*) \leq x$. Let $S_w \subset S$ be the set of signals s such that $p(\theta > \theta^* \mid s) \geq q$, and let F denote the marginal distribution over signals so that $y = F(S_w)$. Then,

$$x \geq p(\theta > \theta^*) = \int p(\theta > \theta^* \mid s)\, dF(s) \geq \int_{S_w} p(\theta > \theta^* \mid s)\, dF(s) \geq \int_{S_w} q dF(s) = qy.$$

For the second claim, let $S_w^* \subset S$ be the set of signals s such that $p(\theta \geq \theta^* \mid s) \geq q$, and let F denote the marginal distribution over signals so that $y = F(S_w^*)$. Then,

$$x = p(\theta \geq \theta^*) = \int p(\theta \geq \theta^* \mid s)\, dF(s) \geq \int_{S_w^*} p(\theta \geq \theta^* \mid s)\, dF(s) \geq \int_{S_w^*} q dF(s) = qy.$$

The third claim is analogous to the second, and omitted. ■

We prove Theorem 3 using the following lemma, which we prove below.

Lemma 2 *Let Δ^k be the simplex in $\mathbf{R}^k$: $\Delta^k = \left\{x \in \mathbf{R}^k : x_i \in [0,1], i = 1, ..., k, \sum x_i = 1\right\}$. Fix $x \in \Delta^k$ and suppose that*

$$\begin{aligned} \sum_{j=i}^{k} x_j &\leq \frac{2}{k}(1 + k - i),\ i = \left\lceil \frac{k+1}{2} \right\rceil, ..., k \qquad \text{and} \\ \sum_{j=1}^{i} x_j &\leq \frac{2}{k} i,\ i = 1, ..., \left\lceil \frac{k-1}{2} \right\rceil \end{aligned}$$

Then there exists a $k \times k$ matrix P, such that $xP = \left(\frac{1}{k}, ..., \frac{1}{k}\right)$, and for p^i the i^{th} row of P, $p^i \in \Delta^k$, $\sum_{j=1}^{i} p_j^i \geq \frac{1}{2}$ and $\sum_{j=i}^{k} p_j^i \geq \frac{1}{2}$.

Proof of Theorem 3. Necessity. To prove necessity, simply apply claims 2 and 3 of Lemma 1 with $q = 1/2$ and $x = i/k$ for $i = 1, 2, .., k$.

Sufficiency. Let P, p^i be as in Lemma 2. Consider the following simple signalling structure, $S = \{1, 2, ..., k\}$, $\Theta = [0, k]$, the distribution of the population is uniform over the type space, and types in $k-$cile j observe signal i with probability $f_j(i) = k p_j^i x_i$, for $i, j = 1, ..., k$.

Note the following:

1)

$$\Pr\left(s^i\right) = \textstyle\sum_j \Pr\left(s^i \mid k - \text{cile } j\right) \Pr(j) = \sum_j k p_j^i x_i \frac{1}{k} = \sum_j p_j^i x_i = x_i \sum_j p_j^i = x_i,$$

so that a fraction x_i of the population see the signal j.

2)

$$\Pr\left(\theta \in k-\text{cile } j \mid s=i\right)=\frac{\Pr\left(s^i \mid j\right)\Pr\left(j\right)}{\Pr\left(s^i\right)}=\frac{kp_j^i x_i \frac{1}{k}}{x_i}=p_j^i,$$

so that an individual who sees the signal i, ascribes the probability p_j^i to being in $k-$cile j.

3) Since $\sum_{j=1}^{i} P_{ij}=\sum_{j=1}^{i} p_j^i \geq \frac{1}{2}$ and $\sum_{j=i}^{k} P_{ij}=\sum_{j=i}^{k} p_j^i \geq \frac{1}{2}$, the median type of an individual that observes signal i is in $k-$cile i.

1), 2), and 3) together imply the sufficiency part of the Theorem. ■

Proof of Lemma 2. Fix an $x \in \Delta^k$. We say that $r \in \Delta^k$ can be justified if there exists a $k \times k$ matrix Q, such that $xQ=r$, and for q^i, the i^{th} row of Q, $q^i \in \Delta^k$, $\sum_{j=1}^{i} Q_{ij} \geq \frac{1}{2}$ and $\sum_{j=i}^{k} Q_{ij} \geq \frac{1}{2}$. Let $\mathcal{R}$ be the set of distributions that can be justified. Note that $\mathcal{R}$ is non-empty, since x itself can be justified by the identity matrix. Furthermore, $\mathcal{R}$ is closed and convex. We need to show that $\left(\frac{1}{k},...,\frac{1}{k}\right) \in \mathcal{R}$.

Suppose not. Then, since $f\left(t\right)=\left\|t-\left(\frac{1}{k},...,\frac{1}{k}\right)\right\|^2$ is a strictly convex function, there is a unique $\left(\frac{1}{k},...,\frac{1}{k}\right) \neq r=\arg\min_{t\in\mathcal{R}} f\left(t\right)$. Let Q be a matrix that justifies r.

Since $r \neq \left(\frac{1}{k},...,\frac{1}{k}\right)$ there exists some $r_i \neq \frac{1}{k}$, and since $r \in \Delta^k$, there must be some i for which $r_i > \frac{1}{k}$, and some i for which $r_i < \frac{1}{k}$. Let $i^*=\max\left\{i : r_i \neq \frac{1}{k}\right\}$ and $i_*=\min\left\{i : r_i \neq \frac{1}{k}\right\}$.

Step 1: We have $r_{i^*}, r_{i_*} < \frac{1}{k}$.

Suppose instead that $r_{i^*} > \frac{1}{k}$ (a similar argument establishes that $r_{i_*} < \frac{1}{k}$). Then, for all $j > i^*$, $r_j=\frac{1}{k}$ and for some $i < i^*$, $r_i < \frac{1}{k}$. Let $\widetilde{i}=\max\left\{i : r_i < \frac{1}{k}\right\}$. We show that for all $i > \widetilde{i}$ (a) for any j such that $j \leq \widetilde{i}$ or $j > i$, either $x_j=0$ or $Q_{ji}=0$; (b) either $x_i=0$ or $\sum_{j=i}^{k} Q_{ij}=\frac{1}{2}$.

To see (a) fix an $i' > \widetilde{i}$ and suppose $x_{j'} > 0$ and $Q_{j'i'} > 0$ for some $j' \leq \widetilde{i}$ or $j' > i'$. Define the matrix $\widetilde{Q}$ by $\widetilde{Q}_{j'\widetilde{i}}=Q_{j'\widetilde{i}}+\varepsilon Q_{j'i'}$, $\widetilde{Q}_{j'i'}=\left(1-\varepsilon\right)Q_{j'i'}$, and for all $(f,g) \notin \left\{\left(j',i'\right),\left(j',\widetilde{i}\right)\right\}$, $\widetilde{Q}_{fg}=Q_{fg}$. We have

$$\left.\begin{array}{l}\text{For } i \neq j',\ \sum_{j=1}^{i}\tilde{Q}_{ij}=\sum_{j=1}^{i}Q_{ij} \geq \frac{1}{2} \text{ and } \sum_{j=i}^{k}\tilde{Q}_{ij}=\sum_{j=i}^{k}Q_{ij} \geq \frac{1}{2} \\ \text{If } j' \leq \widetilde{i},\ \sum_{j=1}^{j'}\tilde{Q}_{j'j}=\sum_{j=1}^{j'}Q_{ij} \geq \frac{1}{2} \text{ and } \sum_{j=j'}^{k}\tilde{Q}_{ij}=\sum_{j=i}^{k}Q_{ij}+\varepsilon Q_{j'i'}-\varepsilon Q_{j'i'} \geq \frac{1}{2} \\ \text{If } i' < j',\ \sum_{j=1}^{j'}\tilde{Q}_{j'j}=\sum_{j=1}^{j'}Q_{ij}+\varepsilon Q_{j'i'}-\varepsilon Q_{j'i'} \geq \frac{1}{2} \text{ and } \sum_{j=j'}^{k}\tilde{Q}_{ij}=\sum_{j=i}^{k}Q_{ij} \geq \frac{1}{2}\end{array}\right\}(i)$$

For ε sufficiently small, define $\widetilde{r}=x\widetilde{Q}$.

We have $\widetilde{r}_{\widetilde{i}}=r_{\widetilde{i}}+x_{j'}\varepsilon Q_{j'i'}$, $\widetilde{r}_{i'}=r_{i'}-x_{j'}\varepsilon Q_{j'i'}$, and for $i \notin \left\{i',\widetilde{i}\right\}$, $\widetilde{r}_i=r_i$. Therefore $\sum_{i=1}^{k}\widetilde{r}_i=\sum_{j=1}^{k}r_i=1$. For small enough ε, $1 \geq \widetilde{r}_{i'} \geq 0$ for all i, since $x_{j'}, Q_{j'i'} > 0$ implies that $r_{i'} > 0$. Hence $\widetilde{r} \in \Delta^k$ and, given (i), $\widetilde{r} \in \mathcal{R}$.

We now show that $f(\tilde{r}) < f(r)$.

$$\begin{aligned} f(\tilde{r}) - f(r) &= \left(r_{\widetilde{i}} + x_{j'}\varepsilon Q_{j'i'} - \frac{1}{k}\right)^2 - \left(r_{\widetilde{i}} - \frac{1}{k}\right)^2 + \left(r_{i'} - x_{j'}\varepsilon Q_{j'i'} - \frac{1}{k}\right)^2 - \left(r_{i'} - \frac{1}{k}\right)^2 \\ &= (x_{j'}\varepsilon Q_{j'i'})^2 + 2x_{j'}\varepsilon Q_{j'i'}\left(r_{\widetilde{i}} - \frac{1}{k}\right) + (x_{j'}\varepsilon Q_{j'i'})^2 - 2x_{j'}\varepsilon Q_{j'i'}\left(r_{i'} - \frac{1}{k}\right) \\ &= 2(x_{j'}\varepsilon Q_{j'i'})\left[x_{j'}\varepsilon Q_{j'i'} + r_{\widetilde{i}} - \frac{1}{k} - r_{i'} + \frac{1}{k}\right] = 2(x_{j'}\varepsilon Q_{j'i'})\left[x_{j'}\varepsilon Q_{j'i'} + r_{\widetilde{i}} - r_{i'}\right] \end{aligned}$$

Recall that $r_{\tilde{\imath}} < \frac{1}{k}$, and since $i' > \widetilde{i}$, $r_{i'} \geq \frac{1}{k}$. Hence, for ε sufficiently small, $[x_{j'}\varepsilon Q_{j'i'} + r_{\widetilde{i}} - r_{i'}] < 0$. We have a contradiction, since, by definition $r = \arg\min_{t\in\mathcal{R}} f(t)$.

To see (b), suppose that for some $i' > \widetilde{i}$ we have $x'_{i'} > 0$ and $\sum_{j=i'}^{k} Q_{i'j} > \frac{1}{2}$. Pick some $j' \geq i'$ with $Q_{i'j'} > 0$. For ε sufficiently small, define $\widetilde{Q}$ by $\widetilde{Q}_{i\widetilde{i}} = Q_{i\widetilde{i}} + \varepsilon Q_{i'j'}$, $\widetilde{Q}_{i'j'} = (1-\varepsilon)Q_{i'j'}$, and for all $(f,g) \notin \left\{(i',j'), \left(i',\widetilde{i}\right)\right\}$, $\widetilde{Q}_{fg} = Q_{fg}$. Define $\widetilde{r} = x\widetilde{Q}$. As before, $\widetilde{r} \in \mathcal{R}$ and $f(\tilde{r}) < f(r)$, a contradiction.

Given (a) and (b), and recalling the defintion of $\widetilde{i}$, we have

$$\begin{aligned} \frac{k-\widetilde{i}}{k} &< \sum_{t=\widetilde{i}+1}^{k} r_t = \sum_{t=\widetilde{i}+1}^{k}\sum_{j=1}^{k} x_j Q_{jt} \\ &= \sum_{t=\widetilde{i}+1}^{k}\sum_{j=\widetilde{i}+1}^{k} x_j Q_{jt} \text{ (by (a), } j \leq \tilde{\imath} \text{ implies } x_j = 0, \text{ or } Q_{jt} = 0) \\ &= \sum_{j=\widetilde{i}+1}^{k} x_j \sum_{t=\widetilde{i}+1}^{k} Q_{jt} = \sum_{j=\widetilde{i}+1}^{k} x_j \sum_{t=j}^{k} Q_{jt} \text{ (by (a), } j > t > \tilde{\imath} \text{ implies } x_j = 0, \text{ or } Q_{jt} = 0) \\ &= \sum_{j=\widetilde{i}+1}^{k} \frac{x_j}{2} \text{ (by (b) either } x_j = 0 \text{ or } \sum_{t=j}^{k} Q_{jt} = \frac{1}{2}) \\ &\leq \frac{k-\widetilde{i}}{k} \text{ (by assumption of the Lemma)} \end{aligned}$$

Thus we have a contradiction.

Step 2: From Step 1 there exists an $\widehat{i}$, $i_* < \widehat{i} < i^*$, such that $r_{\widehat{i}} > \frac{1}{k}$. Since $r_{\widehat{i}} = \sum_{j=1}^{k} x_j Q_{j\widehat{i}}$, for some j^* $Q_{j^*\widehat{i}} > 0$. We now show that this leads to a contradiction.

Consider a small enough ε.

If $j^* < \widehat{i}$, define $\widetilde{Q}$ by $\widetilde{Q}_{j^*\widehat{i}} = (1-\varepsilon)Q_{j^*\widehat{i}}$, $\widetilde{Q}_{j^*i^*} = Q_{j^*i^*} + \varepsilon Q_{j^*\widehat{i}}$, and for all $(j,h) \notin \left\{\left(j^*,\widehat{i}\right),(j^*,i^*)\right\}$, $\widetilde{Q}_{jh} = Q_{jh}$. If $j^* > \widehat{i}$, define $\widetilde{Q}$ by:$\widetilde{Q}_{j^*i_*} = Q_{j^*i_*} + \varepsilon Q_{j^*\widehat{i}}$, $\widetilde{Q}_{j^*\widehat{i}} = (1-\varepsilon)Q_{j^*\widehat{i}}$, and for all $(j,h) \notin \left\{\left(j^*,\widehat{i}\right),(j^*,i_*)\right\}$, $\widetilde{Q}_{jh} = Q$. In either case, for all $i \neq j^*$, $\tilde{Q}_{ij} = Q_{ij}$ so $\sum_{j=1}^{i}\tilde{Q}_{ij} \geq \frac{1}{2}$ and $\sum_{j=i}^{k}\tilde{Q}_{ij} \geq \frac{1}{2}$; for $i = j^*$ if $j^* < \widehat{i}$, $\sum_{j=1}^{i}\tilde{Q}_{ij} = \sum_{j=1}^{i} Q_{ij} \geq \frac{1}{2}$ and $\sum_{j=i}^{k}\tilde{Q}_{ij} = \sum_{j=i}^{k} Q_{ij} - \varepsilon Q_{j^*\widehat{i}} + \varepsilon Q_{j^*\widehat{i}} \geq \frac{1}{2}$; for $i = j^*$ if $j^* > \widehat{i}$, $\sum_{j=1}^{i}\tilde{Q}_{ij} = \sum_{j=1}^{i} Q_{ij} -$

$\varepsilon Q_{j^*\widehat{i}} + \varepsilon Q_{j^*\widehat{i}} \geq \frac{1}{2}$ and $\sum_{j=i}^{k} \tilde{Q}_{ij} = \sum_{j=i}^{k} Q_{ij} \geq \frac{1}{2}$. For $\tilde{r} = x\widetilde{Q}$ it is easy to show (as in Step 1) that $f(\widetilde{r}) < f(r)$ – a contradiction.

If $j^* = \widehat{i}$, $Q_{j^*j^*} > 0$ and so $\sum_{j=1}^{j^*} Q_{j^*j} > 1/2$ or $\sum_{j=j^*}^{k} Q_{j^*j} > 1/2$. Suppose that $\sum_{j=1}^{j^*} Q_{j^*j} > 1/2$ (an analogous argument can be made if $\sum_{j=j^*}^{k} Q_{j^*j} > 1/2$). Define $\widetilde{Q}$ by $\widetilde{Q}_{j^*i^*} = Q_{j^*i^*} + \varepsilon Q_{j^*j^*}$, $\widetilde{Q}_{j^*j^*} = (1-\varepsilon) Q_{j^*j^*}$, and for all $(j,h) \notin \{(j^*,j^*), (j^*,i^*)\}$ $\widetilde{Q}_{jh} = Q_{jh}$. One can then verify that for small enough ε, for all i, $\sum_{j=1}^{i} \tilde{Q}_{ij} \geq \frac{1}{2}$ and $\sum_{j=i}^{k} \tilde{Q}_{ij} \geq \frac{1}{2}$. Defining $\tilde{r} = x\widetilde{Q}$, we obtain $f(\widetilde{r}) < f(r)$ – a contradiction.

This concludes the proof. ■

Proof of Theorem 4. Let $J = \{j \in \{1,...,k\} | x_j > 0\}$ and write $J = \{s_1, ...s_r\}$. We define a signalling model with n people, where $n = ak$ for $a \in \mathbf{N}$, and $a > 3k$, individual i is of type θ_i, and the set of signals is $\{s_1, ...s_r\}$. Define $f_{\theta_{n-r+j}}(s_j) = 1$ for $j = 1,..,r$. For $i = 1,...,n-r$, $j = 1,2,...,r$ define $f_{\theta_i}(s_j)$ so that $\frac{1}{n}\sum_{i=1}^{n} f_{\theta_i}(s_j) = x_{s_j}$. We now describe the types θ_i in greater detail. First we fix any numbers θ_i such that θ_i is strictly increasing in i for $i = 1,...,n-r$ and $\frac{n-1}{n}\theta_{n-r} < \theta_1$. Considering only individuals $1,...,n-k$, for $i = 1,...,k$, let m_i be a median type of k-cile i (the kth k-cile is r members short of $\frac{n}{k}$).

Note that for $j = 1,...,r$, $p(\theta = \theta_{n-r+j} | s = s_j) = \frac{\frac{1}{n}}{x_{s_j}} \geq \frac{1}{n}$. Therefore, $\sum_{i=1}^{n-r} p(\theta = \theta_i | s = s_j)\theta_i \leq \sum_{i=1}^{n-r} p(\theta = \theta_i | s = s_j)\theta_{n-r} = \theta_{n-r}\sum_{i=1}^{n-r} p(\theta = \theta_i | s = j) \leq \theta_{n-r}\frac{n-1}{n} < \theta_1 < m_{s_j}$. Set θ_{n-r+j} so that $E(\theta | s = s_j) = \sum_{i=1}^{n} p(\theta = \theta_i | s = s_j)\theta_i = \sum_{i=1}^{n-r} p(\theta = \theta_i | s = s_j)\theta_i + p(\theta = \theta_{n-r+j} | s = s_j)\theta_{n-r+j} \approx m_{s_j}$ and $\theta_{n-r+j} \neq \theta_i$ for all $i \neq n-r+j$. Note that the types which form each k-cile may have now shifted slightly, since we may have $\theta_{n-r+j} < \theta_{n-r}$ for some j. However, since $n > 3k^2$, the number of types in each $k-$cile is larger than $3k$, so that m_{s_j} is still in k-cile s_j.

A person who sees signal s_j has mean type $E(\theta | s = s_j) \approx m_{s_j}$. Therefore, a person who sees signal s_j has mean type in the jth $k-$cile. Furthermore, a fraction x_{s_j} of the population sees signal s_j, and we are done. ■

Proof of Theorem 5. The proofs of Theorems 1 and 2 used symmetric testing models, in which a signal of 1 is "pass the test". Therefore, those proofs also prove this theorem. ■

Proof of Proposition 2. Suppose data $\left(\varepsilon, \varepsilon, \frac{1}{2}-\varepsilon, \frac{1}{2}-\varepsilon\right) = (x_1, x_2, x_3, 1-x_2-x_3)$ can be α-rationalized by a signalling structure $\sigma = (S, \Theta, f)$ that satisfies mlrp. It can be shown that it is w.l.o.g. to assume that $S = (1,2,3,4)$ and that a person who sees signal $i = 1,2,3,4$ declares himself to be of type i. Since σ satisfies mlrp, $f_1(4) \leq f_2(4) \leq f_3(4)$. Since the signalling structure rationalizes the data, the expected number of people who observe signal j is x_j. Hence

$$\sum_{i=1}^{4} f_i(1) = 4x_1, \quad \sum_{i=1}^{4} f_i(2) = 4x_2 \text{ and } \sum_{i=1}^{4} f_i(3) = 4x_3$$

Since those who observe signal $s = 4$ believe their median type is in quartile 4, we have

$$\begin{aligned}
\frac{1 - f_4(1) - f_4(2) - f_4(3)}{4 - \sum_{i=1}^{4} f_i(1) - \sum_{i=1}^{4} f_i(2) - \sum_{i=1}^{4} f_i(3)} &\geq \frac{1}{2} \\
\sum_{i=1}^{4} f_i(1) + \sum_{i=1}^{4} f_i(2) + \sum_{i=1}^{4} f_i(3) &\geq 2 + 2f_4(1) + 2f_4(2) + 2f_4(3) \qquad (1)
\end{aligned}$$

Since those who observe signal $s = 3$ believe their median type is in quartile 3, we have

$$\begin{aligned}
\frac{f_3(3) + f_4(3)}{\sum_{i=1}^{4} f_i(3)} &\geq \frac{1}{2} \\
2f_3(3) + 2f_4(3) &\geq \sum_{i=1}^{4} f_i(3) \qquad (2) \\
f_3(3) + f_4(3) &\geq f_1(3) + f_2(3) \qquad (3)
\end{aligned}$$

Inequalities (1) and (2) together imply

$$\begin{aligned}
2f_3(3) + 2f_4(3) &\geq \sum_{i=1}^{4} f_i(3) \geq 2 + 2f_4(1) + 2f_4(2) + 2f_4(3) - \sum_{i=1}^{4} f_i(1) - \sum_{i=1}^{4} f_i(2) \\
f_3(3) &\geq \frac{2 + f_4(1) + f_4(2) - \sum_{i=1}^{3} f_i(1) - \sum_{i=1}^{3} f_i(2)}{2}.
\end{aligned}$$

Thus

$$\begin{aligned}
f_3(4) &= 1 - f_3(1) - f_3(2) - f_3(3) \\
&\leq 1 - f_3(1) - f_3(2) - \frac{2 + f_4(1) + f_4(2) - \sum_{i=1}^{3} f_i(1) - \sum_{i=1}^{3} f_i(2)}{2} \\
&= \frac{f_1(1) + f_2(1) - f_3(1) - f_4(1) + f_1(2) + f_2(2) - f_3(2) - f_4(2)}{2}
\end{aligned}$$

Given that $f_1(4) \leq f_2(4) \leq f_3(4)$, we have

$$f_1(4) \leq f_2(4) \leq f_3(4) \leq \frac{f_1(1) + f_2(1) - f_3(1) - f_4(1) + f_1(2) + f_2(2) - f_3(2) - f_4(2)}{2} \qquad (4)$$

From 4, and since $\sum_{j=1}^{4} f_i(j) = 1$ for any i,

$$\begin{aligned}
f_1(3) + f_2(3) &= 2 - (f_1(1) + f_2(1)) - (f_1(2) + f_2(2)) - (f_1(4) + f_2(4)) \\
&\geq 2 - (f_1(1) + f_2(1)) - (f_1(2) + f_2(2)) \\
&\quad - (f_1(1) + f_2(1) - f_3(1) - f_4(1) + f_1(2) + f_2(2) - f_3(2) - f_4(2)) \\
&= 2 - 2(f_1(1) + f_2(1)) - 2(f_1(2) + f_2(2)) + f_3(1) + f_4(1) + f_3(2) + f_4(2) \quad (*)
\end{aligned}$$

Since $\sum_{i=1}^{4} f_i(j) = 4x_j$ for any j we have that

$$\begin{aligned}
(*) &= 2 - 2(4x_1 - f_3(1) - f_4(1)) - 2(4x_2 - f_3(2) - f_4(2)) + f_3(1) + f_4(1) + f_3(2) + f_4(2) \\
&= 2 - 8x_1 - 8x_2 + 3(f_3(1) + f_4(1) + f_3(2) + f_4(2))
\end{aligned}$$

Given 3, we obtain

$$\begin{aligned} f_3(3)+f_4(3) &\geq f_1(3)+f_2(3) \\ &\geq 2-8x_1-8x_2+3\left(f_3(1)+f_4(1)+f_3(2)+f_4(2)\right) \geq 2-8x_1-8x_2. \\ 4x_3 &= \sum_{i=1}^{4} f_i(3) \geq 4-16\left(x_1+x_2\right) \\ 4\left(\frac{1}{2}-\varepsilon\right) &\geq 4-16(2\varepsilon) \\ \varepsilon &\geq \frac{1}{14} \end{aligned}$$

■

Proof of Theorem 6. Let p be the distribution of types in the population and $\Pr(s)$ be the probability that someone sees the signal s. We have

$$\sum_s \Pr(s)\, p(\theta=t|s) = \sum_s \left\{ \sum_\theta p(\theta)\, f_\theta(s) \right\} \frac{f_t(s)\, p(t)}{\sum_\tau f_\tau(s)\, p(\tau)} = \sum_s f_t(s)\, p(t) = p(t)$$

Let $\bar{t}(s) = \sum_t t p(\theta=t|s)$. Then

$$\begin{aligned} E(\bar{t}(s)) &= \sum_s \Pr(s)\, \bar{t}(s) = \sum_s \Pr(s) \left(\sum_t t p(\theta=t|s) \right) \\ &= \sum_t t \sum_s \Pr(s)\, p(\theta=t|s) = \sum_t t p(t) = E(p). \end{aligned}$$

By definition $E(p)=m$. Hence, $E(\bar{t}(s)) = E(p) = m$, and $E\left(\frac{1}{n}\sum_{i=1}^n \bar{t}_i\right) = m$, as was to be shown. ■

Proof of Theorem 7. Necessity is proved by the first claim in Lemma 1. For sufficiency take any q, y and x such that $qy \leq x$. Let the type space be $\Theta = [0,1]$, let the population be distributed uniformly on $[0,1]$, and let $S = \{0,1\}$.

If $y \geq x$ let the signalling structure be such that $f_\theta(1) = \frac{y-x}{1-x}$ for $\theta \leq 1-x$, and $f_\theta(1) = 1$ for $\theta > 1-x$. Recall that a type z is in the top $x\%$ of the population if $p(\theta \leq z) \geq 1-x$, so the worst type in the top $x\%$ is $1-x$ and $p(\theta > 1-x) = x$. Then,

$$\begin{aligned} p(\theta > 1-x \mid s=1) &= \frac{\Pr(1 \mid \theta > 1-x)\, p(\theta > 1-x)}{\Pr(1 \mid \theta > 1-x)\, p(\theta > 1-x) + \Pr(1 \mid \theta \leq 1-x)\, p(\theta \leq 1-x)} \\ &= \frac{x}{x + (1-x)\frac{y-x}{1-x}} = \frac{x}{y} \geq q. \end{aligned}$$

If $y < x$, let the signalling structure be such that $f_\theta(1) = 0$ for $\theta \leq 1-y$, and $f_\theta(1) = 1$ for $\theta > 1-y$. Then, $y\%$ observe signal 1 and

$$p(\theta > 1-x \mid s=1) = 0 + p(\theta > 1-y \mid s=1) = y/y = 1 \geq q$$

as was to be shown. ■

Proof of Proposition 4. Consider a signalling structure $(S, \Theta = (\theta_1, ..., \theta_n), f_\theta)$ and population distribution p with median m and a fraction π of types strictly greater than m. For any $\delta < \frac{1}{2}$ and distribution q over Θ that is δ-close to being degenerate, let $\theta_\delta(q) \in \Theta$ be such that $q(\theta_\delta(q)) \geq 1 - \delta$. Note that $\theta_\delta(q)$ is the median of q. Also, there exists a $\bar{\delta}$ such that if $\delta < \bar{\delta}$ then for any distribution q whose mean is greater than m, we have $\theta_\delta(q) > m$, and for any distribution q whose mean is less than or equal to m, we have $\theta_\delta(q) \leq m$. Set t and $\delta < \overline{\delta}$, such that $\pi < (1-\delta)(\pi + \varepsilon + t - 1)$ and $1 - \pi < (1-\delta)(t - \pi + \varepsilon)$.

Let $(p \mid s)$ denote the posterior of p given s, and define $S^* = \{s \in S : (p \mid s)$ is δ-close to degenerate$\}$, $\hat{S}^m = \{s \in S : \text{median of } (p \mid s) > m\}$, $S^m = \{s \in S^* : \theta_\delta((p \mid s)) > m\}$, $\hat{S}^\mu = \{s \in S : (p \mid s) \text{ has mean } > m\}$ and $S^\mu = \{s \in S^* : (p \mid s) \text{ has mean } > m\}$.

Suppose, by way of contradiction, that $P(S^*) \geq t$ and $P\left(\hat{S}^m\right) = x_m \geq \pi + \varepsilon$. Note that $P(S^m) \geq x_m - (1-t)$. We have

$$\begin{aligned}
\pi &= P\{\theta > m\} = \sum\nolimits_s P(\theta > m \mid s) P(s) \geq \sum\nolimits_{S^m} P(\theta > m \mid s) P(s) \\
&\geq \sum\nolimits_{S^m} (1-\delta) P(s) = (1-\delta) \sum\nolimits_{S^m} P(s) = (1-\delta) P(S^m) \geq (1-\delta)(x_m - 1 + t) \\
&\geq (1-\delta)(\pi + \varepsilon - 1 + t)
\end{aligned}$$

which is a contradiction. A similar argument establishes a contradiction if we assume that $P\left(\hat{S}^\mu\right) = x_\mu \geq \pi + \varepsilon$, (with S^μ playing the role of S^m).

Let $\hat{S}_m = \{s \in S : \text{median of } (p \mid s) \leq m\}$, $S_m = \{s \in S^* : \theta_\delta((p \mid s)) \leq m\}$ and suppose that $P\left(\hat{S}^m\right) = x_m \leq \pi - \varepsilon$. Then,

$$\begin{aligned}
1 - \pi &= P\{\theta \leq m\} = \sum\nolimits_s P(\theta \leq m \mid s) P(s) \geq \sum\nolimits_{S_m} P(\theta \leq m \mid s) P(s) \\
&\geq \sum\nolimits_{S_m} (1-\delta) P(s) = (1-\delta) \sum\nolimits_{S_m} P(s) = (1-\delta) P(s_m) \geq (1-\delta)(1 - x_m - 1 + t) \\
&\geq (1-\delta)(t - x_m) \geq (1-\delta)(t - \pi + \varepsilon),
\end{aligned}$$

which is a contradiction. Again, the argument for x_μ is similar.

The proof of Theorem 8 is a corollary of the following lemma, versions of which are well known (see Wolfstetter (1999) Chapter 4). We will use the strict inequalities in this version of our lemma, and in addition we do not require that the probability distribution has a density. ■

Lemma 3 *The posterior after passing a test fosd the prior, which fosd the posterior after failing the test. Moreover, letting n denote a fail, y a pass, and letting p be the population distribution, for any x such that $1 > p(\theta \leq x) > 0$,*

$$p(\theta \leq x \mid n) > p(\theta \leq x) \quad \textit{and} \quad p(\theta \leq x) > p(\theta \leq x \mid y)$$

Proof of Lemma 3. We only compare the prior to the posterior after n, since the comparison between the prior and the posterior after y is symmetric. From Bayes' Rule we must show that

$$\begin{aligned} p(\theta \leq x \mid n) &= \frac{\int_0^x (1-z)\, dp(z)}{\int_0^1 (1-z)\, dp(z)} \geq \int_0^x dp(z) = p(\theta \leq x) \\ &\Leftrightarrow \int_0^x \left[\frac{1-z}{\int_0^1 (1-z)\, dp(z)} - 1 \right] dp(z) \geq 0 \\ &\Leftrightarrow \int_0^x \frac{E(\theta) - z}{1 - E(\theta)} dp(z) \geq 0 \end{aligned} \tag{5}$$

and that the inequality is strict whenever $p(\theta \leq x) > 0$. First notice that for any $x \leq E(\theta)$, the last inequality is trivially satisfied. The integrand in the last inequality is strictly decreasing in z, and is 0 for $z = E(\theta)$. If for some $x > E(\theta)$ we had

$$\int_0^x \frac{E(\theta) - z}{1 - E(\theta)} dp(z) < 0, \tag{6}$$

then for any $x' > x$ we would also have

$$\int_0^{x'} \frac{E(\theta) - z}{1 - E(\theta)} dp(z) < 0,$$

since $\frac{E(\theta)-z}{1-E(\theta)} < 0$ for all $z > x > E(\theta)$. But this contradicts $\int_0^1 \frac{E(\theta)-z}{1-E(\theta)} dp(z) = \frac{E(\theta)-E(\theta)}{1-E(\theta)} dp(z) = 0$.

Now let us turn to the strict inequalities. Pick any $x \leq E(\theta)$, with $1 > p(\theta \leq x) > 0$. If $x = E(\theta)$, we know that $p(\theta < x) > 0$ (because p is non degenerate), and since the integrand in inequality (5) is strictly positive for all $z < E(\theta)$, we must have that the integral is strictly positive, as was to be shown. If $x < E(\theta)$, the strictly positive integrand and $p(\theta \leq x) > 0$ ensure that the integral is also strictly positive.

Pick any $x > E(\theta)$, and assume that

$$\int_0^x \frac{E(\theta) - z}{1 - E(\theta)} dp(z) \leq 0.$$

Since $1 > p(\theta \leq x)$ holds, we must have $p(\theta > x) > 0$. But then, $\frac{E(\theta)-z}{1-E(\theta)} < 0$ for all $z \geq x > E(\theta)$ and $p(\theta > x) > 0$ imply that

$$\begin{aligned} 0 &= \int_0^1 \frac{E(\theta) - z}{1 - E(\theta)} dp(z) = \int_0^x \frac{E(\theta) - z}{1 - E(\theta)} dp(z) + \inf_{w>x} \int_w^1 \frac{E(\theta) - z}{1 - E(\theta)} dp(z) \\ &\leq 0 + \inf_{w>x} \int_w^1 \frac{E(\theta) - z}{1 - E(\theta)} dp(z) < 0 \end{aligned}$$

which is a contradiction. ■

Proof of Theorem 8. Let y denote passing the test, and n denote failing it. We will develop the arguments for the posterior after y, since those for n are analogous. By Lemma 3, the posterior $p(\cdot \mid y)$ strictly fosd the population distribution, which implies that the posterior mean is strictly larger than the prior mean, which is also the median by symmetry.

Let m be the population median. By Lemma 3, the fraction $E(\theta)$ who observe y consider that their median type is weakly larger than the population median. Suppose first that the population distribution has a density. Then,

$$p(\theta \leq m \mid y) = \frac{\int_0^m \theta dp}{E(\theta)} < \frac{m \int_0^m dp}{E(\theta)} = \frac{mp(\theta \leq m)}{E(\theta)} = p(\theta \leq m) = \frac{1}{2}$$

so that m is no longer the median belief of a person who has seen y. As we know, the person's median belief is weakly larger than m, so in fact it must be strictly larger.

If the distribution does not have a density, assume that $p(m)$ is small enough that

$$p(m)\, m + p(\theta < m)\, E(\theta \mid \theta < m) < m/2$$

Then,

$$p(\theta \leq m) = \int_0^m dp \geq \frac{1}{2} > \frac{p(m)\, m + p(\theta < m)\, E(\theta \mid \theta < m)}{m} = \frac{\int_0^m \theta dp}{E(\theta)} = p(\theta \leq m \mid y)$$

and again which establishes that m is not a median of the posterior after y. ■

References

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

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

Biais, B., D. Hilton, K. Mazurier and S. Pouget, (2005), "Judgemental Overconfidence, Self-Monitoring, and Trading Performance in an Experimental Financial Market," *Review of Economic Studies*, **72(2)**, 287-312.

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

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

Burson, K., R. Larrick, and J. Soll (2005) , Social Comparison and Confidence: When Thinking You're Better than average Predicts Overconfidence, Ross School of Business Working Paper No. 1016.

Camerer, C. (1997), "Progress in Behavioral Game Theory," *Journal of Economic Perspectives*, **11(4)**. pp. 167-88.

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

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.

De Bondt, W. and R. H. Thaler, (1995), "Financial decision-making in markets and firms: a behavioral perspective', in (R. A. Jarrow, V. Maksimovic and W. T. Ziemba, eds), *Finance, Handbooks in Operations Research and Management Science*, vol. 9, pp. 385–410. Amsterdam: North Holland.

Dunning, D. (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)**, 1082-1090

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

Gigerenzer, G., U. Hoffrage, and H. Kleinbolting, (1997) "Probabilistic mental models: A Brunswikian theory of confidence," appearing in Research on Judgnent and Decision Making, Cambridge University Press.

Healy, P.J. and D. Moore, (2007), "Bayesian Overconfidence," mimeo.

Hoelzl, E. and A. Rustichini, (2005), "Overconfident: do you put your money on it?" the *Economic Journal*, **115**, pp. 305-18.

Holland, C., (1993) "Self-bias in older drivers' judgments of accident likelihood", *Accident Analysis and Prevention*, **(25)(4)**, pp. 431-441

Koszegi, B., (2006), "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, **4(4)**, 673-707.

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.

Kruger, J., and D. Dunning (1999), "Unskilled and Unaware of it: How difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, **77(6)**, 1121-1134.

Kyle, 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.

Ledyard, J.O. (1986), "The Scope of the Hypothesis of Bayesian Equilibrium," *Journal of Economic Theory*, **39**, 59-82.

Malmendier, U. and G. Tate (2005), "CEO Overconfidence and Corporate Investment," *Journal of Finance*, **60(6)**, 2661-700.

Marottoli, R. and E. Richardson (1998) , "Confidence in, and self-rating of, driving ability among older drivers," *Accident Analysis and Prevention*, **(30)(3)**, pp. 331-336.

Mathews, M., and A. Moran (1986), "Age differences in male driver's perception of accident risk: the role of perceived driving ability", *Accident Analysis and Prevention*, **(18)(4)**, pp.299-313.

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. (2007), "Not so above average after all: When people believe they are worse than average and its impications for theories of bias in social comparison," Organizational Behavior and Human Decision Processes, **102(1)**, pp 42-58.

Myers Social Psychology, 6th edition

Noth, M. and M. Weber, (2003), "Information Aggregation with Random Ordering: Cascades and Overconfidence," *Economic Journal*, 113(484), 166-89.

Peng, L. and W. Xiong, (2006), "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, **80(3)**, 563-602.

Santos-Pinto, L., and J. Sobel, (2005) "A Model of Positive Self-Image in Subjective Assessments," *American Economic Review,* **95(5)**, 1386–1402.

Scheinkman, J. and W. Xiong, (2003), "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111(6), 1183-1219.

Schwarz, N., Knäuper, B., Hippler, H. J., Norllr-Neumann, E., and Clark, F. (1991), "Rating Scales: Numeric values may change the meaning of scale labels, *Public Opinion Quarterly,* 55, 570-582.

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

Taylor, S. E. and J.D. Brown, J. D. (1988), "Illusion and well-being: A social psychological perspective on mental health," *Psychological Bulletin*, **103**, 193—210.

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

Wang, A. (2001), "Overconfidence, Investor Sentiment, and Evolution," *Journal of Financial Intermediation*, **10(2)**, 138-70.

Walton, D., (1999), "Examining the self-enhancement bias: professional truck drivers' perceptions of speed, safety, skill and consideration," *Transportation Research Part F,* pp. 91-113.

Zabojnik, J. (2004), "A Model of Rational Bias in Self-Assessments," *Economic Theory,*

23(2), 259–82.