UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays on Biased Self Image

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Economics

by

Young Joon Park

Committee in charge:

Professor Joel Sobel, Chair
Professor James Andreoni
Professor Vincent P. Crawford
Professor Uri Gneezy
Professor David Schkade

2009
The dissertation of Young Joon Park is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2009
DEDICATION

To my wife and parents for their love and support.
# TABLE OF CONTENTS

Signature Page ................................................................. iii

Dedication ........................................................................ iv

Table of Contents .............................................................. v

List of Figures .................................................................... viii

List of Tables ....................................................................... ix

Acknowledgments ............................................................... xi

Vita, Publications, and Fields of Study ............................. xii

Abstract of the Dissertation ................................................. xiii

1  (Biased) Self Image in Objective Qualities ................. 1

1. Introduction ................................................................. 1

2. Experiment ................................................................. 5

3. Theories and Hypotheses ................................................. 9

   A. Random Errors .......................................................... 12

   B. Truncated Error with skewed distribution (TE) .......... 13

   C. Egocentric Bayesian Update (EBU) .......................... 14

   D. Signalling Model of Own Attributes (SMOA) .......... 16
1. Introduction .......................................................... 57
2. Field Experiment Design .............................................. 62
3. Hypotheses ............................................................ 67
4. Forecast Bias .......................................................... 69
5. Forecast Precision ...................................................... 71
6. Betting Behavior ...................................................... 73
7. Unskilled and Unaware ............................................... 76
8. Efficiency of Forecasts ................................................. 78
A. False Consensus Effect ............................................. 111
B. Generalized false consensus effect .............................. 111
C. Completeness ......................................................... 113

7. False Consensus Effect in Social Psychology .................... 114

8. Conclusion ............................................................. 118

9. References ............................................................. 119
LIST OF FIGURES

Figure 1.1: Population distribution of the attributes .......................... 53
Figure 1.2: Errors of assessments of own attributes ............................. 54
Figure 1.3: Overestimation of Same-Quality-Group (SQG) - vs. All UCSD 55
Figure 1.4: Overestimation of Same-Quality-Group (SQG) - vs. Same major 56
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Experiment participants’ major</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>Accuracy of own attributes</td>
<td>22</td>
</tr>
<tr>
<td>1.3</td>
<td>Overestimation of the weak-more-than-group (WMTG) depending on answers</td>
<td>25</td>
</tr>
<tr>
<td>1.4</td>
<td>Overestimation of same-quality-group (SQG) depending on answers</td>
<td>26</td>
</tr>
<tr>
<td>1.5</td>
<td>Overestimation of same-quality-group (SQG) - UCSD vs. Same Major</td>
<td>27</td>
</tr>
<tr>
<td>1.6</td>
<td>Overestimation of same-quality-group (SQG) - Answer=0 vs. Answer&gt;0</td>
<td>28</td>
</tr>
<tr>
<td>1.7</td>
<td>Correlation between overestimation and true size of same-quality-group (SQG)</td>
<td>29</td>
</tr>
<tr>
<td>1.8</td>
<td>Individual Categorization</td>
<td>31</td>
</tr>
<tr>
<td>1.9</td>
<td>Individual Heterogeneity</td>
<td>42</td>
</tr>
<tr>
<td>2.1</td>
<td>Distribution of Players’ Forecasts in Tournaments</td>
<td>70</td>
</tr>
<tr>
<td>2.2</td>
<td>OLS Regression Results for Forecast Bias</td>
<td>70</td>
</tr>
<tr>
<td>2.3</td>
<td>OLS Regression Results for Forecast Precision</td>
<td>72</td>
</tr>
<tr>
<td>2.4</td>
<td>Players’ Choices and Earnings from Bets</td>
<td>74</td>
</tr>
<tr>
<td>2.5</td>
<td>OLS Regression Results for Forecast Precision Experience and Elo in Sintra’s Chess Open</td>
<td>77</td>
</tr>
<tr>
<td>2.6</td>
<td>OLS Regression Results for Test of Efficiency of Players’ Forecasts in Sintra’s Chess Open</td>
<td>78</td>
</tr>
<tr>
<td>2.7</td>
<td>UCSD’s 2004 Poker Classic Prizes</td>
<td>95</td>
</tr>
</tbody>
</table>
Table 2.8: Sintra’s Chess Open Prizes ............................... 95

Table 2.9: Players’ Earnings from Forecasts ......................... 96

Table 3.1: Composition of Estimation (Krueger, 1998) ................. 110
ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Joel Sobel. None of my achievement would have been possible without his guidance, patience, and encouragement.

I would also like to thank the members of my dissertation committee, James Andreoni, Vincent P. Crawford, Uri Gneezy, and David Schkade. Their comments and suggestions improved the dissertation substantially.

I want to thank Luís Santos-Pinto for providing valuable research experience in every stage of the year in graduate school and letting me use our joint work as part of my dissertation.

I thank Pedro Rey-Biel for reading the first essay thoroughly and giving me detailed comments. I also thank Hyoung Sik Noh for pointing out a critical mistake of a result in the first essay and suggesting a way to fix it.

Throughout the days in graduate school, I have been indebted to many colleagues. They had to suffer from my abrupt questions which were usually too obvious or pointless. I’d like to thank Youjin Hahn, Yong-Gook Jung, Hiroaki Kaido, Munechika Katayama, Chulyoung Kim, Kwang Kim, Min Seong Kim, Young Do Kim, Dong Jin Lee, Lei Meng, Philip Neary, Adam Sanjurjo, Sarada, Lucas Siga, Suyong Song, Karmen Suen, Hee-Seung Yang, and Michael Yang.

The second chapter “Forecasts of Relative Performance in Tournaments: Evidence from the Field,” in full, is coauthored with Luís Santos-Pinto and has been submitted for publication to the Journal of Theory and Decision.
VITA

2000  B.A., Economics,
      Korea University, Korea

2002  M.A., Economics,
      Korea University, Korea

2009  Ph.D., Economics
      University of California, San Diego.

FIELDS OF STUDY

Major Field: Microeconomics (Behavioral Economics, Experimental Economics, Game Theory)
ABSTRACT OF THE DISSERTATION

Essays on Biased Self Image

by

Young Joon Park

Doctor of Philosophy in Economics
University of California, San Diego, 2009

Professor Joel Sobel, Chair

Theoretical models in economics rely on the assumption that goal-oriented agents optimize with respect to statistically correct beliefs about the environment. Behavioral evidence casts doubt on the assumption of accurate beliefs. There is widespread evidence that agents’ expectations are not just inaccurate, but systematically biased. This bias has potentially large implications in economic settings since most of economic models assume that the economic agents are free of these biases. This dissertation investigates the existence of the biased self image - individuals’ systematically biased view on own behavior or choices compared to others - and attempts to provide rational frameworks on some of them.

I examine a different bias in each chapter. In the first chapter, I study the extent to which agents systematically overestimate (or underestimate) their own attributes compared to others. I run a lab experiment which is designed to test whether agents possess this bias over objective and unambiguous attributes. In the second chapter, I collect survey data from two field tournaments where the participants make forecasts of their relative performance. I study the relationship between the accuracy of their forecasts and the quality of performance. In the
third chapter, I propose a simple model of the false consensus bias that can explain, under mild assumptions, the bias as rational behavior.

In the first essay, “(Biased) Self Image in Objective Qualities”, I attempt to confirm the existence of a prominent systematic bias, identify the contexts in which it is likely to arise, and come up with behavioral assumptions that describe belief formation more accurately than conventional economic assumptions. While a large number of studies confirm the existence of biased self image, the results can often be traced to the ambiguities in the definition of the quality investigated. I run an experiment to find out whether individuals’ beliefs about objectively defined qualities are biased. In particular, I describe data obtained by asking university students about information on their transcripts. Students answered questions about the number of high and low grades they received and the number of classes they took in different departments. They were also asked to rank their own characteristics relative to others. I attempt to test which behavioral model of biased self image explains the data the best.

I found subjects accurately report their characteristics in their transcripts. Their relative assessments, however, are noisy. The fact that I could not find a compelling evidence of systematic bias can raise a question of the existence of biased beliefs in objective qualities. On the other hand, I found a strong evidence of people overestimating the proportion of the group with the same qualities. Also, using within subject analysis, I can categorize individual-level bias. With 5% significance level, almost 60% of the subjects are classified as possessing bias that can be explained by behavioral models of biased self image.

The second essay, “Forecasts of Relative Performance in Tournaments; Evidence from the Field” is coauthored with Luís Santos-Pinto. We use a field survey to investigate the quality of individuals’ beliefs of relative performance in
tournaments. We use data obtained from two field settings, poker and chess, which
differ in the degree to which luck is a factor and also in the information that play-
ers have about the ability of the competition. The main finding of the paper is
that players’ forecasts in both types of tournaments are biased towards overesti-
mation of relative performance. However, the accuracy of the participants’ beliefs
differ between the two settings. We find that poker players’ forecasts of relative
performance are random guesses with an overestimation bias. Chess players also
overestimate their relative performance but make informed guesses. We find sup-
port for the “unskilled and unaware hypothesis” in chess: high skilled chess players
make better forecasts than low skilled chess players. Finally, we find that chess
players’ forecasts of relative performance are not efficient in the sense that they
could enhance their accuracy of the forecast if they used the information about
other participants’ skill.

One of the main findings of the first essay is that subjects do possess the
false consensus bias - tendency for individuals to overestimate the prevalence of
own behavior or choices. Psychological evidence suggests that the bias is frequently
observed in various cases. But the existing analysis of the behavior is abstract.
In the third essay “A Simple Model of the False Consensus Effect”, I propose a
simple model of information acquisition that describes this false consensus effect
and several of its empirically identified qualitative properties. I show that, under
mild assumptions, the tendency to believe one’s own choice to be more informative
about the behavior of a group than that of another’s can be explained as a rati-
onal reaction. This ‘generalized’ false consensus effect has kept being considered
irrational since the observation of Dawes (1989) that provides rational explanation
for the false consensus effect. The paper shows that what is called ‘truly false’ by
Dawes (1989) does not need to be false.
1

(Biased) Self Image in Objective Qualities

1.1 Introduction

Theoretical models in economics rely on the assumption that goal-orient-ed agents optimize with respect to statistically correct beliefs about their envi-ronment. Behavioral evidence casts doubt on the assumption of accurate beliefs. There is widespread evidence that agents’ expectations are not just inaccurate, but systematically biased. This paper attempts to confirm the existence of a prominent systematic bias, identify the contexts in which it is likely to arise, and propose behavioral assumptions that describe belief formation more accurately than the conventional economic assumptions.

I study the existence of bias in assessing one’s own accomplishments relative to others. The social psychology literature argues that the “Lake Wobegone Effect” is typical. Most people report that they are “better than average.” When
asked about driving abilities (Svenson, 1981), contributions on a joint task (Ross & Sicoly, 1979), health (Weinstein, 1980), integrity of marriage (Baker & Emery, 1993), and professional skills among co-workers (Cross, 1977), people tend to rank themselves more highly than others rank them.

This upward bias, or positive self image, has potentially profound implications in economic settings. It may cause excess entry in markets in which entrants overestimate their ability to make profits. It may cause unhappiness in the workplace when a worker observes that someone else received the bonus or raise that he deserved. It may lead to bargaining breakdowns if people overestimate outside opportunities. It may lead to underinsurance if people overestimate their resistance to disease.

While a large number of studies confirm the existence of positive self image, many of these depend on features that make the results poorly suited for economic models. Specifically, the results can often be traced to the ambiguities in the definition of the quality being investigated. For example, there are many variety of ways to define “good driving.” There are strong reasons to believe that the existence of positive self image is not evidence of biased belief formation in these settings. If different individuals define quality in different ways, make investments to maximize their quality, and use their own definition of quality to make relative comparisons, then one would expect the existence of positive self image (in the sense that more than p percent of the population believes that they have more quality than 1-p percent of the population for all p).\(^1\)

There is evidence that positive self image exists even when qualities are defined objectively. Guthrie, Rachlinski, & Wistrich (2001) asked U.S. magistrate judges to compare themselves among other judges in terms of the rate of

\(^1\)See Santos-Pinto & Sobel (2005).
their decisions to get overturned. The overturn is an event that is fully observable to oneself, and one would expect magistrate judges to accurately assess their ranking. However, 56.1% ranked themselves in bottom quartile and 87.7% ranked themselves below the median in rate of overturn. This result suggests the existence of significant bias over objective and unambiguous qualities.

I use experimental evidence of agents’ beliefs on objective qualities to investigate the existence of biased self image. In particular, I describe data obtained by asking university students information regarding their transcripts. Students answered questions about the number of high and low grades they had received and the number of classes they took in different departments. They were also asked to rank their own characteristics relative to other students. Subjects’ answers are compared with their true characteristics. The data permit me to investigate several models of belief formation.

One observation comes through clearly. With few exceptions, students accurately reported their own characteristic. Any bias in beliefs appears to come from the way they evaluated their characteristic relative to others.

Next I investigate whether students’ responses depend on how difficult it is to obtain the characteristic. A simple theory that predicts agents overestimate their relative achievement would not distinguish between an easy or difficult characteristic to obtain. I manipulate difficulty by changing the question. The population distribution of extremely high grades (A+s) has a large mass at 0 and declines steeply. Especially high grades are difficult to achieve. On the other hand, the empirical distribution of passing grades rises steeply. Few students have many failing grades. Although the most naive theory of positive self image makes the same prediction in both cases, there is evidence that the easier it is to obtain the “good” characteristic, the greater the positive self image. In fact, there are ex-
periments that demonstrate negative self image when tests are hard (Camerer & Lovallo (1999), Moore (2002), Moore & Kim (2003), Hoelzl and Rustichini (2005), and Moore & Healy (2008)).

While there is a tendency for subjects to overestimate their relative standing with respect to a neutral characteristic (number of classes taken in a particular discipline), I do not find strong evidence for negative or positive self image in either case. These findings cast doubt on the existence of biased beliefs for objective characteristics.

The data do show some evidence of a different kind of bias. In response to many questions, subjects appear to exhibit a significant bias, wherein they overestimate the proportion of the population who are just like they are. This phenomenon shares similar characteristics with the false consensus effect, a well known behavioral bias in social psychology.

By performing a within subject analysis, I find systematic bias at the individual level. At the 5% significance level, almost 60% of subjects can be classified to possess a specific behavioral bias.

The main advantage of the paper is that subjects’ assessments can be compared directly with their true characteristics. I obtained the population distribution of students’ grade and used it to measure the accuracy of their answers. I could test not only whether subjects possess accurate knowledge of their attributes on school record, but also how accurately they assessed their attributes relative to other students. Hence, I could further perform within subject analysis by measuring individual level bias. Moreover, unlike many of the previous studies on biased self image, subjects did not face strategic or competitive situation. By letting the subjects focus on pure estimation problem, I can avoid possible distortion of the
results.\footnote{In strategic settings, the distortion can arise from various sources such as attitude toward risk or externality effect from other players.}

Another contribution I make to existing literatures is that the result is robust to Benoît & Dubra (2007)’s critique. According to their model, if individuals do not observe own attributes directly, but only observe signals of them, any degree of biased self image can occur even if the individuals follow Bayes rule. Benoît & Dubra point out that much of the existing evidence of biased self image fails to distinguish true bias from this possible ‘apparent’ biased self image. In my experiment, subjects make assessments over objective and unambiguous qualities. The true bias found in this exercise, therefore, is free from their critique.

The paper proceeds as follows. The following Section describes the experimental design. In Section 3, I introduce and propose several possible theories of biased self image and relevant hypotheses. Section 4 describes the results and relates them to the theories. Section 5 concludes.

\subsection*{1.2 Experiment}

72 undergraduate students at the University of California, San Diego (UCSD) from 6 different departments participated in the experiment.\footnote{Subjects were informed that they could earn up to $45 by participating in the experiment that could take about 80 minutes.} Subjects were asked to answer questions about seven of their own attributes related to their school record. They were also asked to assess the relative ranking of the attributes among others. There were four attributes regarding their grades, number of ‘A+’, ‘A- or higher’, ‘C- or lower’, and ‘F’ grades they received in UCSD, and three attributes about the classes, number of ‘Biology’, ‘Philosophy’ and ‘Humanities’
classes they have taken at UCSD. These attributes were selected to address the following features.

- **The attributes are objective and unambiguous.** Each question has a correct answer, and there is little possibility of subjective interpretation of the question. Ordinary students are usually aware of this information related to the school record.\(^4\)

- **The desirability of attributes are various.** Some of the questions about grades are either positive (\(A+\) and \(A-\) or higher) or negative (\(C-\) or lower and \(F\)) in qualities. Class-related questions are neutral in qualities. This selection enables me to possibly separate the simple positive self image due to the psychological reason.\(^5\)

- **The attributes have positively skewed (decreasing) population distribution.** If the relation between biased self image and difficulties of the task is due to the skewed distribution of the outcome, the result will be replicated by attributes with similarly shaped distributions. Due to the lack of examples of attributes that have negatively skewed (increasing) distribution, I chose the attributes that have positively skewed distribution.

- **True attributes can be verified.** I had access to subjects’ personal school record that shows their grades and number of classes. Using the data of all UCSD students, I could obtain the correct population distribution of each attribute. By comparing subjects’ answers and real attributes, I measure errors on knowledge of their school record. Using the actual population distribution, the biases on the self assessment were calculated not only collectively, but also on individual level.

\(^4\)Subjective/ambiguous counterpart examples of these objective/unambiguous questions can be “How well did you perform in class?”, or “How much Biology did you learn in college?”

\(^5\)That is, people simply get utility by assessing one’s achievements or qualities positively.
• **Subjects are rewarded for accurate predictions.** The rewards were based on accuracy of both attributes and relative assessments. The performance-based rewards lessens the possibility of careless responses. Furthermore, the reward does not depend on any strategic component that might alter the result. Thus, I can measure the bias without possible dispute of interpretation.

10 sessions were held and the number of participants in each session ranged from 5 to 11. Participants were asked to read the consent form and the instructions of the experiment. After taking the understanding tests, they answered the main questionnaire and the post experiment survey. The total time of each session ranged from 60 to 70 minutes. Table 1.1 shows the information of subjects’ majors.

<table>
<thead>
<tr>
<th>Major</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economics (Econ)</td>
<td>31 (43.06%)</td>
</tr>
<tr>
<td>Communication (Comm)</td>
<td>3 (4.17%)</td>
</tr>
<tr>
<td>Political Science (Poli)</td>
<td>7 (9.72%)</td>
</tr>
<tr>
<td>Electric and Computer Engineering (ECE)</td>
<td>10 (13.89%)</td>
</tr>
<tr>
<td>Mechanical and Aerospace Engineering (MAE)</td>
<td>9 (12.5%)</td>
</tr>
<tr>
<td>Chemistry and Biochemistry (Chem)</td>
<td>12 (16.67%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>72 (100%)</strong></td>
</tr>
</tbody>
</table>

Four of the questions (A+, F, Biology, and Philosophy) are main questions on attributes with positively skewed distribution. Attributes of three sub questions (A- or higher, C- or lower, Humanities) contain those of three of the main questions (A+, F, and Philosophy). Consequently, population densities of three sub questions shift the corresponding main questions’ densities to the right. All grade related questions were selected to represent either desired or undesired qualities. Figure 62 subjects failed the understanding test. Since both of their mistakes were simple miscalculation, I did not exclude their data from the analysis.
1.1\textsuperscript{7} shows population densities of these attributes.

The questionnaire had three parts. In Part 1, each subject was asked the number of classes they took in UCSD. In Parts 2 and 3, subjects compared themselves to other students. The comparison was made against all UCSD students on Part 2, and students with the same major on Part 3. In order to avoid too much variance on the distribution of attributes, participants were limited for those who completed 90 or more credit hours at UCSD.\textsuperscript{8} The followings are the sample questions from the questionnaire.

\textbf{Example A} Please write down the number of A+ grades from UCSD classes that appear on your transcript.

\textbf{Example B} Please estimate the percentage of UCSD students who have completed at least 90 credit hours at UCSD who have received more, the same number, or fewer Philosophy classes than you.

<table>
<thead>
<tr>
<th>Description</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>The people who have taken more Philosophy classes than you</td>
<td>(1) %</td>
</tr>
<tr>
<td>The people who have taken the same number of Philosophy classes as you</td>
<td>(2) %</td>
</tr>
<tr>
<td>The people who have taken fewer Philosophy classes than you</td>
<td>(3) %</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100 %</td>
</tr>
</tbody>
</table>

The nature of the skewed distribution along with discrete attributes caused large mass near zero attributes. For example, about 62\% of the population had never received an A+. In order to avoid confusion in making comparison, subjects was not required to predict their relative rankings directly. Instead, the questionnaire were designed for subjects to provide their prediction of the proportion of people

\textsuperscript{7}See Appendix.

\textsuperscript{8}By UCSD criteria, students who have taken 90+ credits are classified as juniors or seniors.
who have more number of attributes (more-than-group), the same number of attributes (same-quality-group), and fewer number of attributes (fewer-than-group). The series of the answer was transformed into the size of the population that has weakly more attributes.

\[
\text{Weak-more-than-group (WMTG)} = \text{Size of more-than-group} + \frac{1}{2} \text{Size of same-quality-group}
\]

Weak-more-than-group (WMTG) is measured as above. There is no implication of positive or negative self image with over/under estimation of WMTG because not all attributes used in the experiment are desirable. Overestimation of WMTG on \(A^+\) and \(A^-\) or higher questions correspond to negative self image while overestimation of WMTG on the \(F\) and \(C^-\) or lower questions correspond to positive self image. Moreover, for all class-related questions, bias in WMTG exhibits neither positive nor negative value.

One question from each part was randomly selected to determine the reward. For Part 1 of the experiment, subjects received $5 if the answer they provided was correct, $2.5 if the answer was different by 1, and $0 otherwise. For Parts 2 and 3, they received [[$15-the difference between the reported and the actual weak-more-than-group]], or $0 if the difference was greater than 15%. The average payment per subject was $22.35.

1.3 Theories and Hypotheses

This section introduces several possible theories of biased self image. I also set several hypotheses of the experimental results according to the theories.

I assume individuals’ perceived distribution of population’s attributes is
composed of two distinctive components. The first is general knowledge of the attributes that is commonly shared by individuals, and the second is the type-specific updating scheme that depends on individual bias type. Individuals collect and observe general knowledge about attributes everyday. Particularly, in our example, students can clearly observe how other students perform in classes, which classes are more popular than others, and so on. I further assume the assessment from the general knowledge to be accurate.\footnote{Many possible explanations may lead to a conclusion that the assessment is inaccurate. For example, individuals might obtain biased general knowledge because they observe them from their unrepresentative peer group.} The second part of the individuals’ assessment is the type-specific updating scheme from own signals. Each individual not only collects general information about attributes, but also receives signals about own attributes. I assume individual’s bias type determines how the type-specific update is conducted. The models I introduce vary in how accurate their observations are, and how individuals use own signals to update the assessments.

**Simple Bayesian Problem and Notations**

Throughout this section, the following notation is used. There are $1, ..., I$ individuals and each individual $i$ possesses one of the quantities $x_i = 0, 1, 2, ..., M$ of the attribute. $f(x) = \frac{1}{I} \sum_{i=1}^{I} 1_{x_i=x}$ is the true discrete density for the whole population where $1_{x_i=x}$ has the value 1 when $x_i = x$ and zero otherwise. Suppose $g(x)$ is the type-specific density that will be defined according to the model, and $h(x)$ is subjects’ perceived density. $F(\cdot)$, $G(\cdot)$, and $H(\cdot)$ are cdf’s for $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$ respectively.

The following results depend on the assumption that the true density, $f(x)$, is non increasing.
**Assumption 1.1** The true population density $f(x)$ for four main attributes ($A+$, $F$, Biology, and Philosophy) are non-increasing.

An individual possesses prior belief that assigns probability $\pi$ to the true density derived from general information and $1 - \pi$ to the type-specific density derived from updating type of the individual. If an individual observes own attribute that $x = \alpha$, she will update the density according to Bayes’ rule and the perceived density, $h(x)$ is obtained.

$$ P(\text{True Density}|x = \alpha) = \frac{P(x = \alpha|\text{True Density})P(\text{True Density})}{P(x = \alpha)} = \frac{\pi f(\alpha)}{\pi f(\alpha) + (1 - \pi)g(\alpha)} \equiv \theta $$

Bias is measured in terms of weak-more-than-group (WMTG) and same-quality-group (SQG). The following is a formal definition of WMTG followed by a lemma showing the key characteristic of WMTG.

**Definition 1.1** $F_W(x) = \frac{1}{2}f(x) + \sum_{t>x} f(t)$ is the weak-more-than-group (WMTG) of density $f(\cdot)$.

**Lemma 1.1** For any two distributions $F(x)$ and $G(x)$ with corresponding non-increasing pdfs $f(x)$ and $g(x)$, if $F(x)$ first order stochastically dominates $G(x)$, then $F_W(x) \geq G_W(x)$ for any $x \in \{0, 1, \ldots, M\}$. 
Proof)

\[ F_W(x) = \frac{1}{2} \left[ F(x) - F(x - 1) \right] + 1 - F(x) \]

Therefore,

\[ F_W(x) - G_W(x) = -\frac{1}{2} \left[ (F(x) - G(x)) + (F(x - 1) - G(x - 1)) \right] \geq 0 \]

for any \( x \in \{0, 1, ..., M\} \).

I define biases in WMTG and SQG as follows.

**Definition 1.2** Define \( B_W(x) = H_W(x) - F_W(x) \), the bias in weak-more-than-group (WMTG). \( B_S(x) = h(x) - f(x) \) is the bias in same-quality-group (SQG) where \( h(\cdot) \) and \( f(\cdot) \) are perceived and true pdf of population.

Throughout the paper, \( B_W(x) \) and \( B_S(x) \) are used as the basis to measure bias in estimation of WMTG and SQG. Bias greater (less) than zero means individual overestimated (underestimated) WMTG/SQG.

**1.3.A Random Errors**

If subjects possess bias over only ambiguous attributes since everyone has own subjective interpretation, then they should not possess systematic bias when the qualities to be compared are objective and unambiguous. In that case, subjects should only possess random bias on this experiment. However, if the subjective and ambiguous characteristic of the attributes are not the only reason of bias, individuals might well exhibit systematic bias.
Hypothesis 1.1  Individuals do not show any systematic biased self image.  
\((B_W(x) = B_S(x) = 0, \text{ on average.})\)

1.3.B  Truncated Error with skewed distribution (TE)

Individuals may possess biased self image if they observe own attributes correctly but others’ with error. Specifically, I assume subjects possess correct information about their own attributes, but observe those of others with systematic noise. Then the bias is determined by the shape of population distribution of the attributes.

Suppose each individual observes others’ attribute with an error \(t\) with probability \(e(t)\), and assume \(e(t) = e(-t)\). Define \(h(x)\) and \(H(x^*)\) as the perceived pdf and cdf of attributes respectively.

\[
h(x) = f(x) + \sum_{\substack{t = -x \to M-x}} e(t)(f(t + x) - f(x)) \quad (1.1)
\]

\[
H(x^*) = F(x^*) + \sum_{x=0}^{x^*} \sum_{t=-x}^{M-x} e(t)(f(t + x) - f(x)) \quad (1.2)
\]

**Proposition 1.2**  If \(f(x)\) is strictly increasing (non decreasing), then \(H(x^*) - F(x^*)\) is positive (non-negative), and if \(f(x)\) is strictly decreasing (non increasing), then \(H(x^*) - F(x^*)\) is negative (non-positive).

**Proof**  See Appendix A.

**Corollary 1.3**  An individual with Truncated Error bias overestimates weak-more-than-group.
Intuitively, if the true population distribution is skewed and monotonic, the error structure lessens the degree of skewness of the perceived distribution. Consequently, if the distribution is positively (negatively) skewed, individuals overestimate (underestimate) WMTG.

**Hypothesis 1.2** Individuals overestimate weak more than group ($B_W(x) \geq 0$) for four main questions (A+, F, Biology, and Philosophy).

If subjects possess correct assessment about themselves but receive noisy signals on others’, then they possess TE bias and systematically overestimate WMTG.

### 1.3.C Egocentric Bayesian Update (EBU)

Moore & Healy (2008) provide another possible explanation of the relation between biased self image and difficulties of tasks. It stems from the way individuals update the signal of their own attributes. In their settings, the realization of the attributes is determined by two random variables, the general characteristics of the attributes (or the simplicity of the task), which is same for all players, and an idiosyncratic error term.

Suppose subject $i$’s attribute of school record, $x_i$, is a realization of a random variable $X_i$. Furthermore, $X_i = S + L_i$, where $S$ is overall expected attribute across individuals (general characteristics) with $E(S) = \mu$, and $L_i$ is mean-zero idiosyncratic component. For example, the number of A+ grades each subject receives depends not only on the overall difficulty or rarity of receiving an A+ grade, but also on individuals’ unknown ability or how lucky he or she was.
while taking the exam. Since \( i \) observes \( x_i \) but not \( x_j \) for any other agent, \( j \neq i \), her belief of \( L_j \) does not change after observing own attribute; \( E(X_j|x_i) = E(S|x_i) \)

Suppose an individual observed own attribute \( x_i < \mu \). The individual can ascribe the low value of attribute to either \( S \) (it is less likely to obtain the attribute than she expected) or \( L_i \) (it was randomly determined). Naturally, if the individual puts positive weight on both random variables, then \( E(X_i|x_i) < E(S|x_i) \). Since \( E(S|x_i) = E(X_j|x_i) \), this subject views her attributes lower than others and overestimate WMTG. Also, if the attribute turns out to be more likely to obtain than prior belief, subject are likely to observe \( x_i > \mu \). Consequently, she will underestimate WMTG.

Applying this model to my experiment, if a subject believes receiving an A+ grade is purely determined by luck, the realization of the number of her A+ grades will provide no information about the distribution. This corresponds to the case in which her perceived distribution is a discrete uniform distribution. On the other hand, if the individual fully adjusts to the information she obtained, the perceived distribution should be accurate. Thus, if they put positive weight on both signal and luck, their perceived distribution will be convex combination of true distribution and discrete uniform distribution. Definition 3 summarizes this observation.

**Definition 1.3** An individual possesses **Egocentric Bayesian Update** bias if

\[
h(x) = \theta f(x) + (1 - \theta)g_{EBU}(x), \quad \text{where} \quad g_{EBU}(x) = \frac{1}{M}
\]
Proposition 1.4  *Egocentric Bayesian Updating* individuals overestimate weak-more-than-group.

**Proof**

\[ f(x) - h(x) = (1 - \theta)(f(x) - \frac{1}{M}) \]

\( f(x) - h(x) \) is non-increasing given \( f(x) \) is not increasing.

Therefore, \( H(x) \) first order stochastically dominates \( F(x) \).

**Hypothesis 1.3** Individuals overestimate weak more than group \((B_W(x) \geq 0)\) for four main questions (A+, F, Biology, and Philosophy).\(^{11}\)

1.3.D Signalling Model of Own Attributes (SMOA)

Benoît & Dubra (2007) introduced a signaling model of own attributes and explained how a wide range of (apparent) biased self image can exist with rational agents. The key assumption of the model is that each individual does not possess accurate information about her own attributes (or ‘types’ according to their term). Instead, the individual receives a signal about her attribute and updates her prior. In this case, the posterior on her type depends on the structure of prior belief and probability distribution of signals. Therefore, even though individuals update ‘rationally’, it is possible for more than half of the population to possess positive self image. More precisely, any fraction of population can place itself above the median if individuals do not observe own attribute directly but only observe

\(^{11}\)Note that both TE and EBU subjects will overestimate WMTG for questions with positively skewed attributes.
signals about it. Benoît & Dubra claim that most experiments do not distinguish apparent bias from real bias.

In contrast to Benoît & Dubra (2007), subjects in this experiment directly observe their attributes. Therefore, the result does not apply to their critique. It can still be tested, however, whether individuals possess biased self image while holding the correct information about own attributes.

**Hypothesis 1.4** *Individuals do not have correct assessment about own attributes.*

Hypothesis 4 tests whether Benoît & Dubra’s assumption holds in our experiment.

**Hypothesis 1.5** *Given correct assessment about own attributes, individuals will not possess systematic biased self image.*

If individuals show systematic bias despite possessing correct assessment on own attributes, these biases will be beyond what they called ‘apparent overconfidence’. The result should be robust to Benoît & Dubra’s critique that many results of biased self image can be due to the inaccurate assessment of own attributes.

### 1.3.E Positive Self Image (PSI)

The well known tendency of people to overestimate one’s performance is a good candidate for subjects’ bias. One simple explanation for positive self image is that it simply makes an individual feel better to believe her quality is higher than it really is.

There are other approaches to positive self image which assume individuals have incentive to overestimate one’s quality. Compte & Postlewaite (2004)
suggests if an individual’s performance depends on confidence level, it is optimal to possess positive self image. Hvide (2002) introduced “Pragmatic Beliefs” to explain bias in self assessments. An individual with pragmatic belief does not possess belief according to the true information. Instead, she has an incentive to possess bias. Given the right situation, pragmatic beliefs induce the individual to possess positive self image to maximize her utility where this positive bias becomes self-fulfilling.

In the experiment, the attributes consists of three categories of questions. There are desirable attributes with positive value (‘A+’ and ‘A- or higher’), undesirable attributes with negative value (‘F’ and ‘C- or lower’), and neither desirable, nor undesirable attributes with neutral value (‘Biology’, ‘Philosophy’, and ‘Humanities’). If people simply overestimate their own quality, they should possess positive self image for \( A^+, A^- \text{ or higher}, C^- \text{ or lower} \), and \( F \) questions, and have no bias for \( \text{Biology, Philosophy, and Humanities} \) questions. Therefore, subjects should underestimate the number of people with more \( A^+ \) and \( A^- \text{ or higher} \) grades, and overestimate the number of people with more \( F \) and \( C^- \text{ or lower} \) grades.

**Hypothesis 1.6** Individuals overestimate weak-more-than-group for questions about positive attributes (‘A+’ and ‘A- or higher’), underestimate weak-more-than-group for questions about negative attributes (‘F’ and ‘C- or lower’), and show no bias for questions about neutral attributes (‘Biology’, ‘Philosophy’, and ‘Humanities’).

**1.3.F False Consensus Effect (FCE)**

The false consensus effect has been widely studied among Social Psychologists. Mullen et al. (1985) define the false consensus effect as:
False consensus refers to an egocentric bias that occurs when people estimate consensus for their own behaviors. Specifically, the false consensus hypothesis holds that people who engage in a given behavior will estimate that behavior to be more common than it is estimated to be by people who engage in alternative behaviors.

Ross et al. (1977) initially reported the phenomenon that people overestimate the proportion of people who possess the same quality (opinion, choice, etc). Dawes (1989) pointed out the main reason for the false consensus effect is that people do not discount the fact that they always observer own signal. The false consensus effect is mainly discussed in the context with binary attributes. Applying Dawes’ reasoning to multinomial/discrete case, however, yields a similar result that people overestimate the likelihood of their own attribute.

I propose a simple model in which individuals overestimate the commonness of own attribute. Once an individual observes own attribute $x = \alpha$, she puts weight $\theta$ on true population density and $(1 - \theta)$ on a kernel density function that has mode at $x = \alpha$. Upon observing own signal, if the individual overestimates how others will possess the same signal as she does, her type-specific density will have the most mass on her attribute. For simplicity, I further assume type-specific density to be symmetric and has the specific form as in Definition 4.

**Definition 1.4** An individual who observes own attribute, $x = \alpha$, possesses False Consensus Effect (FCE) bias if

$$h(x) = \theta f(x) + (1 - \theta)g_{FCE}(x; \alpha, \beta)$$

where,

$$g_{FCE}(x; \alpha, \beta) = \frac{\beta^{x-\alpha}}{\sum_{t=0}^{M} \beta^{\lvert t-\alpha \rvert}}$$

for $\alpha \in \{0, 1, ..., M\}$ and $\beta \in [0, 1]$. 
\[ g_{FCE}(x; \alpha, \beta) \] is constructed as a discrete density that has mode at \( x = \alpha \) and varies its shape from a discrete uniform distribution (when \( \beta = 1 \)) to a degenerate density with the spike at \( x = \alpha \) (when \( \beta = 0 \)). I assume that the variance coefficient \( \beta \) increases as the size of comparison group increases. This assumption is true if subjects anticipate the size of the comparison group and discount their own signal as the comparison group becomes larger.

**Proposition 1.5** *Individual with FCE-type bias who possesses attribute, \( x = \alpha \), overestimates same quality group if \( g_{FCE}(\alpha; \alpha, \beta) > f(\alpha) \).*

Proposition 1.5 holds by definition. If \( g_{FCE}(\alpha; \alpha, \beta) = 1 \) (degenerate density function), then the individual will overestimate SQG regardless of own attribute \( x \). On the other hand, if \( g_{FCE}(\cdot; \alpha, \beta) \) is close to uniform, she will overestimate SQG only when her quality is rare (less than uniform). If \( g_{FCE}(\alpha; \alpha, \beta) \) is large enough over the general population, the overestimation should be common. Moreover, the degree of overestimation and size of true SQG should be negatively correlated as individuals with small SQG are more likely to satisfy the assumption of Proposition 1.5.

**Hypothesis 1.7** *Individuals overestimate same quality group.*

**Hypothesis 1.8** *Overestimation of same quality group has negative correlation with the true size of same quality group.*

1.3.G **Individual Heterogeneity**

It is natural to predict that each individual possesses a different type of bias. By comparing subjects' assessments and actual data from their school
records, individual-level bias can be measured. Even though there is no significant
group level bias, the type of bias of each individual subject can be identified.

**Hypothesis 1.9** Different individuals exhibit different biases.

## 1.4 Results

This section describes the results of the experiment. In Subsection 4.1, I
describe subjects’ accuracy of their knowledge on own school record. Subsection
4.2 and 4.3 summarizes subjects’ bias in assessments of WMTG and SQG, respectiv-
ely. In Subsection 4.4, I report the results of within subject analysis. Finally,
in Subsection 4.5, the hypotheses introduced in Section 3 are tested and evaluated
using the outcome of the experiment.

### 1.4.A Accuracy of own attributes

Table 1.2 and Figure 1.2 summarize the accuracy result on questionnaire
part 1. In questionnaire part 1, subjects were asked to provide the information
on their transcript. Thus, this part evaluates how accurately subjects possessed
knowledge on their own attributes. The first five columns of Table 1.2 show how
many subjects had (if any) errors and in which direction they were. The latter
three columns show absolute mean, mean, and standard error of the assessment
errors respectively.

The subjects generally showed very high accuracy for four main questions
(\(A^+\), \(F\), Biology, and Philosophy). Most of the subjects (91.3%) possessed accurate
assessments on own attributes and only small number of them had errors. How-
ever, the accuracy for three additional questions (A- or higher, C- or lower, and Humanities) was lower. Overall, 73.61% of the assessments were exactly accurate, and 86.71% of them were within error of 1. Performing Wilcoxon’s Sign Rank test, the null hypothesis of the errors being zero is rejected with 5% significance level for (A- or higher and Humanities) questions. Inaccuracy of A- or higher question could be due to the fact that many subjects received many A- or higher grades. On average, subjects had 11.25 grades of A- or higher. One of the reasons for errors on the Humanities question was ambiguity of the category. Since there are many closely related classes, subjects could have mistakenly counted irrelevant classes, although they left some Humanities classes out.

For the purpose of calculating overestimation, the reported attributes were used instead of the true attributes. When subjects made their predictions about the other students and compared with them, it was the reported answer which was used in that process. Therefore, it is more appropriate to use the reported answers to assess subjects’ biases.

Table 1.2: Accuracy of own attributes

<table>
<thead>
<tr>
<th>Question</th>
<th>Err≥ 2</th>
<th>Err= 1</th>
<th>Error= 0</th>
<th>Err= −1</th>
<th>Err≤ −2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>1 (1.4%)</td>
<td>3 (4.2%)</td>
<td>62 (86.1%)</td>
<td>4 (5.6%)</td>
<td>2 (2.8%)</td>
</tr>
<tr>
<td>A- higher</td>
<td>8 (11.1%)</td>
<td>6 (8.3%)</td>
<td>25 (34.7%)</td>
<td>15 (20.8%)</td>
<td>18 (25%)</td>
</tr>
<tr>
<td>C- lower</td>
<td>11 (15.3%)</td>
<td>5 (6.9%)</td>
<td>45 (62.5%)</td>
<td>5 (6.9%)</td>
<td>6 (8.3%)</td>
</tr>
<tr>
<td>F</td>
<td>0 (0%)</td>
<td>1 (1.4%)</td>
<td>68 (94.4%)</td>
<td>1 (1.4%)</td>
<td>2 (2.8%)</td>
</tr>
<tr>
<td>Biology</td>
<td>1 (1.4%)</td>
<td>1 (1.4%)</td>
<td>68 (94.4%)</td>
<td>0 (0%)</td>
<td>2 (2.8%)</td>
</tr>
<tr>
<td>Philosophy</td>
<td>0 (0%)</td>
<td>4 (5.6%)</td>
<td>65 (90.3%)</td>
<td>3 (4.2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Humanities</td>
<td>4 (5.6%)</td>
<td>7 (9.7%)</td>
<td>38 (52.8%)</td>
<td>11 (15.3%)</td>
<td>12 (16.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>25 (5.0%)</td>
<td>27 (5.4%)</td>
<td>371 (73.6%)</td>
<td>39 (7.7%)</td>
<td>42 (8.3%)</td>
</tr>
</tbody>
</table>

Wilcoxon’s Sign Rank Test: significant with 10% (*), 5% (**), and 1% (***) level.

12 Humanities classes include all classes from the department of Philosophy, Literature, Linguistics, and History.
1.4.B Bias on weak-more-than-group (WMTG)

By comparing the reported WMTG \( H_W(x) \) and the true WMTG obtained from the university record \( F_W(x) \), I test whether subjects showed any bias of relative ranking. Frequencies of over/underestimation are summarized in Table 1.3. ‘Observations’ columns reflects the number of subjects who either underestimated or overestimated WMTG respectively. The mean and standard deviation of bias are also reported in ‘Overall’ columns. For example, subjects overestimated WMTG for the \( A^+ \) question by 4.34\% on average.

The biases in \( A^+ \), \( A^- \) or higher, and \( C^- \) or lower questions are small. Also, the results for these three questions are different between the two comparison groups. The results for \( F \), Biology, Philosophy, and Humanities show relatively more systematic directions of biases. More subjects underestimated WMTG for \( F \) question, and this is consistent over both comparison groups. For the Biology, Philosophy, and Humanities questions, the majority of the subjects overestimated WMTG for both comparison groups.

Table 1.3 shows the summary statistics of overestimation as well as the Wilcoxon’s sign-rank test results.\(^{13}\) The results are divided into three cases; the overall results, the results for subjects who answered that they have zero attribute and the results for those who reported to have more than zero attributes. Four grade questions show ambiguous results; both frequencies and the average overestimation do not show any systematic tendency. However, three class questions show relatively strong evidence of overestimation of WMTG. Subjects significantly overestimated WMTG of Biology and Philosophy questions for both comparison groups. The average rates of overestimation are more than 10\% for all four cases.

\(^{13}\)The null hypothesis of Wilcoxon’s sign-rank test is that the sample median equals zero.
Subjects who have one or more attributes showed higher overestimation. Comparing the average bias for ‘answer=0’ and ‘answer>0’ groups in Table 1.3, the average biases of the subjects with ‘answer>0’ are larger in all 14 cases.

1.4.C Bias on same-quality-group (SQG)

The bias in SQG is measured by comparing the true size and subjects’ estimation of SQG. Subjects systematically overestimated the size of SQG. The results are summarized in Table 1.4, Figure 1.3 and Figure 1.4. The ‘Observations’ columns on Table 1.4 show how many subjects either overestimated, were accurate, or underestimated SQG.\textsuperscript{14} The majority of subjects overestimated SQG for all questions in both comparison groups. The Wilcoxon’s sign rank test results are shown in the sixth column of Table 1.4. The mean value denotes average overestimation of SQG. For example, subjects overestimated the proportion of people out of all UCSD students who received same number of A+ grades by 0.14% on average. The results are divided into three cases; the overall result, the result for subjects who had zero attribute, and the result for those who had more than zero attributes. Except the Philosophy question, overestimation of WMTG was positive with various degrees of significance. Similarly with the bias on WMTG, there is a difference between the subjects with zero and more than zero attributes. Comparing ‘answer=0’ and ‘answer>0’ columns in Table 1.4, those with one or more attributes are the ones who overestimated SQG the most.

\textsuperscript{14}Answers accurate to second decimal point were classified as exact.
Table 1.3: Overestimation of the weak-more-than-group (WMTG) depending on answers

<table>
<thead>
<tr>
<th>Variable</th>
<th>All UCSD Observations</th>
<th>Overall</th>
<th>Answer=0</th>
<th>Answer&gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (a)</td>
<td>Overall (b)</td>
<td>Mean</td>
<td>Std.Dev</td>
</tr>
<tr>
<td>A+</td>
<td>72</td>
<td>33 39</td>
<td>4.34</td>
<td>19.08</td>
</tr>
<tr>
<td>A- higher</td>
<td>72</td>
<td>40 32</td>
<td>-0.97</td>
<td>17.96</td>
</tr>
<tr>
<td>C- lower</td>
<td>72</td>
<td>25 47</td>
<td>6.95***</td>
<td>15.88</td>
</tr>
<tr>
<td>F</td>
<td>72</td>
<td>43 29</td>
<td>0.68</td>
<td>11.64</td>
</tr>
<tr>
<td>Biology</td>
<td>72</td>
<td>20 52</td>
<td>10.96***</td>
<td>18.82</td>
</tr>
<tr>
<td>Philosophy</td>
<td>72</td>
<td>23 49</td>
<td>13.49***</td>
<td>19.15</td>
</tr>
<tr>
<td>Humanities</td>
<td>72</td>
<td>29 43</td>
<td>4.87**</td>
<td>21.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Same major Observations</th>
<th>Overall</th>
<th>Answer=0</th>
<th>Answer&gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (a)</td>
<td>Overall (b)</td>
<td>Mean</td>
</tr>
<tr>
<td>A+</td>
<td>72</td>
<td>45 27</td>
<td>1.17</td>
</tr>
<tr>
<td>A- higher</td>
<td>72</td>
<td>35 37</td>
<td>1.42</td>
</tr>
<tr>
<td>C- lower</td>
<td>72</td>
<td>37 35</td>
<td>0.14</td>
</tr>
<tr>
<td>F</td>
<td>72</td>
<td>50 22</td>
<td>-3.36***</td>
</tr>
<tr>
<td>Biology</td>
<td>72</td>
<td>27 45</td>
<td>10.34***</td>
</tr>
<tr>
<td>Philosophy</td>
<td>72</td>
<td>26 46</td>
<td>10.31***</td>
</tr>
<tr>
<td>Humanities</td>
<td>72</td>
<td>30 42</td>
<td>3.17</td>
</tr>
</tbody>
</table>

(a): Frequency of underestimation, (b): Frequency of overestimation

Wilcoxon’s Sign Rank Test *: significant at the 10% level, **: significant at the 5% level, ***: significant at the 1% level
<table>
<thead>
<tr>
<th>All UCSD Variable</th>
<th>Observations</th>
<th>Overall</th>
<th>Answer=0</th>
<th>Answer&gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a) (b) (c)</td>
<td>Mean</td>
<td>Std.Dev</td>
<td>Obs</td>
</tr>
<tr>
<td>A+</td>
<td>32 1 39</td>
<td>0.14</td>
<td>26.35</td>
<td>39</td>
</tr>
<tr>
<td>A- higher</td>
<td>5 0 67</td>
<td>13.43***</td>
<td>11.96</td>
<td>2</td>
</tr>
<tr>
<td>C- lower</td>
<td>19 0 53</td>
<td>7.77***</td>
<td>16.86</td>
<td>25</td>
</tr>
<tr>
<td>F</td>
<td>27 0 45</td>
<td>3.83**</td>
<td>16.76</td>
<td>61</td>
</tr>
<tr>
<td>Biology</td>
<td>30 0 42</td>
<td>4.46</td>
<td>18.66</td>
<td>30</td>
</tr>
<tr>
<td>Philosophy</td>
<td>33 0 39</td>
<td>-6.00</td>
<td>24.86</td>
<td>46</td>
</tr>
<tr>
<td>Humanities</td>
<td>14 0 58</td>
<td>10.89***</td>
<td>13.43</td>
<td>9</td>
</tr>
<tr>
<td>Same major</td>
<td>Observations</td>
<td>Overall</td>
<td>Answer=0</td>
<td>Answer&gt;0</td>
</tr>
<tr>
<td></td>
<td>(a) (b) (c)</td>
<td>Mean</td>
<td>Std.Dev</td>
<td>Obs</td>
</tr>
<tr>
<td>A+</td>
<td>20 0 52</td>
<td>5.71***</td>
<td>22.51</td>
<td>39</td>
</tr>
<tr>
<td>A- higher</td>
<td>8 0 64</td>
<td>11.84***</td>
<td>10.75</td>
<td>2</td>
</tr>
<tr>
<td>C- lower</td>
<td>17 0 55</td>
<td>12.17***</td>
<td>16.83</td>
<td>25</td>
</tr>
<tr>
<td>F</td>
<td>20 0 52</td>
<td>8.07***</td>
<td>17.42</td>
<td>61</td>
</tr>
<tr>
<td>Biology</td>
<td>24 0 48</td>
<td>9.65***</td>
<td>25.58</td>
<td>30</td>
</tr>
<tr>
<td>Philosophy</td>
<td>29 1 42</td>
<td>1.34</td>
<td>24.60</td>
<td>46</td>
</tr>
<tr>
<td>Humanities</td>
<td>16 0 56</td>
<td>14.02***</td>
<td>19.07</td>
<td>9</td>
</tr>
</tbody>
</table>

(a): Frequency of underestimation, (b): Frequency of exact estimation, (c): Frequency of overestimation

Wilcoxon’s Sign Rank Test *: significant at the 10% level, **: significant at the 5% level, ***: significant at the 1% level
Figures 1.3 and 1.4 describe the magnitudes of overestimation. The histograms show the distribution of subjects’ attributes, and the black solid lines denote the true population distribution of attributes. The blue long-dashed lines denote subjects’ average estimation of SQG. The difference between average estimation and true population attributes show the degree of overestimation. Clearly, subjects exhibited overestimation of SQG for most of the cases. Also, we can see those with relatively small SQG (i.e. those whose attribute value is rare) overestimate more.

Subjects overestimate the SQG more when the comparison group is students with the same major than when the comparison group is the whole UCSD population. In Table 1.4, the average overestimation in Part 2 are mostly smaller than those of Part 3. The only exception is A- or higher question. Results for Wilcoxon’s matched pair sign-rank test results are summarized in Table 1.5.\textsuperscript{15} Except A- or higher question, subjects showed more overestimation when they were comparing themselves within the students who had the same major.

<table>
<thead>
<tr>
<th>All</th>
<th>answer=0</th>
<th>answer&gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td>z-value</td>
<td>obs</td>
<td>z-value</td>
</tr>
<tr>
<td>A+</td>
<td>-1.462</td>
<td>72</td>
</tr>
<tr>
<td>A- higher</td>
<td>0.533</td>
<td>72</td>
</tr>
<tr>
<td>C- lower</td>
<td>-1.919*</td>
<td>72</td>
</tr>
<tr>
<td>F</td>
<td>-2.998***</td>
<td>72</td>
</tr>
<tr>
<td>Biology</td>
<td>-1.773*</td>
<td>72</td>
</tr>
<tr>
<td>Philosophy</td>
<td>-2.722***</td>
<td>72</td>
</tr>
<tr>
<td>Humanities</td>
<td>-1.1</td>
<td>72</td>
</tr>
</tbody>
</table>

\*Wilcoxon’s Sign Rank Test: significant with 10\% (*), 5\% (**), and 1\% (***) level

Another noticeable fact is that the average overestimation is higher for\textsuperscript{15} The null hypothesis of Wilcoxon’s matched pair sign-rank test is two matched samples have the same median. The positive z-value in Table 5 means subjects overestimated more overestimation when the comparison group is the whole UCSD population.
those who had one or more attributes. Excluding $A$- or higher, $F$, and Humanities questions which are heavily weighted on one side, the degree of overestimation are higher with those with one or more attributes. The only exception is $A+$ question. Table 1.6 summarizes the results of Mann-Whitney two sample rank sum test.\textsuperscript{16} Figure 1.3 and Figure 1.4 show this is mainly due to the size of the actual SQG. The differences between black solid lines (true SQG) and blue long-dashed lines (average value of estimated SQG) grow when true SQG shrink.

Table 1.6: Overestimation of same-quality-group (SQG) - Answer=0 vs. Answer>0

<table>
<thead>
<tr>
<th></th>
<th>vs. UCSD</th>
<th>vs. Same Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A+$</td>
<td>0.441</td>
<td>2.436**</td>
</tr>
<tr>
<td>A- higher</td>
<td>0.411</td>
<td>1.542</td>
</tr>
<tr>
<td>C- lower</td>
<td>-2.893***</td>
<td>-0.869</td>
</tr>
<tr>
<td>F</td>
<td>-1.309</td>
<td>1.534</td>
</tr>
<tr>
<td>Biology</td>
<td>-0.96</td>
<td>-0.377</td>
</tr>
<tr>
<td>Philosophy</td>
<td>-2.933***</td>
<td>-3.072***</td>
</tr>
<tr>
<td>Humanities</td>
<td>-0.979</td>
<td>-1.473</td>
</tr>
</tbody>
</table>

Wilcoxon’s Sign Rank Test: significant with 10\% (*), 5\% (**), and 1\% (***) level

Subjects with zero $A+$ grade showed significantly more overestimation than those with one or more $A+$ grades in Part 3. $C$- or lower question of Part 2 and Philosophy question in both Parts showed significant overestimation within the same department.

\textsuperscript{16}The null hypothesis of the Mann-Whitney two sample rank sum test is that the median values of two samples are equal. The positive z-value means the subjects with zero attributes showed more severe overestimation of SQG than those who had one or more attributes.
Table 1.7 shows evidence of the negative relation between the overestimation and the size of SQG. Each column shows the correlation between overestimation of SQG and the actual size of SQG in sections 1 and 2 respectively. Except A+ question on section 3, all results show negative correlation between overestimation and the actual size of SQG.  

Table 1.7: Correlation between overestimation and true size of same-quality-group (SQG)

<table>
<thead>
<tr>
<th>Variable</th>
<th>vs. UCSD</th>
<th>vs. Same Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>-0.0848</td>
<td>0.0459</td>
</tr>
<tr>
<td>A- higher</td>
<td>-0.0389</td>
<td>-0.1072</td>
</tr>
<tr>
<td>C- lower</td>
<td>-0.3373</td>
<td>-0.1003</td>
</tr>
<tr>
<td>F</td>
<td>-0.1542</td>
<td>-0.0542</td>
</tr>
<tr>
<td>Biology</td>
<td>-0.2077</td>
<td>-0.3791</td>
</tr>
<tr>
<td>Philosophy</td>
<td>-0.4284</td>
<td>-0.4443</td>
</tr>
<tr>
<td>Humanities</td>
<td>-0.2218</td>
<td>-0.2291</td>
</tr>
<tr>
<td>Section Overall</td>
<td>-0.3188</td>
<td>-0.2355</td>
</tr>
<tr>
<td>Questions Overall</td>
<td></td>
<td>-0.2705</td>
</tr>
</tbody>
</table>

1.4.D Individual Heterogeneity

Individuals can possess heterogeneous types of biased self image. As I compare and identify some of the possible theories of bias in aggregate level, I can also consider the bias at the individual level. First of all, I use the answers for eight main questions (A+, F, Biology, and Philosophy question in both Parts) to test whether each subject can be classified as TE/EBU type. In order to test whether a subject can be classified as PSI type, eight questions that are either desired (A+ and A- or higher in both Parts) or undesired (C- or lower and F in both Parts) are used. Finally, subjects were tested whether they are FCE-type using their bias in SQG for all 14 questions.

One can also argue that this effect is due to the regression effect. Suppose $\delta$ is the average of true SQG. Assume an extreme case that all subjects report the same value, $\delta$. Then we will obtain a similar negative correlation. Unfortunately, I could not find a way to separate the regression effect.
Table 1.9 is a summary of the individual level result. The ‘ERROR’ column shows the total number of errors and their signs subjects made in assessing their attributes. The next sub-columns under ‘TE/EBU’, ‘PSI’, and ‘FCE’ columns show how likely are subjects to be classified as one of those possible types. In the ‘BIAS’ column, we can see the number of answers that are supportive or un-supportive to be classified as relevant types. The ‘AVG’ columns pairs average bias with the corresponding p-value result from Wilcoxon’s sign rank test. The final column reports the categorization result.

As we can see in Table 1.9, some subjects’ types are easily distinguishable. For example, subject no. 1153 overestimated WMTG on all 8 main questions. He/She overestimated WMTG by 33.39% on average, and none of the other effects can explain this bias. Therefore, subject no. 1153 can be classified as having the bias consistent with the TE/EBU hypothesis. Subject no. 2105 shows a strong tendency to overestimate SQG while the other data indicate other hypotheses cannot provide persuasive explanation. Thus, subject no. 2105 can be considered as the type who tends to overestimate SQG. Subject no. 8031 can be easily considered as a PSI-type. He/She underestimated the number of people with more high grades and overestimated the people with more lower grades. There were also subjects like no. 8122 who did not show any systematic bias.

Among those who cannot be classified in an obvious way, categorization was made according to the following rules:

- When a subject is significantly consistent with only one bias type, he/she is classified as that specific type
- When a subject is significantly supportive as more than one types, he/she is classified as the type with higher significance level

\[18\] See Appendix.
• When there is a two-way-tie in significance level, he/she is counted as 1/2 for both types

• When a subject behavior is inconsistent with all hypotheses, he/she is considered to possess no bias.\(^{19}\)

Although many of the subjects can be categorized as certain types, some of the subjects cannot be classified into any category. Using the Wilcoxon’s sign-rank test within subject, I classify the individual subjects into their types. Table 1.8 shows the result of individual categorization. Each column shows the number of individuals who were classified as a certain type at the 1%, 5%, and 10% significance levels. Since there are limited number of questions for some bias types, the tests are not powerful. For example, the lowest possible significance based on overestimation of WMTG was 5%.

<table>
<thead>
<tr>
<th>Bias Type</th>
<th>Frequencies for Each Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>No Significant Bias (NB)</td>
<td>62 (86.11%)</td>
</tr>
<tr>
<td>Bias in WMTG (TE/EBU)</td>
<td>NA</td>
</tr>
<tr>
<td>Bias in own quality (PSI)</td>
<td>NA</td>
</tr>
<tr>
<td>Bias in SQG (FCE)</td>
<td>10 (13.89%)</td>
</tr>
</tbody>
</table>


58.33% of the subjects fit into one of the theories of bias with 5% significance level, Raising the significance level to 10% level. 73.61% of the subjects can be categorized to exhibit one of the bias types. Among those who can be classified at the 10% significance level, 22.92% were TE/EBU type, 7.64% were PSI type and 43.06% were FCE type.

\(^{19}\)For example, since I did not have model of negative self image, those who systematically underestimated their relative quality were classified as not applicable.
1.4.E Discussions of Results

In this section, I revisit the hypotheses suggested in Section 4 and evaluate the validity of them with the experimental results.

**Hypothesis 1.1** tests that there will be no systematic bias since the attributes are objective and unambiguous. The experimental results, however, suggest subjects do show systematic bias. Estimating WMTG, the result is not decisive. Subjects show strong overestimation (*Biology* and *Philosophy*), marginal overestimation (*C- or higher* and *Humanities*), marginal underestimation (*F*), or insignificant biases (*A+* and *A- or lower*). The result for biases on SQG is more significant overall; They show systematic bias on all questions except *philosophy* question which has insignificant result. Moreover, analyzing individual level bias, almost 60% of the subjects show systematic bias at the 5% significance level.

Under the assumption of truncated errors with skewed density, **Hypothesis 1.2** tested the subjects to overestimate WMTG for all main questions. If subjects observe their trait accurately but those of others with error, then they will overestimate WMTG. However, the data show mixed results. Subjects show overestimation on *Biology* and *Philosophy* questions, but overestimation of *A+* questions are insignificant. Furthermore, they even underestimate WMTG on *F* question when they were comparing themselves among the students with same major. Consequently, we can not attain a strong result that supports the hypothesis of truncated errors with skewed density.

Like Hypothesis 1.2, **Hypothesis 1.3** tests subjects to overestimate WMTG for all main questions. If subjects follow egocentric Bayesian update rule, and attribute their fewer than expected attribute to both wrong expectation ("It is harder than expected to get A+ grades") and luck ("I was unlucky to receive
fewer A+ grades”), then they will overestimate WMTG for the attributes with positively skewed density. The data does not confirm this hypothesis.

**Hypothesis 1.4** tests individuals do not possess accurate knowledge of attributes on own school record. Benoît & Dubra (2007) suggested many of the existing results of overconfidence are due to the fact that individuals are not aware of their own attributes. Observing only signals of own attributes, ‘apparent’ bias can arise for rational individuals. Subjects mostly possess accurate knowledge about own attributes. 73.61% of the overall answers are correct and 86.71% of them are within an error of 1. This result refutes Hypothesis 1.4.

According to **Hypothesis 1.5** tests subjects will possess no systematic bias to questions for which they do possess accurate information. However, the result is opposite of the hypothesis. For some of the attributes that subjects possess relatively accurate knowledge (Biology and Philosophy questions), they show significant overestimation of WMTG. On the other hand, for two attributes that subjects possess the most inaccurate knowledge (A- or lower and Humanities questions), they do not have huge bias.

**Hypothesis 1.6** tests subjects will possess positive self image over all questions related to desirability. More precisely, if subjects systematically overestimate their relative achievements to those of others, they will overestimate WMTG for undesired grades (F and C- or lower), underestimate WMTG for desired grades (A+ and A- or higher), and show no systematic bias for the class questions (Biology, Philosophy, and Humanities). The data suggest subjects do not overestimate own achievements. The result for A+ and A- or higher questions are insignificant. Subjects overestimate WMTG for C- or lower question when they are assessing themselves among the whole UCSD population. This result is consistent with the hypothesis of positive self image. On the other hand, they show underestimation
of WMTG on $F$ question against the students with same major. This means they underestimated the number of people who received more F grades. If people show bias by overestimating their own quality, they should have overestimated those who received more F grades so that their relative quality is better than they really are. Consequently, the positive self image hypothesis is not supported.

As predicted in **Hypothesis 1.7**, I find strong evidence that subjects overestimated SQG. Collectively, for all 14 questions, there are more subjects who overestimate the SQG than those who underestimate it. In 10 out of 14 cases, subjects show significant overestimation, and there is only one case (Philosophy question in Part 2) where subjects’ average overestimation is negative. They also tend to overestimate more when they compared themselves within the same department. This finding is consistent with the theory since the effect of own signal should be discounted as the comparison group becomes larger.

The data also indicate that subjects who possess common attributes show relatively less overestimation of SQG. Figures 3 and 4 show that the difference between their assessments and the true size of SQG is bigger for those who possess rare qualities. Also, the overestimation is negatively correlated with the actual size of SQG. These observations support **Hypothesis 1.8**.

**Hypothesis 1.9** tests biases vary across individuals. Disaggregating the data demonstrates that individuals do show various systematic bias. Using the Wilcoxon’s sign-rank test within subject, I classify the individual subjects into their types. If a subject shows significant results for more than one type, it is counted only for the category that best described the data. At the 10% significance level, 43.06% were classified as FCE types, 7.64% showed PSI type bias, 22.92% showed TE/EBU type of bias, and 26.39% of subjects could not be categorized
according to one of the suggested theories. The individual level result coincides with the aggregate level result as the most common bias type is FCE bias.

1.5 Conclusion

I tested whether individuals possess biased self image over objective and unambiguous attributes. University students who participated in the experiment made assessments of their school record on transcripts. They also predicted the same attributes of other students compared to own record. By comparing the provided answers and the true attributes, I obtained the bias in their assessments.

Summarizing the results, the first clear outcome is that subjects’ assessment of own attributes are fairly accurate. The margin of error shows subjects did possess accurate knowledge about their school record. The data also clearly refute the hypothesis that subjects possess positive self image. The data suggests subjects did not overestimate one’s quality compared to others. On the contrary, for some of the attributes, they showed the opposite bias, negative self image. Benoît & Dubra’s critique on possibility of apparent biased self image does not apply to my results. Contrary to their assumption, subjects in this experiment possessed accurate knowledge of own attributes, yet they still possessed systematic bias. Moreover, subjects displayed statistically insignificant bias on the questions with the highest error.

I found limited evidence that subjects overestimate weak-more-than-group. They significantly overestimated weak-more-than-group for the questions regarding the classes they have taken in University. On the questions regarding

\(^{20}\)Since there are limited number of questions for some bias types, the tests are not powerful. For example, the lowest possible significance based on overestimation of WMTG is 5%.
their grades, however, they did not show systematic overestimation. The fact that I could not find compelling evidence that subjects overestimate weak-more-than-group raises the question on existence of biased beliefs in objective qualities. However, the data also strongly rejects the possibility that individuals possess random bias. Among the possible theories, the data showed the most support on overestimation of same-quality-group. Subjects systematically overestimated the proportion of people who possess the same attributes with them.

The finding that subjects overestimate the same-quality-group generalizes the false consensus effect of binary attributes. In my experiment, subjects showed overestimation even when the attributes are discretely distributed. This can be considered as ‘generalized’ version of false consensus effect as the subjects overestimate the proportion of the group which possess the same “quantity”. I am not aware of similar reports in the literature. This finding suggests there should be further study on this phenomenon.

The data indicate different individuals exhibit different biases. Conducting within subject analysis, 58.33% of subjects show one of the bias types at the 5% significance level. At the 10% significance level, 73.61% of subjects possess biased belief. Among subjects with significant individual-level bias, false consensus effect bias is the most common type with 43.06% at the 10% significance level.

There are several ways to extend the experiment. The experiment was initially designed to test the overestimation of the weak-more-than-group. Therefore, alternative design of experiment to test overestimation of same-quality-group can provide more concrete evidence. The payment rule in this experiment was not directly correlated with the accuracy of estimating the same-quality-group. The payment was rewarded for the accuracy of estimating the weak-more-than-group, and the weak-more-than-group is proportional to same-quality-group. Also,
one can argue that the result stems from the fact that subjects were given only three categories (more-than, same-quality, and fewer-than groups). Dividing a large domain of attribute into only three regions and assigning one of them to the same-quality-group could artificially increase the measured false consensus effect. Finally, it is possible that the regression effect played a partial role in the outcome. As we can see from Table 1.7, subjects who had rare qualities were the ones who overestimated the most. It is possible that part of this comes from the regression effect. Any kind of overestimation calculated in this experiment was calculated by subtracting actual value from the reported assessment. Therefore, the subjects with low actual attributes are the ones who are more likely to overestimate.\footnote{Two alternative measures of bias were tried to lessen the problem of regression effect. The first alternative is Relative Bias that takes account of the possible room for bias. Relative bias is equal to the actual bias divided by possible room for bias. The second method is Attribute based measure. Attribute based bias measure is calculated as the difference of the true attribute and the ‘would be’ attribute according to their estimation. For both cases the overall results hold with little differences.}
1.6 References


Appendix. Proof, Tables and Figures

Proof of Proposition 1.2

\[ H(x^*) - F(x^*) = \sum_{x=0}^{x^*} \sum_{t=0}^{M-x} e(t) (f(t + x) - f(x)) \]

\[ = \sum_{x=0}^{x^*} \left\{ \sum_{t=1}^{M-x} e(t)(f(t + x) - f(x)) + \sum_{t=-x}^{-1} e(t)(f(t + x) - f(x)) \right\} \]

\[ = \sum_{x=0}^{x^*} \left\{ \sum_{t=1}^{M-x} e(t)(f(t + x) - f(x)) - \sum_{t=1}^{x} e(t)(f(x) - f(x - t)) \right\} \]

substituting \( u = x - t \),

\[ H(x^*) - F(x^*) = \sum_{x=0}^{x^*} \sum_{t=1}^{M-x} e(t)(f(t + x) - f(x)) - \sum_{t=1}^{x^*-t} e(t)(f(t + u) - f(u)) \]

\[ = \sum_{x=0}^{x^*} \left\{ \sum_{t=1}^{M-x} e(t)(f(t + x) - f(x)) - \sum_{u=0}^{x^*-1} e(t)(f(t + u) - f(u)) \right\} \]

\[ = \sum_{x=0}^{x^*} \sum_{t=1}^{M-x} e(t)(f(t + x) - f(x)) - \sum_{u=0}^{x^*-1} \sum_{t=1}^{x^*-u} e(t)(f(t + u) - f(u)) \]

\[ = \sum_{x=0}^{x^*} \left\{ \sum_{t=1}^{M-x} e(t)(f(t + x) - f(x)) - \sum_{t=1}^{x^*-x} e(t)(f(t + x) - f(x)) \right\} \]

\[ = \sum_{x=0}^{x^*} \sum_{t=x^*-x+1}^{M-x} e(t)(f(t + x) - f(x)) \]

This is a non-negatively weighted average of terms of the form of \( f(t + x) - f(x) \), for \( t > 0 \).

If \( f(x) \) is strictly increasing (non decreasing), then the sum is positive (non negative), and if \( f(x) \) is strictly decreasing (non increasing), then the sum is negative (non positive). \( \blacksquare \)
Table 1.9: Individual Heterogeneity

<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR</th>
<th>BIAS</th>
<th>AVG</th>
<th>BIAS</th>
<th>AVG</th>
<th>BIAS</th>
<th>AVG</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(+/-)</td>
<td>(+/-)</td>
<td>(+/-)</td>
<td>p-value</td>
<td>(+/-)</td>
<td>(+/-)</td>
<td>p-value</td>
</tr>
<tr>
<td>1</td>
<td>1016</td>
<td>Chem</td>
<td>1(0/1)</td>
<td>7/1</td>
<td>17.83</td>
<td>2/6</td>
<td>-14.23</td>
<td>9/5</td>
<td>2.88</td>
<td>T/E**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1023</td>
<td>Chem</td>
<td>2(0/2)</td>
<td>2/6</td>
<td>-6.80</td>
<td>2/6</td>
<td>-10.07</td>
<td>10/4</td>
<td>10.78</td>
<td>FCE**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1036</td>
<td>Chem</td>
<td>4(3/1)</td>
<td>6/2</td>
<td>12.72</td>
<td>1/7</td>
<td>-6.91</td>
<td>12/2</td>
<td>6.05</td>
<td>FCE**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1049</td>
<td>Chem</td>
<td>2(1/1)</td>
<td>2/6</td>
<td>-3.30</td>
<td>5/3</td>
<td>4.02</td>
<td>11/3</td>
<td>13.73</td>
<td>FCE***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1052</td>
<td>Chem</td>
<td>2(0/2)</td>
<td>7/1</td>
<td>26.83</td>
<td>6/2</td>
<td>10.48</td>
<td>10/4</td>
<td>3.29</td>
<td>T/E**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1069</td>
<td>Chem</td>
<td>1(0/1)</td>
<td>2/6</td>
<td>-9.00</td>
<td>4/4</td>
<td>4.37</td>
<td>11/3</td>
<td>16.26</td>
<td>FCE***</td>
</tr>
</tbody>
</table>

Continued on next page
Table 1.9 : Individual Heterogeneity, Continued

<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR</th>
<th>TE/EBU BIAS (+/-)</th>
<th>TE/EBU AVG</th>
<th>PSI BIAS (+/-)</th>
<th>PSI AVG</th>
<th>FCE BIAS (+/-)</th>
<th>FCE AVG</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1070</td>
<td>Chem</td>
<td>2(1/1)</td>
<td>5/3</td>
<td>6.76</td>
<td>5/3</td>
<td>3.59</td>
<td>9/5</td>
<td>2.44</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.161</td>
<td>0.575</td>
<td></td>
<td>0.470</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1116</td>
<td>Chem</td>
<td>1(0/1)</td>
<td>5/3</td>
<td>7.57</td>
<td>2/6</td>
<td>-10.45</td>
<td>11/3</td>
<td>5.49</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.208</td>
<td>0.036</td>
<td></td>
<td>0.140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1126</td>
<td>Chem</td>
<td>3(2/1)</td>
<td>6/2</td>
<td>22.43</td>
<td>1/7</td>
<td>-23.28</td>
<td>13/1</td>
<td>27.42</td>
<td>FCE***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.036</td>
<td>0.017</td>
<td></td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1132</td>
<td>Chem</td>
<td>1(0/1)</td>
<td>6/2</td>
<td>9.55</td>
<td>0/8</td>
<td>-7.54</td>
<td>12/2</td>
<td>7.13</td>
<td>T/E*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.093</td>
<td>0.012</td>
<td></td>
<td>0.109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1141</td>
<td>Chem</td>
<td>0(0/0)</td>
<td>3/5</td>
<td>-0.43</td>
<td>5/3</td>
<td>-0.16</td>
<td>12/2</td>
<td>15.98</td>
<td>FCE**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.889</td>
<td></td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1153</td>
<td>Chem</td>
<td>3(0/3)</td>
<td>8/0</td>
<td>33.39</td>
<td>4/4</td>
<td>-2.82</td>
<td>7/7</td>
<td>-1.75</td>
<td>T/E**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.012</td>
<td>0.674</td>
<td></td>
<td>0.925</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2013</td>
<td>Poli</td>
<td>0(0/0)</td>
<td>5/3</td>
<td>2.39</td>
<td>5/3</td>
<td>0.16</td>
<td>8/6</td>
<td>7.22</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.575</td>
<td>0.779</td>
<td></td>
<td>0.198</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR (+/-)</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>TE/EBU</th>
<th>PSI</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>FCE</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>2021</td>
<td>Econ</td>
<td>0(0/0)</td>
<td>8/0</td>
<td>29.68</td>
<td>4/4</td>
<td>-3.94</td>
<td>9/5</td>
<td>-0.08</td>
<td>T/E**</td>
<td>0.012</td>
<td>0.674</td>
<td>0.397</td>
</tr>
<tr>
<td>15</td>
<td>2042</td>
<td>Comm</td>
<td>1(1/0)</td>
<td>7/1</td>
<td>23.13</td>
<td>3/5</td>
<td>-9.68</td>
<td>8/6</td>
<td>6.60</td>
<td>T/E**</td>
<td>0.025</td>
<td>0.208</td>
<td>0.286</td>
</tr>
<tr>
<td>16</td>
<td>2061</td>
<td>Econ</td>
<td>3(1/2)</td>
<td>3/5</td>
<td>0.46</td>
<td>3/5</td>
<td>-6.07</td>
<td>10/4</td>
<td>5.76</td>
<td>-</td>
<td>0.674</td>
<td>0.401</td>
<td>0.198</td>
</tr>
<tr>
<td>17</td>
<td>2075</td>
<td>ECE</td>
<td>4(3/1)</td>
<td>3/5</td>
<td>-0.18</td>
<td>6/2</td>
<td>17.47</td>
<td>8/6</td>
<td>0.37</td>
<td>PSI*</td>
<td>1.000</td>
<td>0.093</td>
<td>0.875</td>
</tr>
<tr>
<td>18</td>
<td>2085</td>
<td>ECE</td>
<td>4(1/3)</td>
<td>4/4</td>
<td>1.37</td>
<td>6/2</td>
<td>5.09</td>
<td>6/8</td>
<td>-6.66</td>
<td>-</td>
<td>0.889</td>
<td>0.161</td>
<td>0.300</td>
</tr>
<tr>
<td>19</td>
<td>2095</td>
<td>ECE</td>
<td>2(2/0)</td>
<td>4/4</td>
<td>6.74</td>
<td>1/7</td>
<td>-11.18</td>
<td>12/2</td>
<td>7.45</td>
<td>FCE**</td>
<td>0.575</td>
<td>0.036</td>
<td>0.022</td>
</tr>
<tr>
<td>20</td>
<td>2105</td>
<td>ECE</td>
<td>2(0/2)</td>
<td>2/6</td>
<td>-5.50</td>
<td>6/2</td>
<td>5.73</td>
<td>14/0</td>
<td>17.59</td>
<td>FCE***</td>
<td>0.161</td>
<td>0.208</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Continued on next page
Table 1.9: Individual Heterogeneity, Continued

<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>p-value</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>p-value</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>p-value</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>2116</td>
<td>MAE</td>
<td>1(0/1)</td>
<td>2/6</td>
<td>-2.67</td>
<td>4/4</td>
<td>1.55</td>
<td>11/3</td>
<td>10.30</td>
<td></td>
<td></td>
<td></td>
<td>FCE**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.674</td>
<td>0.575</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3011†</td>
<td>Econ</td>
<td>1(0/1)</td>
<td>2/6</td>
<td>-6.18</td>
<td>4/4</td>
<td>0.54</td>
<td>12/2</td>
<td>18.76</td>
<td></td>
<td></td>
<td></td>
<td>FCE***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.093</td>
<td>0.889</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3021</td>
<td>Econ</td>
<td>0(0/0)</td>
<td>3/5</td>
<td>1.15</td>
<td>4/4</td>
<td>8.94</td>
<td>12/2</td>
<td>6.14</td>
<td></td>
<td></td>
<td></td>
<td>FCE*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.779</td>
<td>0.263</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3033</td>
<td>Poli</td>
<td>0(0/0)</td>
<td>0/8</td>
<td>-8.59</td>
<td>2/6</td>
<td>-10.64</td>
<td>11/3</td>
<td>11.08</td>
<td></td>
<td></td>
<td></td>
<td>FCE**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.012</td>
<td>0.161</td>
<td>0.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>3041</td>
<td>Econ</td>
<td>4(1/3)</td>
<td>1/7</td>
<td>-2.66</td>
<td>4/4</td>
<td>5.54</td>
<td>12/2</td>
<td>4.04</td>
<td></td>
<td></td>
<td></td>
<td>FCE*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.161</td>
<td>1.000</td>
<td>0.074</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>3056</td>
<td>MAE</td>
<td>1(0/1)</td>
<td>7/1</td>
<td>17.04</td>
<td>2/6</td>
<td>-10.28</td>
<td>9/5</td>
<td>6.91</td>
<td></td>
<td></td>
<td></td>
<td>T/E**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.017</td>
<td>0.069</td>
<td>0.140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>3066</td>
<td>MAE</td>
<td>1(1/0)</td>
<td>4/4</td>
<td>1.02</td>
<td>5/3</td>
<td>4.61</td>
<td>10/4</td>
<td>6.28</td>
<td></td>
<td></td>
<td></td>
<td>FCE*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.401</td>
<td>0.096</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
Table 1.9: Individual Heterogeneity, Continued

<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR</th>
<th>BIAS (+/-)</th>
<th>AVG p-value</th>
<th>TE/EBU BIAS (+/-)</th>
<th>AVG p-value</th>
<th>PSI BIAS (+/-)</th>
<th>AVG p-value</th>
<th>FCE BIAS (+/-)</th>
<th>AVG p-value</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>3075</td>
<td>ECE</td>
<td>1(0/1)</td>
<td>5/3</td>
<td>3.92</td>
<td>7/1</td>
<td>7.21</td>
<td>7/7</td>
<td>-5.82</td>
<td>PSI**</td>
<td>0.484</td>
<td>0.017</td>
</tr>
<tr>
<td>29</td>
<td>3086</td>
<td>MAE</td>
<td>1(0/1)</td>
<td>6/2</td>
<td>16.42</td>
<td>2/6</td>
<td>-12.07</td>
<td>11/3</td>
<td>13.63</td>
<td>T/E**</td>
<td>0.036</td>
<td>0.124</td>
</tr>
<tr>
<td>30</td>
<td>3096</td>
<td>MAE</td>
<td>1(1/0)</td>
<td>4/4</td>
<td>10.28</td>
<td>6/2</td>
<td>2.98</td>
<td>12/2</td>
<td>8.12</td>
<td>FCE**</td>
<td>0.401</td>
<td>0.674</td>
</tr>
<tr>
<td>31</td>
<td>3105</td>
<td>ECE</td>
<td>2(1/1)</td>
<td>8/0</td>
<td>27.00</td>
<td>6/2</td>
<td>15.97</td>
<td>6/8</td>
<td>-13.26</td>
<td>T/E**</td>
<td>0.012</td>
<td>0.124</td>
</tr>
<tr>
<td>32</td>
<td>3113</td>
<td>Poli</td>
<td>3(0/3)</td>
<td>3/5</td>
<td>3.43</td>
<td>4/4</td>
<td>7.64</td>
<td>10/4</td>
<td>6.08</td>
<td>-</td>
<td>0.779</td>
<td>0.263</td>
</tr>
<tr>
<td>33</td>
<td>4011</td>
<td>Poli</td>
<td>1(0/1)</td>
<td>4/4</td>
<td>0.02</td>
<td>4/4</td>
<td>-0.51</td>
<td>10/4</td>
<td>8.19</td>
<td>-</td>
<td>0.889</td>
<td>0.889</td>
</tr>
<tr>
<td>34</td>
<td>4033</td>
<td>Poli</td>
<td>2(1/1)</td>
<td>5/3</td>
<td>8.21</td>
<td>1/7</td>
<td>-18.69</td>
<td>11/3</td>
<td>11.82</td>
<td>FCE***</td>
<td>0.208</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Continued on next page
Table 1.9: Individual Heterogeneity, Continued

<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR (+/⁻)</th>
<th>TE/EBU BIAS (+/⁻)</th>
<th>PSI BIAS (+/⁻)</th>
<th>FCE BIAS (+/⁻)</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>4041</td>
<td>ECE</td>
<td>0(0/0)</td>
<td>7/1 20.61 (±0.025)</td>
<td>4/4 -0.90 (1.000)</td>
<td>6/8 -5.24 (0.730)</td>
<td>T/E**</td>
</tr>
<tr>
<td>36</td>
<td>4051</td>
<td>Econ</td>
<td>0(0/0)</td>
<td>7/1 0.92 (0.161)</td>
<td>4/4 0.57 (1.000)</td>
<td>8/6 1.80 (0.470)</td>
<td>-</td>
</tr>
<tr>
<td>37</td>
<td>4061</td>
<td>Econ</td>
<td>1(0/1)</td>
<td>8/0 17.10 (0.012)</td>
<td>5/3 3.41 (0.484)</td>
<td>8/6 -8.44 (0.331)</td>
<td>T/E**</td>
</tr>
<tr>
<td>38</td>
<td>4071</td>
<td>Econ</td>
<td>1(0/1)</td>
<td>0/8 -10.80 (0.012)</td>
<td>5/3 7.01 (0.093)</td>
<td>11/3 12.71 (0.036)</td>
<td>FCE**</td>
</tr>
<tr>
<td>39</td>
<td>4085</td>
<td>ECE</td>
<td>1(0/1)</td>
<td>1/7 -6.06 (0.161)</td>
<td>5/3 2.71 (0.674)</td>
<td>11/3 15.02 (0.016)</td>
<td>FCE**</td>
</tr>
<tr>
<td>40</td>
<td>4091</td>
<td>Econ</td>
<td>2(2/0)</td>
<td>5/3 4.92 (0.401)</td>
<td>0/8 -15.02 (0.012)</td>
<td>8/6 6.67 (0.433)</td>
<td>-</td>
</tr>
<tr>
<td>41</td>
<td>4105</td>
<td>ECE</td>
<td>2(1/1)</td>
<td>3/5 -2.21 (0.484)</td>
<td>8/0 12.65 (0.012)</td>
<td>13/1 13.59 (0.004)</td>
<td>FCE***</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR (+/-)</th>
<th>TE/EBU BIAS (+/-)</th>
<th>TE/EBU AVG p-value</th>
<th>PSI BIAS (+/-)</th>
<th>PSI AVG p-value</th>
<th>FCE BIAS (+/-)</th>
<th>FCE AVG p-value</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>5011</td>
<td>Poli</td>
<td>2(0/2)</td>
<td>1/7 -12.36</td>
<td>0.025</td>
<td>5/3 6.13</td>
<td>0.208</td>
<td>10/4 10.23</td>
<td>0.084</td>
<td>-</td>
</tr>
<tr>
<td>43</td>
<td>5021</td>
<td>Econ</td>
<td>1(0/1)</td>
<td>2/6 -0.48</td>
<td>0.575</td>
<td>4/4 -2.36</td>
<td>0.014</td>
<td>11/2 5.55</td>
<td>FCE**</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>5033</td>
<td>MAE</td>
<td>1(0/1)</td>
<td>6/2 22.91</td>
<td>0.069</td>
<td>1/7 -12.76</td>
<td>0.025</td>
<td>6/8 -0.39</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>5041</td>
<td>Econ</td>
<td>4(0/4)</td>
<td>4/4 -5.67</td>
<td>0.263</td>
<td>2/6 -12.05</td>
<td>0.124</td>
<td>9/4 9.60</td>
<td>FCE**</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>5056</td>
<td>MAE</td>
<td>2(1/1)</td>
<td>5/3 7.16</td>
<td>0.484</td>
<td>6/2 16.18</td>
<td>0.069</td>
<td>8/6 6.33</td>
<td>PSI*</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>6021</td>
<td>Econ</td>
<td>2(1/1)</td>
<td>2/6 1.17</td>
<td>0.575</td>
<td>5/3 0.23</td>
<td>0.779</td>
<td>10/4 -0.73</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>48</td>
<td>6031</td>
<td>Econ</td>
<td>3(3/0)</td>
<td>8/0 22.37</td>
<td>0.012</td>
<td>3/5 -5.68</td>
<td>0.208</td>
<td>9/5 10.48</td>
<td>T/E**</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR (+/-)</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>p-value</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>p-value</th>
<th>BIAS (+/-)</th>
<th>AVG</th>
<th>p-value</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>6041</td>
<td>Econ 2(1/2)</td>
<td>5/3</td>
<td>7.69</td>
<td>6/2</td>
<td>9.17</td>
<td>9/5</td>
<td>8.20</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>6051</td>
<td>Econ 2(1/1)</td>
<td>6/2</td>
<td>13.29</td>
<td>4/4</td>
<td>0.41</td>
<td>12/2</td>
<td>11.29</td>
<td>FCE***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>6061</td>
<td>Econ 1(0/1)</td>
<td>0/8</td>
<td>-16.40</td>
<td>2/6</td>
<td>-7.29</td>
<td>10/4</td>
<td>12.82</td>
<td>FCE*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>6072</td>
<td>Comm 2(0/2)</td>
<td>4/4</td>
<td>4.23</td>
<td>6/2</td>
<td>4.09</td>
<td>10/4</td>
<td>11.09</td>
<td>FCE*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>6086</td>
<td>MAE 3(2/1)</td>
<td>3/5</td>
<td>4.69</td>
<td>2/6</td>
<td>-9.67</td>
<td>14/0</td>
<td>14.12</td>
<td>FCE***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>6101</td>
<td>Econ 2(1/1)</td>
<td>2/6</td>
<td>-2.59</td>
<td>4/4</td>
<td>1.33</td>
<td>9/5</td>
<td>6.61</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>7011</td>
<td>Econ 2(1/1)</td>
<td>4/4</td>
<td>10.12</td>
<td>3/5</td>
<td>-0.02</td>
<td>8/6</td>
<td>4.80</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
### Table 1.9: Individual Heterogeneity, Continued

<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR</th>
<th>TE/EBU BIAS</th>
<th>TE/EBU AVG</th>
<th>PSI BIAS</th>
<th>PSI AVG</th>
<th>FCE BIAS</th>
<th>FCE AVG</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(+/-)</td>
<td>(+/-) AVG p-value</td>
<td>(+/-) AVG p-value</td>
<td>(+/-) AVG p-value</td>
<td>(+/-) AVG p-value</td>
<td>(+/-) AVG p-value</td>
<td>(+/-) AVG p-value</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>7021</td>
<td>Econ</td>
<td>2(2/0)</td>
<td>8/0   18.86</td>
<td>4/4  1.45</td>
<td>8/6 -3.80</td>
<td>0.012</td>
<td>1.000</td>
<td>0.826</td>
<td>T/E**</td>
</tr>
<tr>
<td>57</td>
<td>7031</td>
<td>Econ</td>
<td>1(0/1)</td>
<td>6/2  2.39</td>
<td>2/6 -7.76</td>
<td>8/6 4.64</td>
<td>0.484</td>
<td>0.050</td>
<td>0.414</td>
<td>-</td>
</tr>
<tr>
<td>58</td>
<td>7043</td>
<td>Poli</td>
<td>2(1/1)</td>
<td>1/7 -9.49</td>
<td>4/4 -0.36</td>
<td>12/2 16.03</td>
<td>0.025</td>
<td>1.000</td>
<td>0.006</td>
<td>FCE***</td>
</tr>
<tr>
<td>59</td>
<td>7056</td>
<td>MAE</td>
<td>2(1/1)</td>
<td>2/6 -0.06</td>
<td>3/5 -4.30</td>
<td>11/3 8.85</td>
<td>0.674</td>
<td>0.484</td>
<td>0.045</td>
<td>FCE**</td>
</tr>
<tr>
<td>60</td>
<td>7061</td>
<td>Econ</td>
<td>3(3/0)</td>
<td>3/5  0.48</td>
<td>5/3  9.54</td>
<td>11/3 11.16</td>
<td>0.779</td>
<td>0.208</td>
<td>0.041</td>
<td>FCE**</td>
</tr>
<tr>
<td>61</td>
<td>7071</td>
<td>Econ</td>
<td>4(0/4)</td>
<td>5/3  10.35</td>
<td>1/7 -4.19</td>
<td>12/2 4.73</td>
<td>0.124</td>
<td>0.017</td>
<td>0.096</td>
<td>FCE*</td>
</tr>
<tr>
<td>62</td>
<td>7081</td>
<td>Econ</td>
<td>2(2/0)</td>
<td>6/2  25.93</td>
<td>3/5 -9.07</td>
<td>11/3 15.04</td>
<td>0.036</td>
<td>0.575</td>
<td>0.048</td>
<td>FCE</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR</th>
<th>TE/EBU</th>
<th>PSI</th>
<th>FCE</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>BIAS</td>
<td>AVG</td>
<td>BIAS</td>
<td>AVG</td>
<td>AVG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(+/-)</td>
<td>p-value</td>
<td>(+/-)</td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>63</td>
<td>8011</td>
<td>Econ</td>
<td>3(2/1)</td>
<td>6/2</td>
<td>3.96</td>
<td>8/0</td>
<td>11.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.484</td>
<td>0.012</td>
<td>0.363</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>8021</td>
<td>Econ</td>
<td>2(0/2)</td>
<td>8/0</td>
<td>17.22</td>
<td>5/3</td>
<td>-3.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.012</td>
<td>0.889</td>
<td>0.246</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>8031</td>
<td>Econ</td>
<td>2(1/1)</td>
<td>6/2</td>
<td>16.63</td>
<td>8/0</td>
<td>21.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.069</td>
<td>0.012</td>
<td>0.246</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>8041</td>
<td>Econ</td>
<td>3(0/3)</td>
<td>1/7</td>
<td>-7.18</td>
<td>4/4</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.025</td>
<td>0.779</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>8051</td>
<td>Econ</td>
<td>2(2/0)</td>
<td>2/6</td>
<td>3.28</td>
<td>5/3</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.674</td>
<td>0.401</td>
<td>0.331</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>8065</td>
<td>ECE</td>
<td>3(0/3)</td>
<td>5/3</td>
<td>6.63</td>
<td>6/2</td>
<td>11.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.575</td>
<td>0.050</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>8081</td>
<td>Econ</td>
<td>2(0/2)</td>
<td>6/2</td>
<td>4.68</td>
<td>5/3</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.093</td>
<td>0.401</td>
<td>0.551</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
Table 1.9: Individual Heterogeneity, Continued

<table>
<thead>
<tr>
<th>NO</th>
<th>EID</th>
<th>MAJOR</th>
<th>ERROR</th>
<th>BIAS (+/-)</th>
<th>AVG p-value</th>
<th>BIAS (+/-)</th>
<th>AVG p-value</th>
<th>BIAS (+/-)</th>
<th>AVG p-value</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>8091</td>
<td>Econ</td>
<td>2(2/0)</td>
<td>5/3</td>
<td>10.09</td>
<td>3/5</td>
<td>-4.45</td>
<td>10/4</td>
<td>-0.19</td>
<td>T/E*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.093</td>
<td>0.327</td>
<td>0.551</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>8101</td>
<td>Econ</td>
<td>2(1/1)</td>
<td>4/4</td>
<td>3.90</td>
<td>5/3</td>
<td>4.48</td>
<td>9/5</td>
<td>5.76</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.779</td>
<td>0.401</td>
<td>0.177</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>8122</td>
<td>Comm</td>
<td>2(0/2)</td>
<td>3/5</td>
<td>-0.93</td>
<td>5/3</td>
<td>8.49</td>
<td>9/5</td>
<td>4.72</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.484</td>
<td>0.327</td>
<td>0.221</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ERROR**: Number of errors on assessments of own attributes

**BIAS (+/-)**: Frequencies of positive and negative biases

**TYPE**: Classified type †: Subjects who failed the understanding test

Wilcoxon’s Sign Rank Test *, **, and ***: Significant at the 10%, 5%, and 1% level.
Figure 1.1: Population distribution of the attributes
Figure 1.2: Errors of assessments of own attributes
Figure 1.3: Overestimation of Same-Quality-Group (SQG) - vs. All UCSD
Note: Histograms denote the subjects’ distribution of attributes. The solid line denotes the true population distribution. The dashed line denotes the median value of the subjects’ estimation.
Figure 1.4: Overestimation of Same-Quality-Group (SQG) - vs. Same major
Note: Histograms denote the subjects’ distribution of attributes. The solid line denotes the true population distribution. The dashed line denotes the median value of the subjects’ estimation.
Forecasts of Relative Performance in Tournaments: Evidence from the Field

2.1 Introduction

This paper uses a field experiment to test the rationality of players’ beliefs about their relative performance in tournaments. We consider two field settings, “Texas Hold’em” poker and chess tournaments, which differ in the degree to which luck is a factor and also in the information that players have about the competition.

The main finding of the paper is that players in real-world poker and chess tournaments overestimate their relative performance. This happens when players are given incentives for correct self-assessments and under two different ways of measuring beliefs of relative performance.
The experiment takes place in two poker tournaments—University of California San Diego's 2004 Winter and Spring Poker Classics both held at Viejas Casino in California—and one chess tournament—Sintra’s 2005 Chess Open, held in Sintra, Portugal. We chose poker and chess because we are interested in the extent to which different degrees of luck, skill and information may lead to different beliefs about outcomes. Skill is the most important factor in both types of tournaments but luck plays a larger role in poker than in chess. Another fundamental difference is that chess players usually have better information about the skills of their competitors than poker players.\footnote{The ultimate experiment would control both luck and information and investigate how varying these factors independently affect forecasts of relative performance.}

Before the start of each tournament we distribute a survey to participants where we ask them, among other things, to provide a point forecast of their relative performance. We observe the actual rank of each player in the tournament. When the tournament is over the forecast error of each player is computed and players are paid according to the precision of their forecasts under a quadratic scoring rule.

We also ask players to choose between receiving a sure payment and nine different bets whose payments are contingent on relative performance being above \( c \) percent of the population, with \( c \in \{10, 20, \ldots, 90\} \). This is a new measure of beliefs of relative performance, based on the observation of choices among alternatives, that can be compared with players’ forecasts.

We test for bias in players’ forecasts and bets. We also test if players’ forecasts and bets are significantly different from random choices. To perform these tests we use a parametric approach that takes into account the fact that incomplete information about relative skill together with the fact that forecasts are restricted to lie in a bounded interval force players near the low end of the scale to overestimate relative performance, on average, and players near the high
end to underestimate.\(^2\)

We find that players’ forecasts of relative performance are biased: on average, a poker player overestimates relative performance by 7 to 10 percentiles and a chess player by 6 to 7 percentiles. Players’ betting behavior is consistent with their forecasts. In the Spring Poker Classic, 78.6% of players chose bets that pay when performance is above the median. In Sintra’s Chess Open 63.8% of chess players chose bets that pay when performance is above median. This finding is consistent with previous studies which suggest that people display a systematic tendency to overestimate their abilities.

Additionally, we find that poker players’ forecasts and bets are not significantly different from random guesses with an overestimation bias. By contrast, chess players’ forecasts and bets are significantly better than random choices. We also find that the forecast errors of low skilled chess players are larger than those of high skilled chess players. This finding is consistent with the “unskilled and unaware” hypothesis proposed by Kruger and Dunning (1990). This hypothesis states that the low skilled players lack the cognitive skills to evaluate their ability and so make worse self-assessments of skill than the high skilled players. Finally, we find that chess players’ forecasts of relative performance are not efficient: chess players could have made better forecasts of relative performance if they had used their knowledge about the quality of the competition to make their forecasts.

This paper is an additional contribution to the literature that documents the existence of behavioral biases in judgment and decision making. The tendency that individuals have to overestimate their abilities was discovered in the field of social psychology.\(^3\) Two seminal contributions are Dunning et al. (1989) and

---

\(^2\) On a methodological level, this is the first study to use a statistical test on the accuracy of players’ forecasts of relative performance that takes into account the boundness of the dependent variable.

\(^3\) According to Myers (1996), a textbook in social psychology, “(...) on nearly any dimension that is both subjective and socially desirable, most people see themselves as better than average.”
Kruger and Dunning (1999) who show that overestimation of skill varies systematically with several factors.\textsuperscript{4} However, there are limitations with the psychological evidence. One of them is that individuals are not provided with incentives to think carefully about their predictions.

There is a growing literature in experimental economics on this topic. Camerer and Lovallo (1999) investigate the impact of overestimation of relative skill on entry in markets. They consider a market entry game where subjects’ payoffs are based on rank, which is determined either randomly or through a test of skill. They find that there is more entry when relative skill determines payoffs, which suggests that individuals overestimated their ability to do well on the test relative to others. They also find that there is more entry when individuals self-select into the experiment knowing that higher skill implies higher earnings. They call this finding reference-group neglect.\textsuperscript{5}

Clark and Friesen (forthcoming) study forecasts of relative performance in two tasks: (1) maximizing a two variable unknown function by moving contiguously from cell to cell on a spreadsheet and (2) decoding five letter words. Forecast precision was rewarded with a quadratic scoring rule in 8 sessions and there were no incentives for precise forecasts in 4 sessions. Clark and Friesen found overestimation of relative performance in 3 out of 12 sessions, underestimation in 2 out of 12 sessions, and lack of bias in 7 out of 12 sessions. The use of a quadratic scoring rule did not reduce either forecast bias or variance over non-incentive forecasts.

\textsuperscript{4}For example, the more ambiguous is the definition of the skill the greater is the overestimation effect, overestimation is higher in tasks that require a greater number of skills, overestimation decreases with task difficulty, and overestimation is higher when individuals think they can control the outcome of a task than when they think that the outcome of a task is mostly determined by chance.

\textsuperscript{5}Moore and Cain (2007) use the same experimental design as Camerer and Lovallo (1999) with the added feature that skill-dependent payoffs are based on either an easy or a difficult test of skill. They found more entry when rank was determined by relative performance on the easy test than when rank was determined randomly. They found less entry when rank was determined by relative performance on the difficult test than when rank was determined randomly.
Ferraro (2003) investigates forecasts of relative performance in three introductory microeconomic classes at Georgia State University. Students in these classes took three non-cumulative multiple-choice exams that made up most of their final grade. Immediately after completing each exam, subjects were asked to forecast their relative performance on the exam. Ferraro found that 80% of the subjects that took the first exam believed they were above the 50th percentile. He also found that overestimation of relative performance was not reduced over time. By the third exam, 83% of all subjects still believed they performed above the 50th percentile.

Hoelzl and Rustichini (2005), Moore (2002), Moore and Kim (2003) identify a subject’s beliefs about relative performance by asking the subject whether a reward should be based on a skill-based test or the outcome of a random device. There is overestimation of relative performance when more than half of the subjects prefer to be rewarded on the basis of their performance on the test than on the basis of a randomization device that selects a winner with probability one half. The experiments find overestimation on easy tests and underestimation on hard tests. Monetary payments significantly reduced overestimation of relative performance but did not improve subjects’ choices.

The paper is organized as follows. The next section explains the experimental design. Section 3 describes the hypotheses to be tested. Section 4 shows that players’ forecasts are biased. Section 5 discusses the precision of players’ forecasts. Section 6 looks at players’ bets. Section 7 shows that the forecast errors of

---

6 Forecast accuracy was rewarded with a quadratic scoring rule. In one class, subjects who were closest to predicting their actual percentile received $25 each. In the other two classes, subjects received $5 if their prediction was within one percentile point accurate, $4 if within two percentile points and $1 within three percentile points.

7 As Hoelzl and Rustichini (2005) point out, the drawback of this measure of beliefs of relative performance is that subjects are facing the choice between a lottery with objective uncertainty–outcome of the random device–and lottery with subjective uncertainty–the outcome of the test of skill. Thus, if subjects suffer from ambiguity aversion, this measure is likely to underestimate the subjective perception that subjects have of their relative performance.
the low skilled chess players are larger than those of the high skilled chess players. Section 8 shows that chess players’ forecasts are not efficient. Section 9 discusses the findings and their implications. Section 10 concludes the paper. The Appendix contains the survey used in Sintra’s Chess Open, theoretical results about optimal point forecasts and bets, and the prize structures of each tournament.

2.2 Field Experiment Design

A Texas Hold’em poker tournament is an elimination tournament where players are randomly assigned to different tables that can sit up to 10 players. Tables are reshuffled as players who ran out of chips drop out. When a final table with 10 players is reached there is no further reshuffling. In Texas Hold’em poker each player gets two cards face down, to be combined with five community cards dealt face up in the middle - the first three simultaneously (called the flop), then a fourth (the turn), then a fifth (the river) - to make the best five-card hand.

At the start of a Texas Hold’em poker tournament the role of luck is very important. High ability players with weak hands can be eliminated by low ability players with stronger hands. This happens because at the start of the tournament players’ earnings are very similar so a high ability player is sometimes forced to bet against a low ability player that has a stronger hand. As the tournament evolves the role of luck becomes less important since the earnings of the high ability players become increasingly larger than the earnings of the low ability players.

Most chess tournaments are neither elimination nor round-robin tournaments (each player playing every other player). Typically a Swiss System is used.\(^8\) According to this system players are initially matched in pairs either drawn at

\(^8\)A description of the Swiss System can be found at http://scichess.org/faq/swiss.html
random or seeded according to Elo ratings. After the first round, players who win receive a point, those who draw receive half a point and losers receive no points. Win, lose or draw, all players proceed to the next round where winners are pitted against winners, losers are pitted against losers, and so on. In the subsequent rounds players face opponents with the same (or almost the same) score. Modifications are made to ensure that no player is paired against the same opponent twice and that each player plays an equal number of games with white and black.

A Swiss tournament can handle many players without requiring an impractical number of rounds. However, the final rankings of a Swiss tournament are usually more random than those of a round-robin tournament, depending on the tiebreakers used. Even though the correct player usually wins, and the correct player usually ends up in the last place, the players in between are only sorted roughly without a good tiebreaker depth. So, although chess is itself a skill-based task, rankings in a Swiss chess tournament involve some luck.

The previous paragraphs show us that skill and luck play different roles in a Texas Hold’em poker tournament and in chess tournaments. Clearly, skill is the most important factor in determining players’ performance in both types of tournaments but luck plays a larger role in a Texas Hold’em poker tournament than in a Swiss chess tournament.

Another fundamental difference between poker and chess tournaments is the amount of information that players have about the quality of the competition. Typically, chess players have better information about the skills of their competi-

---

9The Elo rating system in chess is a means of comparing the relative strengths of chess players, devised by Arpad Elo. Players gain or lose rating points depending on the Elo rating of their opponents. If a player wins a game of chess in a rated tournament, they gain a number of rating points that increases in proportion to the difference between their rating and their opponent’s rating. The central statistical assumption of the ELO system is that any player’s tournament performances, spread over a long enough career, will follow a normal distribution. A detailed description of the formulae and theory behind the system can be found at http://home.clear.net.nz/pages/petanque/ratings/descript.htm.
tors than poker players. This happens because Elo ratings are a very informative measure of skill in chess and players usually know their own Elo rating and the Elo ratings of their opponents.

We performed the field experiment at two “Texas Hold’em” poker tournaments held at Viejas Casino in California. The first tournament—“Winter Poker Classic”—was held on March, 7th, 2004. In this tournament there were 155 players each paying a 10 USD entry fee and receiving 1500 USD worth of chips. Once the player used up all chips, he would be eliminated. The total prize pool was 1670 USD. The second tournament—“Spring Poker Classic”—was held on May, 23rd, 2004. In this tournament there were 167 players each paying an entry fee of 20 USD. The total prize pool was 3000 USD. The prize structure of each tournament is depicted in Table A1 in the Appendix.

To obtain poker players’ forecasts of relative performance we asked them the following question:

Of all the individuals participating in the poker tournament, what percentage do you think will be eliminated before you?

Players were instructed to answer the question by choosing a whole number between 0 and 99. The survey also informed players that numbers close to zero indicate that they predict that they will be among worst players in the tournament, and that numbers close to 99 indicate that they predict that they will be among the best players in the tournament.

Sintra’s Chess Open was held in July, 17th, 2005 in Sintra, a village near Lisbon. There were 93 chess players in the tournament. The entry fee for members of Sintra’s Chess Club was 3 EUR while non-members had to pay 6 EUR. The total
prize pool was 1100 EUR. The prize structure of the tournament is depicted in Table A2 in the Appendix.

Sintra’s Chess Open used the Swiss system. At the start of the tournament players with similar Elo ratings were matched in pairs.\textsuperscript{10} After the first round, players were placed in groups according to their score (winners in the 1 group, those who drew go in the 1/2 group, and losers go in the 0 group) and then matched in pairs inside each group. Each round the same procedure was used. There were a total of 8 rounds each lasting 20 minutes. The relative performance of each chess player in the tournament was calculated by the organization using the Swiss method.

Like in poker tournaments, we asked players in Sintra’s Chess Open to predict their relative performance. The main novelty is that we asked chess players to report their own Elo rating and the percentage of players in the tournament with a smaller Elo rating. This gives us an idea of players’ information about the quality of the competition.

Based on each player’s forecast of relative performance and his actual performance, we calculated the forecast error of each player, $E_i$, defined as $E_i = F_i - P_i$, where $F_i$ is player $i$’s forecast of relative performance and $P_i$ is player $i$’s relative performance, with $F_i$ being an integer between 0 and 99 and $P_i$ being a real number in $[0, 100)$. The monetary reward of player $i$, $R_i$, as a function of player $i$’s forecast error, was determined by the quadratic scoring rule

$$R_i = \begin{cases} M - \left[ \text{Int} \left( |E_i| \right) \right]^2, & \text{if } \text{Int} \left( |E_i| \right) \leq X \\ 0, & \text{if } \text{Int} \left( |E_i| \right) > X \end{cases},$$

where $\text{Int} \left( x \right)$ is the closest integer which is smaller than $x$. In the Winter Poker Classic $M = 10$ USD and $X = 4$, in the Spring Poker Classic $M = 20$ USD and

\textsuperscript{10}The Elo ratings of players were public information since the organizers of the tournament posted them at the entrance of the room where the tournament was played.
\(X = 5\), and in Sintra’s Chess Open \(M = 10\) EUR and \(X = 4\).

We chose the quadratic scoring rule because of its simplicity and the fact that it allows us to test the rationality of players’ forecasts using ordinary least squares regressions. DeGroot (1970) shows that the quadratic scoring rule is incentive compatible for a risk neutral player. Propositions 1 and 2 in the Appendix show that the quadratic scoring rule is also incentive compatible for a player with uniform or an unimodal and symmetric distribution of beliefs, regardless of the player’s preferences towards risk.

We did not use a binary lottery payoff scheme to induce risk neutrality from the part of players due to the lack of control associated with performing a field experiment. Most players left the room where the tournament was being held immediately after being eliminated so there was no way they could observe the lottery being drawn.

The survey also asked players for demographic characteristics such as age, sex, and academic major. On average, players took 10 minutes to read, answer, and return the survey. Each player filled his survey individually and returned it to us right after he finished it. The reply rate in the Winter Poker Classic was 87.1\%, the one in the Spring Poker Classic was 77.2\%, and 64.5\% in Sintra’s Chess Open.

---

11 The rewards for forecast precision were explained in a simple way in the surveys. See the Appendix.
12 Alternatively, we could have chosen a scoring rule where the loss is proportional to the absolute value of the forecast error. De Groot (1970) shows that this scoring rule induces risk neutral players to report the median rather than the mean. To test the rationality of players’ forecasts under this alternative scoring rule we would need to use least absolute deviations regressions–see Basu and Markov (2003). Camerer (1982) uses a scoring rule where individuals are paid something when they are exactly correct and nothing otherwise. This rule induces risk neutral players to report the mode rather than the mean.
13 It is not clear that this procedure works in practice. For example, Selten et al. (1999) find that the binary lottery payoff scheme does not induce risk neutrality, but on the contrary, it leads to stronger deviations from risk neutrality than a direct money payoff scheme.
14 We postpone the discussion of self-selection bias to Section 9.
In the Winter and Spring Poker Classics players were asked for their addresses and their earnings from taking the survey were sent by mail. In Sintra’s Chess Open players had the option of receiving their earnings by mail or at the end of the tournament. Most players chose to receive them at the end of the tournament.

2.3 Hypotheses

The main hypothesis that will be tested in this field experiment is that players’ beliefs of relative performance are rational. By definition, if beliefs are rational, then forecasts must be unbiased, that is, there must be no systematic tendency for overestimation or underestimation of relative performance.

H1a Forecasts are unbiased.

We are also interested in having an idea of how precise players’ forecasts are. Forecasts may be extremely imprecise, that is, they may not be distinguishable from random guesses. Alternatively, forecasts may be significantly better than random guesses.

H2a Forecasts are not random guesses.

Since rankings in a Texas Hold’em poker tournament are more random than those in a Swiss chess tournament and poker players have less information about the quality of the competition than chess players, we expect that poker players’ forecasts of relative performance are less precise than those of chess players.
H3 Forecast errors of poker players are greater than those of chess players.

We use an alternative measure of beliefs of relative performance. To do that we ask players to choose among different bets whose payments depended on their relative performance in the tournament. Thus, we also test if players bets are unbiased and if they are significantly better than random choices.

H1b Bets are unbiased.

H2b Bets are not random choices.

Kruger and Dunning (1999) report a series of experiments that show that high skilled individuals make better self-assessments than low skilled individuals. However, their measure of relative skill is not very good in that it only relies in a single observation and effects of experience or familiarity with the task are not taken into account. In Sintra’s chess tournament we have a very good objective measure of relative skill—the Elo rating—and we know the number of chess tournaments that each player has played before. This allows a more stringent test of Kruger and Dunning’s “unskilled and unaware” hypothesis.

H4 The forecast errors of chess players with a low Elo rating are larger than those with a high Elo rating.

To test H4 we regress the absolute forecast error of chess players on their Elo ratings, the number of tournaments they played, and an interaction term (Elo ratings times the number of tournaments played). If the coefficient on Elo rating is negative and significantly different from zero, then we find support for H4.
By definition, rational forecasts must also be efficient, that is, players must make use of all available information to make their forecasts. In the experiment we ask chess players to provide their best estimate of the percentage of the population in the tournament with a lower Elo rating. This allows us to test for efficiency in players’ forecasts.

**H5** Forecasts of chess players are efficient.

To test H5 we regress the relative performance of chess players on their forecasts and on their estimates of the percentage of the population in the tournament with a lower Elo rating. If the coefficient of the second predictor variable is significantly different from zero, then we find evidence against H5.

### 2.4 Forecast Bias

Table 2.1 displays the distribution of forecasts in each tournament divided into intervals of 10 percentiles starting in the interval [0, 10] and ending in [90, 99].

Inspection of Table 2.1 reveals a clear tendency for overestimation of relative performance in all tournaments. Only 26.1% (Winter Poker Classic), 33.3% (Spring Poker classic) and 38.3% (Sintra’s Chess Open) of players who took the survey forecast to finish below the median.

To test if players’ forecasts of relative performance are unbiased, hypothesis H1a, we run the ordinary least squares (OLS) regression \( E_i = \alpha + \varepsilon_i \), where \( E_i \) is the forecast error of player \( i \) and \( \alpha \) is the intercept.\(^{15}\) The results for each tournament are summarized in Table 2.2.

\(^{15}\)This is a standard test of unbiasedness in forecasts.
Table 2.1: Distribution of Players’ Forecasts in Tournaments

<table>
<thead>
<tr>
<th>Forecasts</th>
<th>Winter Poker Classic</th>
<th>Spring Poker Classic</th>
<th>Sintra’s Chess Open</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>share</td>
<td>cum. share</td>
<td>share</td>
</tr>
<tr>
<td>[0, 9]</td>
<td>0.7</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>[10, 19]</td>
<td>4.4</td>
<td>5.3</td>
<td>3.9</td>
</tr>
<tr>
<td>[20, 29]</td>
<td>8.2</td>
<td>13.5</td>
<td>14.6</td>
</tr>
<tr>
<td>[30, 39]</td>
<td>5.9</td>
<td>19.4</td>
<td>3.9</td>
</tr>
<tr>
<td>[40, 49]</td>
<td>6.7</td>
<td>26.1</td>
<td>10.1</td>
</tr>
<tr>
<td>[50, 59]</td>
<td>14.1</td>
<td>40.2</td>
<td>13.2</td>
</tr>
<tr>
<td>[60, 69]</td>
<td>14.8</td>
<td>55.0</td>
<td>13.2</td>
</tr>
<tr>
<td>[70, 79]</td>
<td>10.4</td>
<td>65.4</td>
<td>12.4</td>
</tr>
<tr>
<td>[80, 89]</td>
<td>16.2</td>
<td>81.6</td>
<td>12.3</td>
</tr>
<tr>
<td>[90, 99]</td>
<td>18.4</td>
<td>100.0</td>
<td>15.4</td>
</tr>
</tbody>
</table>

Table 2.2: OLS Regression Results for Forecast Bias

<table>
<thead>
<tr>
<th></th>
<th>Winter Poker Classic</th>
<th>Spring Poker Classic</th>
<th>Sintra’s Chess Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10.02 (3.02)***</td>
<td>7.13 (2.03)**</td>
<td>6.98 (2.31)**</td>
</tr>
<tr>
<td>n=122</td>
<td>n=116</td>
<td>n=60</td>
<td></td>
</tr>
</tbody>
</table>

Dependent variable: Forecast error. t statistics in parentheses. ***, **, * denotes statistical significance at the 1%, 5%, and 10% level respectively.

We see that the mean forecast error in the Winter Poker Classic is equal to 10.02 percentiles, the mean forecast error in the Spring Poker Classic is 7.13 percentiles and 6.98 percentiles in Sintra’s Chess Open. The mean forecast errors in all tournaments are greater than zero at 5% significance level. This shows that, on average, players’ forecasts in all tournaments are biased towards overestimation of relative performance.
2.5 Forecast Precision

To test if players’ forecasts of relative performance are not random guesses, hypothesis H2a, we need to have an idea of how well players’ forecasts predict relative performance. One way to do that is to run the OLS regression

\[ P_i = \alpha + \beta F_i + \varepsilon_i, \]  

(2.1)

where \( F_i \) is player \( i \)’s forecast and \( P_i \) is player \( i \)’s position in the tournament. If we find that the fit of this regression is good and that the estimate for the slope is significantly greater than zero, then there is evidence that players forecasts are not random guesses. By contrast, if we find that the fit of this regression is bad and that the estimate for the slope is not significantly different from zero, then players’ forecasts are not distinguishable from random guesses.

However, the OLS estimates in (2.1) would be biased. Incomplete information about relative skill together with the fact that relative performance is restricted to lie in a bounded interval force people near the low end of the scale to overestimate relative performance, on average, and people near the high end to underestimate.

To address this problem we use the transformation of variables technique. One way to map the variable \( P_i \), which is bounded by 0 and 100, to the real line is to use a logit transformation. The logit transformation of player \( i \)’s relative performance and forecast are given by

\[ U_i = \ln \left( \frac{P_i}{100 - P_i} \right), \]

and

\[ Z_i = \ln \left( \frac{F_i}{100 - F_i} \right), \]

respectively. The transformation implies \( U_i \) and \( Z_i \) are

\[ 16 \text{This happens because the dependent variable is bounded by 0 and 100. The nature of the bias can be demonstrated as follows. If } 0 \leq P_i \leq 100, \text{ then } -\alpha - \beta F_i \leq \varepsilon_i \leq 100 - \alpha - \beta F_i. \text{ Thus, the fact that we have a limited dependent variable implies that the error term is regulated by an upper and a lower bound that depends on the independent variable. So, the distribution of the error term depends on the value of the independent variable and it is not identically distributed. OLS requires, among other things, that the error term is identically distributed and uncorrelated with the regressor.} \]
unconstrained variables.\textsuperscript{17} We use the transformed series to run the ordinary least squares regression $U_i = \alpha + \beta Z_i + \varepsilon_i$. Table 2.3 displays the results obtained for each tournament.

<table>
<thead>
<tr>
<th></th>
<th>Winter Poker Classic</th>
<th>Spring Poker Classic</th>
<th>Sintra’s Chess Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.14 (0.75)</td>
<td>0.01 (0.03)</td>
<td>-0.22 (-1.04)</td>
</tr>
<tr>
<td>Forecast</td>
<td>0.03 (0.34)</td>
<td>0.11 (0.89)</td>
<td>0.50 (6.07)**</td>
</tr>
<tr>
<td>n</td>
<td>122, $R^2=0.001$</td>
<td>116, $R^2=0.007$</td>
<td>60, $R^2=0.39$</td>
</tr>
</tbody>
</table>

Dependent variable: Logit transformation of relative performance. \textsuperscript{18} t statistics in parentheses. ***, **, * denotes statistical significance at the 1%, 5%, and 10% level respectively.

We see from Table 2.3 that the fit of Winter and Spring Poker Classic regressions is very bad: the $R$-squared is equal to 0.1% in the Winter Poker Classic and 0.7% in the Spring Poker Classic. We also see that in both poker tournaments the estimated coefficients are not significantly different from zero.\textsuperscript{18} Thus, we find evidence against hypothesis H2a for poker tournaments, that is, poker players’ forecasts in both tournaments are random guesses. In contrast, we see from Table 2.3 that the fit of the Sintra’s Chess Open regression is 39%. The estimated coefficient for the slope is 0.5 and is significantly different from zero at 1% significance level. Thus, we find evidence that supports hypothesis H2a for Sintra’s Chess Open: chess players’ forecasts of relative performance are not random guesses.\textsuperscript{19}

As expected, these results provide support for H3, that is, poker players’

\textsuperscript{17}See Zarembka (1974) on the transformation of variables technique. This transformation of variables has also been used by Chen and Giovannini (1992) for testing the rationality of exchange rate forecasts within a band.

\textsuperscript{18}Another aspect that needs to be taken into consideration is that players who forecast their performance to be in the bottom of the scale may make larger forecast errors than players who forecast their performance to be in the top of the scale (the transformation of variables may or may not change this pattern of heteroskedasticity). If there is heteroskedasticity in the transformed model, then the OLS estimates are unbiased but inefficient. To address this possibility we run a robust regression using Stata 7.0. We found that the robust standard errors are essentially identical to the OLS standard errors.

\textsuperscript{19}The lack of accuracy of poker players’ forecasts in both tournaments implied that the earnings from their forecasts were quite low as it can be seen in Table A3 in the Appendix.
forecast errors are larger than those of chess players. The difference in information sets together with the fact that relative performance is more random at poker than at chess are likely to explain the fact that chess players’ forecast are more precise than those of poker players.

2.6 Betting Behavior

In the Spring Poker Classic and in Sintra’s Chess Open players were also asked to choose among different bets whose payments depended on their relative performance in the tournament. For example, in the Spring Poker Classic each player was offered the choice of getting a sure payment of $2.00 or betting on his relative performance. There were nine possible bets whose payments were contingent and a player being above \(c\) percent of the population, with \(c \in \{0, 10, 20, \ldots, 90\}\). The bets paid \(200/(100 - c)\) if a player was eliminated after \(c\) percent of the population and zero dollars otherwise.

Proposition 3 in the Appendix shows that for risk neutral players, the choice of bet question is a more stringent test of overestimation of relative performance than the point forecast question. In the forecasting problem, a risk neutral player who overestimates or underestimates relative performance by the same amount faces the same loss. By contrast, in the betting problem, a risk neutral player who overestimates relative performance by 10% incurs a larger loss than if he underestimates it by 10%. Thus, the optimal bet of a risk neutral player should be smaller than his optimal point forecast.

The answers to the choice of bet question in each tournament are sum-

\(^{20}\)This is confirmed by the mean absolute forecast errors of poker and chess players. The mean absolute forecast error in the Winter Poker Classic was 29.66 percentiles and 31.34 percentiles in the Spring Poker Classic. In contrast, the mean absolute forecast error in Sintra’s Chess Open was 17.03 percentiles.
A quick inspection of Table 2.4 shows us that only 21.4% of players who answered the choice of bet question in the Spring Poker Classic and 36.2% of players who did it in Sintra’s Chess Open chose bets that paid them when their performance was below the median. This suggests that players’ bets also reveal overestimation of relative performance.

To test if poker players’ bets are unbiased, hypothesis H1b, we need to compare poker players’ bets to their ranks in the Spring Poker Classic. Ranking bets from 5 (the sure thing), 15 (the $2.22 bet), to 95 (the $20 bet) we find that the average choice of bet of poker players is 68.33. The average rank is the 51.54th percentile. From the data, we find that the $ statistic for the hypothesis test that the average choice of bet is not significantly different from 51.54 is equal to 6.56 and

---

The first column of Table IV reports the payoff of each bet, the second column the ratio of the number of players that were paid for that choice of bet to the number of players who chose that bet, and the third column reports the share of players in the Spring Poker Classic who chose each bet. The remaining three columns provide similar information for Sintra’s Chess Open.
the critical value, at 5% significance level, is equal to $t_{5\%}(139) \approx 1.645$. Thus, we find evidence against hypothesis H1b, that is, we find that poker players’ bets are biased towards overestimation of relative performance. Using a similar procedure we find the same for chess players’ bets.\(^{22}\)

Table IV also shows us that the average reward for choice of bet of poker players is $2.01. This value is not different, at 5% significance level, from the expected reward of a random choice of bet in the Spring Poker Classic: $2.00.\(^{23}\) Thus, we find evidence against hypothesis H2b, that is, we find that poker players’ bets are random choices. By contrast, the average reward for choice of bet of chess players is $1.65. This value is greater, at 5% significance level, than the average reward of a random choice of bet in Sintra’s Chess Open: 1 euro.\(^{24}\) Thus, we find evidence in favor of hypothesis H2b, that is, that chess players’ bets are not random choices.

Players bets are consistent with their forecasts. The correlation between poker players’ bets and their forecasts is equal to 62.5% and the correlation between chess players’ bets and their forecasts is equal to 77%. Still, poker players’ bets seem to reveal more overestimation of relative performance than their forecasts.\(^{25}\)

\(^{22}\)Ranking bets from 5 (the sure thing), 15 (the $1.11 bet), to 95 (the $10 bet) we find that the average choice of bet of chess players is 53.42. The average rank is the 47.61th percentile. From the data, the \(t\) statistic for the hypothesis test that the average choice of bet is not significantly different from 47.61 is equal to 1.35 and the critical value, at 10% significance level, is equal to \(t_{10\%}(57) \approx 1.28\).

\(^{23}\)If a player’s performance is worse than the 10th percentile and makes a random choice of bet his expected reward is \((1/10) \times $2 + (9/10) \times $0 = $0.2. If a player’s performance is better than the 10th percentile and worse than the 20th percentile and makes a random choice of bet is expected reward is \((1/10) \times $2 + (1/10) \times $2.22 + (8/10) \times $0 = $0.422. Doing the same for players in the other deciles we obtain expected rewards for players in each decile. Thus, if all players made random choices of bet the average reward for choice of bet should be approximately equal to

\[
\frac{1}{10}( \$2 + \$0.422 + \$0.672 + \ldots + \$5.857 ) \approx \$2.
\]

\(^{24}\)From the data, the \(t\) statistic for the hypothesis test that the average reward for choice of bet is not significantly different from 1 euro is equal to 2.02 and the 5% critical value is equal to \(t_{5\%}(57) = 1.645\).

\(^{25}\)In effect, while 78.6% of players chose bets that paid them when their performance was above the median only 62.8% of players forecasted that their performance would be above median.
However, since we did not assess players’ preferences towards risk we cannot claim that poker players reveal more overestimation of relative performance in their bets than in their forecasts. For example, if those who enter into a poker tournament are more risk loving rather than risk neutral, then we would expect this pattern of bets and forecasts.

2.7 Unskilled and Unaware

Kruger and Dunning (1999) report a series of experiments with easy skill-based tasks that support the “unskilled-unaware hypothesis”, that is, that the high skilled individuals are better informed about their skills than low skilled individuals. One possible explanation for this finding is that high skilled players are more experienced than low skilled players and that greater experience implies better information about relative skill.\(^{26}\)

In Sintra’s Chess Open we have a very informative objective measure of relative skill—the Elo ratings—and we know players’ previous experience with chess tournaments.\(^{27}\) This means we can study the impact of relative skill on the precision of chess players’ forecasts while taking into account experience effects. To do that we regress the absolute forecast error of chess players on their Elo ratings, the number of tournaments they played, and an interaction term (Elo ratings times the number of tournaments played). If the coefficient on Elo rating is negative and significantly different from zero, then we find support for H4. Thus, we run the

\(^{26}\) However, Burson et al. (2006) show that for difficult skill-based tasks (where there is underestimation of relative performance) the low skilled players are more accurate in their forecasts than the high skilled players.

\(^{27}\) Of the 93 players that took part in Sintra’s Chess Open, 70 had Elo ratings and 23 did not. The range of Elo ratings was from 1090 points to 2441 points. The average Elo rating was 1865 points. Of the 60 players who took our survey only 49 reported the number of chess tournament they had played before. Of these 49 players, 7 had no previous experience with chess tournaments, 15 had participated in 1 to 10 tournaments, 17 had participated in 11 to 100 tournaments, and 10 had participated in 150 to 400 tournaments.
OLS regression

\[ |E_i| = a + b_1 C\text{Exp}_i + b_2 C\text{Elo}_i + b_3 (C\text{Elo}_i \times C\text{Exp}_i) + \varepsilon_i. \]

The results obtained for this regression are reported in Table V.

Table 2.5: OLS Regression Results for Forecast Precision Experience and Elo in Sintra’s Chess Open

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>22.1738 (5.31)***</td>
<td></td>
</tr>
<tr>
<td>CExp</td>
<td>0.1073 (1.57)</td>
<td></td>
</tr>
<tr>
<td>C Elo</td>
<td>-0.0289 (-3.11)***</td>
<td></td>
</tr>
<tr>
<td>CExp \times C Elo</td>
<td>-0.0001 (-1.45)</td>
<td></td>
</tr>
</tbody>
</table>

n=49, R^2=0.23

Dependent variable: Absolute forecast error

The results obtained for this regression are reported in Table V.

Table 2.5 shows that the relation between experience and forecasts is insignificant. However, the coefficient for skill–Elo rating–is negative and significant at 1% significance level. This means that, on average, the higher the Elo rating of a player, the smaller is that player’s absolute forecast error. Thus, we find support for H4, that is, the forecast errors of low skilled chess players are greater than those of high skilled chess players. This finding is consistent with Kruger and Dunning’s (1999) “unskilled-unaware hypothesis” and can not be explained by the fact that high skilled chess players are more experienced in chess tournaments than low skilled chess players.

\[ C\text{Exp}_i = \text{Exp}_i - \text{mean(Exp}_i), \text{ and } C\text{Elo}_i = \text{Elo}_i - \text{mean(Elo}_i). \]

Independent variables were centered to avoid multicolinearity problem. Also, by centering the variables, we can interpret \( b_1 \) and \( b_2 \) as average effects of \( \text{Exp}_i \) and \( \text{Elo}_i \) on \( |E_i| \) respectively.
2.8 Efficiency of Forecasts

For chess players’ forecasts of relative performance to be rational they must be unbiased and efficient. We already know that chess players’ forecasts are biased. Can we say anything about efficiency? If a chess player makes an efficient forecast of relative performance then he must use all the available information that he has about his relative skill to make that forecast. Since we asked players to provide their best estimate of the percentage of the population in the tournament with a lower Elo rating we can use this variable to test for efficiency in chess’ players forecasts—hypothesis H5.

To do that we run the OLS regression $U_i = \alpha + \beta_1 Z_i + \beta_2 W_i + \varepsilon_i$, where $W_i = \ln \left( \frac{Lelo_i}{100 - Lelo_i} \right)$, with $Lelo_i$ being player $i$’s assessment of the percentage of the population that has a lower Elo rating. If $\beta_2$ is significantly different from zero, then there is evidence that chess players’ forecasts are not efficient. The results obtained for this regression are displayed in Table 2.6.

Table 2.6: OLS Regression Results for Test of Efficiency of Players’ Forecasts in Sintra’s Chess Open

<table>
<thead>
<tr>
<th></th>
<th>Regression 1</th>
<th>Regression 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.07 (-0.48)</td>
<td>Constant 0.14 (0.90)</td>
</tr>
<tr>
<td>Forecast</td>
<td>0.60 (7.07)**</td>
<td>Forecast 0.35 (3.13)**</td>
</tr>
<tr>
<td></td>
<td>Lower Elo 0.24 (3.15)**</td>
<td></td>
</tr>
<tr>
<td>n=44</td>
<td></td>
<td>n=44</td>
</tr>
<tr>
<td>$R^2$=0.54</td>
<td></td>
<td>$R^2$=0.63</td>
</tr>
<tr>
<td>Adjusted $R^2$=0.53</td>
<td></td>
<td>Adjusted $R^2$= 0.62</td>
</tr>
</tbody>
</table>

Dependent variable: Logit transf. of relative performance
$ t $ statistics in parentheses. $***$, $**$, $*$ denotes statistical significance at the 1%, 5%, and 10% level respectively.

The results from the two regressions in Table 2.6 show us that chess
players' forecasts are not efficient. The model with the explanatory variable Lower Elo (regression 2) has a better fit than the model without it (regression 1). Thus, we find support against H5: chess players could have made better forecasts if they had taken into consideration their own subjective assessments of the percentage of the population with a smaller Elo rating.29

2.9 Discussion

In this section we discuss the findings and implications of the paper.

2.9.A Self-Selection, Incentives, Risk and Information

The main finding of this paper is that poker and chess players' forecasts of relative performance are biased towards overestimation of relative performance. We can not rule out the possibility that the overestimation observed in the data is due to a self-selection bias. We had a reply rate of 87.1% in the Winter Poker Classic, 77.2% in the Spring Poker Classic, and 64.5% in Sintra’s Chess Open. If the players who did not answer our survey have a tendency to underestimate their relative performance, then poker and chess players’ forecasts might not be biased.30

The size of the biases in this paper is modest when compared to those often found in the psychology literature.31 The use of financial incentives for

---

29The correlation between chess players’ subjective assessments of the percentage of the population with a smaller Elo rating and the actual percentage with smaller Elo ratings was 87%.
30In future research on this topic we plan to give players a monetary reward just for filling in and returning the survey. This will improve reply rates and reduce the role of self-selection biases.
31For example, Svenson (1981) reports that between 81% and 90% of Americans think they are safer drivers than the median driver. Myers (1996) cites a study according to which: “In Australia, 86 percent of people rate their job performance as above average, 1 percent as below average.” Baker et al. (1998) cite a survey of General Electric Company employees according to which: “58 percent of a sample of
precision in forecasts could be an explanation for this finding. This is consistent with Hoelzl and Rustichini (2005), who show that monetary incentives can reduce overestimation bias (but do not improve precision).\textsuperscript{32} In fact, the main limitation of our paper might be the fact that we only provide modest monetary incentives for precision in forecasts and bets. If monetary incentive are modest, then a player’s expected rewards from taking the survey are small and do not depend much on his forecasts or bets. We cannot rule out the possibility that the biases would disappear if players would have been given greater monetary incentives.

The tendency to overestimate relative performance found might be due to a positive correlation between risk preferences and skill. If low skilled players are risk averse and high skilled players are risk seeking, then forecasts may be biased towards the positive side even though there are no biases in beliefs. Since we did not control for players’ risk attitudes we cannot rule out this explanation.\textsuperscript{33}

The finding of a bias in chess player’s beliefs of relative performance is particularly surprising since chess players have good information about the quality of the competition.\textsuperscript{34} This finding can be interpreted from two different perspectives. From the perspective of advocates of rational expectations the bias in chess players’ forecasts of relative performance is small and so we should not worry about it. By contrast, from the perspective of advocates of behavioral biases in human judgment, the fact that the bias persists when there is plenty of information about the quality of the competition constitutes strong evidence against rational expectations. Moreover, since in most tasks it is hard to find measures

---

\textsuperscript{32}It is not clear if using monetary incentives improves the precision of individuals’ forecasts. For a good discussion of this topic see Camerer and Hogarth (1999).

\textsuperscript{33}We intend to do this in future work on this topic using Holt and Laury’s (2002) procedure.

\textsuperscript{34}Previous studies could not address this issue since they lack reliable objective measures of relative skill like the Elo rating.
of relative performance as informative as the Elo rating is for chess, advocates of behavioral biases, would argue that biases in judgments of relative performance are widespread.

2.9.B Difference in Biases in Poker and Chess

According to the psychology literature individuals are more overconfident when they think that they have control over the outcome of the task. Consistent with this view, Camerer and Lovallo (1999) find more entry in the skill-based treatment than in the luck-based treatment of the market entry game.

As we have seen, skill is the most important factor in determining players’ performance in poker and chess tournaments but luck plays a larger role in a Texas Hold’em poker tournament than in Swiss chess tournament. However, we found that the overestimation bias is larger in poker than in chess. Does this finding contradict previous studies? Not necessarily.

An explanation for this finding might be a self-selection effect due to the different nature of the poker and chess tournaments in this experiments. In any tournament the players who overestimate their skills the most are the ones that are more likely to enter. Now, suppose that individuals are attracted to tournaments not only for the utility they can get from the money they win net of the entry fee but also from the utility from playing. Participants in the poker tournaments were eliminated when they lost all their chips. If a poker player thinks that he is one of the worst players, then he expects to play only a few games of poker for the $10 or $20 entry fee. In this sense it is not rational to participate in poker tournaments with this kind of beliefs. In contrast, a chess player who thinks he is one of the worst could still play interesting games even after having lost some games (against
opponents of approximately the same strength). So, playing a chess tournament without winning a prize may be more satisfying than leaving a poker tournament early and without a prize. This could lead to a smaller self-selection effect in chess tournaments.\(^{35}\)

It might also be that, in the subjective view of players, poker players perceive that the role of skill in a poker tournament is greater than it actually is. If players forecasts are based on this view, then overestimation of relative performance in poker can be greater than overestimation in chess.\(^{36}\)

The smaller overestimation bias found in chess might also be due to cultural differences. Most players in Sintra’s Chess Open from Portugal whereas most players in the Winter and Spring poker classics are from the US. If individuals from the US have a greater tendency to overestimate their abilities than individuals from Portugal, then the bias in chess should be smaller than the bias in poker.

2.9.C Implications

We now discuss the implications of the findings for the parties involved in gambling and workplace tournaments. In settings where skill matters for making decisions overestimation of relative performance may have important implications for behavior. One such setting is a tournament. The decision to participate in a tournament or the choice of how much effort to put in depends on correct beliefs

\(^{35}\)Differences in risk aversion between those who self-select into a poker tournament and those who self-select into a chess tournament might also explain why the bias is larger in poker than in chess. I am thankful to a referee for pointing this out.

\(^{36}\)In fact, we found some support for this possibility. We asked players in the survey how they thought their position in the tournament would be determined. Players could chose among seven options that ranged from “Only by relative skill” to “Only by luck” with 5 other options in between. On average, players in both poker and chess tournaments thought that “skill is more important than luck but that luck plays a large role in determining relative performance.” So, on average, there were no significant differences in poker and chess players’ perceptions of the role of luck in poker and in chess, respectively.
about skill.

If gamblers have a tendency to overestimate their relative performance in tournaments, then their utility will decrease due to their misguided choices (e.g., putting in too much or too little effort). However, managers of casinos can gain if they take this into account when they design tournaments. To do that they should either increase entry fees or reduce tournament prizes.

In workplace settings, overestimation of relative skill leads workers to overestimate the probability of favorable outcomes. If this is the case, then managers of firms should, on average, prefer incentive schemes featuring payments contingent on relative performance (e.g., rank-order tournaments or incentive schemes composed partly by fixed pay and partly by variable pay dependent on the magnitude of relative performance) to individualistic incentive schemes (e.g., fixed salary plans or piece rates).

### 2.10 Conclusion

This paper shows that players in real-world tournaments tend to overestimate relative performance. The bias is present in poker as well as in chess tournaments. On average, a poker player overestimates his relative performance by 7 to 10 percentiles, and a chess player by 6 to 7 percentiles. We also find that poker and chess players are willing to bet on their overly favorable views of relative performance.

This chapter “Forecasts of Relative Performance in Tournaments: Evidence from the Field,” in full, is coauthored with Luís Santos-Pinto and has been submitted for publication to the Journal of Theory and Decision.


2.11 References


Academic Medicine, vol. 76, no. 10, S87–S89.


2.12 Appendix

2.12.A Sintra’s 2005 Chess Open Survey

You are about to answer a survey that, among other things, asks you to make a prediction of your relative position in this chess tournament. Depending upon how well you make your prediction you may be able to earn up to 10 euros (Question 1). The survey also asks you to choose between different lotteries whose prizes depend on your relative position in the tournament (Question 2). Depending on your choice of lottery and how well you perform in the tournament you may earn up to an additional 10 euros. We will send you your payment by mail if you provide us your name and address. If you prefer, you can provide us only your e-mail address and we will tell you your payment by e-mail and then you can give us your address if you wish to receive it by mail. This survey is confidential.

Name:_____________  E-mail:____________

Address:____________________________

Zip code:_____________  Age:_______  Sex:_____

Q1: Please read the following question carefully: Of all the individuals participating in this chess tournament what percentage do you think will be ranked below you?

Before you answer note that, after the tournament is over, we will compare your prediction with the ratio of the actual number of players ranked below you to the total number of players. We will then pay you (in euros) for your prediction as follows:
10 if the prediction is less than 1% away from your position;

9 if the prediction is more than 1% and less than 2% away from your position;

6 if the prediction is more than 2% and less than 3% away from your position;

1 if the prediction is more than 3% and less than 4% away from your position;

0 otherwise.

Now, answer the question by choosing a whole number between 0 and 99 (recall that the number you choose represents your best estimate of what percentage of people will be ranked below you. Numbers close to zero indicate that you predict that you will be among worst players in the tournament, numbers close to 99 indicate that you predict that you will be among the best players in the tournament).

**Q2:** Consider the 10 lotteries below, whose prizes (in euros) depend on your ranking in the tournament. Choose one of the options:

We pay you 1.00 for sure

We pay you 1.11 if at least 10% of players are ranked below you and 0 otherwise

We pay you 1.25 if at least 20% of players are ranked below you and 0 otherwise

We pay you 1.43 if at least 30% of players are ranked below you and 0 otherwise

We pay you 1.67 if at least 40% of players are ranked below you and 0 otherwise

We pay you 2.00 if at least 50% of players are ranked below you and 0 otherwise

We pay you 2.50 if at least 60% of players are ranked below you and 0 otherwise
We pay you 3.33 if at least 70% of players are ranked below you and 0 otherwise

We pay you 5.00 if at least 80% of players are ranked below you and 0 otherwise

We pay you 10.00 if at least 90% of players are ranked below you and 0 otherwise

Q3: What is your Elo rating? If you don’t know the answer to this question, then choose between: a) I don’t have an Elo rating or b) I have an Elo rating but I can’t recall it.

Q4: What is your best estimate of the percentage of players in this tournament who have an Elo rating less than yours?

Q5: How many chess tournaments have you played before? Consider that a chess tournament involves monetary prizes and at least 20 players.

Q6: How do you think your position in this tournament will be determined?
Choose one

Only by your relative skill at playing chess

More by your relative skill than by luck, and luck plays a small role

More by your relative skill than by luck, and luck plays a large role

As much by your relative skill as by luck

More by luck than by your relative skill, and relative skill plays a large role

More by luck than by your relative skill, and relative skill plays a small role

Only by luck
2.12.B Forecasting Problem

Suppose that an individual’s beliefs of relative performance are a continuous random variable $X$. Let beliefs have density $g(x)$, continuous and with support in $[a, b]$, with $0 \leq a < b \leq 1$. Suppose this individual has initial wealth $\bar{w}$ and utility of wealth $U(w)$. Let $f$ represent the individual’s point forecast, with $f \in [0, 1]$. This individual’s wealth—a continuous version of the discrete quadratic scoring rule—is given by $w = \bar{w} + [w_0 - (x - f)^2]$, with $w_0 \geq 1$. The optimal point forecast of this individual is given by

$$\max_{f \in [0, 1]} \int_a^b U(\bar{w} + w_0 - (x - f)^2)g(x)dx.$$  \hspace{1cm} (2.2)

We will call (2.2) the point forecast problem. The first-order condition to (2.2) is given by

$$\int_a^b U'(\bar{w} + w_0 - (x - f^*)^2)2(x - f^*)g(x)dx = 0.$$

and the second-order condition by

$$\int_a^b [U''(\bar{w} + w_0 - (x - f^*)^2)2(x - f^*)^2 - U'(\bar{w} + w_0 - (x - f^*)^2)] g(x)dx < 0.$$

If an individual is risk averse we have $U' > 0$ and $U'' < 0$ and the second-order condition is verified. If an individual is risk neutral we have $U' > 0$ and $U'' = 0$ and the second-order condition is also satisfied. If an individual is risk seeking we have $U' > 0$ and $U'' > 0$ and we can’t tell if the second-order condition is satisfied or not.

It is a well known result that optimal point forecast of a risk neutral individual is his mean belief of relative performance.\textsuperscript{37} Proposition 1 shows that the

optimal point forecast of an individual with uniform beliefs of relative performance is his mean belief of relative performance, regardless of his preferences towards risk.

**Proposition 2.1** If an individual’s beliefs of relative performance have the uniform distribution with support \([a, b]\), with \(0 \leq a < b \leq 1\), then \(f^* = E(X)\).

**Proof.** Using integration by parts, the first-order condition to the point forecast problem is equivalent to

\[
U(\bar{w} + w_0 - (b - f^*)^2)g(b) - U(\bar{w} + w_0 - (a - f^*)^2)g(a) = \int_a^b U(\bar{w} + w_0 - (x - f^*)^2)g'(x)dx.
\]

If beliefs have the uniform distribution, then \(g'(x) = 0\) for all \(x\) and \(g(a) = g(b)\), so the above condition reduces to \(U(\bar{w} + w_0 - (b - f^*)^2) = U(\bar{w} + w_0 - (a - f^*)^2)\), or \(f^* = (a + b)/2 = E(X)\), that is, the optimal point forecast of an individual with uniform beliefs of relative performance is his mean belief of relative performance.

Proposition 2.2 shows that the optimal point forecast of an individual with unimodal and symmetric beliefs of relative performance is his mean belief of relative performance, regardless of his preferences towards risk.

**Proposition 2.2** If an individual’s beliefs of relative performance are unimodal and symmetric, then \(f^* = E(X)\).

**Proof.** Let the distribution of beliefs have support in \([a, b]\). Using integration by
parts, the first-order condition to the point forecast problem is equivalent to

\[ U(\bar{w} + w_0 - (b - f^*)^2)g(b) - U(\bar{w} + w_0 - (a - f^*)^2)g(a) = \int_a^b U(\bar{w} + w_0 - (x - f^*)^2)g'(x)dx. \]

or,

\[ U(\bar{w} + w_0 - (a - f^*)^2)g(a) + \int_a^{E(X)} U(\bar{w} + w_0 - (x - f^*)^2)g'(x)dx = \]

\[ U(\bar{w} + w_0 - (b - f^*)^2)g(b) + \int_{E(X)}^b U(\bar{w} + w_0 - (x - f^*)^2)(-g'(x))dx. \]

If \( a \leq f^* < E(X) \) and \( g \) is symmetric and unimodal, then the first term in the LHS is greater than the first term in the RHS and the value of the integral in the LHS is greater than the value of integral in the RHS. But then the value of the LHS is greater than the value of the RHS, a contradiction. If \( E(X) < f^* \leq b \) and \( g \) is symmetric and unimodal, the first term in the LHS is smaller than the first term on the RHS and the value of the integral in the LHS is smaller than the value of integral in the RHS. But then the value of the LHS is smaller than the value of the RHS, a contradiction. Thus, it must be that \( f^* = E(X) \)

\[ \text{(2.12.C Betting Problem)} \]

Suppose that an individual has beliefs of relative performance given by the density \( g(x) \), with support in \([a, b]\), with \( 0 \leq a < b \leq 1 \). Suppose this individual has initial wealth \( \bar{w} \) and utility of wealth given by \( U(w) \). Let \( c \) represent the choice of bet, with \( c \in [0, 1] \). This individual’s wealth—a continuous version of our discrete bets choice—is given by

\[
w = \begin{cases} 
\bar{w} + \frac{w_0}{1-c}, & x \geq c \\
\bar{w}, & x < c 
\end{cases}
\]
with \( w_0 \geq 1 \). The optimal bet of this individual is the solution to

\[
\max_{c \in [0,1]} G(c)U(\bar{w}) + [1 - G(c)] U\left(\bar{w} + \frac{w_0}{1 - c}\right).
\] (2.3)

We will call (2.3) the betting problem. We can state the following result.

**Proposition 2.3** If an individual is risk neutral and his beliefs of relative performance are

(i) uniform with support \([a, 1]\), with \(0 \leq a\), then his optimal bet is any \(c^* \in [a, 1]\);

(ii) uniform with support \([a, b]\) with \(0 \leq a < b < 1\) then \(c^* = a < E(X) = f^*\);

(iii) unimodal and symmetric, then \(a < c^* < \text{Mode}(X) = E(X) = f^*\);

(iv) unimodal and positively skewed, then \(a \leq c^* \leq \text{Mode}(X) < E(X) = f^*\).

**Proof:** Let start by proving (i). If an individual is risk neutral and has uniform beliefs with support in \([a, 1]\) then the objective function of the betting problem is \(\bar{w} + w_0/(1 - a)\). Since this individual’s utility does not depend on his choice of bet he must be indifferent between any bet in \([a, 1]\).

Let us show (ii). If an individual is risk neutral and has uniform beliefs with support in \([a, b]\) with \(0 \leq a < b < 1\), then the objective function problem of the betting problem is \(\bar{w} + \frac{b - c}{b - a} \frac{w_0}{1 - c}\). It is clear that for this case the optimal bet is \(c^* = a\).

Let us show (iii). If an individual is risk neutral and has unimodal and symmetric beliefs, then the first-order condition to the betting problem becomes \(-g(c^*)\frac{w_0}{1 - c^*} + [1 - G(c^*)] \frac{w_0}{(1 - c^*)^2} = 0\) or \(1 - G(c^*) = g(c^*)(1 - c^*)\). This is equivalent to

\[
\int_{c^*}^{1} g(x)dx = \int_{c^*}^{1} g(c^*)dx.
\] (2.4)
If we can show there exists an $x_0$ strictly greater than $c^*$ such that \( g(c^*) < g(x_0) \) then it must be that $c^* < \text{Mode}(X)$ since $\text{Mode}(X) = \max g(x)$. Suppose, by contradiction that: (1) for all $x > c^*$ we have $g(x) \leq g(c^*)$ and (2) that there exists an $x_0 > c^*$ such that $g(x_0) \leq g(c^*)$. By the well know result that one can integrate inequalities, assumptions (1) and (2) imply that \( \int_{c^*}^{1} g(x)dx < \int_{c^*}^{1} g(c^*)dx \), which contradicts (2.4). Thus, we must either have that (a) $g(x) = g(c^*)$ for $x \geq c^*$, or (b) there exists an $x_0 > c^*$ such that $g(c^*) < g(x_0)$. Case (a) is a degenerate case. If case (b) holds then we know that $c^* < \text{Mode}(X)$. So, for a unimodal and symmetric density of beliefs we have that $c^* < \text{Mode}(X) = E(X)$. To finish the proof we still need to show that the second-order condition to the betting problem is satisfied. This condition is given by

\[
-g'(c^*) \frac{w_0}{1-c^*} - 2g(c^*) \frac{w_0}{(1-c^*)^2} + 2 [1 - G(c^*)] \frac{w_0}{(1-c^*)^3},
\]

which simplifies to \(-g'(c^*) \frac{w_0}{1-c^*}\). We see that the second-order condition is satisfied whenever $g'(c^*) > 0$. But, if $c^* < \text{Mode}(X) = E(X)$ and the distribution is unimodal and symmetric, then it must be that $g'(c^*) > 0$.

Finally, let us show (iv). When $g'(a) > 0$ the proof is similar to that of (iii) with the exception that for a unimodal and positively skewed density of beliefs we have that $\text{Mode}(X) < E(X)$. Note that when $g'(a) > 0$ the second-order condition is satisfied and $a < c^* < \text{Mode}(X) < E(X)$. When $g'(a) < 0$ we have a corner solution: $c^* = \text{Mode}(X) = a$.  

\[\blacksquare\]
2.12.D Prize Structures and Earnings from Forecasts

Table 2.7: UCSD’s 2004 Poker Classic Prizes

<table>
<thead>
<tr>
<th>Rank</th>
<th>Prize</th>
<th>Rank</th>
<th>Prize</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st place</td>
<td>$447 (27%)</td>
<td>1st place</td>
<td>$792 (27%)</td>
</tr>
<tr>
<td>2nd place</td>
<td>$209 (12%)</td>
<td>2nd place</td>
<td>$370 (12%)</td>
</tr>
<tr>
<td>3rd place</td>
<td>$164 (10%)</td>
<td>3rd place</td>
<td>$290 (10%)</td>
</tr>
<tr>
<td>4th place</td>
<td>$149 (9%)</td>
<td>4th place</td>
<td>$264 (9%)</td>
</tr>
<tr>
<td>5th place</td>
<td>$134 (8%)</td>
<td>5th place</td>
<td>$238 (8%)</td>
</tr>
<tr>
<td>6th place</td>
<td>$119 (7%)</td>
<td>6th place</td>
<td>$211 (7%)</td>
</tr>
<tr>
<td>7th place</td>
<td>$104 (6%)</td>
<td>7th place</td>
<td>$185 (6%)</td>
</tr>
<tr>
<td>8th place</td>
<td>$89 (5%)</td>
<td>8th place</td>
<td>$158 (5%)</td>
</tr>
<tr>
<td>9th place</td>
<td>$75 (4%)</td>
<td>9th place</td>
<td>$132 (4%)</td>
</tr>
<tr>
<td>10th-18th places</td>
<td>$20 (12%)</td>
<td>10th-18th places</td>
<td>$40 (12%)</td>
</tr>
<tr>
<td>Sum</td>
<td>$1670 (100%)</td>
<td>Sum</td>
<td>$3000 (100%)</td>
</tr>
</tbody>
</table>

Table 2.8: Sintra’s Chess Open Prizes

<table>
<thead>
<tr>
<th>Rank</th>
<th>Monetary Prize</th>
<th>Symbolic Prize</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st place</td>
<td>300 euro (27%)</td>
<td>Trophy</td>
</tr>
<tr>
<td>2nd place</td>
<td>180 euro (16%)</td>
<td>Trophy</td>
</tr>
<tr>
<td>3rd place</td>
<td>120 euro (11%)</td>
<td>Trophy</td>
</tr>
<tr>
<td>4th place</td>
<td>75 euro (7%)</td>
<td>Medal</td>
</tr>
<tr>
<td>5th place</td>
<td>50 euro (5%)</td>
<td>Medal</td>
</tr>
<tr>
<td>6th-10th places</td>
<td>30 euro (14%)</td>
<td>Medal</td>
</tr>
<tr>
<td>11th-15th places</td>
<td>25 euro (11%)</td>
<td>Medal</td>
</tr>
<tr>
<td>16th-20th places</td>
<td>20 euro (9%)</td>
<td>Medal</td>
</tr>
<tr>
<td>Sum</td>
<td>1100 euro (100%)</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2.9: Players’ Earnings from Forecasts

<table>
<thead>
<tr>
<th></th>
<th>Winter Poker Classic</th>
<th></th>
<th>Spring Poker Classic</th>
<th></th>
<th>Sintra’s Chess Open</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reward</td>
<td>Paid</td>
<td>Reward</td>
<td>Paid</td>
<td>Reward</td>
<td>Paid</td>
<td>Reward</td>
</tr>
<tr>
<td>$ 0</td>
<td>107</td>
<td>$ 0</td>
<td>105</td>
<td>0 euro</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>$ 1</td>
<td>4</td>
<td>$ 4</td>
<td>3</td>
<td>1 euro</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$ 6</td>
<td>6</td>
<td>$11</td>
<td>2</td>
<td>6 euro</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$ 9</td>
<td>4</td>
<td>$16</td>
<td>1</td>
<td>9 euro</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>$10</td>
<td>1</td>
<td>$19</td>
<td>1</td>
<td>10 euro</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Rewards $86  Rewards $149  Rewards $110 euro

Players 122  Players 116  Players 60

Average $0.70  Average $1.28  Average 1.83 euro
3

A Simple Theory of the False Consensus Effect

3.1 Introduction

In most strategic environments, it is important to be able to estimate the characteristics and behavior of other. For example, optimal bidding in auctions depends on not only one’s valuation of the good but also the valuation of other. In a bargaining situation, players’ optimal strategies will depend on the valuations or time preferences of their opponents. Estimating the taste of potential customers is a key to estimate the demand for a new business. If the characteristics of other players are not known, estimating them is a critical task.

Psychological evidence suggests individuals possess systematic bias when they compare their attributes to others. In particular, the False Consensus Effect states that individuals overestimate the number of the people who possess the same attributes as they do. When there are two possible positions for people to
choose, A and B, the false consensus effect exists when the average estimation of the fraction of people who chose position A made by those who hold position A is greater than the average estimation the fraction of people who chose position A made by those who hold position B. To the extent that the false consensus effect demonstrates an inability of individuals to process information rationally, it suggests that conventional economic analysis of strategic behavior, which assumes unbiased expectation formation, is deficient. This paper studies the extent to which the false consensus bias is consistent with a conventional model of information processing.

Ross et al. (1977) first reports the false consensus bias. The authors conducted a series of experiments that asked college students if they were willing to perform certain actions. Both the group who agreed to perform the action and the group who were unwilling to perform the action overestimated the size of the group who responded as they did. For example, in one of the experiments, college undergraduates were asked whether they would wear a sandwich sign saying “Eat at Joe’s”. Subjects’ reaction were divided evenly\(^1\), but their consensus estimation was biased toward overestimating their own cohort group. When asked to estimate the proportion of responders who agreed to wear the sign, the average subjects who agreed to wear the sign estimated that 64.6% of the subjects would wear the sign. The average estimate of the group by the subjects who would not wear the sign was 31.2%.

Dawes (1989) points out that the seemingly biased behavior can be described as a rational response given the fact that individuals are always able to observe their own attributes. This specific nature of own observation makes people project others’ behavior from their own. He claimed that it is rational to use own endorsement as part of the information to estimate others’ behavior. To il-

\(^1\)53% answered that they would wear the sign and 47% answered they would’t.
Suppose individuals have a prior belief that everyone is equally likely to choose either alternative. If they only observe the choice they made and update the prior regarding their choice as a single sample, they will all consider themselves to belong to the majority group. Consequently, the population possesses the false consensus effect. This is called the induction hypothesis. With this observation, Dawes claims that the false consensus effect should not be considered irrational unless people consider their own choice to be more informative than a choice made by another person in the process of estimating the overall composition of population choice. This `generalized’ false consensus effect has been considered as an example of irrational behavior even after Dawes’s finding.

Although the induction hypothesis can explain the tendency for people to possess the false consensus effect, it overly simplifies agents’ process of selecting their choices. Classical approaches by social psychologists focused on the estimation procedure once individuals’ endorsements are already determined. Thus the analysis on how individuals reach the decision was overlooked. However, it is reasonable to believe that individuals have different preferences and beliefs even though they make the same choice. This observation motivates introducing another framework that describes individuals’ procedures of making own decisions.

I propose a simple model of information acquisition in which individuals overestimate the size of the cohort group due to the imperfect information. In my model, an individual observes partial information about the true state of the world that determines the overall distribution of people’s choices. This information is used not only to make her own decision but also to estimate the true state of the world and others’ decisions. The estimate will be biased toward overestimating own cohort group because people will overestimate the probability of others receiving similar information. Consequently, the result suggests that the generalized false consensus effect can arise as the consequence of a rational individual’s reaction to
a limited information.

The rest of the paper is organized as follows. Section 2 reviews related literatures. I propose the model and its implications in Section 3. Section 4 summarizes several preliminary results that are used in following analysis. The stylized facts of the bias are introduced in Section 5. Section 6 relates the main results to the stylized facts. I attempt to illustrate the classical hypotheses of the false consensus effect from social psychology and compare their different implications on several situations in Section 7. Section 8 is the conclusion.

3.2 Related literature

Since the seminal finding of Ross et al. (1977), the false consensus effect was studied both experimentally and theoretically in social psychology. The experimental evidence indicates people possess this bias on various questions related to personal traits and characteristics (being optimistic, having brown or blue eyes (Ross et al, 1977), attitude on women’s rights, (Judd and Johnson, 1981)), personal preference (eating white or brown bread (Gilovich et al, 1983)), personal activities (housekeeping activities (van der Pligt, 1984)), political expectations (preferred presidential candidate (Brown, 1982), performance of the political leaders (Manstead, 1982), prediction on the result of national vote (Baker et al, 1995)), factual information (knowledge of trivia questions (Mullen, 1983), the induction problem on the identification of suicide notes (Sherman et al, 1984)), and so on.

Mullen et al. (1985) reviewed 115 experiment results and tested the robustness of the false consensus effect. By combining the results of multiple

\footnote{Subjects were provided two possible suicide notes, in which only one of them is an actual note. Subjects guessed the correct one and asked to estimate the proportion of people who made the same choice.}
studies, they found the evidence of the false consensus effect in various attributes. Furthermore, their results indicate that the broadness of the reference group (eg. whether the estimation was made over the whole population or fellow students) and the difference between alternative choices in actual consensus (the difference between the agreed and the disagreed) do not have significant effect on the degree of the false consensus effect.\(^3\)

Following Dawes (1989), who introduced a rational explanation of the false consensus effect, there were several studies conducted to confirm the existence of generalized false consensus effect. Alicke and Largo (1995, study 2) ran an experiment that lets subjects take a test consisted with a series of social sensitivity questions. After taking the test, subjects were informed the bogus result whether they passed or failed the test. Also, the subjects were informed about the result of 4 other subjects, which was manipulated to form a combination of either 4 pass and 1 fail result or 1 pass and 4 fail result. Then subjects were asked to provide the estimate of subjects who passed the test out of 1000 imaginary subjects. They measured the degree of the generalized false consensus effect by comparing the answers of subjects who share the same overall information sets, but distinguished by their own outcomes. The result indicates that subjects put higher weight on their own (fake) endorsement than those of others.

Engelmann and Strobel (2000) test the generalized false consensus effect by conducting an experiment where subjects estimate the consensus after observing randomly selected choices of others. They share the view of Dawes (1989) that the false consensus effect can arise as a result of rational reaction due to the nature of own observation and improve the experimental design of Alicke and Largo (1995) by providing monetary incentives and eliminating the manipulation.

\(^3\)On the other hand, they found the number of items in each experiment and the order of the report (whether the estimate is made first, or making of own choice is made first) has significant effect of the degree of the false consensus effect.
On various questions about own attributes, subjects in groups of 16 individuals answered their own endorsements. Then, they observed endorsements of 4 other randomly selected subjects. Finally, they estimated the consensus of the remaining 11 subjects. By comparing the estimate between the subjects who share the same overall observations yet opposite own choice, the authors test whether subjects consider their own choices more informative. The result shows the false consensus effect remains: after observing others’ endorsements, the estimate of the prevalence of each alternative was still greater from the insider. However, the outcome does not indicate subjects possess the generalized false consensus effect.\footnote{Engelmann and Strobel (2004) later found that the generalized false consensus effect reappears when individuals needs to put marginal effort to obtain information of others.} In fact, out of 26 pairings, there were only 11 cases in which the results support the hypothesis of the generalized false consensus effect. Also, only 2 out of 11 supporting results were considered ‘substantial’ (significant with 5% level and at least 5 subjects from each group) while there were 10 substantial cases that indicates the opposite result.

Krueger and Clement (1994) introduced the egocentrism hypothesis to explain the phenomenon that individuals consider own decisions more informative than others’. They introduce the egocentric aspect of human behavior as a possible explanation of generalized false consensus effect. They emphasize that people tend to use different mental procedures when they process their own behavior rather than that of others. Thus people might consider own information more importantly. Consequently, the generalized false consensus effect can be explained under their assumptions.

Goeree and Großer (2007) and Vanberg (2008) introduce simple models that provide rationale of the false consensus effect. Their models characterize the observation of Dawes (1989) and use the false consensus effect to illustrate the behaviors of voters and experimental subjects respectively. Their simple charac-
terization, however, does not fully describe the actual behavior as the simplified assumptions generate the extreme result such that everyone believes that she belongs to the majority. Agents receive a binary signal which is correlated with the binary state of the world. Given agents have uniform prior belief on the states, all agents put higher posterior probability on the state corresponding to their signal. Eventually, everyone believes there are more people who received the same signal they received than those who received the opposite signal.

3.3 Setup

The state of the world is $\theta \in [0, 1]$. There is a continuum of agents. They are characterized by a preference parameter $\lambda$ and a signal $s$. We assume that $\lambda$ is distributed independently of $\theta$ and $s$. The density of $\lambda$ on $[0, 1]$ is $g(\cdot)$ with cdf $G$. We interpret $s \in [0, 1]$ to be an idiosyncratic estimate of the true state $\theta$. We assume that the signals received by individual agents are identically distributed and conditionally independent given $\theta$. Let $f(s, \theta)$ be the density function of signals and states; $f(\cdot)$ describes the information structure. Assume that $s$ is an unbiased estimate of $\theta$ so that the conditional expectation of $\theta$ given $s$ is equal to $s$. All distribution functions are common knowledge.

We assume that $f$ satisfies the Monotone Likelihood Ration Property (MLRP). More precisely for $s < s'$ and $\theta < \theta'$,

$$\frac{f(s'|\theta')}{f(s'|\theta)} \leq \frac{f(s'|\theta)}{f(s|\theta)}.$$

A well known consequence of this property (Krishna, 2009) is

1. $F(s|\theta)$ is decreasing in $\theta$.

2. $F(\theta|s)$ is decreasing in $s$. 
Given her characteristics, an agent makes a binary choice, picking either $h$ or $l$. Assume that preferences are such that an $(s, \lambda)$ type chooses $h$ if and only if $s \geq \lambda$.\(^5\) Let $\mathcal{H}$ be the set of agents in the economy who select $h$.

Let $H(\theta) \equiv \int_0^1 G(s)f(s|\theta)ds$. $H(\theta)$ is the fraction of the population that selects $h$ when the true state of the world is $\theta$. We are interested in how an agent’s estimate of $H$ depends on her membership in $\mathcal{H}$. Assume agents’ prior belief of the true $H$ is $\tau$. i.e. $\int_0^1 H(\theta')f_\theta(\theta')d\theta' = \tau$ where $f_\theta$ is the marginal density of $\theta$.

Agents can also observe other agents’ endorsements. In this case, it is useful to define the conditional distribution of the state $\theta$ given all available information.

**Definition 3.1** Let $\Gamma(\theta|s, n, k)$ be the conditional distribution of $\theta$ when an agent receives the signal $s$, and observes $n$ numbers of the population where $k$ of them endorses $h$. The conditional distribution will be denoted as

$$\Gamma(\theta|s, n, k) = \frac{\int_0^\theta f(s, t)H(t)^k(1-H(t))^{n-k}dt}{\int_0^1 f(s, r)H(r)^k(1-H(r))^{n-k}dr} \quad (3.1)$$

with the corresponding pdf, $\gamma(\theta|s, n, k)$.

Note $\Gamma(\theta|s, n, k)$ can be viewed as a generalization of the conditional distribution $F(\theta|s)$. $\Gamma$ corresponds to the case in which individuals observe not only their own signal, but also the endorsements of the subset of the population.

Also notice that, regardless of agents’ own signal, $s$, it is assumed that there is no bias observing other agents’ endorsement. i.e. each agent observes the same number of endorsements and the probability of observing a certain number of $h$ endorsers is the same for everyone. Also, assume agents do not change their own

\(^5\)For example, an agent’s preferences may be represented by $(\theta - \lambda)(x_h - x_l)$ where $x_i = 1$ if the agent takes action $i$ and 0 otherwise.
endorsement after they observe other agents’ behavior. The observation of other agents’ endorsement changes the belief of the true state. Therefore, agents might have incentive to change their own endorsement depending on the characteristics of the problem.\footnote{These two assumptions are revisited in Section 7 and the implications are analyzed when the assumptions are relaxed.}

### 3.4 Preliminary Results

It is useful to state some preliminary results that are commonly used in the analysis. First of all, the higher the true state is, the more people choose $h$.

**Lemma 3.1** $H(\theta)$ is increasing in $\theta$.

Proof) for $\theta < \theta'$,

$$H(\theta) - H(\theta') = \int_0^1 G(s)[f(s|\theta) - f(s|\theta')]ds$$

$$= -\int_0^1 g(s)[F(s|\theta) - F(s|\theta')]ds$$

where the second equation is obtained by integration by parts. The expression is non positive due to the MLRP of $F$. \[\blacksquare\]

In order to compare the differences of estimates between two endorsement groups, one can consider the probability distribution of signal $s$ of agents in each endorsement group.

**Definition 3.2** Define the $\Phi_i(s)$ as the distribution of $s$ given that the agent is in category $i$:

$$\Phi_i(s) = \int_0^1 \int_0^s G(r)f(r, \theta)dr \frac{d\theta}{\int_0^1 G(t)f(t|\theta)dt}$$

(3.2)
and
\[
\Phi_l(s) = \int_0^1 \frac{\int_0^s (1 - G(r)) f(r, \theta) dr}{\int_0^1 (1 - G(t)) f(t, \theta) dt} d\theta.
\] (3.3)

Also denote \(\phi_i(s)\) for \(i \in \{h, l\}\) as the corresponding density function. \(\Phi_h\) first order stochastically dominates \(\Phi_l\).

**Lemma 3.2** \(\Phi_h(s) \leq \Phi_l(s)\) for all \(s\).

Proof
\[
\Phi_h(s) - \Phi_l(s) = \int_0^1 \left[ \int_0^s G(t) f(t, \varphi) dt - \int_0^s (1 - G(t)) f(t, \varphi) dt \right] d\varphi
= \int_0^1 \left[ \int_0^s G(t) f(t|\varphi) dt - F(s|\varphi) \int_0^1 G(t) f(t|\varphi) dt \right] f_\theta(\varphi) d\varphi
\]

The numerator of the integrand can be rearranged to
\[
\int_0^s G(t) f(t|\varphi) dt - F(s|\varphi) \int_0^1 G(t) f(t|\varphi) dt
= (1 - F(s|\varphi)) \int_0^s G(t) f(t|\varphi) dt - F(s|\varphi) \int_0^1 G(t) f(t|\varphi) dt
= G(s) f_\theta(\varphi) \left( (1 - F(s|\varphi)) \int_0^s \frac{G(t)}{G(s)} f(t|\varphi) dt - F(s|\varphi) \int_s^1 \frac{G(t)}{G(s)} f(t|\varphi) dt \right)
\] (3.4)

since
\[
\int_0^s \frac{G(t)}{G(s)} f(t|\varphi) dt \leq F(s|\varphi), \quad \text{and} \quad \int_s^1 \frac{G(t)}{G(s)} f(t|\varphi) dt \geq 1 - F(s|\varphi),
\]
expression (3.4) is non-positive.

Lemma 3.2 states that, on average, agents in \(H\) group receive higher signals compared to agents in \(L\) group.
In order to compare the average estimate of $H$ made by two endorsement groups, we further need to describe agents’ estimation rule. There are two possible cases depending on the information set available to the agents. The first case is when individuals cannot observe any of the other agents’ endorsements. The situation can be interpreted as the initial response to the task. In this situation, the estimate of the others’ endorsements should be determined only by the signal of the true state, $s$.

**Definition 3.3** Define $A(s)$ as the expected estimation of $H$ for an agent who received the signal $s$. More precisely,

$$A(s) = \int_0^1 H(\theta)f(\theta|s)d\theta$$  \hspace{1cm} (3.5)

Another possible scenario is when agents are able to observe the endorsements of a subset of the population. After all agents determine their initial endorsements, they observe other agents’ behaviors. This case is a generalization of the case where no observation is possible. The formal definition of agents’ estimation rule follows.

**Definition 3.4** Define $\tilde{A}(s, n, k)$ as the expected estimation of $H$ for an agent who received the signal $s$, and observed $n$ number of other agents’ endorsement where $k$ of their endorsements are $h$. More precisely,

$$\tilde{A}(s, n, k) = \int_0^1 H(\theta)\gamma(\theta|s, n, k)d\theta$$  \hspace{1cm} (3.6)

If she observes $n$ members of the population and $k$ of them endorses $h$, her estimation of the fraction of the agents in $H$ is denoted as $\tilde{A}(s, n, k)$. For simplicity, assume the number of the observed endorsement, $n$, is same for all agents.
3.5 Stylized Facts

This section describes several experimentally observed stylized facts about the false consensus effect in the language of the model.

3.5.A Population possess the false consensus effect

In multiple experiments with various settings, subjects systematically show that they overestimate the size of their own cohort group. This is the false consensus effect.

**Definition 3.5 (False Consensus Effect)** Given the true state and available information $s$, the population possesses the false consensus effect if

$$
\int_0^1 A(s)\phi_h(s)ds \geq \int_0^1 A(s)\phi_l(s)ds
$$

with a strict inequality if the signal in non-trivial.

The average estimation of $H$ made by agents in $\mathcal{H}$ is compared to the average estimation of $H$ made by agents in $\mathcal{L}$. If the estimation made by agents in $\mathcal{H}$ is larger, then we call the population possess the false consensus effect.

3.5.B Generalized false consensus effect

According to induction hypothesis, individuals might possess the false consensus because they consider their own endorsement decision as a single sample in the process of estimating the overall behavior of the population. The underlying logic is that there is no reason for individuals to undervalue the information drawn
from their own decision. This observation further suggests that there should be no overvaluing of one's own endorsement either. Thus, once individuals are exposed to more observations of other individuals’ choices, they should evaluate another individual’s endorsement in the same manner as they evaluate their own information. In various experiments (Sherman et al. (1984), Krueger and Clement (1994), and Alicke and Largo (1995)), however, it has shown that subject put more weight on own choices than that of another when they estimate the behaviors of the population. Engelmann and Strobel (2000) report the result that this tendency disappears under the setting in which the subjects had proper incentives and was not deceived by the experimenter. However, they also found the result is fragile (Engelmann and Strobel, 2004) as the generalized false consensus effect reappears when subjects are required to put marginal effort to obtain the information of other subjects.

**Definition 3.6 (Generalized False Consensus effect)** Given the true state and available information \((s, n, k)\), the population possesses the generalized false consensus effect if

\[
\int_0^1 \tilde{A}(s, n, k) \phi_h(s) ds \geq \int_0^1 \tilde{A}(s, n, k) \phi_l(s) ds \tag{3.8}
\]

with a strict inequality if the signal in non-trivial.

Similar to the case of the false consensus effect, the generalized false consensus effect is defined by comparing the average estimation of \(H\) made by agents in \(\mathcal{H}\) is compared to the average estimation of \(H\) made by agents in \(\mathcal{L}\). If the estimation made by agents in \(\mathcal{H}\) is larger, then we say that the population possess the generalized false consensus effect.
3.5.C Completeness

Another commonly observed phenomenon in the experimental results of the false consensus effect is that regardless of the endorsement group, there are agents who believe they are majorities as well as agents who believe they are minorities. Krueger (1998) reports a more detailed result from the experiment of Dawes and Mulford (1996) in terms of the subjects’ accuracy of the estimation. Among 68% of subjects who belonged to the majorities, about 23% of them estimated that they belong to the minorities. Also, 17% out of 32% who belong to the minorities believe they belong to the majorities.

Table 3.1: Composition of Estimation (Krueger, 1998)

<table>
<thead>
<tr>
<th>Actual Consensus</th>
<th>Estimated Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Majority</td>
</tr>
<tr>
<td>Majority</td>
<td>0.62</td>
</tr>
<tr>
<td>Minority</td>
<td>0.45</td>
</tr>
<tr>
<td>Majority</td>
<td>0.68</td>
</tr>
<tr>
<td>Minority</td>
<td>0.17</td>
</tr>
<tr>
<td>Minority</td>
<td>0.32</td>
</tr>
</tbody>
</table>

This evidence cannot be generated by the models of Goeree and Großer (2007) or Vanberg (2008) as they did not allow heterogeneity of agents in their models. Also, the signal agents receive is characterized in a simple binary way. Hence, according to their settings, every agent believes she belongs to the majority.

Definition 3.7 (Completeness) The population is called complete if given \( \theta \), the lowest possible signal \( \underline{s} \) and the highest possible signal \( \overline{s} \) satisfy \( \tilde{A}(\underline{s}, n, k) < \frac{1}{2} \) and \( \tilde{A}(\overline{s}, n, k) > \frac{1}{2} \), and there exist \( \lambda^*, \lambda^{**} \in \Lambda \) such that \( \lambda^* < \underline{s} \) and \( \lambda^{**} \geq \overline{s} \).

\(^7\)Subjects were asked to estimate the consensus over 128 questions from the California Psychological Inventory. Since the authors had access to the data collected from a large population, the actual consensus of the population could be measured.
The existence of $\lambda^*$ implies that there is an agent who endorses $h$ and believes herself to be a minority. The existence of $\lambda^{**}$ implies that there is an agent who endorses $l$ and believes herself to be a minority. Note that the existence of agents who believe they belong to the majority is automatically achieved.

### 3.6 Results

In this section, I introduce the implications of the model on the stylized facts described in the previous section.

#### 3.6.A False Consensus Effect

**Proposition 3.3** The population possess the false consensus effect.

The formal proof is omitted as it is a special case of Proposition (3.5). The agents in $\mathcal{H}$ group tend to receive higher signals and the higher signals lead them to make a higher estimate the size of $\mathcal{H}$. Thus, on average, the $\mathcal{H}$ group has a higher estimate of the size of the $\mathcal{H}$ group than does the $\mathcal{L}$ group.

#### 3.6.B Generalized false consensus effect

Now consider the case in which agents can observe some of the others’ endorsements. $\Gamma$ is the conditional distribution of $\theta$ given agents received their own signal, $s$ and observed $n$ other agents’ choices where $k$ of these choices are $h$. It can be shown that the MLRP of the conditional distribution $F$ is preserved with extra information, $(n, k)$. 
Lemma 3.4 \( \Gamma(\theta|s, n, k) \geq \Gamma(\theta|s', n, k) \) for \( s < s' \).

Proof) For \( \theta < \theta' \),
\[
\frac{\gamma(\theta'|s, n, k)}{\gamma(\theta|s, n, k)} = \frac{f(\theta'|s) H(\theta')^k (1 - H(\theta'))^{n-k}}{f(\theta|s) H(\theta)^k (1 - H(\theta))^{n-k}}
\]
is decreasing in \( s \) due to the MLRP of \( F \). Hence for \( s < s' \),
\[
\frac{\gamma(\theta'|s, n, k)}{\gamma(\theta|s, n, k)} \geq \frac{\gamma(\theta'|s', n, k)}{\gamma(\theta|s', n, k)}.
\]
Thus \( \Gamma \) also satisfies the MLRP with respect to \( s \).

Lemma (3.4) demonstrates that even with the observations of others’ endorsements, the relationship between \( s \) and \( \theta \) remains the same as the case without the observations. With the result that \( \Gamma \) behave in similar ways as \( F \), I can show that population possess the generalized false consensus effect.

Proposition 3.5 The population possess the Generalized False Consensus Effect.

Proof) \( \tilde{A}(s, n, k) \) increases in \( s \) as for \( s < s' \),
\[
\tilde{A}(s, n, k) - \tilde{A}(s'n, k) = \int_0^1 H(\theta) \left[ \gamma(\theta|s, n, k) - \gamma(\theta|s'n, k) \right] d\theta
\]
\[
= - \int_0^1 H'(\theta) \left[ \Gamma(\theta|s, n, k) - \Gamma(\theta|s'n, k) \right] d\theta
\]
\[
\leq 0.
\]
The result follows as
\[
\int_0^1 \tilde{A}(s, n, k)\phi_h(s)ds - \int_0^1 \tilde{A}(s, n, k)\phi_l(s)ds = \int_0^1 \tilde{A}(s, n, k)[\phi_h(s) - \phi_l(s)]ds
\]
\[
= - \int_0^1 \tilde{A}_1(s, n, k)[\Phi_h(s) - \Phi_l(s)]ds
\]
\[
\geq 0
\]
where $\tilde{A}_1(s, n, k)$ is the partial derivative of $\tilde{A}$ with respect to $s$.

The result is derived directly by the MLRP assumption. On average, the agents in $H$ receives higher $s$ than the ones in $L$. Therefore, their average estimation of the fraction of the population who endorses $h$ is larger among the members in $H$.

3.6.C Completeness

The population is complete if, in both endorsement groups, there are agents who believe they are majorities as well as agents who believe they are minorities. Intuitively, if agents’ preferences are sufficient diverse, there will be diversity in agents’ belief.

**Proposition 3.6** Define $\underline{\eta} = \max\{\underline{\lambda}, \underline{\mathbf{s}}\}$ and $\overline{\eta} = \min\{\overline{\lambda}, \overline{\mathbf{s}}\}$ where $\underline{\lambda}$ and $\overline{\lambda}$ are lowest and highest preference, and $\underline{\mathbf{s}}$ and $\overline{\mathbf{s}}$ are lowest and highest signal given $\theta$ respectively. The population is complete if for all $s' \in \{s : A(s) < 1/2\}$ and $s'' \in \{s : A(s) > 1/2\}$, the following two conditions hold.

1. $\underline{\eta} < s'$
2. There exists the preference value $\lambda^*$ such that $s'' < \lambda^* < \overline{\eta}$.

Proof) An agent whose preference is $\underline{\eta}$ and receives the same signal chooses $h$ and believes herself to be a minority. Also, an agent whose preference is $\lambda^*$ and receives the signal $\overline{\eta}$ chooses $l$ and believes herself to be a minority. Also, note that the results with the observations, $(n, k)$, is achieved automatically since any extreme observation is possible.
3.7 False Consensus Effect in Social Psychology

In this section, I compare some of the main explanations of the false consensus effect in social psychology with my model and describe their implications. I describe the extent the different predictions could be separated from each other. Krueger (1998) summarizes the hypotheses regarding the false consensus effect and related topics. There are five different categories of hypothesis: sampling bias, motivational bias, induction, reversal of causality, and overconfidence in one’s own belief.

The hypothesis of sampling bias states that individuals receive or recognize biased samples about others’ behavior. The reason for this may be ‘selective exposure’- the tendency to associate with similar people. In this case, individuals could overestimate the prevalence of own choice as the available sample is drawn from a subset with more similar members. Also, psychological evidence shows that the similar behavior of others comes to mind more easily than dissimilar behavior. Individuals tend to view own attributes as more salient than the alternative. If individuals notice similar attributes more easily than dissimilar ones, then it is possible for them to systematically overestimate the number of agents who behave as they do.

The hypothesis of motivational bias assumes that people want to consider their own behaviors to be rational and appropriate responses to the situation. One mechanism by which the bias could arise is if agents believe that if others select an action, then there is a rational basis for selecting the action. Thus, individuals possess an intrinsic preference to overestimate the prevalence of own decisions in the population because behaving like others is a sign of rationality. The hypothesis is closely related to the implication of the self-serving bias - the tendency for people to attribute their successes to their own abilities but their failures to chance. People
with a self-serving bias may overestimate the prevalence of their own bad behavior, but underestimate the prevalence when the behavior is desirable (Suls et al., 1988). They protect their own ego by perceiving that their own undesirable traits are more common and enhance their own ego by believing that their own desirable traits are rare. Also people tend to possess greater false consensus bias when the reference group is attractive than when the reference group is unattractive (Marks & Miller, 1982).

The *induction* hypothesis (Dawes, 1989) assumes that the information that forms the basis for an individual’s choice will lead to systematic biases in the assessments of the behavior of others. If the information that induces one’s own decisions is correlated with the information that motivates the decisions of others, some version of the false consensus effect will arise. When estimating an unknown attribute (others’ endorsement), it is natural (and rational) to use the available information. This behavior is called projection, and the false consensus effect is well described with this behavior.

The *reversal of causality* hypothesis attributes the false consensus effect to people’s tendency to choose their own endorsement following the majority or normative behavior. Most of the results of the false consensus effect measure the correlation between agents’ endorsements and estimates. The causal direction is usually assumed but verified only with carefully conducted experiment. Hence it is possible that individuals think that the population is more like they are because the group’s behavior determined their own behavior. Agents who sample different parts of the group may therefore end up behaving differently, but still maintain the belief that their behavior is representative.

The hypothesis of *overconfidence of own belief* assumes that people overestimate the precision of their own belief compared to that of others. This hy-
The hypothesis shares the view of the induction hypothesis that people use their own endorsement as well as others’ to estimate the overall behavior of the population. The hypothesis is introduced to provide rationale to the generalized false consensus effect. To illustrate, consider three cases that differ in the degree of how agents weigh their own endorsement compared to others’. In one extreme case, agents do not consider own endorsement as informative at all, and there will be no false consensus effect given agents observe an unbiased sample of the population. The intermediate situation is the case where agents consider their own endorsement as informative as any other endorsement of another agent. This is the case illustrated by induction hypothesis, and the false consensus effect arises in this case. In the other extreme case, agents do not consider others’ endorsement as informative, and there is the highest possible false consensus effect. In each endorsement group, all agents use the information that led them to that specific endorsement to predict the overall behavior of the population. The hypothesis of overconfidence of own belief assumes that the agents’ behavior is between the second and the third case illustrated above. Krueger and Clement (1994) suggest several assumptions that could give rise to this behavior. Firstly, people may not understand the nature of sampling procedure. If people understand how they make their own decision, but not how others make theirs, then individuals may view their own endorsement to be more informative than that of another since they understand the underlying logic of reaching their decision but not those of others. Secondly, the order of the observations can also explain the bias. The primacy effect indicates the tendency of people to put more weight on the information obtained earlier. Self-related information is an initial observation. Thus, it is possible for individuals to view other’s information obtained later as less important. Finally, own information can be processed more automatically than that of others. The heuristic process takes less effort for own information. Thus in the process of estimating the behavior of the population, agents can lose some information obtained from other agents’
endorsements due to the higher cost to recall them. This difference of mental cost in processing information might be able to explain the discrepancy between own-others’ information.

The alternative hypotheses described above have their own merits. However, each of them predicts different outcomes in some situations depending on the characteristics of the attributes.

My model does not distinguish the bias between behaviors with different desirability. However, the hypothesis of motivational bias predicts that people might overestimate the consensus if the comparison is made on undesirable attributes and underestimate the consensus if the comparison is made on desirable ones. Also, my model does not assume any difference in reference groups. Thus, the result of heterogenous reference group is out of the paper’s scope.

The hypothesis of reversed causal relation predicts that the false consensus effect arises not because people possess bias in assessing others’ behavior, but because they choose their behavior as what the majority does. This result as well as the primacy hypothesis can possibly be tested by reversing the order of information acquisition. Suppose I run an experiment and tell subjects that they sampled a subset of a population and learned that \( x \% \) chose option \( h \). Then tell the subjects what option \( h \) is. The result supports the hypothesis of the reversed causal relation if subjects show behaviors that increase the false consensus bias when the treatment use the reversed order. The result supports the hypothesis of primacy if the subjects report different estimate in different order treatment.
3.8 Conclusion

In this paper, I introduce a simple model of information acquisition that describes the false consensus effect. The model extends the induction hypothesis (Dawes, 1989) that provides a rational explanation of the false consensus effect and describes the false consensus effect in its general form. I suggest that the generalized false consensus effect—people’s tendency to overestimate the information of own choice—can arise if individuals can only observe partial information about the true state of the world.
3.9 References


