Overconfidence and Elitism

Hui-Wen Ng

November 1, 2019

Please click here to access my website for the latest version.

Abstract

I propose a behavioral foundation for the practice of elitism and show that it can give rise to inaccurate statistical discrimination against candidates who lack elite credentials. My theoretical model demonstrates that hirers’ overconfidence in own ability, coupled with personal success, can cause them to overvalue elite credentials beyond what these credentials signal about candidates’ underlying productivity. Results from a controlled laboratory experiment confirm the model’s predictions: on average, subjects with a higher level of confidence have a higher relative willingness to pay for candidates with similar credentials to themselves. By simulating the competitive labor market outcome using the experimental data, I find that an increase in hirer confidence raises elite firms’ tendency to hire candidates with elite credentials, thereby leading to a greater degree of labor market sorting in equilibrium. Given that overconfidence has been widely documented, the results suggest that this channel may be present in hiring decisions in practice.

*I am indebted to my advisors Pedro Dal Bó, Louis Putterman, Geoffroy de Clippel, and Neil Thakral for their support and continuous feedback in the development of this project. I am also thankful to Nisvan Erkal, Jack Fanning, Boon Han Koh, and seminar participants at Brown University for their helpful comments and suggestions. I gratefully acknowledge support from The James M. and Cathleen D. Stone Wealth and Income Inequality Project and Brown University Social Science Research Institute.

†Department of Economics, Brown University.
1 Introduction

The existence of labor market discrimination, the phenomenon where workers with the same productivity are treated differently depending on certain observable characteristics, has been consistently documented (Bertrand and Mullainathan, 2004; Charles and Guryan, 2008; Blau and Kahn, 2017). Theories of discrimination have been developed to explain this phenomenon, either through innate taste-based preferences (Becker, 1957) or statistical discrimination (Phelps, 1972; Arrow, 1973). However, within the literature, a less commonly examined issue is the role of inaccurate beliefs about a worker’s expected productivity, which can give rise to inaccurate statistical discrimination (Bohren et al., 2019).

In this paper, I propose a behavioral foundation for inaccurate statistical discrimination in the context of a particular type of hiring behavior: elitism. In particular, I show that overconfidence, a behavioral bias in which individuals overestimate their own ability, can distort how hirers form beliefs about the informativeness of credentials regarding a candidate’s underlying productivity. Overconfidence causes hirers who experienced personal success, such as attending an elite institution, to place too much emphasis on credentials when evaluating candidates. As a consequence, high ability candidates who lack credentials suffer a disadvantage in the hiring process greater than that caused by regular statistical discrimination alone.

In investigating the impact of overconfidence on hiring outcomes, I proceed in three steps: First, I develop a theoretical model to explain how overconfidence affects hirers’ valuation of candidates. Second, I test the predictions of the model in a laboratory experiment with undergraduates from an Ivy League university. Finally, I analyze the equilibrium in a competitive labor market to study how this channel affects the tendency of firms to hire workers of their own type, which I refer to as the level of sorting in the market.

The theoretical model establishes a link between hirers’ own outcomes and their valuation of candidates. A key idea of the model is that there is uncertainty regarding how informative credentials are about an agent’s underlying ability: an agent’s credentials (in this case, their school) can either be a highly informative or a noisy signal of the agent’s ability. In this setting, hirers form beliefs about the informativeness of school credentials using their personal experience, namely the school they attended, and their beliefs about their own ability. Overconfidence biases a hirer’s posterior beliefs: as a direct consequence of Bayesian updating, an overconfident hirer who attended an elite school will overestimate the probability that school credentials are informative about an agent’s underlying ability.

\footnote{There have been many studies documenting evidence of overconfidence and studying its impact on different aspects of economic activity. See Camerer and Lovallo (1999); Barber and Odean (2001); Hoelzl and Rustichini (2005); Moore and Healy (2008); Merkle and Weber (2011); Benoît and Dubra (2011).}
During the hiring process, this causes the hirer to form an inaccurate assessment about candidates’ underlying abilities based on their credentials, and as a result be willing to pay a larger-than-rational premium for candidates from an elite school.\footnote{While elitism is the main focus of this project, my model also generates predictions for non-elite firms, where the opposite effect happens. Overconfidence can cause hirers who are from non-elite universities to underestimate the informativeness of school prestige about ability, and as a result also be too inclined to hire non-elite candidates.}

As it is difficult to obtain data on hirer overconfidence in naturally occurring datasets, let alone isolate the impact of overconfidence from other factors that may cause elitism, such as regular statistical discrimination, taste-based preferences, or network effects, I run a laboratory experiment to test the predictions of the model. A laboratory experiment is particularly suitable in this setting as it allows me to exogenously vary the degree of confidence among participants and study the causal impact of overconfidence on hiring decisions.

The experimental design closely parallels the model. Subjects first complete a real-effort task consisting of trivia questions. They are then assigned to either the Top or Bottom group based on one of two possible group assignment mechanisms. In the first mechanism, subjects’ groups are \textit{highly} reflective of their underlying task performance, and in the second, groups are \textit{mildly} reflective of task performance. Given uncertainty about the group assignment mechanism, subjects express their willingness to pay to hire a Top group member and a Bottom group member, and receive a payoff based on the “hired” subject’s underlying task performance. The main treatment for the experiment varies the level of confidence among subjects by exogenously varying the difficulty of the task between sessions. This is motivated by previous literature which suggests that subjects are generally overconfident on easy tasks and underconfident on difficult tasks (Merkle and Weber 2011; Moore and Healy 2008; Barron and Gravert 2018).

My main findings from the experiment are as follows: First, I confirm that variation in the task difficulty between treatments generates a large and statistically significant difference in subjective beliefs regarding performance. In the easy-task treatment, subjects are overconfident and report a higher average belief that they are in the top half of performing subjects (0.67) relative to the hard-task treatment (0.43), where they are underconfident on average.

Second, the experimental results confirm the main testable prediction of the model: subjects in the easy-task treatment report a higher willingness to pay to hire a member of their group compared to the hard-task treatment. Since I exogenously vary the level of overconfidence between the two treatments, this effect is causal. Subjects’ reported beliefs about the group assignment mechanism are consistent with the channel suggested by the model: Top group subjects in the easy-task treatment overestimate the probability that
group assignment is highly informative of underlying performance, while Top group subjects in the hard-task treatment underestimate this probability.

The results from the baseline model and the experiment pertain to the microfoundation of hiring decisions, hirers’ willingness to pay for candidates. I conduct a labor market analysis to study how this translates into actual hiring outcomes in a labor market equilibrium. I simulate the labor market equilibrium using the experimental data to study the equilibrium outcome if the experimental subjects had participated in a competitive labor market. I find that an increase in the level of confidence among hirers leads to a greater tendency of firms to hire candidates from their own group in equilibrium. In other words, as confidence increases, Top group hirers are more likely to hire Top group candidates and less likely to hire Bottom group candidates in equilibrium, leading to an elitist bias in hiring decisions.

My findings have important implications for social welfare: given that overconfidence is a widely documented behavioral bias, the results suggest that elitist behavior in hiring may stem from this channel and constitute a form of inaccurate statistical discrimination. Since studies have shown that a student’s chance of attending an elite university is highly correlated with family income (Chetty et al. 2017), this can contribute to inefficiently high levels of income inequality and low levels of intergenerational social mobility.

The rest of the paper is organized as follows. Section 2 describes a model of how overconfidence affects hiring decisions. Section 3 describes the design of the experiment, the identification strategy and hypotheses, followed by the results of the experiment. Section 4 extends the results by studying how hiring outcomes are determined in a labor market equilibrium. Section 5 concludes.

2 Model

I present a model of how overconfidence affects hiring decisions. In the baseline version of the model, I assume that there is a single representative hirer, as this allows for a clear illustration of the intuition of the model. I later relax this assumption, allowing for there to be heterogeneity between hirers, and show that the main result carries through in expectation.

2.1 Model Setup

There are two types of agents: hirers and candidates. Each agent $i$ is born with either high or low ability $a_i \in \{H, L\}$. Let $q \equiv \Pr(a_i = H) \in [0, 1]$ denote the probability that each agent is born with high ability. For simplicity, I assume that $q = \frac{1}{2}$. Each agent has a public outcome in society, $s_i \in \{T, B\}$, corresponding to the top and bottom outcome respectively.
Agents observe their own public outcome $s_i$ but not their ability $a_i$. We may think of the public outcome as one’s level of income, the ranking of the university that one attended, one’s occupation, or any other outcome that represents an observable measure of success in society. For simplicity, let us think of $s_i$ as the ranking of the university that one attends. I will refer to $s_i$ as “school” for brevity.

There are two possible states of the world governing how an agent’s school $s_i$ depends on their underlying ability $a_i$, a low noise and a high noise state: $\theta \in \{LN, HN\}$. Depending on the state of the world, the relationship between ability $a_i$ and school $s_i$ is:

$$s_i = \begin{cases} a_i & \text{with probability } p^\theta \\ \bar{a}_i & \text{with probability } 1 - p^\theta \end{cases}$$

where $\bar{a}_i$ is the opposite of $a_i$, and $p^{LN} > p^{HN} > \frac{1}{2}$ are the probabilities that an agent’s school reflects their underlying ability in each state. In this way, regardless of the state of the world, one’s school is always an informative signal of their ability. However, the states differ in how informative school is of ability: in the low noise state, one’s school is a more informative signal of ability than in a high noise state. Let $r \equiv \Pr(\theta = LN)$ denote the beliefs about “informativeness of school”, and assume that there is a common prior on $r$ that $r = \frac{1}{2}$.

There is a representative hirer who seeks to hire candidates to work at their firm: they can hire a top candidate with $s_i = T$, a bottom candidate with $s_i = B$, one of each type, or no candidates at all. As an agent in the model, the hirer themself has attended a $T$ or $B$ school. The hirer observes the entire pool of candidates and their respective schools. Henceforth, let all variables related to the hirer be denoted by subscript $h$, and all variables related to candidates be denoted by subscript $c$.

Once hired, assume that the candidate becomes a worker and produces $f(a_c)$ for the firm, such that what the hirer cares about is a candidate’s ability, not their school. Assume that $f(H) > f(L)$. Each firm earns a profit of the following form from hiring a candidate:

$$\pi = f(a_c) - w_{sc}$$

---

3 We can motivate this by noting that in real life, agents receive signals about their ability e.g. test scores, job offers, competition outcomes, but they do not actually know their true ability level.

4 Assuming this simplifies the labor market equilibrium analysis later on. This assumption is roughly analogous to assuming that each type of candidate has diminishing marginal productivity at the firm.

5 Intuitively, one may think of the hirer as an agent who used to be a candidate and who was hired by the firm in the previous period. However, in this paper, I shall focus on the static setting and take the hirer’s school as exogenous.
Assume that the representative hirer has correct beliefs in expectation about the ability of all candidates, i.e. that $E(q_e) = \frac{1}{2}$ for all candidates, and updates beliefs according to Bayes’ rule.

### 2.2 Overconfidence

Within the literature, overconfidence can be classified into three types: *overestimation*, in which one overestimates their actual ability, *overplacement*, in which one overestimates their ability relative to others, and *overprecision*, which refers to excessive certainty regarding the accuracy of one’s beliefs (Moore and Healy, 2008). This paper studies the first type of overconfidence, overestimation, and how it may interact with personal experience to affect hiring decisions.

To define overconfidence in the context of the model, I shall introduce some notation for subjective beliefs. Let all belief variables with superscript $s$ denote subjective beliefs. For example, $q^s_h$ denotes the hirer’s subjective belief that they are high ability. I assume that $q^s_h$ is an exogenous parameter and represents the agent’s belief about their own ability prior to being admitted to a school. I then define overconfidence in one’s own ability as follows:

**Definition 1.** An agent $i$ is overconfident in their own ability if $q^s_i > q = \frac{1}{2}$.

In other words, an agent is overconfident if they overestimate their probability of being born with high ability, or, in other words, their subjective belief that they are of high ability is greater than the objective probability. In addition, let the difference $q^s_h - q$ denote an agent’s degree of overconfidence. The greater $q^s_h - q$ is above zero, the greater the agent’s overconfidence.

### 2.3 Hirer’s Beliefs About Informativeness of School

Given their prior belief $r$ about the informativeness of school and their belief $q^s_h$ about their ability, after observing their school $s_h$, the hirer can update their beliefs about the informativeness of school. Define $\hat{r}(s_h, q^s_h) \equiv \Pr(\theta = LN|s_h, q^s_h)$, the hirer’s posterior belief about the informativeness of education given their own school $s_h$ and their subjective belief about own ability $q^s_h$. The following claim provides properties of the $\hat{r}(s_h, q^s_h)$ function. All proofs are relegated to the Appendix.

**Claim 1.** The following are properties of $\hat{r}(s_h, q^s_h)$ for $q^s_h \in [0,1]$:

1. $\hat{r}(T, q^s_h) = \hat{r}(B, 1 - q^s_h)$
2. $\hat{r}(T, q^s_h)$ is increasing in $q^s_h$. 


Property 1 establishes the symmetry of the \( \hat{r} \) function around school and beliefs about own ability. Consider the case when \( q_h^s = 1 \). Then \( \hat{r}(T, 1) = \hat{r}(B, 0) \). This says that if the hirer is fully certain that they are high ability and attends a top school, they would have the same posterior belief as if they were fully certain that they are low ability and attend a bottom school.

Property 2 says that if the hirer attends a top school, their belief about the informativeness of school is increasing in their belief that they are high ability. Intuitively, if the hirer is very certain that they are of high ability and is admitted to a top school, then their school “matches” their beliefs about their own ability, and they will update their beliefs that school is highly informative of underlying ability. Note that combining Properties 1 and 2 imply that conversely, \( \hat{r}(B, q_h^s) \) is decreasing in \( q_h^s \). In other words, if the hirer instead attends a bottom school, they will update their beliefs that school is less informative of underlying ability.

Applying the definition of overconfidence to Claim 1 illustrates how being overconfident affects the hirer’s beliefs regarding the informativeness of school:

**Corollary 1.** If the hirer is overconfident, then attending a top school will cause them to overestimate the informativeness of school, while attending a bottom school will cause them to underestimate the informativeness of school. In addition, the degree to which the hirer misjudges the informativeness of school is increasing in their level of overconfidence.

In this way, biased beliefs regarding their own ability can distort how a hirer uses their own personal experience to update beliefs regarding the informativeness of school.

### 2.4 Valuation of Candidates

I now study how this in turn affects the hirer’s valuation of candidates. Recall that the hirer only observes candidates’ schools \( s_c \) but not their ability \( a_c \). Let me define \( WTP_{s_c}^{sh}(q_h^s) \) to be the maximum willingness-to-pay of a hirer with school \( s_h \) for a candidate with school \( s_c \), given that the hirer has subjective beliefs \( q_h^s \) about their own ability. Since the firm’s profit from hiring one candidate is \( \pi_h = f(a_c) - w_{sc} \), \( WTP_{s_c}^{sh}(q_h^s) \) is equal to the expected productivity of the candidate \( c \) given the candidate’s school \( s_c \) and the hirer’s school \( s_h \), which I will denote by \( E[f(a_c)|s_c, s_h, q_h^s] \).

In addition, let \( \Delta WTP^{sh}(q_h^s) \) denote the hirer’s willingness-to-pay premium for top school candidates over bottom school candidates, given the hirer’s school \( s_h \), such that \( \Delta WTP^{sh}(q_h^s) \equiv E[f(a_c)|T, s_h, q_h^s] - E[f(a_c)|B, s_h, q_h^s] \). For brevity, I shall refer to \( \Delta WTP^{sh}(q_h^s) \) as the willingness-to-pay premium.

---

6There is a slight abuse of notation here since \( q_h^s \) is a parameter taken as exogenous by the hirer.
Proposition 1. The hirer’s willingness to pay for candidates has the following properties:

a. The hirer is willing to pay a higher premium for top school candidates if they attended a top school compared to if they attended a bottom school if and only if they are overconfident:
$$\Delta WTP^T(q^*_h) > \Delta WTP^B(q^*_h) \iff q^*_h > q = \frac{1}{2}.$$  

b. The hirer’s willingness-to-pay premium is independent of the hirer’s own school if and only if the hirer has objective beliefs about their own ability:
$$\Delta WTP^T(q^*_h) = \Delta WTP^B(q^*_h) \iff q^*_h = q = \frac{1}{2}.$$  

The proof in Appendix 6.2 demonstrates that $$\Delta WTP^T(q^*_h) > \Delta WTP^B(q^*_h)$$ can be expressed as a product of 3 terms:

$$\left[ f(H) - f(L) \right] \left[ \hat{r}(T, q^*_h) - \hat{r}(B, q^*_h) \right] \left[ p^{LN} - p^{HN} \right] > 0$$  \hspace{1cm} (1)

The first term in the product represents the marginal increase in productivity from hiring a high ability rather than low ability worker. The second term, which represents the marginal increase in the hirer’s posterior belief about the informativeness of school if the hirer himself attended a top school, is positive if and only if $$q^*_h > \frac{1}{2}$$ by Claim 1. The third term, $$p^{LN} - p^{HN}$$, which represents the increase in informativeness of the low noise versus the high noise state of the world, is positive by assumption. Thus, the inequality holds if and only if $$q^*_h > \frac{1}{2}$$.

In this simple setting, I am able to show how the hirer’s willingness-to-pay for candidates from different schools depends on their own personal experience (their school) and their beliefs about their own ability. The hirer’s belief about their own ability affects how they update their beliefs about the informativeness of the school signal after observing their own school. Based on how informative they believe school to be, the hirer then forms valuations of candidates from different schools.

If the hirer is overconfident, then upon being admitted to a top school, they update their beliefs that school is more informative of ability. As a consequence, they are willing to pay a larger premium for candidates from a top school, thereby leading to elitism. Conversely, if they are admitted to a bottom school, they update their beliefs that school is less informative of ability, and are willing to pay a smaller premium for top candidates. This is despite the fact that if the hirer has objective prior beliefs about their own ability $$a_h$$, i.e. $$q^*_h = \frac{1}{2}$$, then the premium that the hirer is willing to pay for top candidates is independent of the hirer’s own school.

Let $$\Delta WTP^R = \Delta WTP^T(\frac{1}{2}) = \Delta WTP^B(\frac{1}{2})$$ denote the rational willingness to pay premium for a top candidate.

Corollary 2. The following is true if and only if the hirer is overconfident:
a. If the hirer attended a top school, he is willing to pay a larger premium for a candidate from a top school: \( q_{h}^{*} > q = \frac{1}{2} \Leftrightarrow \Delta WTP^{T}(q_{h}^{*}) > \Delta WTP^{R} \).

b. If the hirer attended a bottom school, he will be willing to pay a smaller-than-rational premium for a top candidate: \( q_{h}^{*} > q = \frac{1}{2} \Leftrightarrow \Delta WTP^{B}(q_{h}^{*}) < \Delta WTP^{R} \).

Another way to interpret this result is that hirers who are overconfident about their ability are willing to pay a greater amount for a candidate of their own type compared to if they had rational beliefs. In this way, overconfidence in one’s ability can serve as a microfoundation for homophily, causing hirers to demonstrate a preference for others similar to themselves. In addition, the greater the level of overconfidence, the more hirers are willing to overpay for a candidate of their own type. This preference may arise solely due to mistaken beliefs regarding own ability, even if the hirer updates beliefs in a Bayesian manner.

2.5 Generalizing the Model: Heterogeneous Agents

In this section, I relax the assumption that there is a single representative hirer with probability of being high ability \( q = \frac{1}{2} \). In reality, agents may receive private signals of their own ability that are unobservable by other agents. In this case, agents’ objective probability of success incorporating their private signals may not all be a half. To account for this, I assume that there is a continuum of agents and allow for \( q \) to vary across agents, such that each agent may have an individual-specific objective probability of being high ability, \( q_{i} \). \( q_{i} \) is distributed with CDF \( F(q_{i}) \) and PDF \( f(q_{i}) \). However, at the population level, half of the population is high ability and half of the population is low ability, such that \( E(q_{i}) = \frac{1}{2} \), consistent with the representative hirer case. In addition, let \( F^{*}(q_{i}^{*}) \) and \( f^{*}(q_{i}^{*}) \) denote the CDF and PDF of agents’ subjective beliefs that they are born with high ability.

Accordingly, I modify the definition of overconfidence to correspond to average beliefs at the population level rather than at an individual level.

**Definition 2.** There is overconfidence at the population level if \( E(q_{i}^{*}) > \frac{1}{2} \).

This definition says that there is overconfidence among agents at the population level if the expectation of agents’ subjective beliefs that they are of high ability is greater than the population proportion of a half. Intuitively, since at the population level, only half of the agents can be of high ability, if in expectation, agents’ subjective beliefs that they are of high ability is greater than half, some agents must be possess overconfident beliefs. Analogous to the representative hirer case, I define \( E(q_{i}^{*} - \frac{1}{2}) \) to be the degree of overconfidence in the population. The greater it is above zero, the greater the level of overconfidence.
Definition 3. $F^s(q^s)$ confidence dominates $F^{s'}(q^s)$ if $F^s(q^s)$ first order stochastically dominates $F^{s'}(q^s)$.

Intuitively, if $F^s(q^s)$ confidence dominates $F^{s'}(q^s)$, then the population of agents is more confident if they have distribution of subjective beliefs $F^s(q^s)$ compared to if they had distribution of subjective beliefs $F^{s'}(q^s)$. Note that if $F^s(q^s)$ first order stochastically dominates the distribution of objective probabilities $F(q)$, then there is necessarily overconfidence at a population level.

The following result is the analog of Proposition 1 in the heterogeneous agents version of the model:

Proposition 2. If $F^s(q^s)$ first order stochastically dominates $F(q)$, then:

$$E_{F^s} \left[ \Delta WTP^T(q^s_h) - \Delta WTP^B(q^s_h) \right] > E_{F} \left[ \Delta WTP^T(q_h) - \Delta WTP^B(q_h) \right]$$

(2)

The difference $E \left[ \Delta WTP^T(q^s_h) - \Delta WTP^B(q^s_h) \right]$ represents the average amount each group is willing to overpay for their own type relative to the opposite type of hirer. If the distribution of agents’ subjective beliefs dominates the distribution of objective probabilities, then on average, agents are willing to pay too much for candidates from their own school and too little for candidates from the other school relative to these candidates’ expected productivity.

3 Experiment

To examine whether overconfidence may contribute to the persistence of elitism through the channel suggested by the model, I test the main prediction of the model, Proposition 2, by running a laboratory experiment. In naturally occurring datasets, it is difficult to isolate the impact of overconfidence from other factors that may cause elitism. Observing elite firms in the real world demonstrating a higher willingness to pay for elite candidates may reflect elite candidates being of higher ability, elite firms having better information about elite candidates due to a similar background, or simply taste-based preferences. A laboratory experiment is particularly suitable in this setting, as it allows me to exogenously vary the degree of confidence among participants and thereby study the causal impact of overconfidence on hiring decisions.

I design the experiment to closely parallel the more general version of the model with heterogeneous agents. In order to generate subjects’ underlying “ability” as in the model, subjects undergo a real-effort task. Subjects may differ in their innate ability to perform
the task, and this innate ability, which corresponds to an agent’s probability of being high ability in the model, is unobservable.

3.1 Experimental Design

Subjects begin the experiment by doing a trivia task. The purpose of this task is to assign a level of performance to subjects: half of the subjects are classified as “High performance” and half of the subjects are classified as “Low performance”, corresponding to high and low ability in the model. The task consists of 25 multiple-choice general-knowledge questions, and subjects have 6 minutes to complete the task. Sample questions from the trivia task can be found in Appendix Section 6.3.1 Subjects in each session faced one of two versions of the task: an easy version and a hard version, representing the two treatments of the experiment, more details of which will follow.

After completing the task, subjects are ranked according to the number of questions they answered correctly and assigned a performance level depending on their relative performance in that session: the top half performing subjects in that session are High performance subjects, while the bottom half performing subjects are Low performance subjects. If there is a tie at the cutoff between the top and the bottom half, then the tie is broken randomly. Subjects are told all of the above the information, but not their performance level.

I then elicit subjects’ beliefs about the probability that they are a High performance subject, or, in other words, among the top half performing subjects in that session. I denote this by \( q_i^s \). The belief elicitation mechanism is as in Grether (1992) which incentivizes truthful reporting of subjective beliefs independent of risk preferences.\(^7\)

Each subject is then assigned to a group: Top or Bottom. There are two group assignment mechanisms: A and B. In Mechanism A, a subject’s group assignment reflects their underlying performance with 95% chance, and reflects the opposite of performance with 5% chance. In other words, if a subject is a High performance subject, then there is a 95% chance that they are assigned to the Top group, and a 5% chance that they are assigned to the Bottom group. In Mechanism B, group assignment reflects underlying performance with 55% chance, and the opposite of performance with 45% chance. Hence, a subject’s group is an informative signal of performance in both mechanisms, but group assignment is more informative about performance in Mechanism A than in Mechanism B.

\(^7\)The belief incentivization mechanism works as follows: Each subject is asked to report their subjective belief \( p^s \in [0, 100] \) (in a percentage form) about their probability of earning \( H \) in the task. A random number \( n \in [0, 100] \) is drawn, and if \( n < p^s \), then the subject is paid 4 bonus tokens if they earned \( H \) in the task and 0 otherwise. If \( n \geq p^s \), then the subject faces a gamble that pays 4 tokens with \( n\% \) chance and 0 otherwise. See also: Holt (2007); Karni (2009), pages 384-385 of the former.
Subjects know that there is an equal chance that Mechanism A or B may be used to assign groups, but they do not know which mechanism is randomly selected by the computer. They do, however, observe which group they are assigned to. This parallels the feature of the model where hirers do not know which state of the world has occurred nor their own ability, but observe their own school. To ensure subjects understand how groups were assigned, they are asked to complete several comprehension questions. The comprehension questions do not affect their payment, but they must correctly answer all questions to continue.

Subjects then face a set of hiring decisions: one for a member of the Top group and one for a member of the Bottom group. They are told that if they “hire” an individual, what matters for their payoff is the individual’s underlying performance: a High performance individual will yield them a payoff of 200 tokens, and a Low performance individual will yield them a payoff of 40 tokens. However, they only observe the individual’s group. The set of hiring decisions entail expressing their willingness to pay for a randomly drawn individual from the Top group and a randomly drawn member of the Bottom group in that session. Let \( WTP_i^w(g, c) \) denote subject \( i \)'s willingness-to-pay for a group \( w \) worker, where subject \( i \) belongs to group \( g \) and treatment \( c \). The willingness to pay decisions are incentivized using the Becker-Degroot-Marschak mechanism [Becker et al., 1964]. Before proceeding to the actual decisions, subjects are once again asked several comprehension questions.

After the hiring decisions were complete, I elicited subjects’ beliefs about the probability that Mechanism A was used to assigned groups, and denote this belief by \( r_i^a \). At the end of the experiment, one of the following was decisions was randomly chosen to determine subjects’ payment: their performance in the trivia task, one of the two belief elicitation questions, or one of the “hiring” decisions. The experiment was then completed with a brief questionnaire, which elicited responses such as subjects’ gender and perception of task difficulty. The instructions for the experiment and other experimental materials can be found in Appendix Section 6.3.

### 3.1.1 Treatments

My main treatment variable of interest is relative overconfidence at a population level. Since subjects may vary in their innate ability in the task, each subject may have a different

---

8In particular, the incentivization mechanism is as follows: Each subject is asked to report their willingness to pay \( WTP_T \in [0, 200] \) for a worker from the Top group. A random integer number \( n \in [0, 200] \) is drawn, and if \( n \leq WTP_T \), then a randomly drawn member of the Top group is chosen. The subject’s payoff is \( f(T) - n \), where \( f(T) \) is 200 if the randomly drawn member is a High performance subject and \( f(T) \) is 40 if the randomly drawn member is a Low performance subject. If \( n > WTP_T \), then the subject’s payoff is 0. A similar process is used to elicit subjects’ willingness to pay for a Bottom group member. To ensure that subjects do not end up receiving a negative payoff from each of the hiring decisions, they are given an endowment of 160 tokens for each decision.
objective probability of success. However, since each subject’s objective probability of success is unobservable, I use the sample analog of overconfidence at a population level as in the general version of the model as it is a more robust measure of overconfidence.

I define the confidence level in the sample to be the average reported beliefs of subjects that they belong to the top half, which I denote by \( \bar{q}_s^i \). To test whether or not there is overconfidence, I test whether the data is rationalizable at the population level or not ([Benoît and Dubra](#)). I classify the data to be not rationalizable and there to be overconfidence at the population level if the average \( q_s^i \) is significantly greater than \( \frac{1}{2} \). Intuitively, since only half of the subjects can be in the top half, if the average subjective belief of being in the top half is greater than \( \frac{1}{2} \), then the population has overconfident beliefs. Furthermore, the greater the difference \( E(\bar{q}_s^i) - \frac{1}{2} \), the greater the level of overconfidence in the population. Otherwise, if the average \( q_s^i \) is not significantly different from \( \frac{1}{2} \), the data are rationalizable and there is no overconfidence. Finally, if the average \( q_s^i \) is significantly less than \( \frac{1}{2} \), I classify the session to be underconfident at the population level.

My main treatment for the experiment is to exogenously vary the level of confidence by varying the difficulty of the task between sessions: in the high-confidence treatment, subjects do an easy version of the task, while in the low-confidence treatment, subjects do a difficult version of the task. This is motivated by previous literature that suggests that individuals are generally overconfident on easy tasks and underconfident on difficult tasks ([Moore and Healy](#), [Merkle and Weber](#), [Barron and Gravert](#)).

I check that two conditions are satisfied to confirm that the treatment works: first, there is overconfidence in the high confidence treatment: \( E(q_s^i|HighC) > \frac{1}{2} \). Second, the distribution of subjective beliefs in the high-confidence treatment first-order stochastically dominates that of the low-confidence treatment: \( F_{HighC}(q_s^i) < F_{LowC}(q_s^i) \forall q_s^i \).

### 3.2 Hypothesis

For each subject \( i \), we have the following outcome variables: \( WTP_T^i(g,c) \) and \( WTP_B^i(g,c) \), which represent subject \( i \)’s willingness to pay for a \( T \) group member and \( B \) group member respectively, where subject \( i \) belongs to group \( g \in \{T,B\} \) and is in treatment \( c \in \{HighC,LowC\} \).

Analogous to the model, let me define \( \Delta WTP^i(g,c) \equiv WTP_T^i(g,c) - WTP_B^i(g,c) \) to be subject \( i \)’s willingness-to-pay premium for a Top group member over a Bottom group member. My main hypothesis is the analog of Proposition 2, the main testable implication of the model:

**Hypothesis 1.** The average gap in WTP premium between Top and Bottom groups is greater
in the easy-task (high-confidence) treatment than the hard-task (low-confidence) treatment.

\[ E[\Delta WTP(T, H)] - E[\Delta WTP(B, H)] > E[\Delta WTP(T, L)] - E[\Delta WTP(B, L)] \]

Intuitively, this says that in the easy-task (high-confidence) treatment, on average, hirers are willing to pay more for a worker of their own type compared to in the hard-task (low-confidence) treatment.

The hypothesis can be tested by running the following difference-in-difference regression:

\[ \Delta WTP_i = \alpha + \beta_1 g + \beta_2 c + \beta_3 g c + \varepsilon_i \]

For Hypothesis 1, we test whether the interaction term between the high-confidence and Top group dummies \( \beta_3 > 0 \).

### 3.3 Results

I conducted experimental sessions in April and May of 2019 at the Brown University Social Science Experimental Laboratory (BUSSEL), using undergraduate students as subjects. I ran 14 sessions of the experiment: 7 sessions of the easy-task (high-confidence) treatment, and 7 sessions of the hard-task (low-confidence) treatment. There was a total of 297 subjects. The experiment lasted approximately 50 minutes. The average payment was $22.86, including a $5 show up fee. The experiment was programmed using z-Tree (Fischbacher, 2007). The exchange rate from the experimental currency, tokens, to US dollars was $0.09 per token.

#### 3.3.1 Descriptive Statistics

I first present subjects’ performance in the trivia task by treatment. Figure 1 presents the histograms of number of correct answers for the easy-task (high-confidence) and hard-task (low-confidence) treatment. The histogram of the high-confidence treatment lies to the right of that of the low-confidence treatment.

Table 1 provides the average number of questions that subjects answered correctly in the two treatments. In the easy-task (high-confidence) treatment, subjects scored an average of 20.46 questions out of a total of 25 questions correct, while in the hard-task (low-confidence) treatment, subjects scored an average of 7.84 questions correct. As can be seen from both the histogram and the table, on average, subjects scored significantly fewer questions correct in the low-confidence treatment than in the high-confidence treatment, which confirms that the version of the task in the high-confidence treatment is more difficult than that in the low-confidence treatment.
Figure 1: Trivia Score by Treatment

![Bar chart showing trivia scores by treatment]

Table 1: Number of Correct Answers by Treatment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Average</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Confidence (n = 148)</td>
<td>7.84</td>
<td>2.33</td>
</tr>
<tr>
<td>High-Confidence (n = 149)</td>
<td>20.46</td>
<td>2.40</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td><strong>12.62</strong></td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>&lt; 0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Number of questions answered correctly in the trivia task out of 25 questions.
To study how a difference in the difficulty of the task affects subjects’ level of confidence in their task performance, I next present subjects’ beliefs about their performance in the trivia task. Figure 2 illustrates the histogram of subjects’ reported beliefs of being in the top half of task performance by treatment. As can be seen in the figure, the subjective beliefs of being in the top half are skewed to the left in the high-confidence treatment, with most of the mass of being concentrated in the top half. On the other hand, in the low-confidence treatment, the subjective beliefs are skewed to the right, with more mass being concentrated in the bottom half.

Figure 2: Subjective Belief of Being in the Top Half by Treatment

Figure 3 presents the CDF of subjective beliefs $q_i^s$ by treatment. Consistent with what we would expect, the CDF for the high-confidence treatment first order stochastically dominates the CDF for the low-confidence treatment.

Table 2 presents the confidence levels of the two treatments, and the results confirm what we observe in the histograms. In the high-confidence treatment, the confidence level is 0.67, while in the low-confidence treatment, it is 0.43. This difference is significant at the 1% level. In addition, the confidence level in the high-confidence treatment is significantly different from the population proportion of $\frac{1}{2}$ with a p-value < 0.0001.

Taken together, this confirms that a difference in the task difficulty affects the overall level of confidence among subjects: subjects who faced an easy version of the task generally demonstrate a higher level of confidence, while those who faced the difficult version of the task generally demonstrate a lower level of confidence.

Hence, the two criteria for the treatment are satisfied. First, the CDF of subjective beliefs in the easy-task (high-confidence) treatment first order stochastically dominates the
Figure 3: CDF of $q^*$ by Treatment

![CDF of $q^*$ by Treatment](image)

Table 2: Average $q^*_i$ by Treatment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Average</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Confidence ($n = 148$)</td>
<td>0.43</td>
<td>0.23</td>
</tr>
<tr>
<td>High-Confidence ($n = 149$)</td>
<td>0.67</td>
<td>0.20</td>
</tr>
<tr>
<td>Difference</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>&lt; 0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Average belief of subjects that they scored High performance in the task.
CDF in the hard-task (low-confidence) treatment. Second, in the easy-task (high-confidence) treatment, subjects display overconfidence at the population level: $E(q^*_t|HighC) > \frac{1}{2}$.

### 3.3.2 Main Analysis

Table 3 depicts the results from the main difference-in-difference estimation. The first column represents the basic regression without clustering at the session level nor adding controls. The second column clusters standard errors at the session level, while the remaining two columns include controls.

The coefficient on the interaction term between the dummies for high-confidence treatment and Top group is positive and is statistically significant at the 5% level in all specifications. This confirms the main testable implication of the model: on average, as confidence increases, hirers have a higher willingness to pay for workers of their own type (relative to hirers from the other group). We can interpret the magnitude of the coefficient as follows: in the high-confidence treatment, there is a 25.79 token greater relative willingness to pay for a worker of their own group.

Table 3: Impact of Treatment on WTP Premium

<table>
<thead>
<tr>
<th></th>
<th>WTP Premium</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>(7.150)</td>
<td>(10.24)</td>
<td>(10.36)</td>
<td>(13.35)</td>
<td></td>
</tr>
<tr>
<td>Top Group</td>
<td>-10.39</td>
<td>-10.39</td>
<td>-11.80</td>
<td>-11.74</td>
</tr>
<tr>
<td>(7.657)</td>
<td>(10.04)</td>
<td>(10.13)</td>
<td>(9.924)</td>
<td></td>
</tr>
<tr>
<td>High Confidence*Top Group</td>
<td>25.79**</td>
<td>25.79**</td>
<td>26.76**</td>
<td>28.07**</td>
</tr>
<tr>
<td>(10.73)</td>
<td>(11.69)</td>
<td>(11.70)</td>
<td>(11.66)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>70.80***</td>
<td>70.80***</td>
<td>80.70***</td>
<td>70.40***</td>
</tr>
<tr>
<td>(4.916)</td>
<td>(7.961)</td>
<td>(12.49)</td>
<td>(19.67)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>297</td>
<td>297</td>
<td>297</td>
<td>290</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.021</td>
<td>0.021</td>
<td>0.027</td>
<td>0.034</td>
</tr>
<tr>
<td>Cluster at Session Level</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Gender Control</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty of Task Control</td>
<td>YES</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

7 observations dropped in Column (4) due to non-responses in the Difficulty of Task question.

Since the difference in confidence levels $\bar{q}_s^e$ between the high and low-confidence treatment is 24 percentage points (p.p.), a back-of-the-envelope calculation suggests that going from being fully unsure whether they had high or low performance ($q^*_i = 0.5$) to being fully
sure that they are a high performance subject \( q_i^* = 1 \) results in a 53.73 token increase in a hirer’s relative willingness to pay for a worker of their group. Alternatively, since the standard deviation of the WTP premium variable is 46.10, moving from the high-confidence treatment to the low-confidence treatment increases the average WTP premium by 0.56 standard deviations.

Table 4 presents the average WTP premium by treatment and group. The average WTP premiums are larger in the diagonals than the off-diagonals, as is consistent with the hypothesis.

Table 4: Average WTP by Treatment and Group

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Bottom</th>
<th>Top</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>70.80</td>
<td>60.41</td>
<td>66.52</td>
</tr>
<tr>
<td>High</td>
<td>61.60</td>
<td>77.00</td>
<td>68.94</td>
</tr>
<tr>
<td>Total</td>
<td>66.45</td>
<td>69.33</td>
<td>67.73</td>
</tr>
</tbody>
</table>

Mean of WTP Premium for different treatments and groups.

3.3.3 Session-Level Analysis

I also study the data by conducting a session-level analysis. To do this, I regress \( \Delta WTP_i \) on a dummy for the Top group separately for each session. In other words, I run one regression for each session, rather than on the full sample. This allows me to obtain a separate coefficient for the interaction term for each session.

Figure 4 plots the estimated coefficients by high or low-confidence treatment. For strong evidence of the hypothesis, we would expect to see the points for the high-confidence treatment to the right of those for the low-confidence treatment. This generally appears to be the case, though there is an outlier on the right and some overlap in the coefficients close to zero.

Table 5 below presents the output regressing the session level coefficients on a high-confidence dummy. The estimated coefficient is very similar to that of the regression on the whole sample, which suggests that analysis at the session level is consistent with the analysis on the full sample.

---

9This assumes that the effect is linear.
Table 5: Impact of Confidence on WTP Premium (Session Level)

<table>
<thead>
<tr>
<th></th>
<th>Estimated WTP Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Confidence</td>
<td>24.45*</td>
</tr>
<tr>
<td></td>
<td>(12.47)</td>
</tr>
<tr>
<td>Constant</td>
<td>-8.419</td>
</tr>
<tr>
<td></td>
<td>(8.815)</td>
</tr>
<tr>
<td>Observations</td>
<td>14</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.243</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1
3.3.4 Checking Posterior Beliefs

As a further test that overconfidence affects subjects’ willingness to pay through the channel suggested by the model, I study subjects’ reported beliefs about the informativeness of group, in other words, subjects’ reported beliefs $r_i^s$ that Mechanism A was used to assign groups. To confirm that reported beliefs are in alignment with the hypothesis, I run the main difference-in-difference regression, replacing the dependent variable with $r_i^s$, the results of which are presented in Table 6.

Table 6: Impact of Confidence on $r_i^s$

<table>
<thead>
<tr>
<th>Subjective Belief that Groups Assigned by Mechanism A</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Confidence</td>
<td>-0.117***</td>
<td>-0.117**</td>
<td>-0.119**</td>
<td>-0.130**</td>
</tr>
<tr>
<td></td>
<td>(0.0337)</td>
<td>(0.0438)</td>
<td>(0.0415)</td>
<td>(0.0511)</td>
</tr>
<tr>
<td>Top Group</td>
<td>-0.117***</td>
<td>-0.117***</td>
<td>-0.103***</td>
<td>-0.110***</td>
</tr>
<tr>
<td></td>
<td>(0.0361)</td>
<td>(0.0346)</td>
<td>(0.0325)</td>
<td>(0.0358)</td>
</tr>
<tr>
<td>High Confidence*Top Group</td>
<td>0.313***</td>
<td>0.313***</td>
<td>0.303***</td>
<td>0.305***</td>
</tr>
<tr>
<td></td>
<td>(0.0506)</td>
<td>(0.0511)</td>
<td>(0.0531)</td>
<td>(0.0562)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.615***</td>
<td>0.615***</td>
<td>0.511***</td>
<td>0.530***</td>
</tr>
<tr>
<td></td>
<td>(0.0232)</td>
<td>(0.0339)</td>
<td>(0.0476)</td>
<td>(0.0577)</td>
</tr>
<tr>
<td>Observations</td>
<td>297</td>
<td>297</td>
<td>297</td>
<td>290</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.126</td>
<td>0.126</td>
<td>0.153</td>
<td>0.153</td>
</tr>
<tr>
<td>Cluster at Session Level</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Gender Control</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty of Task Control</td>
<td></td>
<td></td>
<td></td>
<td>YES</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Again, our coefficient of interest is the interaction term between high-confidence and Top Group. In all specifications of the model, the coefficient is positive and statistically significant at the 1% level, which is consistent with the implications of the model.

Overall, there is strong evidence supporting the main testable implication of the model, that overconfidence causes hirers to have a greater relative willingness to pay for a candidate in their own group. This is because overconfidence distorts how subjects update their beliefs regarding how informative group membership is about underlying performance.

4 Labor Market Analysis

In the baseline model and the experiment, I have shown that hirer overconfidence affects the microfoundation underlying hiring decisions: their willingness to pay for candidates. I now
study how this effect on willingness to pay translates to actual hiring outcomes by extending the analysis to include a labor market equilibrium. In particular, I am interested in how hirer overconfidence affects the tendency for firms to hire candidates of their own type in equilibrium, which I will refer to as the *level of sorting* in the market.

For simplicity, I consider the case that the labor market is perfectly competitive, and there are two markets: one for top candidates and one for bottom candidates. Each type of firm can participate in both markets and hire at most one top and one bottom candidate. I define a labor market equilibrium as follows:

**Definition 4.** Let \( D_{s_c}(w_{s_c}) \) and \( S_{s_c}(w_{s_c}) \) represent the demand and supply functions in the market for candidates with school \( s_c \), and let \( w_{s_c} \) denote the wage for a candidate with school \( s_c \). The labor market equilibrium for the markets for Top and Bottom candidates is characterized by a set of wages \((w^*_T, w^*_B)\) such that:

1. Supply and demand are equal in each market: \( D_T(w^*_T) = S_T(w^*_T) \) and \( D_B(w^*_B) = S_B(w^*_B) \).
2. Firms and candidates are price takers: at any given wage \( w_T \) and \( w_B \), they decide if they are willing to participate in the market or not.

In this section, I continue with the generalized version of the model, where hirers may vary in \( q_h \), their objective probability of being born with high ability, and \( q_h \) is distributed with CDF \( F(q_h) \) and PDF \( f(q_h) \).

### 4.1 Supply and Demand Functions

Before analyzing the labor market equilibrium, I first derive the supply and demand functions in each market.

#### 4.1.1 Supply Function

For simplicity, I assume that candidates do not have any outside option and will accept any wage \( w_{s_c} \geq 0 \). In this way, Top and Bottom candidates supply labor inelastically.

#### 4.1.2 Demand Function

Denote \( E(\pi_{s_c}) \) to be the expected profit of a firm who hires a candidate with school \( s_c \). A firm will demand a candidate if \( WTP_{s_c}(q_h^*) \geq w_{s_c} \), i.e. if their willingness to pay for a

\[10\]I will use the terms “hirer” and “firm” interchangeably.

\[11\]If I assume that a firm can only hire one candidate, there will be various dependencies between the two markets that might cause a firm to switch between workers they prefer depending on the model’s parameters. In order to avoid this, I assume that a firm can hire more than one candidate. Assuming that they can hire one \( T \) and one \( B \) candidate is similar in spirit to assuming diminishing marginal productivity.
candidate with school $s_c$, conditional on their subjective beliefs about own ability $q^*_h$ and their own school $s_h$, exceeds the wage for that candidate.

To derive the demand functions, I first derive properties of the willingness to pay function $WTP_{s_h}^{s_c}(q^*_h)$. Note that hirers may vary in three different dimensions: first, their objective probability of being born with high ability $q$, second, their subjective belief of being born with high ability $q^*_s$, and third, their own school $s_h$.

**Claim 2.** The willingness to pay function $WTP_{s_h}^{s_c}(q^*_h)$ has the following properties:

1. For all $q^*_h \in [0, 1]$ and $s_h \in \{T, B\}$, $WTP_{s_c}^{s_h}(q^*_h) > WTP_{s_B}^{s_h}(q^*_h)$
2. $WTP_{s_c}^{T}(q^*_h) = WTP_{s_c}^{B}(1 - q^*_h) \forall q^*_h \in [0, 1]$
3. $WTP_{s_c}^{T}(q) = WTP_{s_B}^{B}(q)$ is increasing in $q \in [0, 1]$.

Property 1 says that regardless of how hirers vary in the above three dimensions, all hirers always value a Top candidate over a Bottom candidate. Intuitively, this is because independent of whether the state of the world is low noise or high noise, the probability that school reflects underlying ability, $p^\theta$, is always greater than a half. As such, school is always an informative signal of ability, and as a consequence, a candidate from a top school always has a higher expected productivity than a candidate from a bottom school.

Property 2 says that in the market for Top candidates, a Top hirer with subjective belief $q$ has the same willingness to pay as a Bottom hirer with subjective belief $1 - q$. Hence there is some form of symmetry in willingness to pay by hirers from opposite schools. This is because these two hirers will have the same posterior belief regarding the informativeness of school $\hat{r}(T, q^*_h) = \hat{r}(B, 1 - q^*_h)$ by Claim 1. Since a hirer’s willingness to pay for a candidate is a 1:1 mapping to their belief about the informativeness of school, they have the same willingness to pay.

The third property of the willingness to pay function is that the greater is one’s belief in their own ability, the greater they are willing to pay for a candidate of their own type, and, combining the second and third property, the less they are willing to pay for the candidate of the opposite type.

The following corollary gives us the impact of being overconfident on the valuation of individual candidates.

**Corollary 3.** Overconfident hirers ($q^*_h > q_h$) overvalue candidates of their own type and undervalue candidates of the opposite type.

Figure 5 depicts how the willingness to pay of different firms for different candidates depends on the type of the firm, the type of the candidate, and the hirers’ beliefs. The solid
Figure 5: Willingness to Pay for Candidates by Firm Type

The two curves on the top illustrate the willingness to pay for a top candidate. As a firm’s level of confidence $q_{bh}$ increases, if they are a top firm, their willingness to pay for a top candidate $WTP_T^T(q_{bh})$ increases. On the other hand, if they are a bottom firm, their willingness to pay for a top candidate $WTP_T^B(q_{bh})$ decreases in $q_{bh}$.

The two curves at the bottom illustrates the willingness to pay for a bottom candidate. In this case, the firm’s willingness to pay is increasing in $q_{bh}$ if the firm is a bottom firm and is decreasing in $q_{bh}$ if the firm is a top firm. Note that since a top candidate is always valued more than a bottom candidate (as shown in Claim 3), the willingness to pay curve for a Top candidate always lies strictly above the willingness to pay curve for a Bottom candidate.

Using the properties of the willingness to pay function above, I derive the demand functions for Top and Bottom candidate. Details of the derivation are provided in the Appendix. The demand function for Top candidates is given by:

$$D_T = D_T^T + D_T^B$$

$$= (1 - p^\theta) \left( 1 - F^s (\hat{q}) \right) + \left( 2p^\theta - 1 \right) \left[ \frac{1}{2} - E(\hat{q} | q_{bh}^s \leq \hat{q}) F^s (\hat{q}) \right]$$

Demand by Top Hirers

$$+ p^\theta F^s (1 - \hat{q}) + (1 - 2p^\theta) E(\hat{q} | q_{bh}^s \leq 1 - \hat{q}) F^s (1 - \hat{q})$$

Demand by Bottom Hirers
where \( \hat{q} \equiv WTP_T^{-1}(w_T) \). The total demand for top candidates at a given wage is the mass of all hirers who have a maximum willingness to pay at least equal to that wage. As previously shown, hirers’ maximum willingness to pay for a top candidate is monotonically increasing in how informative they believe school to be about underlying ability. For hirers who attended top schools, this is increasing in their confidence regarding their ability. This corresponds to the first term in the demand function, representing the demand by top hirers. In particular, all top hirers who believe there is at least a \( \hat{q} \) probability of them being high ability will believe school to be a sufficiently informative signal of ability that they are willing to pay wage \( w_T \) for a top worker.

Conversely, among hirers who attended bottom schools, those who are less confident about their own ability will believe that school is a more informative signal of underlying ability, and therefore have a higher willingness to pay for a top worker. This corresponds to the second term in the demand function. All bottom hirers who believe there is less than a \( 1 - \hat{q} \) probability of them being high ability will believe school to be a sufficiently informative signal of ability that they are willing to pay wage \( w_T \) for a top worker.

Analogously, total demand for Bottom candidates is:

\[
D_B = D_B^T + D_B^B = p^\theta \left[ 1 - F^s(\hat{q}) \right] + \left( 1 - 2p^\theta \right) \left[ E(q_h) - E(q_h|q_h^s \leq \hat{q}) F^s(\hat{q}) \right] + \left( 1 - p^\theta \right) F^s(1 - \hat{q}) + \left( 2p^\theta - 1 \right) E(q_h|q_h^s \leq 1 - \hat{q}) F^s(1 - \hat{q})
\]

where \( \tilde{q} \equiv WTP_B^{-1}(w_B) \).

### 4.2 Labor Market Equilibrium

I assume that there is a unit mass of firms and a unit mass of candidates. By the assumption that the probability that a random hirer is high ability is \( \frac{1}{2} \) and the law of iterated expectations, half of the firms are \( T \) firms and half are \( B \) firms. Analogously, half of the candidates are \( T \) candidates and half are \( B \) candidates.

#### 4.2.1 Rational Benchmark

I begin with the case that all firms in the market are rational, i.e. \( F^s(q_h) = F(q_h) \) for all \( h \), and as a consequence \( E(q_h^s) = \frac{1}{2} \). If all firms are rational with \( F^s(q_h^s) = F(q_h) \), in the labor market equilibrium, all firms will express their rational willingness to pay for a Top and
Bottom candidate conditional on the hirer’s objective belief of being high ability $q_h$ and their own school $s_h$. In this case, the equilibrium wages for both $T$ and $B$ candidates are equal to the marginal firm’s willingness to pay for each worker: $(w^*_T, w^*_B) = (WTP^R_T(q), WTP^R_B(q))$.

To study whether firms are more likely to hire candidates of their own type in equilibrium, let me define the degree of sorting, $d \in [0, 1]$, to be:

$$d = \frac{(\text{Mass of Top hirers hiring Top candidates}) + (\text{Mass of Bottom hirers hiring Bottom candidates})}{\text{Total mass of candidates}}$$

If $d = \frac{1}{2}$, then there is no sorting, and firms do not demonstrate a preference for hiring candidates of their own type. If $d \in (\frac{1}{2}, 1]$, then there is positive sorting, and firms demonstrate a preference for hiring candidates of their own type. Finally, if $d \in [0, \frac{1}{2})$, then there is negative sorting, which corresponds to the case where firms demonstrate a preference for hiring candidates of the opposite type. The following proposition illustrates that if all hirers have objective beliefs about their own ability and the distribution of abilities is symmetric, then there will be no sorting in equilibrium. All proofs are relegated to the Appendix.

**Proposition 3.** If $f(q)$ is symmetric and $f^s(q^s) = f(q)$, then $\forall \theta \in \{LN, HN\}$, there is no sorting in equilibrium: $d^* = \frac{1}{2}$.

I now study how the degree of sorting in the labor market equilibrium changes when all hirers are overconfident.

**Definition 5.** All hirers are overconfident if $q^*_{s_h} \geq q_h$ for all $q_h$.

In other words, all hirers are overconfident if they weakly overestimate their probability of being high ability.

**Proposition 4.** If all hirers are overconfident, there will be a greater degree of sorting than if hirers had objective beliefs about their own ability.

If all firms are overconfident, then as in Corollary 3, hirers will be willing to pay too much for candidates of their own type and willing to pay too little for candidates of the opposite type relative to a candidate’s expected productivity. This implies that in the market for Top candidates, Top hirers will have a higher willingness to pay than rational while Bottom hirers will have a lower willingness to pay than rational. Since firms on the demand curve are ordered in descending order of their willingness to pay, relative to the case when all firms are rational, this means that there will be a greater mass of Top hirers who demand Top candidates at higher wages, and a greater mass of Bottom hirers who demand Top candidates at lower wages.
Conversely, in the market for Bottom candidates, Top hirers will have a lower willingness to pay than rational, and Bottom hirers will have a hirer willingness to pay than rational. Thus, relative to the rational case, there will be a greater mass of Bottom hirers in the left upper segment of the demand curve, and a greater mass of Top hirers in the right lower segment.

4.3 Simulating Labor Market Outcomes with Experimental Data

Using the experimental data, I simulate the hiring outcomes in equilibrium if the experiment subjects had participated in a competitive labor market. Using subjects’ reported willingness to pay, I derive the demand curves for Top and Bottom candidates. I plot the demand curves by group and confidence treatment, which leads to a $2 \times 2$ cell analysis: High and Low confidence treatment, and Top and Bottom candidates. As the realized proportion of subjects that are Top and Bottom group members in the data is not identical, I normalize the demand curves so that the horizontal axis is the fraction of firms who demand the worker at a given wage. Figure 6 depicts the normalized demand curves.

Figure 6: Demand Curves
In each cell, I plot the demand by Top and Bottom hirers separately, as this will allow me to study how overconfidence affects the degree of sorting. The solid curve represents the demand curve for Top hirers, while the dashed curve represents the demand curve for Bottom hirers with a reversed horizontal axis. At any given wage, the distance from the left vertical axis to the solid curve represents the mass of Top hirers who demand the worker, while the distance from the right vertical axis to the dashed curve represents the mass of Bottom hirers who demand the worker. The sum of these two distances represents total demand for a type of worker at a given wage $w$.

To simulate the equilibrium outcome, I assume that all experiment subjects are hirers and that the market is balanced: in other words, that there are an equal number of workers of each type. In this case, the supply of workers of a given type when the market is balanced is the total length of the horizontal axis. In this case, the competitive equilibrium is simply the intersection between the two demand curves on the diagram, since this is the wage $w$ in the market where total demand is equal to supply. At this given wage, we can find the fraction of Top and Bottom hirers employing workers in that market.

To study the impact of overconfidence on the degree of sorting, we compare the two graphs on the left with the two on the right. If the intersection of the two demand curves is exactly in the mid point of the horizontal axis, this means that an equal proportion of Top and Bottom hirers employ that type of candidate in equilibrium. If the intersection is to the left of the mid point ($< 0.5$), this means that a greater proportion of Bottom hirers hire that type of candidate, while an intersection towards the right of the midpoint ($> 0.5$) implies that a greater proportion of Top hirers hire that type of candidate in equilibrium. As such, if the intersection is to the right of the midpoint in the market for Top candidates, or the intersection is to the left of the midpoint in the market for Bottom candidates, this implies more sorting as a greater proportion of Top hirers are hiring Top candidates, or Bottom hirers hiring Bottom candidates.

The diagrams suggest that there is a greater level of sorting in hiring outcomes in the simulated labor market equilibrium in the high confidence treatment. In the market for Top candidates, the intersection in the high confidence treatment (0.53) is further to the right than the low confidence treatment (0.43), suggesting a greater proportion of Top hirers matching with Top candidates when there is overconfidence. In the market for Bottom candidates, the picture is less clear, as there is some bunching in reported willingness to pay around $w_B = 40$. The right most point of the intersection in the high and low confidence treatments are similar, but the left most point of the intersection in the high confidence treatment is further left than the low confidence treatment. From the free-form feedback in the questionnaire, this bunching in reported willingness to pay around $w_B = 40$ is attributable
to loss aversion.

5 Conclusion

In this paper, I study how hirer overconfidence in own ability, a behavioral bias, affects their hiring decisions. I develop a simple theoretical model that shows that overconfidence in their own ability can cause hirers to overvalue candidates from a similar educational background as themselves: overconfident hirers who attend elite schools are willing to pay too much of a premium for candidates from elite schools, thereby propagating elitism in hiring decisions. I run a laboratory experiment with Ivy League undergraduates where I randomly assign subjects to an easy or difficult version of a real effort task to vary the level of overconfidence in the session.

The experimental results confirm the main hypotheses: in sessions where subjects faced the easy version of the task, there was overconfidence at the session level, and subjects reported a higher relative willingness to pay for a member of their own group. On the other hand, in the difficult task treatment, subjects were underconfident at the session level and reported a lower relative willingness to pay for a member of their own group. Extending the model to a labor market equilibrium and simulating equilibrium outcomes using the experimental data, I find that the impact of hirer overconfidence on valuation of candidates translates into a greater degree of sorting in the labor market compared to if hirers had objective beliefs about their own ability: there is greater incidence of hirers hiring candidates of their own group in equilibrium.

There are several policy implications of this study. First, the research findings can help firms and institutions develop better hiring procedures to give greater equality of opportunity to candidates from all backgrounds. Less discretion and more structure in the hiring process may be an improvement, as this limits room for behavioral biases to affect hiring decisions. For example, studies have shown that evaluating candidates based on work sample tests and assessments related to actual job responsibilities leads to employing better workers on average (Hoffman et al., 2017). Another method that may improve hiring procedures is to replace face-to-face free-form interviews with structured ones, where each candidate for the position is asked the same set of questions. The former, despite being a popular method of assessing candidates, has consistently been shown to be a poor predictor of job performance (Dana et al., 2013; McDaniel et al., 1994).

Understanding how overconfidence, and more generally behavioral biases affect hiring decisions at the top of the income distribution can help inform the policy debate on inequality in society and social mobility in the US. This will also have policy implications for how
universities can modify their admissions processes to give greater opportunity to candidates from low income backgrounds.

This project also provides an explanation for why people often have a preference for others similar to themselves, a phenomenon known as homophily (McPherson et al., 2001; Currarini et al., 2009). Homophily is commonly treated as a type-based preference. However, my model predicts that people who are more overconfident may exhibit a stronger inclination for others similar to themselves. The findings from this study may thus help us understand potential behavioral microfoundations or causes leading to homophily. These results may have implications not only for labor market hiring, but also group formation, development of political preferences, and information dissemination.
References


6 Appendix

6.1 Overconfidence by Gender

Males are significantly more overconfident than females.

Figure 7: Subjective Belief of Being in the Top Half by Gender

Figure 8: CDF of $q^*$ by Gender
Table 7: Average subjective belief of being in the top half by gender

<table>
<thead>
<tr>
<th>male</th>
<th>Average</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (n=195)</td>
<td>0.52</td>
<td>0.23</td>
</tr>
<tr>
<td>1 (n=102)</td>
<td>0.60</td>
<td>0.26</td>
</tr>
<tr>
<td>Total (n=297)</td>
<td>0.55</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Average subjective belief of being in the top half by gender. This difference is statistically significant to a 1% level.

6.2 Proofs

Proof of Claim 1. I first prove Property 1: $\hat{r}(T, q_h^s) = \hat{r}(B, 1 - q_h^s)$.

\[
\hat{r}(T, q_h^s) = \Pr(\theta = LN | s_h = T, q_h^s) = \frac{\Pr(s_h = T | \theta = LN, q_h^s) \Pr(\theta = LN)}{\Pr(s_h = T | \theta = LN, q_h^s) \Pr(\theta = LN) + \Pr(s_h = T | \theta = HN, q_h^s) \Pr(\theta = HN)} = q_h^s p_{LN} + (1 - q_h^s) (1 - p_{LN})
\]

\[
\hat{r}(B, 1 - q_h^s) = \Pr(\theta = LN | s_h = B, 1 - q_h^s) = \frac{\Pr(s_h = B | \theta = LN, 1 - q_h^s) \Pr(\theta = LN)}{\Pr(s_h = B | \theta = LN, 1 - q_h^s) \Pr(\theta = LN) + \Pr(s_h = B | \theta = HN, 1 - q_h^s) \Pr(\theta = HN)} = (1 - q_h^s) (1 - p_{LN}) + q_h^s p_{LN} + (1 - q_h^s) (1 - p_{HN}) + q_h^s p_{HN} = \hat{r}(T, q_h^s)
\]
I next show Property 2. We are interested in the impact of \( q_h^s \) on \( \hat{r}(T, q_h^s) \):

\[
\hat{r}(T, q_h^s) = \frac{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN})}{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN}) + q_h^s p^{HN} + (1 - q_h^s)(1 - p^{HN})}
\]

\[
\hat{r}(T, q_h^s) = \frac{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN})}{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN}) + q_h^s p^{HN} + (1 - q_h^s)(1 - p^{HN})}
\]

\[
\hat{r}(T, q_h^s) = 1 + \frac{q_h^s p^{HN} + (1 - q_h^s)(1 - p^{HN})}{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN})} \cdot \frac{q_h^s p^{HN} + (1 - q_h^s)(1 - p^{HN})}{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN})}
\]

\[
\hat{r}(T, q_h^s) = 1 + \frac{q_h^s p^{HN} + (1 - q_h^s)(1 - p^{HN})}{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN})}
\]

\[
\hat{r}(T, q_h^s) = \frac{1}{1 + C}
\]

where \( C \equiv \frac{q_h^s p^{HN} + (1 - q_h^s)(1 - p^{HN})}{q_h^s p^{LN} + (1 - q_h^s)(1 - p^{LN})} \) and \( \frac{\partial \hat{r}(T, q_h^s)}{\partial C} \) < 0. Let \( x \equiv \frac{q_h^s}{1 - q_h^s} \), and note that \( \frac{\partial x}{\partial q_h^s} > 0 \). Then,

\[
C = \frac{xp^{HN} + 1 - p^{HN}}{xp^{LN} + 1 - p^{LN}}
\]

In order to find the impact of \( q \) on \( \hat{r}(T, q_h^s) \):

\[
\frac{\partial \hat{r}(T, q_h^s)}{\partial q_h^s} = \frac{\partial \hat{r}(T, q_h^s)}{\partial C} \frac{\partial C}{\partial x} \frac{\partial x}{\partial q_h^s}
\]

Differentiating \( C \) with respect to \( x \):

\[
\frac{\partial C}{\partial x} = \frac{p^{HN}(xp^{LN} + 1 - p^{LN}) - (xp^{HN} + 1 - p^{HN})p^{LN}}{(xp^{LN} + 1 - p^{LN})^2}
\]

\[
= \frac{xp^{HN} - p^{LN}}{(xp^{LN} + 1 - p^{LN})^2} < 0
\]

since \( p^{LN} > p^{HN} \). Thus:

\[
\frac{\partial \hat{r}(T, q_h^s)}{\partial q_h^s} = \frac{\partial \hat{r}(T, q_h^s)}{\partial C} \frac{\partial C}{\partial x} \frac{\partial x}{\partial q_h^s} > 0
\]

\[
\square
\]

34
Proof of Proposition 1. A hirer’s willingness to pay for a candidate is:

\[
WTP^{s_h}(q^*_h) = E[f(a_c) | s_c, s_h, q^*_h] = \Pr(a_c = H | s_c, s_h, q^*_h)f(H) + \Pr(a_c = L | s_c, s_h, q^*_h)f(L)
\]  

(3)

Then,

\[
\Delta WTP^T(q^*_h) > \Delta WTP^B(q^*_h)
\]

(4)

\[\iff \quad f(H)[\Pr(a_c = H|T, T, q^*_h) - \Pr(a_c = H|B, T, q^*_h) - \Pr(a_c = H|T, B, q^*_h) + \Pr(a_c = H|B, q^*_h)] > f(L)[\Pr(a_c = L|T, B, q^*_h) - \Pr(a_c = L|B, q^*_h) - \Pr(a_c = L|T, q^*_h) + \Pr(a_c = L|B, T, q^*_h)] \]

(5)

Substituting \(\Pr(a_c = L|\cdot, \cdot) = 1 - \Pr(a_c = H|\cdot, \cdot)\) and simplifying yields:

\[
\left[f(H) - f(L)\right] \left[\Pr(a_c = H|T, T, q^*_h) - \Pr(a_c = H|B, T, q^*_h) - \Pr(a_c = H|T, B, q^*_h) + \Pr(a_c = H|B, q^*_h)\right] > 0
\]

(6)

Let \(B \equiv \Pr(a_c = H|T, T, q^*_h) - \Pr(a_c = H|B, T, q^*_h) - \Pr(a_c = H|T, B, q^*_h) + \Pr(a_c = H|B, q^*_h)\).

Noting that:

\[
\Pr(a_c|s_c, s_h) = \Pr(\theta = LN|s_h, q^*_h)\Pr(a_c|s_c, LN) + \Pr(\theta = HN|s_h, q^*_h)\Pr(a_c|s_c, HN)
\]

\[= \hat{r}(s_h, q^*_h)\Pr(a_c|s_c, LN) + (1 - \hat{r}(s_h, q^*_h))\Pr(a_c|s_c, HN)\]

(7)

B can be rewritten as:

\[B = \left[\hat{r}(T, q^*_h) - \hat{r}(B, q^*_h)\right] \left[\Pr(a_c = H|T, LN) - \Pr(a_c = H|B, LN)\right]
\]-\left[\Pr(a_c = H|T, HN) - \Pr(a_c = H|B, HN)\right]\]

(8)

Substituting (8) back into (6) yields:

\[
\left[f(H) - f(L)\right]\left[\hat{r}(T, q^*_h) - \hat{r}(B, q^*_h)\right] \left[\Pr(a_c = H|T, LN) - \Pr(a_c = H|B, LN)\right]
\]-\left[\Pr(a_c = H|T, HN) - \Pr(a_c = H|B, HN)\right]\] > 0

(9)
where the conditional probabilities are:

\[
\Pr(a_c = H|T, LN) = \frac{\Pr(s_c = T|a_c = H, \theta = LN) \Pr(a_c = H|\theta = LN)}{\Pr(s_c = T|\theta = LN)} = \frac{p_{LN} q_c}{q_c p_{LN} + (1 - q_c)(1 - p_{LN})}
\]

\[
\Pr(a_c = H|B, LN) = \frac{\Pr(s_c = B|a_c = H, \theta = LN) \Pr(a_c = H|\theta = LN)}{\Pr(s_c = B|\theta = LN)} = \frac{q_c (1 - p_{LN})}{q_c (1 - p_{LN}) + (1 - q_c)p_{LN}}
\]

\[
\Pr(a_c = H|T, HN) = \frac{\Pr(s_c = T|a_c = H, \theta = LN) \Pr(a_c = H|\theta = HN)}{\Pr(s_c = T|\theta = HN)} = \frac{p_{HN} q_c}{q_c p_{HN} + (1 - q_c)(1 - p_{HN})}
\]

Note that \( f(H) - f(L) > 0 \) by assumption, and \( \hat{r}(T, q_{h}) - \hat{r}(B, q_{h}) > 0 \) iff \( q_h > \frac{1}{2} \) by Claim 1. We want to check what conditions are necessary for the term in the big brackets in (8) to be positive. In other words, we have to check whether:

\[
\frac{p_{LN} q_c}{q_c p_{LN} + (1 - q_c)(1 - p_{LN})} > \frac{q_c (1 - p_{LN})}{q_c (1 - p_{LN}) + (1 - q_c)p_{LN}} - \frac{q_c (1 - p_{HN})}{q_c p_{HN} + (1 - q_c)(1 - p_{HN})}
\]

The term on the LHS and the term on the RHS are the same except that on the left the \( p \)'s correspond to \( p_{LN} \) and on the right they correspond to \( p_{HN} \). Thus, the inequality is true if:

\[
\frac{pq_c}{q_c p + (1 - q_c)(1 - p)} > \frac{q_c (1 - p)}{q_c (1 - p) + (1 - q_c)p}
\]

increases as \( p \) increases.

Since the hirer has a uniform prior on candidates’ abilities: \( q_c = 1 - q_c = \frac{1}{2} \), this term
simplifies to:

\[
\frac{pq_c}{q_c p + (1 - q_c)(1 - p)} - \frac{q_c(1 - p)}{q_c(1 - p) + (1 - q_c)p} = \frac{p}{p + (1 - p)} - \frac{1 - p}{(1 - p) + p} = 2p - 1
\]

Thus, Equation (9) can be rewritten as:

\[
[f(H) - f(L)] \left[ \hat{r}(T, q_h^*) - \hat{r}(B, q_h^*) \right] \left[ p^{LN} - p^{HN} \right] > 0
\]

By assumption, \( f(H) - f(L) > 0 \) and \( p^{LN} - p^{HN} > 0 \). As was shown in an earlier claim, \( \hat{r}(T, q_h^*) - \hat{r}(B, q_h^*) > 0 \) iff \( q_h^* > \frac{1}{2} \). Thus, the claim is true iff \( q_h^* > \frac{1}{2} \).

Proof of Proposition 1. Analogous to the proof of Proposition 1, for the case where the hirer attends a top school \( s_h = T \), the inequality above can be rewritten as:

\[
[f(H) - f(L)] \left[ \hat{r}(T, q_h^*) - \hat{r}\left( T, \frac{1}{2} \right) \right] \left( p^{LN} - p^{HN} \right) > 0
\]

By assumption, \( f(H) - f(L) > 0 \) and \( p^{LN} - p^{HN} > 0 \). From Claim 1, \( \hat{r}(T, q_h^*) - \hat{r}\left( T, \frac{1}{2} \right) > 0 \). Hence, the inequality holds true.

Analogously, if the hirer attends a bottom school \( s_h = B \), the inequality above can be rewritten as:

\[
[f(H) - f(L)] \left[ \hat{r}(B, q_h^*) - \hat{r}\left( B, \frac{1}{2} \right) \right] \left( p^{LN} - p^{HN} \right) > 0
\]

From Corollary 1, \( \hat{r}(B, q_h^*) - \hat{r}\left( B, \frac{1}{2} \right) < 0 \). Hence, the inequality is proven.

Proof of Proposition 2. The claim follows immediately if \( \Delta WTP^T(q^*) - \Delta WTP^B(q^*) \) is an increasing function of \( q^* \).

From the proof of Proposition 1,

\[
\Delta WTP^T(q^*) - \Delta WTP^B(q^*) = [f(H) - f(L)] \left[ \hat{r}(T, q_h^*) - \hat{r}(B, q_h^*) \right] \left[ p^{LN} - p^{HN} \right]
\]

By assumption, \( f(H) - f(L) > 0 \) and \( p^{LN} - p^{HN} > 0 \) and do not depend on \( q^* \). From Claim 1, we know that \( \hat{r}(T, q_h^*) \) is increasing in \( q^* \) and \( \hat{r}(B, q_h^*) \) is decreasing in \( q_h^* \). Hence, \( \hat{r}(T, q_h^*) - \hat{r}(B, q_h^*) \) and therefore \( \Delta WTP^T(q^*) - \Delta WTP^B(q^*) \) is also increasing in \( q^* \).
Proof of Claim 2. I first prove Property 1. I first show that the claim is true for hirers with \( s_h = B \):

\[
WTP_T^B(q_h^s) - WTP_B^B(q_h^s) = E\left[f(a_c)|T, B, q_h^s\right] - E\left[f(a_c)|B, B, q_h^s\right]
\]

\[
= \Pr(a_c = H|T, B, q_h^s)f(H) + \Pr(a_c = L|T, B, q_h^s)f(L)
\]

\[
- \Pr(a_c = H|B, B, q_h^s)f(H) - \Pr(a_c = L|B, B, q_h^s)f(L)
\]

\[
= \left[f(H) - f(L)\right] \left(\Pr(a_c = H|T, B, q_h^s) - \Pr(a_c = H|B, B, q_h^s)\right)
\]

where:

\[
\Pr(a_c = H|T, B, q_h^s) - \Pr(a_c = H|B, B, q_h^s)
\]

\[
= \hat{r}_h(B, q_h^s) \Pr(a_c = H|T, LN) + \left(1 - \hat{r}_h(B, q_h^s)\right) \Pr(a_c = H|T, HN)
\]

\[
- \hat{r}_h(B, q_h^s) \Pr(a_c = H|B, LN) - \left(1 - \hat{r}_h(B, q_h^s)\right) \Pr(a_c = H|B, HN)
\]

\[
= \hat{r}_h(B, q_h^s) \left(\Pr(a_c = H|T, LN) - \Pr(a_c = H|B, LN)\right) +
\]

\[
\left(1 - \hat{r}_h(B, q_h^s)\right) \left(\Pr(a_c = H|T, HN) - \Pr(a_c = H|B, HN)\right)
\]

\[
= \hat{r}_h(B, q_h^s)(2p^{LN} - 1) + \left(1 - \hat{r}_h(B, q_h^s)\right)(2p^{HN} - 1)
\]

Hence,

\[
WTP_T^B(q_h^s) - WTP_B^B(q_h^s) = \left[f(H) - f(L)\right] \left(\hat{r}_h(B, q_h^s)(2p^{LN} - 1) + \left(1 - \hat{r}_h(B, q_h^s)\right)(2p^{HN} - 1)\right)
\]

By assumption, \( f(H) - f(L) > 0 \). Hence, if the term in the second big bracket is postitive, then the claim for \( B \) hirers is proven. Note that:

\[
2(\hat{r}_h(B, q_h^s)p^{LN} + (1 - \hat{r}_h(B, q_h^s))p^{HN}) - 1 > 0 \iff \hat{r}_h(B, q_h^s)p^{LN} + (1 - \hat{r}_h(B, q_h^s))p^{HN} > \frac{1}{2}
\]

Since \( p^{LN} \) and \( p^{HN} > \frac{1}{2} \) and \( \hat{r}(B, q_h^s) \in [0, 1] \), \( \hat{r}_h(B, q_h^s)p^{LN} + (1 - \hat{r}_h(B, q_h^s))p^{HN} > \frac{1}{2} \). Hence:

\[
WTP_T^B(q_h^s) - WTP_B^B(q_h^s) = \left[f(H) - f(L)\right] \left(2(\hat{r}_h(B, q_h^s)p^{LN} + (1 - \hat{r}_h(B, q_h^s))p^{HN}) - 1\right) > 0
\]

Combining this result with Proposition \( \Box \) proves the case for hirers with \( s_h = T \).

I next prove Property 2. I prove the first equality by showing that \( WTP_T^T(q) - WTP_T^T(1-q) = 0 \). By expanding the function and factoring terms as was done in the proof of Proposi-
\[ WP_T^T(q) - WP_B^T(1 - q) = \left[ f(H) - f(L) \right] \left[ \hat{r}_h(T, q) - \hat{r}_h(B, 1 - q) \right] (p_{LN} - p_{HN}) \]

Since \( f(H) - f(L) > 0 \) and \( p_{LN} - p_{HN} > 0 \), \( WP_T^T(q) - WP_B^T(1 - q) = 0 \) if and only if \( \hat{r}_h(T, q) - \hat{r}_h(B, 1 - q) = 0 \), which was shown in Claim 1.

The proof of \( WP_B^B(q) - WP_B^B(1 - q) = 0 \) is analogous.

Finally, I prove Property 3. I first show that \( WP_T^T(q) \) is increasing in \( q \).

\[
WP_T^T(q) = E[f(a)|s_c = T, s_h = T, q] = \Pr(a_c = H|s_c = T, s_h = T, q)f(H) + \Pr(a_c = L|s_c = T, s_h = T, q)f(L)
\]

\[
= \Pr(a_c = H|s_c = T, s_h = T, q)f(H) + \left( 1 - \Pr(a_c = H|s_c = T, s_h = T, q) \right) f(L)
\]

\[
= f(L) + \left[ f(H) - f(L) \right] \left[ \hat{r}_h(T, q) \Pr(a_c = H|s_c = H, \theta = LN) + \left( 1 - \hat{r}_h(T, q) \right) \Pr(a_c = H|s_c = H, \theta = HN) \right]
\]

Since by Claim 1 I showed that \( \hat{r}(T, q) \) is increasing in \( q \), \( WP_T^T(q) \) is increasing in \( q \) if and only if \( \Pr(a_c = H|s_c = T, \theta = LN) > \Pr(a_c = H|s_c = T, \theta = HN) \).

From the proof of Proposition 1,

\[
\Pr(a_c = H|s_c = T, \theta = LN) = \frac{p_{LN}q_c}{q_c p_{LN} + (1 - q_c)(1 - p_{LN})}
\]

and

\[
\Pr(a_c = H|s_c = T, \theta = HN) = \frac{p_{HN}q_c}{q_c p_{HN} + (1 - q_c)(1 - p_{HN})}
\]

Hence,

\[
\Pr(a_c = H|s_c = T, \theta = LN) > \Pr(a_c = H|s_c = T, \theta = HN) \iff \frac{p_{LN}q_c}{q_c p_{LN} + (1 - q_c)(1 - p_{LN})} > \frac{p_{HN}q_c}{q_c p_{HN} + (1 - q_c)(1 - p_{HN})} \iff p_{LN} > p_{HN}
\]

39
Analogously,
\[
WTP^B_T(q) = E[f(a)|s_c = T, s_h = B, q]
= \Pr(a_c = H|s_c = T, s_h = B, q)f(H) + \Pr(a_c = L|s_c = T, s_h = B, q)f(L)
= \Pr(a_c = H|s_c = T, s_h = B, q)f(H) + \left(1 - \Pr(a_c = H|s_c = T, s_h = B, q)\right)f(L)
= f(L) + \frac{f(H) - f(L)}{2} \hat{r}_h(B, q) \Pr(a_c = H|s_c = T, \theta = LN)
+ (1 - \hat{r}_h(B, q)) \Pr(a_c = H|s_c = T, \theta = HN)
\]

By Claim \[1\], \(\hat{r}(B, q_h^*)\) is decreasing in \(q_h^*\), it is also the case that \(WTP^B_T(q)\) is decreasing in \(q\).

\[\Box\]

**Proof of Proposition 3.** In the market for top candidates, \(d^* = \frac{1}{2} \iff D^T_T = D^B_T\).

\[
D^T_T = D^B_T
\iff \frac{1}{2} - (1-p^\theta)F^s(\hat{q}) - (2p^\theta - 1)E(q_h|q_h^* \leq \hat{q})F^s(\hat{q})
= p^\theta F^s(1 - \hat{q}) + (1 - 2p^\theta)E(q_h|q_h^* \leq 1 - \hat{q})F^s(1 - \hat{q})
\]

By symmetry of the distribution of \(q\), \(1 - F(q) = F(1 - q)\), so the equality can be rewritten as:

\[
F(\hat{q})\left[1 - E(q_h|q_h \leq \hat{q}) - E(q_h|q_h \leq 1 - \hat{q})\right] = \frac{1}{2} - E(q_h|q_h \leq 1 - \hat{q})
\]

Using the fact that \(E(q_h|q_h^* \leq \hat{q}) + E(q_h|q_h^* \geq 1 - \hat{q}) = 1\), we can further simplify the above:

\[
F(\hat{q})\left[1 - E(q_h|q_h \geq 1 - \hat{q}) - E(q_h|q_h \leq 1 - \hat{q})\right] = \frac{1}{2} - E(q_h|q_h \leq 1 - \hat{q})
\]

\[
\frac{1}{2} = F(\hat{q})E(q_h|q_h \geq 1 - \hat{q}) + F(1 - \hat{q})E(q_h|q_h \leq 1 - \hat{q})
\]

The equality above always holds true by the law of iterated expectations. Hence, the claim is proven. 

\[\Box\]

**Proof of Proposition 4.** If \(q_h^* \geq q_h \Rightarrow \Pr(q_h^* \leq \hat{q}|s_h = T) \leq \Pr(q_h \leq \hat{q}|s_h = T)\).
The demand for Top candidates by Top hirers can be written as:
\[
D_T^T = \frac{1}{2}(1 - \Pr(q_h^* \leq \hat{q}|s_h = T))
\]

Hence, the difference in demand by Top hirers due to overconfidence is:
\[
D_T^T(q_h^*) - D_T^T(q_h) = \frac{1}{2}\left[\Pr(q_h \leq \hat{q}|s_h = T) - \Pr(q_h^* \leq \hat{q}|s_h = T)\right] \geq 0
\]

Analogously, the demand for Top candidates by Bottom hirers can be written as:
\[
D_B^T = \frac{1}{2}\Pr(q_h^* \leq 1 - \hat{q}|s_h = B)
\]

Hence, the difference in demand by Bottom hirers due to overconfidence is:
\[
D_T^B(q_h^*) - D_T^B(q_h) = \frac{1}{2}\left[\Pr(q_h \leq 1 - \hat{q}|s_h = B) - \Pr(q_h^* \leq 1 - \hat{q}|s_h = B)\right] \leq 0
\]

There is a greater mass of T firms demanding a T worker when firms are overconfident, and there is a smaller mass of B firms demanding a B candidate relative to if all firms are rational. Together, this leads to a greater degree of sorting.

6.3 Experiment Materials

6.3.1 Sample of Questions in Trivia Task

Easy Treatment:
1. Which of the following dogs is the smallest?
   (a) Dachshund
   (b) Poodle
   (c) Pomeranian
   (d) Chihuahua
2. From what trees do acorns grow?
   (a) Oak
   (b) Maple
   (c) Walnut
   (d) Beech
3. What color are emeralds?
   (a) Blue
   (b) Green
4. Who is the patron saint of Ireland?
   (a) St David  
   (b) St Andrew  
   (c) St George  
   (d) St. Patrick

5. What is the capital of Australia?
   (a) Sydney  
   (b) Melbourne  
   (c) Canberra  
   (d) Perth

**Hard Treatment:**

1. Boris Becker contested consecutive Wimbledon men’s singles finals in 1988, 1989, and 1990, winning in 1989. Who was his opponent in all three matches?
   (a) Michael Stich  
   (b) Andre Agassi  
   (c) Ivan Lendl  
   (d) Stefan Edberg

2. Suharto held the office of president in which large Asian nation?
   (a) Malaysia  
   (b) Japan  
   (c) Indonesia  
   (d) Thailand

3. Who was Henry VIII’s wife at the time of his death?
   (a) Catherine Parr  
   (b) Catherine of Aragon  
   (c) Anne Boleyn  
   (d) Jane Seymour

4. What do you most fear in hormephobia?
   (a) Saliva  
   (b) Shock  
   (c) Worms  
   (d) Silence

5. For what did Einstein get the Nobel prize in Physics?
   (a) Quantum mechanics
Instructions 1: Introduction
Welcome to the experiment. This is an experiment on decision-making. All information that you are told in this experiment is truthful and involves no deception, and all decisions that you make are anonymous.

No communication between participants will be permitted during the remainder of the session. You are also not permitted to use any electronic devices or programs other than the designated experiment software to communicate with others or to look up information.

The experiment will consist of 2 phases: Phase 1 and Phase 2. In Phase 1, you will be asked to answer a set of trivia questions. This will be followed by Phase 2, where you will have to make decisions related to the outcome in Phase 1. More information about Phase 2 will follow after Phase 1.

In both phases, you may earn tokens depending on your actions, the actions of others, and chance. The money you will be paid at the end of today’s experiment depends on the number of tokens you have earned. The rate at which tokens convert to dollars is $0.09 per token.

Experiment Payment
Your payment for this experiment will consist of two parts:
  1. A show up fee of $5
  2. A payment which will depend on either your performance in the trivia questions in Phase 1 or one of the decisions you make in Phase 2. Which one is selected for payment is random, equally likely, and only determined at the end of the experiment.

Trivia Questions Instructions
You will be asked to answer a set of 25 multiple choice general knowledge questions. You will have 6 minutes to complete the questions. There will be a timer that shows you how much time is remaining.
There are 5 questions on each page. Your answers on each page will be saved when you click a button to leave the page. For as long as there is time remaining, you can go forward and backward between the pages and modify your answers to the questions freely.

**How your Payoff is Determined if Your Trivia Score is Selected for Payment:**
Your score is the number of trivia questions you answered correctly. If your score is among the top half of all participants in today’s experiment, then you will earn 250 tokens. If your score is among the bottom half of all participants in today’s experiment, then you will earn 100 tokens. If there is a tie in the score at the cut-off between the top and the bottom half, then the tie will be broken randomly.

**Instructions 2**
Welcome to Phase 2 of the experiment. In this phase, you will make four decisions.

**Classification of Performance**
You will now be classified as a High or Low performance participant depending on your trivia score.
- If your score is among the top half of all participants in today’s experiment, you are a **High performance** participant.
- If your score is among the bottom half of all participants in today’s experiment, you are a **Low performance** participant.

If there is a tie in the score at the cut-off between the top and the bottom half, then the tie will be broken randomly.

**Decision 1: Beliefs about Performance**
In the next screen, you will be asked what you think are the chances you think you are a High performance participant. Recall that a High performance participant is one whose trivia score is in the top half of all participants in today’s experiment.

You will be asked to report a number between 0 and 100, where 0 corresponds to no chance and 100 corresponds to full certainty that you are a High performance participant.

If this question is selected for payment, you have the opportunity to earn 100 or 250 tokens depending on your reported belief and whether you are a High performance participant. The closer your guess to the correct answer, the higher your chance of receiving the higher payment of 250 tokens. Otherwise, you will earn 100 tokens. The payment mechanism is such
that *reporting your true belief about your performance will maximize your chance of earning the higher payment of 250 tokens.*

**Instructions 3: Assignment to Groups**

Now, you and all other participants in this experiment will be assigned into groups. There will be two groups: **Top** and **Bottom.**

All participants will be assigned to groups based on their trivia performance and chance. We will say that a participant’s group is “reflective of his/her/their performance” if a High trivia performance participant is assigned to the Top group, or a Low trivia performance participant is assigned to the Bottom group. On the other hand, we will say that a participant’s group is “not reflective of his/her/their performance” if a High trivia performance participant is assigned to the Bottom group, or a Low trivia performance participant is assigned to the Top group.

To determine how groups are assigned, the computer will perform a simulation of one fair coin flip with a 50% chance of heads and a 50% chance of tails, and *this outcome applies to all participants in today’s session:*

- If the coin comes up **heads, every participant** in today’s session has a:
  - 95% chance to be assigned to a group that **reflects his/her/their performance,** and a
  - 5% chance to be assigned to a group that **does not reflect his/her/their performance**

- If the coin comes up **tails, every participant** in today’s session has a:
  - 55% chance to be assigned to a group that **reflects his/her/their performance,** and a
  - 45% chance to be assigned to a group that **does not reflect his/her/their performance**

After the computer has flipped the coin and determined the chance that each participant is assigned to a group that reflects their performance, whether each participant’s group ends up being reflective of their performance is resolved individually.

Once the coin flip has taken place and participants are assigned to groups, you will be informed which group you are assigned to, but you will not be informed whether your group reflects your performance or what the outcome of the coin flip was.

The figure below illustrates an example of how groups are assigned for the following four
Summary

To determine group assignment, the computer flips one coin for all participants where heads or tails is equally likely. If the coin comes up heads, a participant’s group is highly likely to reflect their trivia performance. If the coin comes up tails, a person’s group is mildly likely to reflect their trivia performance. Note that in both cases, there is some chance that group reflects performance, and a chance that it does not.

In the next screen, you will face several practice questions to help you understand the experiment better. The practice questions do not affect how much you are paid, but you must correctly answer all questions to continue.

Instructions 4: “Hiring” of Top and Bottom group members

For the following two decisions, you have an endowment of 160 tokens.

You will have the opportunity to “hire” a randomly drawn participant from each group in today’s experiment, and receive a reward depending on their trivia performance. If you “hire” a High performance participant, your reward will be 200 tokens. If you “hire” a Low performance participant, your reward will be 40 tokens. As such, what determines your reward is the underlying performance of the participant. However, you cannot see what their performance is, but instead you observe which group they are from.

To “hire” a participant, you have to pay a hiring cost that has not yet been determined. You will be asked what is the maximum cost you are willing to pay to “hire” a randomly drawn
participant from each group: one from the Top group and one from the Bottom group. For each of the decisions, you may end up paying that amount or less, but never more.

If this question is selected for payment, you have the opportunity to earn up to a maximum of 360 tokens including your endowment. On the next page, we explain the payment mechanism in more detail, but what is the most important is that truthfully reporting the maximum hiring cost you are willing for a member of each group will maximize your expected earnings from each of the decisions.

Cost Determining Mechanism and Your Payoff
Specifically, the cost determining mechanism is as follows: After you have decided on the maximum hiring cost you are willing to pay for a random member of each group, the computer will first determine whether you hire a Top group member. The computer randomly draws a hiring cost between 0 and 200 tokens for a random Top group member. All numbers are equally likely.

- If the computer drawn hiring cost is lower than or the same as the maximum you stated, you hire a Top group member and pay the computer drawn hiring cost.
- If the computer drawn hiring cost is greater than the maximum you stated, you do not hire a Top group member.

If you hire a Top group member, then one member of the Top group is randomly selected. Your payoff, in addition to your endowment of 160 tokens, will be your reward based on their trivia performance (either 200 or 40 tokens) less the computer drawn hiring cost. In the event that your payoff is negative, this will be deducted from your endowment of 160 tokens. If you do not hire a Top group member, your payoff from this decision will be your endowment of 160 tokens.

Separately, the computer repeats a similar process to determine if your hire a Bottom group member and your payoff to that decision. Note that the computer will randomly draw a new hiring cost for a Bottom group member.

Note: The hiring cost that you pay is not given to another participant as a “wage”. All participants face the same decision as you. If one of these decisions is selected for payment, their earnings are determined by their “hiring” decision as was described above.

The figure below is a graphical summary of the “hiring” decision for a member of one group (either Top or Bottom):
Summary
You have an endowment of 160 tokens. You can gain a reward that depends on the trivia performance of another participant in today’s session, but you can only observe the group they are from. You are asked to choose the maximum you are willing to pay for a randomly drawn participant from the Top group and a randomly drawn participant from the Bottom group. The cost determining mechanism is such that truthfully reporting the maximum hiring cost you are willing to pay for a member of each group will maximize your expected payoff from this decision.

In the next screen, you will once again be asked to answer several practice questions.

Phase 2: Decision 4
This is your final decision for the experiment. In the next screen, you will be asked what you think the chances are that the coin flip came up heads. Recall that if the coin flip came up heads, there is a 95% chance that your group reflects your performance, and a 5% chance that it does not. If the coin flip came up tails, there is a 55% chance that your group reflects your performance, and a 45% chance that it does not.

If this question is selected for payment, you have the opportunity to earn a fixed payment of either 100 or 250 tokens depending on your reported belief and the outcome of the
coin flip. Your payment will be determined using the same procedure as the other question about beliefs earlier in today’s experiment. The closer your guess to the correct answer, the higher your chance of receiving the higher payment of 250 tokens. The best way to maximize your chance of receiving the higher payment of 250 tokens is to truthfully report your best estimate of the answer.

6.4 Deriving the Demand Function for Top Candidates

We begin with the demand function for Top candidates $D_T$. The total demand for Top candidates is the sum of demand by Top hirers $D^T_T$ and demand by Bottom hirers $D^B_T$:

$$D_T = D^T_T + D^B_T$$

I first derive the demand for Top candidates by Top hirers $D^T_T$. A Top hirer will demand a Top candidate if:

$$WTP^T_T(q^*_h) \geq w_T$$

$$q^*_h \geq WTP^T_T^{-1}(w_T) \equiv \hat{q}$$

At a given wage $w_T$ in the market for top candidates, we may think of $\hat{q}$ as the subjective belief of the Top firms with the marginal willingness to pay at that wage. We will refer to these firms as the marginal Top firms at wage $w_T$. Hence, the demand for Top candidates by Top hirers $D^T_T$ is:

$$D^T_T = \Pr(s_h = T) \Pr(q^*_h \geq \hat{q}|s_h = T)$$

$$= \frac{1}{2} \left[ 1 - \Pr(q^*_h \leq \hat{q}|s_h = T) \right]$$

$$= \frac{1}{2} \left[ 1 - \frac{\Pr(s_h = T|q^*_h \leq \hat{q}) F^s(\hat{q})}{\Pr(s_h = T)} \right]$$

$$= \frac{1}{2} \left[ 1 - 2 \Pr(s_h = T|q^*_h \leq \hat{q}) F^s(\hat{q}) \right]$$

(10)
where:

$$\Pr(s_h = T|q^*_h \leq \hat{q}) = \int \Pr(s_h = T|q_h = m, q^*_h \leq \hat{q}) \, dF_q(m|q^*_h \leq \hat{q})$$

$$= \int_0^1 \left[ mp^\theta + (1 - m)(1 - p^\theta) \right] \, dF_q(m|q^*_h \leq \hat{q})$$

$$= \int_0^1 \left[ (1 - p^\theta) + m(2p^\theta - 1) \right] \, dF_q(m|q^*_h \leq \hat{q})$$

$$= (1 - p^\theta) + (2p^\theta - 1)E(q_h|q^*_h \leq \hat{q})$$

(11)

Substituting (11) back into (10):

$$D^*_T = \frac{1}{2} \left[ 1 - 2 \left[ (1 - p^\theta) + (2p^\theta - 1)E(q_h|q^*_h \leq \hat{q}) \right] \right] F^s(\hat{q})$$

$$= \frac{1}{2} - (1 - p^\theta)F^s(\hat{q}) - (2p^\theta - 1)E(q_h|q^*_h \leq \hat{q})F^s(\hat{q})$$

$$= (1 - p^\theta) \left[ 1 - F^s(\hat{q}) \right] + (2p^\theta - 1) \left[ \frac{1}{2} - E(q_h|q^*_h \leq \hat{q})F^s(\hat{q}) \right]$$

Analogously, Bottom hirers will demand Top candidates if:

$$WTP^B_T(q^*_h) \geq w_T$$

$$WTP^T_T(1 - q^*_h) \geq w_T$$

(by Claim 2)

$$1 - q^*_h \geq WTP^T_T^{-1}(w_T)$$

$$q^*_h \leq 1 - \hat{q}$$
Hence, the demand by Bottom hirers for Top candidates $D^B_T$ is:

$$D^B_T = \Pr(s_h = B) \Pr(q_h^* \leq 1 - \hat{q}|s_h = B)$$
$$= \frac{1}{2} \Pr(q_h^* \leq 1 - \hat{q}|s_h = B)$$
$$= \frac{1}{2} \frac{\Pr(s_h = B|q_h^* \leq 1 - \hat{q}) F^*(1 - \hat{q})}{\Pr(s_h = B)}$$
$$= \Pr(s_h = B|q_h^* \leq 1 - \hat{q}) F^*(1 - \hat{q})$$

Thus, total demand for Top candidates is:

$$D_T = D^T_T + D^B_T$$
$$= (1 - p^\theta) \left(1 - F^*(\hat{q}) \right) + (2p^\theta - 1) \left[\frac{1}{2} - E(q_h|q_h^* \leq 1 - \hat{q}) F^*(\hat{q}) \right]$$

7 Deriving Demand for Bottom candidates $D_B$

I next derive the demand function for $B$ candidates, $D_B$, in a similar way.

$$D_B = D^B_B + D^T_B$$
Bottom hirers will demand a Bottom candidate if:

\[ WTP_B(q_h^s) \geq w_B \]
\[ q_h^s \geq WTP_B^{-1}(w_B) \equiv \tilde{q} \]

Hence, the demand for Bottom candidates by Bottom hirers \( D_B \) is:

\[ D_B = \Pr(s_h = B) \Pr(q_h^s \geq \tilde{q} | s_h = B) \]
\[ = \frac{1}{2} \left[ 1 - \Pr(q_h^s \leq \tilde{q} | s_h = B) \right] \]
\[ = \frac{1}{2} \left[ 1 - \frac{\Pr(s_h = B | q_h^s \leq \tilde{q}) F^s(\tilde{q})}{\Pr(s_h = B)} \right] \]
\[ = \frac{1}{2} \left[ 1 - 2 \Pr(s_h = B | q_h^s \leq \tilde{q}) F^s(\tilde{q}) \right] \quad (12) \]

where:

\[ \Pr(s_h = B | q_h^s \leq \tilde{q}) = \int_0^1 \Pr(s_h = B | q_h = m, q_h^s \leq \tilde{q}) dF_q(m | q_h^s \leq \tilde{q}) \]
\[ = \int_0^1 \left[ m(1 - p^\theta) + (1 - m)p^\theta \right] dF_q(m | q_h^s \leq \tilde{q}) \]
\[ = \int_0^1 \left[ m(1 - 2p^\theta) + p^\theta \right] dF_q(m | q_h^s \leq \tilde{q}) \]
\[ = p^\theta + (1 - 2p^\theta) E(q_h | q_h^s \leq \tilde{q}) \quad (13) \]

Substituting (13) back into (12):

\[ D_B = \frac{1}{2} \left[ 1 - 2 \left[ p^\theta + (1 - 2p^\theta) E(q_h | q_h^s \leq \tilde{q}) \right] F^s(\tilde{q}) \right] \]
\[ = \frac{1}{2} - p^\theta F^s(\tilde{q}) - (1 - 2p^\theta) E(q_h | q_h^s \leq \tilde{q}) F^s(\tilde{q}) \]
\[ = p^\theta \left( 1 - F^s(\tilde{q}) \right) + (1 - 2p^\theta) \left[ \frac{1}{2} - E(q_h | q_h^s \leq \tilde{q}) F^s(\tilde{q}) \right] \]
Analogously, Top hirers will demand Bottom candidates if:

\[ WTP_T^B(q_h^s) \geq w_B \]
\[ WTP_B^B(1 - q_h^s) \geq w_B \]
\[ 1 - q_h^s \geq WTP_B^B^{-1}(w_B) \]
\[ q_h^s \leq 1 - \tilde{q} \]

Hence, the demand by Top hirers for Bottom candidates \( D_T^B \) is:

\[
D_T^B = \Pr(s_h = T) \Pr(q_h^s \leq 1 - \tilde{q} | s_h = T) \\
= \frac{1}{2} \Pr(q_h^s \leq 1 - \tilde{q} | s_h = T) \\
= \frac{1}{2} \left[ \Pr(q_h^s \leq 1 - \tilde{q} | s_h = T) \right] \\
= \Pr(q_h^s \leq 1 - \tilde{q}) F^s(1 - \tilde{q}) \\
= \left[ \int_0^1 \Pr(q_h^s \leq 1 - \tilde{q}) dF_q(m|q_h^s \leq 1 - \tilde{q}) \right] F^s(1 - \tilde{q}) \\
= \left[ \int_0^1 \left[ mp^\theta + (1 - m)(1 - p^\theta) \right] dF_q(m|q_h^s \leq 1 - \tilde{q}) \right] F^s(1 - \tilde{q}) \\
= \left[ \int_0^1 \left[ m(2p^\theta - 1) + (1 - p^\theta) \right] dF_q(m|q_h^s \leq 1 - \tilde{q}) \right] F^s(1 - \tilde{q}) \\
= \left[ (1 - p^\theta) + (2p^\theta - 1)E(q_h|q_h^s \leq 1 - \tilde{q}) \right] F^s(1 - \tilde{q}) \\
= (1 - p^\theta)F^s(1 - \tilde{q}) + (2p^\theta - 1)E(q_h|q_h^s \leq 1 - \tilde{q})F^s(1 - \tilde{q})
\]

Thus, total demand for Bottom candidates is:

\[
D_B = D_B^B + D_T^B \\
= p^\theta \left[ 1 - F^s(\tilde{q}) \right] + (1 - 2p^\theta) \left[ \frac{1}{2} - E(q_h|q_h^s \leq \tilde{q})F^s(\tilde{q}) \right] \\
\text{Demand by Bottom Hirers} \\
+ (1 - p^\theta)F^s(1 - \tilde{q}) + (2p^\theta - 1)E(q_h|q_h^s \leq 1 - \tilde{q})F^s(1 - \tilde{q}) \\
\text{Demand by Top Hirers}
\]