Essays on Matching Theory and Behavioral Market Design

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Siqi Pan

Graduate Program in Economics

The Ohio State University

2017

Dissertation Committee:

Professor Paul J. Healy, Advisor
Professor Lucas Coffman
Professor John Kagel
Professor Huanxing Yang
Abstract

This dissertation focuses on the design and implementation of matching markets where transfers are not available, such as college admissions, school choice, and certain labor markets. The results contribute to the literature from both a theoretical and a behavioral perspective, and may have policy implications for the design of some real-life matching markets.

Chapter 1, “Exploding Offers and Unraveling in Two-Sided Matching Markets,” studies the unraveling problem prevalent in many two-sided matching markets that occurs when transactions become inefficiently early. In a two-period decentralized model, I examine whether the use of exploding offers can affect agents' early moving incentives. The results show that when the culture of the market allows firms to make exploding offers, unraveling is more likely to occur, leading to a less socially desirable matching outcome. A market with an excess supply of labor is less vulnerable to the presence of exploding offers; yet the conclusion is ambiguous for a market with a greater degree of uncertainty in early stages, which depends on the specific information structure. While a policy banning exploding offers tends to be supported by high quality firms and workers, it can be opposed by those of lower quality. This explains the prevalence of exploding offers in practice.

Chapter 2, “Constrained School Choice and Information Acquisition,” investigates a common practice of many school choice programs in the field, where the length of
students’ submitted preference lists are constrained. In an environment where students have incomplete information about others’ preferences, I theoretically study the effect of such a constraint under both a Deferred Acceptance mechanism (DA) and a Boston mechanism (BOS). The result shows that ex-ante stability can only be ensured under an unconstrained DA, but not under a constrained DA, an unconstrained BOS, or a constrained BOS. In a lab experiment, I find that the constraint also affects students’ information acquisition behavior. Specifically, when faced with a constraint, students tend to acquire less wasteful information and distribute more efforts to acquire relevant information under DA; such an effect is not significant under BOS. Overall, the constraint has a negative effect on efficiency and stability under both mechanisms.

Chapter 3, “Targeted Advertising on Competing Platforms,” is jointly written with Huanxing Yang. We investigate targeted advertising in two-sided markets. Each of the two competing platforms has single-homing consumers on one side and multi-homing advertising firms on the other. We focus on how asymmetry in platforms’ targeting abilities translates into asymmetric equilibrium outcomes, and how changes in targeting ability affect the price and volume of ads, consumer welfare, and advertising firms' profits. We also compare social incentives and equilibrium incentives in investing in targeting ability.

Chapter 4, “The Instability of Matching with Overconfident Agents: Laboratory and Field Investigations,” focuses on centralized college admissions markets where students are evaluated and allocated based on their performance on a standardized exam. A single exam’s measurement error causes the exam-based priorities to deviate from colleges'
aptitude-based preferences: a student who underperforms in one exam may lose her placement at a preferred college to someone with a lower aptitude. The previous literature proposes a solution of combining a Boston algorithm with pre-exam preference submission. Under the assumption that students have perfect knowledge of their relative aptitudes before taking the exam, the suggested mechanism intends to trigger a self-sorting process, with students of higher (lower) aptitudes targeting more (less) preferred colleges. However, in a laboratory experiment, I find that such a self-sorting process is skewed by overconfidence, which leads to a welfare loss larger than the purported benefits. Moreover, the mechanism introduces unfairness by rewarding overconfidence and punishing underconfidence, thus serving as a gender penalty for women. I also analyze field data from Chinese high schools; the results suggest similar conclusions as in the lab.
To

Fang, Ming, and Hengli
Acknowledgments

I am deeply indebted to my advisor Paul J. Healy for all the support, patience, and encouragement throughout my graduate studies. I would like to thank John Kagel, Lucas Coffman, Huanxing Yang, Katie Baldiga Coffman, and Yan Chen for helpful comments and inspiration. I am grateful to Hengli Zhang, Ming Pan, and Qiquan Xu for their help in the data collection process. The projects in Chapters 2 and 4 were supported by JMCB Grants for Graduate Student Research in Economics. All remaining errors are mine.
Vita

2009.................................................................B.A. Finance, Shanghai Jiao Tong University

2011...............................................................Research Master, Economics, Tilburg University

2012...............................................................M.A. Economics, The Ohio State University

2012 to present .............................................Graduate Teaching Associate, Department of Economics, The Ohio State University

Fields of Study

Major Field: Economics
# Table of Contents

Abstract ............................................................................................................... ii  

Acknowledgments ................................................................................................. vi  

Vita.............................................................................................................................. vii  

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

List of Figures ......................................................................................................... x  

Chapter 1: Exploding Offers and Unraveling in Two-Sided Matching Markets .......... 1 

Chapter 2: Targeted Advertising on Competing Platforms ................................... 17 

Chapter 3: Constrained School Choice and Information Acquisition .................. 40 

Chapter 4: The Instability of Matching with Overconfident Agents: Laboratory and Field Investigations ................................................................. 51 

References ............................................................................................................. 90 

Appendix A: Proofs ............................................................................................... 93 

Appendix B: Additional Analysis .......................................................................... 101 

Appendix C: Supplementary Tables and Figures .................................................... 108
List of Tables

Table 1. Score Distributions and Aptitudes ................................................................. 59
Table 2. Treatment Design .......................................................................................... 67
Table 3. Predicting Factors of Overconfidence (OLS) .............................................. 74
Table 4. First Choice in Preference Submission (Ordered Logit) ............................. 78
Table 5. Individual Welfare Distortion (OLS) .............................................................. 81
Table 6. Overconfidence in Shanghai and Xinjiang (OLS) ...................................... 87
Table 7. Individual Welfare Distortion (OLS) .............................................................. 107
Table 8. Standardized Exams in Representative Countries ................................. 108
Table 9. Distribution of Score Rankings (Example 1) ............................................. 108
Table 10. Descriptive Statistics of Key Variables ..................................................... 109
Table 11. Ordered Logit Marginal Effects for First Choice .................................... 109
Table 12. Individual Welfare Distortion (Ordered Logit) ........................................ 111
Table 13. Ordered Logit Marginal Effects for Individual Welfare Distortion .......... 112
List of Figures

Figure 1. Timings of Preference Submission ......................................................... 67
Figure 2. Score-Stability and Aptitude-Stability .................................................. 70
Figure 3. Score-Stably and Aptitude-Stably Matched Pairs .................................. 71
Figure 4. Distribution of OverconfidenceEA, OverconfidenceIN, and ExamError ..... 72
Figure 5. Truth-telling in Preference Submission .................................................. 75
Figure 6. Preference Submission under BOS ....................................................... 76
Figure 7. Distributions of Individual Welfare Distortion from Aptitude-Stability .... 79
Figure 8. Overconfidence and Individual Welfare Distortion .............................. 79
Figure 9. Gender Penalty under PreExam-BOS and Halfway-BOS ..................... 80
Figure 10. Levels of Overconfidence and the Exam's Measurement Error .......... 86
Figure 11. Three Observations of OverconfidenceEA and OverconfidenceIN ....... 101
Figure 12. The Means by AptitudeRank .............................................................. 102
Figure 13. Levels of Overconfidence with Additional Information .................... 104
Figure 14. Preference Submission under BOS .................................................... 104
Figure 15. Aptitude-Stably Matched Pairs .......................................................... 105
Figure 16. Distributions of Individual Welfare Distortion from Aptitude-Stability ... 106
Figure 17. Overconfidence and Individual Welfare Distortion .......................... 106
Figure 18. Biased Beliefs and Non-guess-based Sorting Strategies .................... 110
Chapter 1

Exploding Offers and Unraveling in Two-Sided Matching Markets

1 Introduction

Many two-sided matching markets exhibit the tendency to unravel in time with transactions occurring earlier and earlier. In these markets, with participants’ qualities gradually revealed over time, early transactions can lead to significant efficiency losses: due to the lack of information in early stages, a higher probability of mismatch often leads to instability and thus costly rematching procedures afterwards.\(^1\) In order to effectively halt such an unraveling process, the previous literature has identified several factors that may influence early moving incentives; one of them is the use of exploding offers.

An exploding offer is an offer that comes with a time limit. The offer has to be accepted within the time limit, or it is considered rejected. In contrast, an open offer can be held until the end of the market. According to Niederle and Roth (2009), exploding offers are prevalently used in many markets facing serious unraveling problems, such as the market for new gastroenterologists in the US. In a lab experiment, they are able to reproduce the facilitating effect of exploding offers on early transactions. However, not all markets that use exploding offers suffer from unraveling. For example, in the job market for junior economists, offers often come with short time limits, yet little tendency towards unraveling has been observed so far.

This paper aims to reconcile the above phenomena from a theoretical perspective. In a two-period decentralized matching market, the true qualities of workers are not fully revealed

\(^1\)Roth and Xing (1994) provide a detailed overview of various evidence for market unraveling. See also Mongell and Roth (1991), Haruvy, Roth, and Ünver (2006), Avery, Fairbanks, and Zeckhauser (2009), Avery et al. (2001), and Fréchette, Roth, and Ünver (2007).
until the second period; firms and workers only observe a signal in the first period. By comparing markets with or without a banning policy on exploding offers, I show that when firms are allowed to make exploding offers, equilibria without unraveling are supported by a smaller parameter space and thus a stable matching is less likely to be achieved. Intuitively, a worker tends to exploit an open offer by holding it until the last period. In response, a firm uses an exploding offer to eliminate the risk of being rejected and remaining unmatched at end of the market. The time limit forces the worker to balance the cost of rejecting the current offer and the likelihood of receiving a better offer in the future. As a result, an early exploding offer is more easily accepted than an early open offer, and thus the market is more likely to unravel when exploding offers are allowed. In addition, I identify the sufficient conditions under which an equilibrium without unraveling never exists, and the sufficient and necessary condition for the market to fully unravel.

The above results indicate that the use of exploding offers is only a necessary but not a sufficient condition for unraveling to occur.\(^2\) An important question we should ask is what characteristics of a market can make it less likely to be affected by exploding offers? This study highlights two findings on comparative statics. First, a market tends to be less vulnerable when there is an excess supply of labor, that is, when workers outnumber job vacancies in the market. In this case, low quality workers remain unmatched in an equilibrium without unraveling, and a firm incurs the risk of hiring these workers by moving before the resolution of uncertainty. Therefore, the firm may be unwilling to deviate even when an early exploding offer would be accepted. Similar intuitions are also present in Niederle, Roth, and Ünver (2013), in which they find unraveling tends to occur in a market with comparable demand and supply.\(^3\)

Second, signal accuracy has an ambiguous effect on how vulnerable a market is to exploding offers; the conclusion depends on the specific information structure. As mentioned above, in the first period, market participants only observe a signal suggesting each worker’s potential quality or type. As the signal becomes more accurate, a firm is more willing to make an early exploding offer to a high type worker while the worker is less willing to accept; a firm is less willing to make an early exploding offer to a low type worker while the worker is more willing to accept. In other words, the early moving incentives always change in opposite directions for different sides of the market. The overall effect of signal accuracy hinges on how uncertainty is resolved in a specific market. Under the same framework, I show that opposite effects can be produced simply by altering the information structure of

---

\(^2\) In a different environment, Fainmesser (2013) also shows that the use of exploding offers is a necessary condition for unraveling to occur.

\(^3\) On a related note, Niederle and Roth (2003) suggest that the unraveling in the gastroenterology market might have been triggered by a demand shock. McKinney, Niederle, and Roth (2003) conduct an experimental research on demand and supply imbalances tailored to this market.
the market. Such a finding reconciles the different conclusions from two previous studies. Roth and Xing (1994) find that unraveling tends to be impeded if the uncertainty in early stages is sufficiently large; yet a non-monotonic relationship is shown in Fainmesser (2013): an increase in signal accuracy generates greater market unraveling when the signal is inaccurate enough, while the effect reverses when the signal is sufficiently accurate. This paper suggests that the information structure is a key element that can affect the comparative statics on signal accuracy.

Given the facilitating effect of exploding offers on early transactions, a natural policy consideration is whether exploding offers should be allowed in a market. I investigate the welfare aspect by asking the following questions. Suppose there is a proposal on a ban towards exploding offers. Who would support? Who would oppose? The result indicates that such a banning policy can be supported by high quality firms and workers but opposed by agents of lower qualities. This provides a theoretical support for the experimental findings in Niederle and Roth (2009). In their exploding offer treatment, some higher (lower) quality firms and workers receive significantly lower (higher) payoffs than in the open offer treatment. The conclusion also explains the prevalence of exploding offers in practice and why it is sometimes difficult to achieve consensus on how to solve the unraveling problem in real markets.\footnote{For example, see a discussion regarding the US market for new gastroenterologists by Niederle and Roth (2005).}

Besides the use of exploding offers and the market characteristics discussed above, the previous literature has identified several other factors that could affect unraveling, including the stability of centralized matching algorithms (Roth, 1984, 1991; Kagel and Roth, 2000; Ünver, 2001, 2005), market congestion (Roth and Xing, 1997), the quality distribution over participants on each side (Niederle, Roth, and Ünver, 2013), similarity of preferences (Halaburda, 2010), social network structures (Fainmesser, 2013), and strategic complementarities (Echenique and Pereyra, 2016). The motives behind unraveling are also investigated in some different environments. Under the framework of competitive markets, Li and Rosen (1998), Li and Suen (2000, 2004), and Suen (2000) show how unraveling can occur as a form of insurance in the absence of complete markets. In a model with asymmetric information, but without evolving uncertainty, Lee (2009) explains early contacting as a way to avoid adverse selection.

The subsequent analysis proceeds as follows. Section 2 describes the model. Section 3 provides the equilibrium predictions. In Section 4, I give a simple welfare analysis and discuss some policy issues. Section 5 concludes.
2 The Model

Consider a two-sided matching market with $F$ firms and $W$ workers, where $F \geq 3$ and $W \geq 3$. Let $\mathcal{F} = \{f_1, \ldots, f_F\}$ be the set of firms, and $\mathcal{W} = \{w_1, \ldots, w_W\}$ be the set of workers. Each firm has the capacity to hire at most one worker, and each worker can work for at most one firm. A market with $W < F$ (or $W > F$) is said to have excess demand (or excess supply) of labor.

All workers agree on the same ranking of firms: $f_F \succ f_{F-1} \succ \ldots \succ f_1$, and all firms agree on the same ranking of workers: $w_W \succ w_{W-1} \succ \ldots \succ w_1$. The ranking of firms is common knowledge to the entire market. The true ranking of workers is revealed over time. Let $\mathcal{R}$ be the set of all possible strict rankings of workers, in which each ranking/state is realized with equal probability $\frac{1}{W!}$. Denote the true ranking/state as $\succ \in \mathcal{R}$.

In terms of utility, all firms value a match with the $i$-th ranked worker in the true state ($w_i$) as $v_i = i$, and all workers value a match with the $j$-th ranked firm ($f_j$) as $u_j = j$. Unmatched market participants derive zero utility: $v_0 = u_0 = 0$. Therefore, any match is preferable to remaining unmatched. Notice in this setting, a firm’s utility from a match depends only on the worker’s rank, and a worker’s utility depends only on the firm’s rank.

The two functions $v_i$ and $u_j$ indicate worker quality and firm quality respectively.

The outcome of a matching market, that is, a matching is said to be stable if and only if there is no worker–firm pair in which each prefers one another to her current match. Since the existence of such pairs often leads to costly rematching procedures afterwards, stability is used as a central criterion to evaluate market outcomes by the two-sided matching literature.

In the current environment with strict rankings and aligned preferences, it is easy to see that the assortative matching in the true state $\succ$ constitutes the unique stable matching.

The market lasts for two hiring periods, with the true ranking $\succ$ revealed in Period 2. At the beginning of Period 1, a public signal $\hat{\succ} \in \mathcal{R}$ is observed by both firms and workers. With probability $\alpha$, $\hat{\succ}$ is the same as $\succ$. Otherwise, $\hat{\succ}$ is a uniform random draw from $\mathcal{R}$. The parameter $\alpha \in (0, 1)$ measures signal accuracy: a larger value of $\alpha$ indicates a smaller uncertainty faced by the market. Denote a worker’s type in Period 1 as $\hat{r}$, which is her rank in $\hat{\succ}$. Denote a worker’s true rank as $r$, which is her rank in $\succ$.

---

5Halaburda (2010) considers the similarity of firms’ preferences over workers as a comparative statics parameter while having all workers agree on the same ranking of firms. The result shows that similarity of preferences is an important factor driving unraveling.

6Normalizing the utility range to be between 0 and 1 would not change the main results of the paper.

7An example that allows non-linear utility distributions can be found in the working paper version (https://www.dropbox.com/s/xtb9zv5tuuyhi4e/Exploding_Offers_WorkingPaper0801.pdf?dl=0). It suggests that a market is less likely to be affected by exploding offers if the quality distribution over firms is more convex, or the quality distribution over workers is more concave. On a related note, Niederle, Roth, and ¨Unver (2013) show in a lab experiment that unraveling only occurs when demand and supply are comparable, that is, when there exist excess workers, but a shortage of high quality workers.
The game proceeds as follows. In Period 1, a public signal $\hat{\succ}$ is observed. Next, each firm simultaneously makes an offer to at most one worker. Finally, each worker simultaneously chooses at most one offer to accept from those available to her. A similar procedure takes place in the second period, except that the true ranking of workers $\succ$, instead of a signal, is observed at the beginning of the period. All actions of firms and workers are publicly observed.

I focus my discussion on two types of offers: exploding offers and open offers.

**Definition 1.** An *exploding offer* is an offer that comes with a time limit. It can only be accepted within the time limit. Otherwise, it is rejected.

**Definition 2.** An *open offer* is an offer that can be held until the last period.

In the current two-period model, an exploding offer has to be accepted immediately, in the same period in which it is made. However, if a worker receives an open offer in the first period, she could choose to hold it until Period 2. An open offer made in Period 2 is equivalent to an exploding offer.

Regarding the culture or norms of the market, I make the following two assumptions.

**Assumption 1.** (Binding acceptances) Once a worker accepts an offer, the acceptance is binding. A worker cannot renege on her acceptance.

**Assumption 2.** (Binding rejections) Once a worker rejects an offer, the rejection is binding. A firm will not make an offer to the same worker again.

Following Niederle and Roth (2009), Assumption 1 is made to ensure the validity of exploding offers. Assumption 2 is an important and reasonable addition because (i) it reflects the norms of some real-life two-sided matching markets such as the market for judicial clerks; (ii) it increases the power of exploding offers by raising workers’ rejection costs. Hence, without such an assumption, the effects of exploding offers on market outcomes can be largely underestimated.  

---

8 The reality in some markets is less stringent than Assumption 2. For example, in the job market for junior economists, although when rejecting an exploding offer, a candidate typically does not consider the possibility that she may receive an offer from the same employer again, the phenomenon of nonbinding rejections is still observed in some situations. In this case, a more realistic setting is to have each firm decide whether to raise its leverage by attaching a commitment of binding rejection when making an early exploding offer. For these markets, although Assumption 2 significantly simplifies the analysis, it can lead to an overestimation of the effects of exploding offers.
3 Equilibrium Analysis

In this section, I start with a baseline case where firms can only make open offers, and then relax the constraint by allowing exploding offers. The discussion mainly concerns two types of subgame perfect Nash equilibria in weakly undominated pure strategies: those without unraveling, and those with full unraveling.

The following characterization of equilibria focuses on the timing of offers made by “relevant” firms, which include all firms as \( W \geq F \) and only firms \( f_F, f_{F-1}, \ldots, \) and \( f_{F-W+1} \) as \( W < F \). This is because in either type of equilibria, \( f_{F-W}, f_{F-W-1}, \ldots, \) and \( f_1 \) as \( W < F \) are indifferent among all possible strategies and their actions do not affect the equilibrium outcome.

**Definition 3.** An *equilibrium without unraveling* is an equilibrium where no relevant firms make any offers until the last period.

In an equilibrium without unraveling, no actions are taken by relevant firms in Period 1. In Period 2, two cases are considered separately: (i) when \( W \geq F \), \( f_F \) makes an offer to \( w_W, f_{F-1} \) to \( w_{W-1}, \ldots, f_1 \) to \( w_{W-F+1} \), and all offers are accepted; (ii) when \( W < F \), \( f_F \) makes an offer to \( w_W, f_{F-1} \) to \( w_{W-1}, \ldots, f_{F-W+1} \) to \( w_1 \), and all these offers are accepted.

**Definition 4.** An *equilibrium with full unraveling* is an equilibrium where every relevant firm makes an early offer in Period 1.

In an equilibrium with full unraveling, after \( \succ \) is revealed in Period 1, two cases are considered separately: (i) when \( W \geq F \), \( f_F \) makes an offer to type \( \hat{r} = W \), \( f_{F-1} \) to \( \hat{r} = W - 1, \ldots, f_1 \) to \( \hat{r} = W - F + 1 \), and all offers are accepted; (ii) when \( W < F \), \( f_F \) makes an offer to type \( \hat{r} = W \), \( f_{F-1} \) to \( \hat{r} = W - 1, \ldots, f_{F-W+1} \) to \( \hat{r} = 1 \), and all these offers are accepted.

In both types of equilibria, when \( W < F \), firms \( f_{F-W}, f_{F-W-1}, \ldots, \) and \( f_1 \) may adopt any strategy; their offers (if made) are not accepted by any workers. Clearly, the unique outcome of equilibria without unraveling is the assortative matching according to the true ranking of workers \( \succ \), which is the unique stable matching in the current setting. The unique outcome of equilibria with full unraveling is the assortative matching according to the signal-suggested ranking of workers \( \hat{\succ} \), which is only stable when the signal is correct.

In addition, an equilibrium is said to have *partial unraveling* if some relevant firms make offers in Period 1, and some make offers in Period 2.
3.1 Open Offers Only

Consider the case where firms can only make open offers due to the culture, norms, or policies in a market environment. The following proposition describes the equilibrium outcome.

**Proposition 1.** *When firms are not allowed to make exploding offers, there only exist equilibria without unraveling; the stable matching is the unique equilibrium outcome.*

While the full proof is provided in the appendix, the basic intuition is clear. In Period 1, every type of worker has a positive probability of having the highest quality in the true state. Therefore, as long as the best firm $f_F$ moves in Period 2, a worker strictly prefers to hold any early open offer until Period 2. Knowing this, $f_F$ strictly prefers to wait until the last period, so that all workers will stay in the market and the one of the highest quality can be perfectly identified. Since no offer is accepted in Period 1, the other firms cannot make themselves better off by moving early; instead, they incur the risk of being rejected in the last period and remaining unmatched.

Proposition 1 shows that in a market where firms only use open offers, an equilibrium without unraveling always exists, while an equilibrium with full unraveling never does. This provides us with a very clean baseline, so that the effects of exploding offers can be easily identified from the change in the parameter spaces supporting these two types of equilibria.

3.2 Exploding and Open Offers

Now I consider the case where both open offers and exploding offers can be made in a market.

**Lemma 1.** *In an equilibrium in undominated strategies, firms $f_{F-1}, f_{F-2}, \ldots, f_1$ never make an open offer in Period 1 when they are allowed to make exploding offers.*

The best firm $f_F$ is indifferent between an open offer and an exploding offer since neither of them will be rejected by any worker. In Period 2, every firm is indifferent because an open offer is equivalent to an exploding offer. Hence, in the subsequent analysis of equilibria in weakly undominated strategies, Lemma 1 allows us to consider only exploding offers without loss of generality.

After observing a signal $\succ$ in Period 1, both firms and workers update their beliefs. Posteriors on the true state $\succ$ are given by

$$
\Pr(\succ \mid \succ) = \alpha + \frac{1 - \alpha}{W!},
$$

---

9For example, exploding offers are publicly discouraged in the US market for new graduate students. The Council of Graduate Schools has published a resolution stating that students are under no obligation to respond to offers of financial support prior to April 15 (http://cgsnet.org/april-15-resolution).
and
\[ \Pr(\succ' | \hat{\succ}) = \frac{1 - \alpha}{W}, \quad \forall \succ' \neq \hat{\succ}. \]

Posterior on the true rank \( r \) of a type-\( \hat{r} \) worker are given by
\[ \Pr(r | \hat{r}) = \alpha + \frac{1 - \alpha}{W}, \]
and
\[ \Pr(r' | \hat{r}) = \frac{1 - \alpha}{W}, \quad \forall r' \neq \hat{r}. \]

The posteriors described above involve a spike at the point \( \hat{\succ} \), which indicates a high probability of the true state being the signal-suggested ranking, and an equally low probability of any other ranking being realized.\(^{10}\) Therefore, the assortative pairs in the state \( \hat{\succ} \) will be frequently used in the subsequent analysis. These “signal-suggested pairs” are formally defined as follows.

**Definition 5.** The signal-suggested type of a firm \( f_j \) is a function defined as
\[ \hat{r}(j) \equiv \begin{cases} j + W - F & \text{if } j > F - W, \\ 0 & \text{if } j \leq F - W. \end{cases} \]

The signal-suggested firm of a type \( \hat{r} \) is a function defined as
\[ j(\hat{r}) \equiv \begin{cases} \hat{r} - W + F & \text{if } \hat{r} > W - F, \\ 0 & \text{if } \hat{r} \leq W - F. \end{cases} \]

A firm and its signal-suggested type, or a type of worker and its signal-suggested firm, are called a signal-suggested pair.

### 3.2.1 Equilibria without Unraveling

As mentioned in the introduction, in practice not all markets that use exploding offers suffer from unraveling. Proposition 2 provides sufficient conditions for an equilibrium without unraveling to sustain even when exploding offers are allowed in a market, while Proposition 3 provides sufficient conditions for such an equilibrium never to exist.

Recall a market with \( W < F \) (or \( W > F \)) is said to have excess demand (or excess supply) of labor. I further identify the case of extreme excess supply if \( W \geq 2F \), and the case of moderate excess supply if \( F < W \leq 2F - 1 \).

\(^{10}\)In Section 3.3, I provide an example of a different information structure, under which the probability of a ranking gradually decreases as its Kendall \( \tau \) distance to the signal-suggested ranking increases.
Proposition 2. When firms are allowed to make exploding offers, an equilibrium without unraveling always exists if

(i) the market has an extreme excess supply of labor; or

(ii) the market has a moderate excess supply of labor and a sufficiently inaccurate signal.

Mathematically speaking, an equilibrium without unraveling always exists if (i) \( W \geq 2F \), or (ii) \( F + 2 < W \leq 2F - 1 \) and

\[ \alpha \leq \frac{(W - F)^2 - W + F - 2}{(W - F)^2 + W + F - 2}. \]

A deviation from such an equilibrium involves both sides of the market: a firm should want to make an early exploding offer in Period 1 to a worker who wants to accept it. Hence, the condition for an equilibrium without unraveling to sustain requires that, for each worker type \( \hat{r} \), the firms whose offers would be accepted should not be willing to offer. In most cases, the signal-suggested firm-worker pairs have the strongest incentive to deviate, thus affecting the binding constraint that drives the results in Proposition 2. Below I use an example to explain why.

Example 1. In a market with 3 firms and 3 workers, an equilibrium without unraveling yields the following assortative matching:

\[
\begin{array}{ccc}
  w_3 & w_2 & w_1 \\
  f_3 & f_2 & f_1 \\
\end{array}
\]

Consider a deviation between \( f_2 \) and a worker of type \( \hat{r} = 3 \). Given the posteriors on this type in Period 1, the offer will be accepted if

\[
2 \geq (\alpha + \frac{1 - \alpha}{3}) \times 3 + \frac{1 - \alpha}{3} \times 1 + \frac{1 - \alpha}{3} \times 1.
\]

While the LHS gives the worker’s utility when accepting the offer, the RHS is the worker’s expected utility when rejecting. The second part of the RHS indicates a rejection cost of \( \frac{1 - \alpha}{3} \): after rejecting \( f_2 \), the worker can only receive an offer from \( f_1 \) even when her true rank turns out to be \( r = 2 \) in Period 2. On the other hand, if \( f_2 \) makes an early offer to its signal-suggested type \( \hat{r} = 2 \), the offer will be accepted when

\[
2 \geq \frac{1 - \alpha}{3} \times 3 + (\alpha + \frac{1 - \alpha}{3}) \times 1 + \frac{1 - \alpha}{3} \times 1.
\]

Again, the second part of the RHS indicates a rejection cost of \( (\alpha + \frac{1 - \alpha}{W}) \).

Example 1 shows that a deviation between a signal-suggested pair tends to succeed more easily: although a firm always prefers higher types, its signal-suggested type is more likely to accept its early offer due to the higher rejection cost \( (\alpha + \frac{1 - \alpha}{3}) > (\alpha + \frac{1 - \alpha}{W}) \). This is quite intuitive since it indicates that a worker is more reluctant to reject what appears to be “a good match” in Period 1 — a firm that is most likely to be her match in a stable matching.

Therefore, for an equilibrium without unraveling to sustain, the binding condition in most cases requires the best firm that would be accepted by its corresponding type not be willing to offer.

Condition (ii) of Proposition 2 is mainly driven by the workers’ side of the market. As the signal in Period 1 becomes less accurate, the cost of rejecting a signal-suggested firm \( (\alpha + \frac{1 - \alpha}{3}) \), or more generally, \( (\alpha + \frac{1 - \alpha}{W}) \), decreases. Thus, a worker is more likely to reject
an exploding offer and an equilibrium without unraveling is more likely to sustain.\footnote{Although the cost of rejecting a non-signal-suggested firm $\frac{1-\alpha}{\alpha}$ increases as the signal becomes less accurate, here the elimination of such a deviation is not the binding condition and does not drive the results. In Section 3.3, I provide an example with a different information structure, under which the non-signal-suggested pairs can also influence the binding condition, and thus signal accuracy may have the opposite effect on market unraveling.} On the other hand, condition (i) stems from a boundary solution on the firms’ side, in which case no firm is willing to make an early offer to its signal-suggested type even if it is always accepted. It arises when there is an extreme excess supply of labor ($W \geq 2F$), which means in an equilibrium without unraveling, even the worst firm $f_1$ is matched with an above-average worker. This significantly increases a firm’s risk in making an offer before the resolution of uncertainty since the worker is more likely to have a lower quality compared to the firm’s match in equilibrium.

**Proposition 3.** When firms are allowed to make exploding offers, an equilibrium without unraveling never exists if

(i) the market has an excess demand of labor and a sufficiently accurate signal; or
(ii) the market has a moderate excess supply of labor and a sufficiently accurate signal.

Mathematically speaking, an equilibrium without unraveling never exists if (i) $W \leq F$ and $\alpha > \frac{W}{2W-1}$, or (ii) $F < W < 2F - 1$ and $\alpha > \frac{(W-F)^2 + W + F - 2}{(W-F)^2 + 3W + F - 2}$. For a deviation to occur, there must exist a type of worker and a firm in Period 1 such that, the firm is willing to offer and the worker is willing to accept. Again, the binding condition here hinges on the deviations between signal-suggested pairs. Proposition 3 shares the similar intuition with Proposition 2: a more accurate signal encourages a deviation by increasing the worker’s rejection cost, while an excess demand of labor decreases the firm’s risk of being worse off in a deviation.

There exists a gap between the two parameter spaces identified in Propositions 2 and 3. Let $\alpha_1$ and $\alpha_2$ be functions of $W$ and $F$ such that $\alpha_1(W, F) \equiv \frac{(W-F)^2 - W + F - 2}{(W-F)^2 + W + F - 2}$ and $\alpha_2(W, F) \equiv \frac{(W-F)^2 + W + F - 2}{(W-F)^2 - W + F - 2}$. For some cases of moderate excess supply ($F + 2 < W < 2F - 1$), an equilibrium without unraveling always exists when $\alpha \leq \alpha_1$ and never exists when $\alpha > \alpha_2$. It is easy to confirm that $\alpha_1 < \alpha_2$.

Such a gap is due to an integer problem, which makes conditions in Propositions 2 and 3 sufficient but not necessary. Denote $j^A$ as the lowest quality firm whose early offer is accepted by its signal-suggested type and $j^O$ as the highest quality firm willing to make such an offer given the acceptance. For an equilibrium without unraveling to exist, a sufficient and necessary condition only requires no integer to be in the range between $j^A$ and $j^O$, which depends on the specific parameter values of $\alpha$, $W$, and $F$. In order to draw a general conclusion, we need to have the range completely empty, that is, $j^A \geq j^O$. Similarly,
although an equilibrium without unraveling does not exist as long as there is an integer in the range between $j^A$ and $\overline{j^O}$, for a general conclusion, we need to set the range larger than 1, that is, $\overline{j^O} - j^A > 1$.

For cases within the gap, whether unraveling will occur depends on the specific values of $\alpha$, $F$, and $W$. To illustrate, I provide the following examples where $F + 2 < W < 2F - 1$ and $\alpha \in (\alpha_1, \alpha_2]$. An equilibrium without unraveling exists in case (1) but not in case (2).

**Example 2.** (1) When $\alpha = 0.51$, $F = 5$, and $W = 8$, there does not exist an equilibrium without unraveling: given $j^A \approx 0.7$ and $\overline{j^O} = 1.5$, we know that the signal-suggested pair $f_1$ and $\hat{r} = 4$ has an incentive to deviate.

(2) When $\alpha = 0.51$, $F = 7$, and $W = 12$, an equilibrium without unraveling can sustain since there is no integer between $j^A \approx 1.2$ and $\overline{j^O} = 1.5$.

### 3.2.2 Equilibria with Full Unraveling

Now we discuss an extreme case, where the market fully unravels with every relevant firm making an exploding offer in Period 1.\(^{12}\)

**Proposition 4.** When firms are allowed to make exploding offers, an equilibrium with full unraveling exists if and only if $W \leq F$ and $\alpha \geq \frac{W - 2}{W}$, that is, the market has an excess demand and a sufficiently accurate signal.

When $W > F$, an equilibrium with full unraveling can never sustain because after Period 1, there are $W - F$ workers left in the market. Then every firm has an incentive to deviate by waiting, in which case it becomes the only available firm in Period 2 and its choice set is expanded by $W - F$ workers after the resolution of uncertainty. When $W \leq F$, however, a firm has no incentive to wait as long as in its deviation, no worker would reject an offer from her signal-suggested firm and become available in Period 2. With a more accurate signal, it is more costly for a worker to reject her offer in equilibrium, and thus a deviation is less likely to occur.

### 3.2.3 Equilibria with Partial Unraveling

In an equilibrium with partial unraveling, only some of the relevant firms make offers in Period 1, while the others choose to wait until Period 2. In contrast to the previous two types of equilibria, partial unraveling may take various forms. In this section, I first rule out some impossible structures of such equilibria, and then demonstrate some possibilities in examples.

\(^{12}\)Recall that relevant firms include all firms as $W \geq F$, and only firms $f_F$, $f_{F-1}$,..., and $f_{F-W+1}$ as $W < F$. 

11
Remark 1. There cannot exist an equilibrium with partial unraveling where every firm that moves early has a higher quality than all the firms that choose to wait.

For example, with three firms in a market, there cannot exist an equilibrium with \( f_3 \), or \( f_3 \) and \( f_2 \) being the only early moving firm(s). Such a structure cannot sustain because there exists a profitable deviation for every early moving firm in equilibrium: by deviating to Period 2, it becomes the highest quality firm with an expanded choice set after the resolution of uncertainty.

Combining Propositions 3 and 4, we can identify two parameter spaces where neither an equilibrium without unraveling nor an equilibrium with full unraveling exists: (i) \( F < W < 2F - 1 \) and \( \alpha > \frac{(W-F)^2+W+F-2}{(W-F)^2+3W+F-2} \); (ii) \( 4 \leq W \leq F \) and \( \frac{W-1}{2W-1} < \alpha < \frac{W-2}{W} \). Only equilibria with partial unraveling may exist in these cases. However, as shown in the following example, for some values of \( W, F, \) and \( \alpha \), there does not exist any equilibrium.

Example 3. Suppose there are 4 workers and 3 firms in the market.

1. When \( \alpha \in \left( \frac{3}{7}, \frac{5}{11} \right] \), there only exists the following equilibrium: in Period 1, \( f_2 \) makes an exploding offer to type \( \hat{i} = 4 \) and \( f_1 \) makes an exploding offer to type \( \hat{i} = 2 \); in Period 2, \( f_3 \) makes an offer to the best worker left in the market; all offers are accepted.

2. When \( \alpha \in \left( \frac{5}{11}, \frac{1}{2} \right) \), there does not exist any equilibrium.

3. When \( \alpha \geq \frac{1}{2} \), there only exist the following two equilibria. (i) In Period 1, \( f_1 \) makes an exploding offer to type \( \hat{i} = 2 \); in Period 2, \( f_3 \) makes an offer to the best worker left in the market and \( f_2 \) makes an offer to the second best; all offers are accepted. (ii) In Period 1, \( f_2 \) makes an exploding offer to type \( \hat{i} = 3 \) and \( f_1 \) makes an exploding offer to type \( \hat{i} = 2 \); in Period 2, \( f_3 \) makes an offer to the best worker left in the market; all offers are accepted.

The above example provides us with some interesting intuitions. In case (1), the highest type worker \( \hat{i} = 4 \) in Period 1 is “stolen” by an early moving firm \( f_2 \). But since the signal is not accurate enough, the best firm \( f_3 \) still prefers to wait for the resolution of uncertainty. In case (2), as the signal becomes more accurate, \( \hat{i} = 4 \) becomes more attractive but is still willing to accept an early offer from \( f_2 \). As a result, \( f_3 \) is forced to move early as well so as to prevent such an early transaction; the equilibrium in case (1) can no longer sustain. In case (3) with an even more accurate signal, \( \hat{i} = 4 \) is no longer willing to accept \( f_2 \) in Period 1; both the highest type worker and the highest quality firm choose to wait. Clearly, in the analysis of equilibria with partial unraveling, signal accuracy plays an important role in determining how or whether the early moving incentives for one firm are affected by early offers of other firms, because it tells the high quality firms whether it is worthwhile to “fight” for workers of high types.
3.3 The Ambiguous Effect of Signal Accuracy

According to the above analysis, exploding offers has a facilitating effect on early transactions; such an effect tends to be less salient in a market with an excess supply of labor or a less accurate signal. In this section, I provide an example to show that signal accuracy may have an opposite influence under a different information structure.

**Example 4.** There are three workers in the market. Recall \( \mathcal{R} \) is the set of all possible strict rankings of workers. The Kendall \( \tau \) distance, denoted as \( K(\succ', \succ'') \), is a function that counts the number of pairwise disagreements between two rankings \( \succ' \) and \( \succ'' \), \( \forall \succ', \succ'' \in \mathcal{R} \). In other words, \( K(\succ', \succ'') \) measures the distance between \( \succ' \) and \( \succ'' \).

For instance, between two rankings of three workers, \( K(w_1 \succ w_2 \succ w_3, w_1 \succ w_3 \succ w_2) = 1 \), \( K(w_1 \succ w_2 \succ w_3, w_3 \succ w_1 \succ w_2) = 2 \), and \( K(w_1 \succ w_2 \succ w_3, w_2 \succ w_3 \succ w_1) = 3 \).

The signal \( \hat{\succ} \) in Period 1 equals a ranking \( \succ' \) with probability \( \frac{1}{6} \left[ 1 + 3\beta - 2\beta K(\succ', \hat{\succ}) \right] \), \( \forall \succ' \in \mathcal{R} \), which is decreasing in the distance between \( \succ' \) and the true state \( \succ \). The parameter \( \beta \in (0, \frac{1}{3}) \) measures signal accuracy: a larger value of \( \beta \) indicates a smaller uncertainty faced by the market. The posteriors are thus given by

\[
\Pr(\succ = \succ' | \hat{\succ}) = \frac{1}{6} \left[ 1 + 3\beta - 2\beta K(\succ', \hat{\succ}) \right] \quad \forall \succ' \in \mathcal{R}.
\]

That is, the probability of a ranking being the true state gradually decreases as its distance to the signal \( \hat{\succ} \) increases. This is different from the single spike at \( \hat{\succ} \) in the posteriors described in Section 3.2.

Consider the simple case with three firms. An equilibrium without unraveling exists if and only if \( \beta \geq \frac{1}{4} \). This result stems from the binding condition that prevents a worker of type \( i = 3 \) from accepting \( f_2 \) in Period 1, with \( f_2 \) always willing to make such an exploding offer if accepted. In contrast to Proposition 2, here the condition puts a lower bound to signal accuracy, that is, it requires the signal to be sufficiently accurate for the market not to unravel.

The intuition lies in the differences between the two information structures. The posteriors described in Section 3.2 put a high probability on the signal-suggested ranking, and an equally low probability on any other ranking. As a result, the binding condition for no unraveling mainly concerns the deviations between signal-suggested pairs. In such a pair, a more accurate signal increases the worker’s rejection cost without affecting the firm’s early moving incentive (given the worker’s acceptance of its early exploding offer). Overall,

---

\(^{13}\)See Kendall (1938) and Kemeny (1959).

\(^{14}\)See equations (9) and (11) in the appendix for the expressions of \( j^A \) (the lowest quality firm whose early offer is accepted by its signal-suggested type) and \( j^O \) (the highest quality firm willing to make such an offer given the acceptance) respectively. Signal accuracy \( \alpha \) only enters the former but not the latter.
a more accurate signal makes the market more vulnerable to the use of exploding offers. Under the information structure in Example 4, however, non-signal-suggested pairs can also influence the binding condition. For these pairs, as the signal becomes more accurate, a high type worker in Period 1 becomes more attractive; firms are more willing to make an early exploding offer while the worker is less willing to accept. Meanwhile, a low type worker becomes less attractive; firms are less willing to offer while the worker is more willing to accept. In other words, the early moving incentives always change in opposite directions for different sides of the market. Then in certain cases such as Example 4, when the binding condition prevents a high type worker from accepting a non-signal-suggested firm, the existence of an equilibrium without unraveling requires the signal to be sufficiently accurate. In contrast to the predictions in Section 3.2.1, here a more accurate signal can make a market less likely to be affected by exploding offers.

The above analysis suggests that the effect of signal accuracy may hinge on how uncertainty is resolved in a specific market, which is different from the prediction of Roth and Xing (1994). Like the basic model of this paper, their model indicates a definitive effect: unraveling tends to be impeded if the uncertainty in early stages is sufficiently large, that is, if the signal is inaccurate enough. However, Example 4 provides an information structure under which unraveling can be facilitated by a sufficiently inaccurate signal.

4 Welfare Analysis

Knowing that the use of exploding offers tends to facilitate unraveling and that unraveling hurts the stability of two-sided matching markets, a natural policy consideration is whether a market should allow firms to make exploding offers. In this part, I investigate the welfare aspect by asking the following questions. Suppose there is a proposal on a ban towards exploding offers. Who would support? Who would oppose? In Section 3, I have identified two types of equilibria that are of the most interest in this environment: those without unraveling, and those with full unraveling. A comparison between these two types of equilibria can shed some light on these questions.

Suppose $W$, $F$, and $\alpha$ are such that $4 \leq W \leq F$ and $\alpha \geq \frac{W-2}{W}$. According to previous results, when exploding offers are banned from the market, the market does not unravel (Proposition 1); when exploding offers are allowed, the market always unravels (Proposition 3) and an equilibrium with full unraveling exists (Proposition 4).

Comparing to the assortative matching, $f_j$ is better off in a fully unraveled market if

$$ j + W - F < \alpha (j + W - F) + \frac{(1 - \alpha)(W + 1)}{2}, $$

(1)
where \( j = F - W + 1, F - W + 2, \ldots, F \).\(^{15}\) That is,

\[
W - F + 1 \leq j < \hat{j} \equiv \frac{2F - W + 1}{2}.
\] (2)

From \( W \leq F \), we have

\[
\hat{j} \geq \frac{F + 1}{2}.
\] (3)

The LHS of (1) is \( f_j \)'s payoff in an equilibrium without unraveling, since it is matched with worker \( w_{j+W-F} \). In an equilibrium with full unraveling, \( f_j \) is matched with its signal-suggested type \( \hat{r}(j) = j + W - F \). The RHS gives the expected payoff. The function \( \hat{j} \) is defined as the cutoff firm.\(^{16}\) Firms ranked lower than \( \hat{j} \) are better off (or indifferent) in full unraveling, while those ranked higher are worse off. Inequality (3) indicates that full unraveling tends to benefit medium or low quality firms while hurting high quality firms.

On the workers’ side, \( w_i \) is better off in full unraveling if

\[
i - W + F < \alpha (i - W + F) + (1 - \alpha) \left[ \frac{(W + 1)}{2} - W + F \right],
\] (4)

or equivalently,

\[
i < \hat{i} \equiv \frac{W + 1}{2}.
\] (5)

The LHS of (4) is \( w_i \)'s payoff in an equilibrium without unraveling, since it is matched with firm \( f_{i+W-F} \). In an equilibrium with full unraveling, \( w_i \) is matched with firm \( j(\hat{r}) = \hat{r} - W + F \), which depends on her type \( \hat{r} \) in Period 1. The RHS gives the expected payoff. The function \( \hat{i} \) is defined as the cutoff worker, which suggests that full unraveling tends to benefit low quality workers while hurting high quality workers.

Now we consider a banning policy on exploding offers. Suppose every participant holds the common belief that the market does not unravel if exploding offers are banned but fully unravels if they are allowed. Then clearly such a policy will be supported by high quality firms and workers, and opposed by firms of low or medium qualities and workers of low qualities. Similar conclusion is drawn from the experimental results in Niederle and Roth (2009). They find that while early matches are costly for the highest quality firms and workers, some lower quality firms and applicants tend to gain from them. The conclusion also explains the prevalence of exploding offers in practice and why it is sometimes difficult to achieve consensus on how to solve the unraveling problem in real markets (Niederle and Roth, 2005).

\(^{15}\)Firms \( f_1, f_2, \ldots, f_{F-W} \) are indifferent since they are unmatched in both cases.

\(^{16}\)There is an abuse of language here since \( \hat{j} \) is not necessarily an integer.
5 Conclusion

Many two-sided matching markets tend to unravel in time with transactions occurring earlier and earlier. Using a two-period decentralized model, this paper shows that when a market culture allows firms to make exploding offers, such unraveling is more likely to take place and lead to an unstable matching due to the lack of information in the early stages.

An excess supply of labor makes a market less vulnerable to the presence of exploding offers, while the effect of signal accuracy is ambiguous, depending on how uncertainty is resolved in a market. Therefore, although exploding offers in general tends to facilitate early transactions, it is only a necessary but not a sufficient condition for unraveling to occur. The policy regarding exploding offers should be tailored to the specific environment of interest, while taking into consideration such market characteristics as labor supply and demand and information structure. More importantly, the welfare analysis in this paper indicates that a ban towards exploding offers may benefit high quality firms and workers but hurt agents of low qualities. This suggests the need for policymakers to balance the costs and gains when addressing the unraveling problem in a market.

Admittedly, the model in this paper only provides a simple and tractable benchmark for the analysis of exploding offers. Further extensions into environments like heterogeneous preferences or asymmetric information may give us some additional insights regarding how the use of exploding offers can affect the outcome of a matching market.
Chapter 2

Targeted Advertising on Competing Platforms\textsuperscript{1}

1 Introduction

The Internet has revolutionized the advertising industry. One distinguishing feature of online advertising is that online platforms are able to provide customized advertisements to relevant consumers. In other words, online platforms have high targeting abilities. This is achieved mainly because online platforms are able to track consumers’ web browsing activities. For instance, social network websites, such as Facebook, can reasonably estimate the current interests of a consumer by tracing his activities on the network, and provide the ads of relevant products that might interest him. The high targeting ability of online advertising brings two potential benefits. First, for advertising firms, fewer advertising messages get lost, as ads are sent to more relevant consumers on average. Second, for consumers, who usually do not like irrelevant ads, on average they are less likely to encounter irrelevant ads. Probably for these reasons, the aggregate spending of advertising on the Internet has increased dramatically in recent years, while that of traditional media (TV, newspaper, radio) has declined steadily.

Two recent papers studied the impacts of targeted advertising on advertising markets. Bergemann and Bonatti (2011) model advertising markets as perfectly competitive markets, focusing mainly on the effects of changes in targeting ability on equilibrium ad prices. In their model, consumers are passive and do not incur any nuisance cost by viewing irrelevant ads. Moreover, platforms do not play any active role as the advertising markets are assumed to be perfectly competitive. In Johnson (2013), consumers incur nuisance costs by viewing irrelevant ads, and they might not participate if there are too many irrelevant ads. In

\textsuperscript{1}This chapter is coauthored with Huanxing Yang.
this setting, he investigates the impacts of increasing targeting ability on market outcomes. However, in his model there is no advertising price and platforms play no active role. In the real world, platforms, such as Google and Facebook, play an active role in advertising markets. They not only bring together consumers and advertising firms for potential match, but they also actively set ad prices and actively develop new technologies and methods to improve targeting ability.

The goal of this paper is to study targeted advertising in two-sided markets, with platforms playing an active role in identifying relevant consumers and setting prices. Moreover, platforms are competing with each other and have different targeting abilities. In particular, we ask the following questions. How does asymmetry in targeting ability translate into asymmetric behavior of two platforms? How do changes in targeting ability affect the prices of ads, the number of advertising firms, and the total volume of ads? Will consumers always benefit from increases in targeting ability? Will increases in targeting ability affect different types of firms differently? Do platforms invest too little or too much in targeting ability relative to the socially optimal level?

Specifically, based on Anderson and Coate (2005), we develop a model with two competing platforms acting as bridges between consumers and advertising firms. Consumers’ tastes about the two platforms’ contents are horizontally differentiated a la Hotelling. Advertising firms (simply firms sometimes) are heterogeneous in terms of the profitability of each product sold. While firms can be multi-homing, that is, they can participate on both platforms, consumers are single-homing, which means each consumer only participates on one platform. With some probability, a consumer is interested in (or relevant for) a firm’s product, or, in other words, there is a potential match within the consumer-firm pair. The role of advertising is to turn potential matches into actual purchases: a sale is realized if and only if there is a potential match within the consumer-firm pair and the consumer receives an advertisement from the firm. For each consumer-firm pair on a platform, the platform generates a binary and informative signal regarding whether there is a potential match. The accuracy of the signals indicates the targeting ability of a platform, and two platforms have different targeting abilities. Consumers incur nuisance costs in viewing any ads, but also derive benefits from viewing relevant ads. Overall, we assume that ads impose negative externalities on consumers. In terms of timing, first the two platforms simultaneously set ad prices per impression. Then firms decide whether to participate on each platform, and at the same time consumers decide which platform to join.

We show that there is a unique equilibrium. Our first set of main results illustrate how asymmetry in targeting ability translates into asymmetric behavior of two platforms. First,

\footnote{One can consider the one with a higher targeting ability as an online platform and the other one as an offline platform.}
compared to the platform with the lower targeting ability, in equilibrium the platform with the higher targeting ability always has more advertising firms, attracts more consumers, and is more profitable. However, the advantaged platform could charge a higher or a lower ad price, and has a higher or lower total number of ads, than the other platform. Note that these predictions are potentially testable (we can interpret the platform with a higher targeting ability as an online platform, and the other as an offline platform). Second, consumers on the advantaged platform (with the higher targeting ability) always incur a lower nuisance cost than those on the other platform; and any firm that advertises on both platforms earns a higher profit on the advantaged platform.

The driving force behind the above results are as follows. Given the negative externality imposed by advertising firms on consumers, each platform faces a tradeoff between the number of participating firms and the number of participating consumers. From a platform’s point of view, the number of participating firms and the number of participating consumers are complements. Thus naturally the platform with the higher targeting ability “spends” its advantage in accommodating more participating firms as well as attracting more consumers. The ambiguous conclusions on the price and total volume of ads result from two opposing effects: a mix effect (the advantaged platform having a lower proportion of irrelevant ads) and a volume effect (the advantaged platform having more advertising firms). Specifically, due to the mix effect, the advantaged platform tends to have a higher price and a smaller volume of ads than the disadvantaged one. But in the meantime, the volume effect works in the opposite direction: it leads to a lower ad price and a bigger ad volume on the advantaged platform. The overall effect can go either way. As to consumers’ net nuisance costs, the mix effect favors consumers on the advantaged platform, while the volume effect favors consumers on the other platform. However, the fact that consumers on the advantaged platform always incur a lower nuisance cost means that overall the mix effect dominates the volume effect in terms of consumer welfare.

Our second set of main results illustrate how an increase in the targeting ability of the advantaged platform affects the equilibrium outcome. First, the advantaged platform will have more advertising firms, attract more consumers, and become more profitable. The disadvantaged platform will have fewer advertising firms, attract fewer consumers, have fewer ads in total, increase its ad price, and become less profitable. However, the ad price of the advantaged platform could either increase or decrease, and the same pattern holds for its total volume of ads. Again, these predictions are potentially testable. In particular, the predictions regarding the disadvantaged platform are largely consistent with the anecdotal evidence mentioned earlier: offline media have been suffering from the competition with new Internet media. Second, all consumers, on either platform, are unambiguously better off. Moreover, some less profitable firms, who were excluded previously but start to advertise
on the advantaged platform after the change, are better off; all firms will be better off if their profitability follows a uniform or an exponential distribution.³

Most of the above results are intuitive. As the advantaged platform becomes even more advantaged, in order to protect its consumer base, the disadvantaged platform accommodates fewer firms. Again, since participating firms and consumer base are complements, the advantaged platform “spends” its additional advantage on both accommodating more firms and attracting more consumers. All the changes regarding the disadvantaged platform are monotonic, because its targeting ability remains the same but its number of participating firms decreases. As to the advantaged platform, underlying the non-monotonicity of the ad price and total number of ads are two countervailing effects: a mix effect and a volume effect. An increase in the targeting ability reduces the proportion of irrelevant ads. Such a mix effect tends to reduce the total number of ads, and thus the platform can now charge a higher ad price. But the volume effect, which refers to an increase in the number of advertising firms on the platform, tends to increase the total number of ads and reduce the ad price. The overall effect can go either way.

The mix effect and the volume effect also apply to the analysis of consumer welfare on the advantaged platform. The former tends to benefit consumers while the latter tends to hurt them. However, the mix effect always dominates so that consumers are always better off.⁴ This is because the consumers on the disadvantaged platform benefit from fewer advertising firms, which means that the consumers on the advantaged platform must benefit even more since otherwise the advantaged platform would have lost some consumers. This result is different from Johnson (2013), in which consumers might be worse off as the targeting ability of ads increases.

Finally, we compare social incentives and private incentives to invest in targeting abilities. In particular, two platforms are symmetric and they both make investment decisions in the first stage. It turns out that in equilibrium platforms could underinvest as well as overinvest in targeting ability. Relative to the social optimum, the fact that platforms do not care about consumers’ nuisance costs per se and cannot fully appropriate firm surplus imply that platforms tend to underinvest in targeting ability. On the other hand, the business stealing effect under private incentives implies that platforms tend to overinvest in targeting ability, since a higher targeting ability means a bigger consumer base and a bigger profit at the expense of the other platform. Quantitatively, underinvestment in targeting ability is much more likely to occur, while overinvestment occurs only under very special conditions. This result is somewhat surprising, as one would have thought that, as long as

---

³With these distributions, the combined total number of relevant ads on the two platforms increases as well.

⁴We also show that, when the targeting ability of the disadvantaged platform increases, all consumers are still better off.
the two platforms' contents are homogenous enough, the business stealing effect under the individual incentive will be magnified sufficiently to dominate and overinvestment will occur. However, because homogeneity in contents intensifies competition, in equilibrium two platforms will accommodate fewer firms and their profits are reduced. This will dampen platforms’ incentive to invest in targeting ability as they will get a lower profit any way. For overinvestment to occur, the nuisance cost of ads has to decrease correspondingly as the two platforms’ contents become more homogenous, in order to prevent the competition between two platforms from being too fierce.

This paper is related to the literature on informative advertising (Butters, 1977; Grossman and Shapiro, 1984). In this literature, recently there has been an increasing interest in the role of targeting ability. One line of research (Esteban et al., 2001; Iyer et al., 2005; Galeotti and Moraga-Gonzlez, 2008; and Gal-Or et al., 2010) focuses on how targeted advertising affects the competition and equilibrium prices in product markets, instead of analyzing the volume and price of advertising. Taylor (2011) not only studies the effect of targeting accuracy on competition in product markets, but also considers its impact on media’s choices regarding content differentiation. Athey and Gans (2010) studies how targeting affects the volume and price of advertising. In particular, they identify a supply-side impact of targeting: it allows a more efficient allocation of scarce advertising space, and the resulting increase in the supply of ads space might push down the price of advertising.\footnote{In a consumer search model, de Corniere (2013) studies how a search engine’s targeting ability of keywords search affects the fee of advertising. Yang (2013) develops a model of targeted search, analyzing how the quality of search affects the variety of goods offered, prices, and consumer welfare.}

As mentioned earlier, in the literature on targeted advertising this paper is most closely related to Bergemann and Bonatti (2011) and Johnson (2013), both focusing on the impacts of targeted advertising on the advertising markets. And we already pointed out earlier the main differences between our paper and their papers. In the second part of Bergemann and Bonatti (2011) they study competing platforms. But in their model consumers are multi-homing, so ads on two platforms are substitutes. Moreover, platforms do not play any active role as the equilibrium prices on both platforms are determined by demand and supply.\footnote{More differences in results between our paper and these two papers are discussed in the text later.}

To the best of our knowledge, this paper provides the first analysis of targeted advertising under the framework of competing two-sided platforms. Hence, it also contributes to the rapidly expanding literature on competition in two-sided markets (see Rochet and Tirole, 2003, 2006; and Armstrong, 2006). The most relevant model to our paper is the model of “competitive bottlenecks” (Armstrong, 2006), in which one side of the market is single-homing, while the other is multi-homing. The major insight of Armstrong (2006) is that, due to the competitive pressure on the single-homing side, platforms might be forced to
transfer some of the surplus towards this side, which leads to too few multi-homing agents in equilibrium.

In terms of modeling, our paper is closely related to Anderson and Coate (2005), who study advertising on competing (media) platforms. The difference is that in their model two platforms are symmetric and advertising cannot be targeted. In our model, platforms are able to target advertising to relevant consumers, and the two platforms are asymmetric in that they have different targeting abilities. In a media competition model, Peitz and Valletti (2008) introduce the possibility that consumers may pay for their access to content, and the results show that advertising intensity is higher in free-to-air television than in pay-tv stations. Ambrus and Reisinger (2006) extend Anderson and Coate (2005), by allowing consumers/viewers to be multi-homing, and then compare this to the original setting with single-homing consumers. Athey et al. (2013) extend Anderson and Coate (2005) to a setting in which consumers switch media outlets stochastically (multi-homing, but not fully), and study how the tracking technologies of platforms affect the equilibrium outcomes in the advertising market. None of these papers studies targeted advertising.

The rest of the paper is organized as follows. Section 2 sets up the model. In Section 3 we characterize the equilibrium outcome. Section 4 conducts comparative statics regarding changes in targeting ability. In Section 5 we endogenize the targeting abilities of platforms. Section 6 offers conclusions and discussions. All the proofs can be found in Appendix A.

2 The Model

There are two platforms, $A$ and $B$, who bring together consumers and advertising firms (simply firms sometimes). Consumers consume the content of a platform, and firms participate on platforms in order to send ads to consumers. Both consumers and firms are of unit mass.\footnote{It is not essential that consumers and firms are of the same measure.} The two platforms’ contents are horizontally differentiated. Using Hotelling’s location model, we assume that platforms $A$ and $B$ are located at 0 and 1, respectively. Consumers’ tastes about the platforms’ contents are also differentiated. Specifically, consumers are uniformly distributed on $[0, 1]$. Let $d$ be the location of a consumer. Consumers are single homing, meaning that a consumer will join only one platform. Firms are potentially multi-homing: each firm can join neither, one, or both platforms.

For any given consumer-firm pair, with probability $q \in (0, 1/2)$ there is a match. That is, the consumer is interested in the firm’s product, and we call such a consumer a relevant consumer, and an ad between such a pair a relevant ad. The probability that a match exists is i.i.d. across all consumer-firm pairs. A consumer will buy one unit of product from a firm if and only if the consumer is relevant to the firm and he receives an ad from that firm. So
in our model advertising is purely informative. Denote $S \in \{0, 1\}$ as the state indicating whether a consumer is relevant to a firm, with 1(0) denoting that the consumer is relevant (irrelevant). For any consumer-firm pair on platform $i$, the platform generates a signal $s \in \{0, 1\}$ regarding the possible match. The signals are also conditionally independent across consumer-firm pairs. The information structure is as follows:

$$\Pr\{s = 1|S = 1\} = \alpha_i; \quad \Pr\{s = 0|S = 0\} = \alpha_i.$$  

The parameter $\alpha_i \in (1/2, 1]$, which measures the accuracy of the signals, captures the targeting ability of platform $i$. We assume that $\alpha_A$ and $\alpha_B$ are common knowledge. The posteriors that a consumer is relevant to a firm can be calculated as follows:

$$\Pr\{S = 1|s = 1\} = \frac{\alpha_i q}{\alpha_i q + (1 - \alpha_i)(1 - q)};$$

$$\Pr\{S = 1|s = 0\} = \frac{(1 - \alpha_i)q}{(1 - \alpha_i)q + \alpha_i(1 - q)}.$$ 

We assume $q < 1/2$ is small enough such that $\Pr\{S = 1|s = 0\}$ is also small enough, so that it never pays for any firm to send ads to consumers when the signal is 0. For any firm, the set of consumers with signal 1 can be considered as that firm’s targeted set of consumers, to which the firm might send ads. Note that the size of the targeted set of consumers is $\alpha_i q + (1 - \alpha_i)(1 - q)$, which is increasing in $q$, and is decreasing in $\alpha_i$ as $q < 1/2$. Also note that, although the targeted sets of consumers are different for different firms, for each firm on the same platform the targeted set of consumers is of the same size, since the probability of being relevant is i.i.d. across consumer-firm pairs and the signals are conditionally independent. The two platforms are different in targeting abilities. Specifically, $\alpha_B > \alpha_A$. One can consider platform $A$ as a traditional offline media and platform $B$ as an online media. In case that both platforms are online media, the difference in targeting abilities could result from different technologies in tracing consumers.

Each consumer incurs a nuisance cost $\gamma > 0$ by viewing an ad. And each consumer gets a gross payoff $\lambda > 0$ if she buys a relevant product. With this specification, a consumer incurs a net nuisance cost of $\gamma$ by viewing an irrelevant ad, and a net nuisance cost of $\gamma - \lambda$ by viewing a relevant ad (could be net benefit if $\lambda > \gamma$). A location-$d$ consumer gets a utility of $\beta - td$ minus the net nuisance cost of ads if participating on platform $A$, and she gets a utility of $\beta - t(1 - d)$ minus the net nuisance cost of ads if participating on platform $B$. The parameter $\beta$ captures the gross utility of a consumer joining either platform by consuming the content provided by that platform. We assume $\beta$ is high enough so that all consumers

---

$^8$Johnson (2013) makes a similar assumption. $\lambda$ can be considered as the surplus enjoyed by a consumer engaging in trade.
participate. The parameter $t$ is the transportation cost in standard Hotelling models, which indicates the degree of horizontal differentiation between two platforms’ contents.

Firms are heterogeneous in terms of profitability. Denote $v$ as a firm’s profit per sale. Firms’ types $v$ are distributed on $[0, \bar{v}]$ with cumulative distribution function $F(v)$ and density function $f(v)$, with $f(\cdot) > 0$ everywhere in the support, differentiable, and strictly logconcave.

The timing is as follows. In the first stage, the two platforms set advertising prices (per impression), $p_A$ and $p_B$, simultaneously. In the second stage, observing $p_A$ and $p_B$, consumers simultaneously decide which platform to join, and firms, at the same time, decide whether to participate on each platform simultaneously. All agents have rational expectations.

Our model is based on Anderson and Coate (2005). The difference is that we add the aspect of advertisement targeting, and two platforms are asymmetric in that they have different targeting abilities. The model also resembles Armstrong’s (2006) model of “competitive bottlenecks,” in which platforms are actively competing for single-homing consumers, while there is no competition for multi-homing firms. The two-sided market only has an one-way externality: firms exert externalities on consumers by posting ads.

Several real-world examples largely fit the main assumptions of our model. In the market of social media, Facebook and Myspace compete for users and they begin to rely more heavily on advertising revenue. Moreover, users are typically single-homing. In the magazine market, Time and The Economist are competing for readers. In the local news markets, local newspapers and local TV stations are competing for viewers. In the latter two examples, some consumers might be multi-homing. But it does not qualitatively affect the results of our model as long as some consumers are single-homing (as if platforms only compete for the single-homing consumers). Even if all consumers are multi-homing, it is reasonable to think that each consumer’s attention allocation between two media is sensitive to the amount of irrelevant ads on them: other things equal, a consumer tends to spend more time on a medium with less irrelevant ads. This model of multi-homing consumers with attention allocation is quite similar to our model with single-homing consumers, as in both cases platforms are competing for consumer attention by restricting the amount of ads.

3 Equilibrium with Competing Platforms

Given an ad price $p_i$, $i = A, B$, a firm of type $v$ will get the following profit per ad if advertising on platform $i$: $\frac{\alpha_i q v}{\alpha_i q + (1-\alpha_i)(1-q)} - p_i$. Thus a firm of type $v$ will advertise on
platform \( i \) if and only if
\[
v \geq \frac{\alpha_i q + (1 - \alpha_i)(1 - q)}{\alpha_i q} \quad p_i \equiv \hat{v}_i.
\]
The term \( \hat{v}_i \) is the cutoff or marginal type of firms for platform \( i \): firms with types above \( \hat{v}_i \) will advertise on platform \( i \) and those with types below \( \hat{v}_i \) will not. Given that firms are multi-homing, a firm’s decision to join platform \( i \) is independent of its decision to join platform \( j \). Moreover, since firms pay prices per impression, a firm’s decision as to whether to join either platform does not depend on consumers’ platform choices. Let \( \mu_i \) be the fraction of firms that advertise on platform \( i \). Setting the ad price \( p_i \) is equivalent to choosing the cutoff firm type \( \hat{v}_i \) or the fraction of participating firms \( \mu_i \). In particular, \( \mu_i = 1 - F(\hat{v}_i) \). Setting the ad price \( p_i \) is equivalent to choosing the cutoff firm type \( \hat{v}_i \) or the fraction of participating firms \( \mu_i \). In particular, \( \mu_i = 1 - F(\hat{v}_i) \).

The relationship between \( p_i \) and \( \mu_i \) can be expressed as follows:
\[
p_i(\mu_i) = \frac{\alpha_i q}{\alpha_i q + (1 - \alpha_i)(1 - q)} \quad (1)
\]
Note that a higher price \( p_i \) leads to a higher cutoff type \( \hat{v}_i \) and a smaller fraction of participating firms \( \mu_i \).

Let \( m_i \) be the fraction of consumers joining platform \( i \). Sometimes we call \( m_i \) the consumer base of platform \( i \). Recall that consumers are single-homing. Because potential matches are i.i.d. and signals are conditionally i.i.d. across all consumer-firm pairs, all consumers joining platform \( i \) will receive the same number of ads. A consumer on platform \( i \) incurs the following expected net nuisance cost, \( y_i \), by viewing each ad:
\[
y_i = \gamma(1 - q)(1 - \alpha_i) - (\lambda - \gamma)\alpha_i q.
\]
And he incurs a total net nuisance cost of \( y_i \mu_i \). Define \( \hat{d} \) as the location of the consumer who is indifferent between joining the two platforms. In particular,
\[
\hat{d} = \frac{1}{2} + \frac{\gamma(1 - q)}{2t} [(1 - \alpha_B)\mu_B - (1 - \alpha_A)\mu_A] - \frac{(\lambda - \gamma)q}{2t} [\alpha_B\mu_B - \alpha_A\mu_A].
\]
Consumers with \( d \leq \hat{d} \) will join platform \( A \) and those with \( d \geq \hat{d} \) will join platform \( B \).

Thus, the consumer bases can be written as
\[
m_A(\mu_A, \mu_B) = \frac{1}{2} + \frac{\gamma(1 - q)}{2t} [(1 - \alpha_B)\mu_B - (1 - \alpha_A)\mu_A] - \frac{(\lambda - \gamma)q}{2t} [\alpha_B\mu_B - \alpha_A\mu_A], \quad (2)
\]
\[9\]For each firm on platform \( i \), with probability \( (1 - q)(1 - \alpha_i) \) \( (\alpha_i q) \) the consumer will be in that firm’s targeted set (signal is 1 for the consumer-firm pair) and the consumer is actually irrelevant (relevant) for the firm’s product.

\[10\]Recall that firms and consumers move simultaneously. But this does not matter as consumers have rational expectations: given the ad prices, consumers can predict how many firms will advertise on each platform and how many ads will be sent on each platform.
Let $x_i \equiv y_i/2t$. Then the consumer bases can be more compactly written as

$$m_i(\mu_i, \mu_j) = \frac{1}{2} + x_j \mu_j - x_i \mu_i.$$ 

Since consumers do not pay any participation fee to either platform, their choice of platform depends solely on the (net) nuisance costs of ads. The term $x_i$ captures how sensitive the consumer base is to changes in $\mu_i$: bigger $x_i$ and $x_j$ (a bigger net nuisance cost and a smaller transportation cost $t$) imply stronger competition for consumers. By $q < 1/2$ and $\alpha_A < \alpha_B$, it can be easily verified that $y_A > y_B$ and $x_A > x_B$. Intuitively, since platform $B$ has a higher targeting ability, a consumer on platform $B$ in expectation incurs a lower net nuisance cost from viewing an ad than a consumer on platform $A$.

**Condition 1.** $y_B = -(\lambda - \gamma)q\alpha_B + \gamma(1 - q)(1 - \alpha_B) > 0$.

For the rest of the paper, we assume that Condition 1 holds. Condition 1 says that the net expected nuisance cost of viewing an ad for a consumer on platform $B$ is positive. Since $\alpha_B > \alpha_A$, it implies that, $y_A$, the net expected nuisance cost of viewing an ad for a consumer on platform $A$, is positive as well. It further implies that $x_A$ and $x_B$ are both positive. Intuitively, Condition 1 means that ads overall impose negative externalities on consumers. Note that Condition 1 is always satisfied if $\lambda < \gamma$. If $\lambda > \gamma$, then it requires either $q$ be small enough or $\alpha_B$ be sufficiently below 1. We believe Condition 1 is a reasonable assumption, as in many real-world applications ads by large impose negative externalities on consumers. This is also consistent with what is commonly assumed in the literature (Anderson and Coate, 2005, for instance).\(^1\)

Platform $i$’s profit, $\Pi_i(\mu_i, \mu_j)$, can be computed as

$$\Pi_i(\mu_i, \mu_j) = [\alpha_i q + (1 - \alpha_i)(1 - q)]p_i(\mu_i)m_i(\mu_i, \mu_j)\mu_i.$$ 

The term $m_i \mu_i$ is the total number of consumer-firm pairs on platform $i$. Among those, a $\alpha_i q + (1 - \alpha_i)(1 - q)$ fraction of consumer-firm pairs have signal 1 and ads will be sent. Therefore, the total number of impressions is $\alpha_i q + (1 - \alpha_i)(1 - q)$ times $m_i \mu_i$. Multiplying by the ad price per impression $p_i$, we get the above expression of $\Pi_i$. An important observation from the above expression is that, for each platform, the fraction of advertising firms $\mu_i$ and its consumer base $m_i$ are complements, as $\Pi_i$ is proportional to $m_i \mu_i$.

---

\(^1\)At the end of Section 4, we will briefly discuss what will happen if Condition 1 is not satisfied.
Using (1) to get rid of \( p_i \), the profit function can be written as

\[
\Pi_i(\mu_i, \mu_j) = \alpha_i q F^{-1}(1 - \mu_i)m_i(\mu_i, \mu_j)\mu_i = \alpha_i q\mu_i\tilde{v}_i m_i(\mu_i, \mu_j). \tag{4}
\]

In (4), the term \( \alpha_i q\mu_i\tilde{v}_i \equiv \alpha_i qR(\mu_i) \) is platform \( i \)'s revenue per consumer. From (4), we can see that an increase in \( \alpha_i \) will directly increase platform \( i \)'s profit. Intuitively, an increase in targeting ability will enable platform \( i \) to identify more relevant consumers and rule out more irrelevant consumers, which enables the platform to charge a higher price without affecting the fraction of participating firms.

Denote an equilibrium as \( (\mu_A^*, \mu_B^*) \). Differentiating \( \Pi_i(\mu_i, \mu_j) \) with respect to \( \mu_i \), we get

\[
\frac{\partial \Pi_i}{\partial \mu_i} = \alpha_i q\{[\tilde{v}_i] - \frac{\partial q}{\partial \mu_i}\} \tag{5}
\]

According to (5), an increase in \( \mu_i \) affects \( \Pi_i \) through two channels. First, it affects the marginal revenue per consumer, which is captured by the term \( R'(\mu_i) = \tilde{v} \). Second, it will change platform \( i \)'s consumer base (the business stealing effect), which is captured by the second term in (5).

By the logconcavity of \( f \), the marginal revenue function \( R'(\mu) \) is strictly decreasing in \( \mu \).\(^{12}\)
Moreover, \( R'(\bar{\mu}) = 0 \) and \( R'(0) = 0 - \frac{1-F(0)}{F(0)} < 0 \). Thus there is a unique \( \mu \) \( \in (0, 1) \) such that \( R(\mu) = 0 \). Note that \( \mu \) uniquely maximizes revenue per consumer \( R(\mu) \), and a monopoly platform will choose \( \mu \).

**Lemma 1.** (i) \( \frac{\partial \Pi_i}{\partial \mu_i} < 0 \) if \( \mu \in (\bar{\mu}, 1] \). (ii) \( \mu_i^* \in (0, \bar{\mu}) \). (iii) For \( \mu_i \in [0, \bar{\mu}) \), \( \frac{\partial \mu_i}{\partial \mu_j} < 0 \) and \( \frac{\partial^2 \Pi_i}{\partial \mu_i \partial \mu_j} > 0 \).

By part (ii) of Lemma 1, in search for equilibrium we can restrict our attention to the domain \( [0, \bar{\mu}] \), which we will do in the subsequent analysis. The underlying reason is that the business stealing effect is always negative: an increase in \( \mu_i \) reduces the consumer base of platform \( i \). This means that to satisfy the first-order condition, the marginal revenue per consumer has to be positive. Part (iii) shows that, in the relevant domain the second-order condition is satisfied, which implies that the first-order condition is sufficient for the best response function \( \mu_i^*(\mu_j) \). The fact that \( \frac{\partial \mu_i}{\partial \mu_j} > 0 \) implies that \( \mu_A \) and \( \mu_B \) are strategic complements.

Following previous analysis, an equilibrium \( (\mu_A^*, \mu_B^*) \) is characterized by the following first-order conditions:

\[
x_B \mu_B^* = x_A \mu_A^* + \frac{x_A \mu_A^*}{\tilde{v}_A - \frac{1-F(\tilde{v}_A)}{F(\tilde{v}_A)}} - \frac{1}{2}, \tag{6}
\]

\[
x_A \mu_A^* = x_B \mu_B^* + \frac{x_B \mu_B^*}{\tilde{v}_B - \frac{1-F(\tilde{v}_B)}{F(\tilde{v}_B)}} - \frac{1}{2}. \tag{7}
\]

\(^{12}\)Since \( f \) is logconcave, \( \frac{1-F(v)}{f(v)} \) is strictly decreasing in \( v \) (see Bagnoli and Bergstrom, 2005, for details). Thus \( v - \frac{1-F(v)}{f(v)} \) is strictly increasing in \( v \), which implies that \( R'(\mu) \) is strictly decreasing in \( \mu \).
Lemma 2. There is a unique equilibrium.

After establishing the existence of a unique equilibrium, we compare the two platforms’ equilibrium behavior in the following proposition (how the asymmetry in targeting abilities translates into asymmetric outcomes).

Proposition 1. In the unique equilibrium the following properties hold. (i) Compared to platform \( A \), platform \( B \) has more participating firms, a bigger consumer base, and a higher equilibrium profit: \( \mu_B^* > \mu_A^* \), \( m_A(\mu_A^*, \mu_B^*) < 1/2 < m_B(\mu_A^*, \mu_B^*) \), and \( \Pi_A^* < \Pi_B^* \). (ii) Platform \( B \) charges a higher ad price (\( p_A^* < p_B^* \)) if
\[
\frac{1 + \frac{1 - q}{\alpha_B} \frac{1 - q}{\alpha_A}}{1 + \frac{1 - q}{\alpha_B} \frac{1 - q}{\alpha_A}} \geq \frac{y_A}{y_B}.
\]
(iii) Platform \( B \) has less ads in total if \( t \) is big enough.

The results in part (i) of Proposition 1 are intuitive. Since platform \( B \) has a higher targeting ability, each ad on platform \( B \) imposes less negative externality on consumers (\( y_B < y_A \) and hence \( x_B < x_A \)). This means the consumer base is less sensitive to \( \mu_B \) than to \( \mu_A \). As a result, in equilibrium platform \( B \) should have a lower marginal revenue per ad (closer to the monopoly solution in which the marginal revenue is zero), which implies more participating firms. However, in equilibrium, platform \( B \), who has a natural advantage in attracting more consumers, will not accommodate too many firms such that its consumer base falls below 1/2. This is because, from each platform’s point of view, the number of participating firms and the number of participating consumers are complements. Thus, in equilibrium platform \( B \) has a bigger consumer base. The fact that platform \( B \) has a higher profit is easy to understand: by mimicking platform \( A \)’s strategy, platform \( B \) can always guarantee a higher profit than platform \( A \)’s since it will have a bigger consumer base.

However, part (ii) of Proposition 1 indicates that it is not clear whether the platform with a higher targeting ability will charge a higher ad price. This is because there are two effects working in opposite ways. On the one hand, a higher targeting ability means that platform \( B \) can charge a higher ad price to attract the same cutoff type of firms (the mix effect). On the other hand, platform \( B \) has more advertising firms or a lower cutoff type of firms (\( \hat{v}_A^* > \hat{v}_B^* \)) (the volume effect), which tends to make the price charged by platform \( B \) relatively lower.\(^{13}\) The overall effect can go either way. In the proposition, we identified a sufficient condition for the mix effect to dominate. The LHS of the condition is the ratio of the signal to noise ratio on platform \( B \) to that on platform \( A \), which roughly measures the strength of the mix effect. The RHS of the condition is the ratio of the negative externality imposed on consumers by each ad on platform \( A \) to that on platform \( B \). This roughly measures the strength of the volume effect, as it cannot be surpassed by the ratio of the number of participating firms on platform \( B \) to that on platform \( A \).

\(^{13}\)In general, the volume effect is hard to quantify, as it depends on how sensitive \( \hat{v}_A^* \) and \( \hat{v}_B^* \) are to \( \alpha_A \) and \( \alpha_B \), which in turn depends on the distribution of firm types.
Thus, the condition specifies a scenario in which the mix effect dominates. Here we provide an example in which $p^*_A > p^*_B$ and the volume effect dominates. Suppose $v$ is uniformly distributed on $[0,1]$, $q = 0.3$, $t = 0.1$, $\gamma = 3$, $\lambda = 2.9$, $\alpha_A = 0.88$, and $\alpha_B = 0.98$. Then $0.6062 = p^*_A > p^*_B = 0.5440$.

For similar reasons, platform $B$ could have more total ads or fewer total ads than platform $A$. In particular, the volume effect (more participating firms on platform $B$) tends to increase the total number of ads on platform $B$ relative to platform $A$. However, the mix effect means that the proportion of irrelevant ads on platform $B$ is smaller, which tends to decrease the total number of ads on platform $B$ relative to platform $A$. The overall effect can go either way. Part (iii) of Proposition 1 identifies a condition under which platform $A$ has more ads in total.

The predictions of Proposition 1 are potentially empirically testable. Using the interpretation of an online platform ($B$) competing with an offline one ($A$), Proposition 1 implies that the online platform will have more advertising firms, attract more consumers, and is more profitable, relative to the offline platform. Moreover, relatively low-profit firms will only advertise on the online platform. However, the online platform could charge a higher or a lower ad price than the offline platform, and the total number of ads could be higher or lower on the online platform relative to the offline one.

The next proposition compares consumer welfare and firms’ profits across two platforms.

**Proposition 2.** In the unique equilibrium the following properties hold. (i) Compared to consumers on platform $A$, each consumer on platform $B$ incurs a lower total net nuisance cost. (ii) Any firm that advertises on both platforms earns a higher profit on platform $B$.

Part (i) of Proposition 2 is surprising. Note that the mix effect favors consumers on platform $B$, as its proportion of irrelevant ads is smaller. However, the volume effect favors consumers on platform $A$, since it has fewer participating firms. Nevertheless, Consumers on platform $B$ always incur a lower total net nuisance cost than those on platform $A$, which means that the mix effect dominates. Actually this result is directly implied by the fact that platform $B$ has a bigger equilibrium consumer base. If consumers on platform $B$ were incurring a higher total net nuisance cost than those on platform $A$, then platform $B$ would have a smaller consumer base than platform $A$. More fundamentally, this result is due to the fact, for each platform, the number of participating firms and the number

---

14 Platform $B$ has more relevant ads in total than platform $A$ because the former has more participating firms, a bigger consumer base, and a higher targeting ability, which means more relevant consumers are identified. Formally, the total relevant ads on platform $i$ is $\alpha_i q \hat{\mu}_i m^*_i$. Since $\alpha_A < \alpha_B$, $\hat{\mu}_A < \hat{\mu}_B$, and $m^*_A < m^*_B$, we have $\alpha_A q \hat{\mu}_A m^*_A < \alpha_B q \hat{\mu}_B m^*_B$.

15 In the limit, when $t$ goes to infinity, both platforms act like local monopolists and both accommodate the same number of firms, $\bar{\mu}$. In this limiting case, the volume effect disappears and the mixed effect implies that platform $A$ has more ads in total.
of participating consumers are complements. Thus, in equilibrium platform \( B \) controls the size of the volume effect such that it still has a bigger consumer base than platform \( A \).

Part (ii) of Proposition 2 is easy to understand. Since platform \( B \) has more participating firms and a lower marginal firm, each firm on platform \( B \) must earn a higher profit per consumer than it can earn on platform \( A \). Each participating firm’s total profit should also be higher on platform \( B \) as it has a bigger consumer base. To summarize, all parties on platform \( B \) (consumers, firms, and the platform itself) benefit from its higher targeting ability relative to the corresponding parties on platform \( A \).

Since firms’ ads impose negative externalities on consumers, an increase in the nuisance cost \( \gamma \), just like a decrease in the transportation cost \( t \), means the competition for consumers becomes fiercer. As a result, both platforms reduce ad volume per consumer by accommodating fewer firms. The profit of platform \( A \) decreases as \( \gamma \) increases. However, as \( \gamma \) increases the profit of platform \( B \) could either decrease or increase. This is because, although platform \( B \) suffers from intensified competition, the increase in \( \gamma \), in the mean time, amplifies its advantage in attracting consumers. The following example shows that the profit of platform \( B \) could increase in \( \gamma \). Suppose \( v \) is uniformly distributed on \([0, 1]\), \( q = 0.1 \), \( t = 0.3 \), \( \lambda = 0.5 \), \( \alpha_A = 0.6 \), and \( \alpha_B = 0.8 \). As \( \gamma \) increases from 0.2 to 0.6, \( m^*_B \) increases from 0.5329 to 0.5594, and \( \Pi^*_B \) increases from 0.0107 to 0.0111. However, as \( \gamma \) increases from 1.5 to 2, \( m^*_B \) decreases from 0.5557 to 0.5462, and \( \Pi^*_B \) decreases from 0.0104 to 0.0098.

4 Comparative Statics

In this section, we investigate how changes in platforms’ targeting abilities affect the equilibrium outcome. We begin with an increase in platform \( B \)’s targeting ability.

**Proposition 3.** Suppose \( \alpha'_B > \alpha_B \) and \( \alpha_A \) remains the same. Let the superscript ‘ denote the endogenous variables in the equilibrium under \((\alpha_A, \alpha'_B)\). Then, (i) the number of advertising firms on platform \( A \) decreases, and that on platform \( B \) increases: \( \mu^*_A < \mu^*_A \) and \( \mu^*_B > \mu^*_B \); (ii) platform \( A \)’s consumer base shrinks while Platform \( B \)’s consumer base expands: \( m^*_A < m^*_A \) and \( m^*_B > m^*_B \); (iii) platform \( A \)’s profit decreases while platform \( B \)’s profit increases: \( \Pi^*_A < \Pi^*_A \) and \( \Pi^*_B > \Pi^*_B \); (iv) platform \( A \)’s ad price increases, \( p^*_A < p^*_A \), while platform \( B \)’s ad price could either increase or decrease; in particular, \( p^*_B > p^*_B \) if \( \frac{\nu}{\gamma} \leq \frac{(1+\frac{1-\alpha_B}{\alpha_B})}{\gamma} \); (v) the total number of ads on platform \( A \) decreases; the total number of ads on platform \( B \) could increase or decrease; if the distribution of \( v \) is either uniform or exponential, then the total combined relevant ads on two platforms increase.

The intuition for Proposition 3 is as follows. As the targeting ability of the advantaged
platform \((B)\) increases, platform \(B\) becomes more advantaged. Thus, naturally platform \(B\)'s profit increases while that of platform \(A\) decreases. Since ad revenue per consumer and consumer base are complements, platform \(B\) will “spend” the additional advantage in increasing both the fraction of participating firms and its consumer base (which means that the increase in the number of participating firms is low relative to the increase in targeting ability). To protect its own consumer base, platform \(A\) responds by reducing the fraction of participating firms.

The result that platform \(A\) increases its ad price when \(\alpha_B\) increases seems somewhat counter-intuitive, as one would expect platform \(A\) to reduce its price when the other platform becomes more advantaged. However, in the current model platforms are not competing for firms, instead they are competing for consumers who suffer from negative externalities from firms’ ads. As \(\alpha_B\) increases, platform \(A\) responds by reducing its fraction of participating firms. This means the marginal advertising firm of platform \(A\) now has a higher profitability, which implies a higher ad price. The fact that the ad price of platform \(B\) is not monotonic in \(\alpha_B\) is again due to two opposite effects. In particular, according to the mix effect, a higher targeting ability enables platform \(B\) to charge a higher ad price for the same marginal firm. On the other hand, an increase in \(\alpha_B\) triggers a volume effect: platform \(B\) accommodates more firms, which implies a downward shift in the type of the marginal firm and a lower ad price. The overall effect can go either way. In the proposition, we identified a sufficient condition for the mix effect to dominate. Roughly speaking, the LHS measures to what extent an increase in \(\alpha_B\) improves the signal to noise ratio on platform \(B\), the size of the mix effect. Similarly, the RHS measures to what extent an increase in \(\alpha_B\) reduces the negative externalities imposed on consumers by each ad on platform \(B\), which provides an upper bound for the size of the volume effect.\(^{16}\) As shown in the following numerical example, with a relatively big \(\gamma\), \(p_B^*\) could be increasing in \(\alpha_B\) when \(\alpha_B\) is relatively small and be decreasing in \(\alpha_B\) when \(\alpha_B\) is close to 1. Suppose \(v\) is uniformly distributed on \([0, 1]\), \(q = 0.3\), \(t = 0.3\), \(\gamma = 6\), \(\lambda = 5.9\), and \(\alpha_A = 0.85\). As \(\alpha_B\) increases from 0.86 to 0.9, \(p_B^*\) increases from 0.5136 to 0.5213. As \(\alpha_B\) increases from 0.95 to 0.99, \(p_B^*\) decreases from 0.5170 to 0.5100.\(^{17}\)

Part (v) of Proposition 3 shows that, while the total number of ads on platform \(A\) decreases, the change in the total number of ads on platform \(B\) is ambiguous.\(^{18}\) The non-monotonicity of the total number of ads on platform \(B\) is again caused by the two

\(^{16}\)In two empirical studies, Chandra (2009) on newspapers and Chandra and Kaiser (2010) on magazines, the ad prices are found to be higher in markets with more homogeneous subscribers (more segmented or a higher targeting ability).

\(^{17}\)Bergemann and Bonatti (2011) also show that the equilibrium ad prices are non-monotonic in targeting ability. But their underlying mechanism is quite different from ours, as they focus on perfectly competitive ad markets while we focus on two-sided markets with platforms making strategic decisions.

\(^{18}\)Note that on platform \(A\) both the number of participating firms and the consumer base decrease.
countervailing effects. As $\alpha_B$ increases, the mix effect tends to reduce the number of irrelevant ads on platform $B$, while the volume effect tends to increase the number of ads. The following examples show that either effect could dominate: the total number of ads on platform $B$ is increasing in $\alpha_B$ when $\alpha_B$ is relatively small and is decreasing in $\alpha_B$ when $\alpha_B$ is close to 1. Suppose $v$ is uniformly distributed on $[0, 1]$, $q = 0.1$, $t = 0.3$, $\gamma = 4$, $\lambda = 3.9$, and $\alpha_A = 0.8$. As $\alpha_B$ increases from 0.82 to 0.9, the total number of ads on platform $B$ increases from 0.0350 to 0.0380. As $\alpha_B$ increases from 0.96 to 0.99, the total number of ads on platform $B$ decreases from 0.0383 to 0.0365.

As to the combined total number of relevant ads, there are two opposite effects as well. As $\alpha_B$ increases, the total number of relevant ads on platform $A$ decreases, while that on platform $B$ increases.\footnote{Since platform $B$ accommodates more firms and its consumer base increases, the total number of relevant ads on platform $B$ increases.} Moreover, platform $B$ gains market share, which tends to increase the combined total number of relevant ads. For general distributions of $v$, it is hard to compare the relative magnitudes of the first and the second effect, as how sensitive $\mu_A^*$ and $\mu_B^*$ are to $\alpha_B$ depends on the specific distribution. When the distribution is either uniform or exponential, we are able to show that the combined total number of relevant ads is increasing in $\alpha_B$.

Note that most of the predictions of Proposition 3 are potentially empirically testable. Again we use the interpretation of an online platform ($B$) competing with an offline one ($A$). Proposition 3 predicts that, as the targeting ability of the online platform increases, the online platform will have more advertising firms, attract more consumers, and become more profitable. The offline platform will have fewer advertising firms, attract fewer consumers, have fewer ads in total, increase its ad price, and become less profitable (these are largely consistent with the anecdotal evidence mentioned in the introduction). However, the ad price of the online platform could either increase or decrease, and the same pattern holds for the total volume of ads on the online platform.

**Proposition 4.** Suppose $\alpha_B' > \alpha_B$ and $\alpha_A$ remains the same. Let the superscript $'$ denote the endogenous variables in the equilibrium under $(\alpha_A, \alpha_B')$. Then, (i) each consumer incurs a lower total net nuisance cost; (ii) firms with $v \in (\hat{v}_B^*, \tilde{v})$ are strictly better off; if the distribution of $v$ is either uniform or exponential, then firms participating on both platforms $(v \geq \hat{v}_A^*)$ are also better off.

Part (i) of Proposition 4 indicates that each consumer, on either platform, is unambiguously better off from an increase in $\alpha_B$. Intuitively, as $\alpha_B$ increases, platform $A$ reduces ad volume per consumer, and thus consumers on platform $A$ are better off. Since platform $B$ gains more consumers, consumers remaining with platform $B$ must gain more than those on platform $A$, because otherwise platform $B$ would have lost consumers to platform $A$.\footnote{Since platform $B$ accommodates more firms and its consumer base increases, the total number of relevant ads on platform $B$ increases.}
This result is different from Johnson (2013), in which consumers might incur a higher total nuisance cost when the targeting ability of an implicit monopolist platform increases. In his model, an increase in targeting ability also has two effects: the mix effect benefiting consumers as the fraction of irrelevant ads decreases, and the volume effects hurting consumers as firms will send more ads. While in his model the overall effect is ambiguous, in our model the mix effect always dominates the volume effect (on platform $B$) so that overall consumers benefit from a higher targeting ability. The underlying reason for the difference is that in our model the two platforms are competing with each other for consumers, while in Johnson (2013) there is just a single platform (hence no competition). With competition, each platform has to worry about its consumer base. As a result, when a platform’s targeting ability increases, in order to gain more consumer base, it will restrict the increase in the number of participating firms so that the volume effect is smaller than the mix effect.

To summarize, an increase in $\alpha_B$ induces both the mix effect and the volume effect on platform $B$, which work in opposite directions. The overall effects on the ad price and the total number of ads on platform $B$ are ambiguous, depending on the distribution of firm types. However, the overall effect on consumer welfare is always positive. The fact that consumers are better off means that the volume effect cannot be relatively too large. But this restriction is not tight enough to ensure that the mix effect dominates the volume effect as to ad price or as to the total number of ads.

On the firms’ side, an increase in $\alpha_B$ induces some new firms, who initially did not participate on either platform, to participate on platform $B$. Those firms are clearly better off as now they have access to relevant consumers. As to firms who initially participated only on platform $B$, they are also better off with a bigger $\alpha_B$, since platform $B$ is now appealing to even lower types. Note that these two groups of firms are the relatively low-profit ones. This result can be considered as one manifestation of the “long tail of advertising”: the increase in the targeting ability of the advantaged platform enables less profitable firms, who were previously excluded, to have access to the advertising market and hence consumers. As to firms participating on both platforms, an increase in $\alpha_B$ makes them worse off on platform $A$ (paying a higher ad price), and better off on platform $B$. When the distribution of $v$ is either uniform or exponential, the gain on platform $B$ outweighs the loss on platform $A$, which makes those firms overall better off.

---

20 Another difference is that in Johnson (2013), the platform is passive in the sense that the per-impression ad price is exogenously given.

5 Investment in Targeting Ability

In this section, we investigate the platforms’ incentives as well as the social planner’s incentives to invest in targeting ability. For that purpose, we add an investment stage to the basic model, in which the targeting abilities of platforms are chosen. Specifically, suppose in the investment stage, each platform \( i, i = A, B \), simultaneously chooses its own targeting ability \( \alpha_i \). Two platforms have the same cost function \( C(\alpha_i) \), where \( C'(\cdot) \geq 0 \) and \( C''(\cdot) > 0 \). We further assume that \( C'(\tfrac{1}{2}) \) and \( C'(\overline{\alpha}) = \infty \), where \( \overline{\alpha} \) is defined as \( \overline{\alpha} = \min\{1, \tfrac{\gamma(1-q)}{(\lambda-\gamma)q+\gamma(1-q)}\} \). These conditions ensure that the resulting targeting abilities for both platforms will be interior: \( \alpha_i \in (1/2, \overline{\alpha}) \). At the end of the investment stage, the chosen \( \alpha_A \) and \( \alpha_B \) become publicly observable. Then the regular stage of the basic model begins. Note that the two platforms now become symmetric, which can be interpreted as two online platforms competing with each other.

Given \( \alpha_A \) and \( \alpha_B \), the equilibrium in the regular stage has been characterized in Section 3. Denote \( \mu_i^*(\alpha_A, \alpha_B) \) and \( \mu_B^*(\alpha_A, \alpha_B) \) as the equilibrium fractions of advertising firms. Now consider the choice of targeting ability in the investment stage. We will focus on symmetric equilibrium in which the two platforms choose the same targeting ability. Denote the (symmetric) equilibrium level of targeting ability as \( \alpha^* \). Suppose platform \( j \) chooses the equilibrium targeting ability: \( \alpha_j = \alpha^* \). Then platform \( i \)'s profit evaluated at the investment stage is given by

\[
\Pi_i(\alpha_i, \alpha^*) = \alpha_i q \hat{v}_h(\alpha_i, \alpha^*) \mu_i^*(\alpha_i, \alpha^*) \left\{ \frac{1}{2} + x(\alpha^*) \mu_j^*(\alpha_i, \alpha^*) - x(\alpha_i) \mu_i^*(\alpha_i, \alpha^*) \right\} - C(\alpha_i).
\]

Taking the derivative with respect to \( \alpha_i \), using the Envelope Theorem, and imposing symmetry \( \alpha_i^* = \alpha^* \), we get the following first order condition

\[
q \mu^* \hat{v}^* \times \left\{ \frac{1}{2} + x(\alpha^*) \alpha^* \frac{\partial \mu_i^*(\alpha^*, \alpha^*)}{\partial \alpha_i} + k \alpha^* \mu^* \right\} = C'(\alpha^*),
\]

where \( k = \frac{\partial x}{\partial \alpha} = \frac{1}{2\alpha} [\lambda q + \gamma (1 - 2q)] > 0 \), and the equilibrium \( \mu^* = \mu^*(\alpha^*, \alpha^*) \) is given by

\[
Z^* = \frac{\mu^* \hat{v}^*}{\hat{v}^* -} = \frac{1}{2x(\alpha^*)}.
\]

\( ^{22} \)In a two-sided market setting, Belleflamme and Peitz (2010) show that sellers’ incentive to invest depend on whether competing platforms charge access fees, whether sellers are single-homing or multi-homing, whether buyers are single-homing or multi-homing, and the nature of the investment (whether the investment brings direct benefit to buyers). Our setting is different in that we consider platforms’ incentive to invest in targeting ability.

\( ^{23} \)Similar to Condition 1, \( \alpha < \overline{\alpha} \) ensures that ads overall impose negative externalities on consumers.
The term $\frac{\partial \mu_j^*(\alpha^*, \alpha^*)}{\partial \alpha_i}$ can be calculated as follows:

$$\frac{\partial \mu_j^*(\alpha^*, \alpha^*)}{\partial \alpha_i} = k \frac{dZ^*}{d\mu^*} \mu^* - Z^* \frac{dZ^*}{d\mu^*} \mu^* \mu^*.$$  

$$dZ^* = \frac{(\mu^*)^2 + (\hat{v}^*)^2 + (\mu^*)^2 \hat{v}^* f'(\hat{v}^*)}{f(\hat{v}^*)\hat{v}^*}.$$  

From previous results, $\frac{\partial \mu_j^*(\alpha^*, \alpha^*)}{\partial \alpha_i} < 0$.

An individual platform’s incentive to invest in targeting ability, represented by the LHS of (8), can be separated into two effects. The first effect is the profit margin effect, captured by the first term in the braces. An increase in $\alpha$ means a platform can charge a higher price per impression without reducing the number of participating firms. The second effect is the business stealing effect, captured by the last two terms in the braces. An increase in $\alpha_i$ would increase platform $i$’s equilibrium consumer base in the regular stage. While the third term captures the direct impact of changes in $\alpha_i$ on the consumer base, the second term is the indirect impact through the changes in platform $j$’s ad volume.

Now consider the socially optimal level of targeting ability. Suppose a social planner chooses the targeting abilities for both platforms in the investment stage, and then lets the two platforms compete by choosing volumes of ads. The socially optimal solution must be symmetric: the two platforms have the same targeting ability. Given the equilibrium in the later stage $\mu^*(\alpha, \alpha)$ (or $\hat{v}^*(\alpha, \alpha)$), the social surplus generated by both platforms evaluated at the investment stage is given by

$$SS(\alpha) = \tilde{\beta} - \frac{1}{4} t - y(\alpha)\mu^*(\alpha, \alpha) + \alpha q \int_{\hat{v}^*(\alpha, \alpha)}^{\mathcal{V}} vf(v)dv - 2C(\alpha).$$

According to the above expression, the social surplus has five terms. The first term $\tilde{\beta}$ is consumers’ basic utility of joining platforms. The second term is the total transportation cost incurred by consumers. The third term is the total net nuisance cost suffered by consumers. The fourth term captures the firm surplus generated on two platforms, and the last term is the total investment cost. The socially optimal $\alpha^o$ solves

$$\frac{1}{2} [2tk\mu^* - y(\alpha^o) \frac{d\mu_i^*}{d\alpha}] +$$

24In an alternative setting, the social planner could choose both the targeting abilities in the first stage and the ad volumes in the second stage. In this setting, the social and equilibrium incentive to invest are further away from each other compared to the setting we are considering.

35
where
\[ \frac{d\mu^*}{d\alpha} = \frac{k}{2[x(\alpha^o)]^2 \frac{dZ^*}{d\mu^*}}, \quad Z^*(\alpha^o) \equiv \frac{\mu^*}{\tilde{v}^* - \frac{1-F(\tilde{v}^*)}{f(\tilde{v}^*)}} = \frac{1}{2x(\alpha^o)}. \]

The LHS of (10) represents the social incentive to invest in targeting ability. The first term (including the two terms in the brackets) is the consumer surplus effect, which measures how an increase in \( \alpha \) impacts consumer surplus. The first term in the brackets is the direct effect: an increase in \( \alpha \) reduces consumers’ chance of viewing irrelevant ads. The second term in the brackets is an indirect effect: as \( \alpha \) increases, both platforms will increase ad volume, which decreases consumer welfare. Overall, the consumer surplus effect is positive (this will be shown later). The second term is the direct effect on firm surplus: an increase in \( \alpha \) means that existing firms can be matched with more relevant consumers. The last term is the indirect effect on firm surplus: platforms with a higher \( \alpha \) will accommodate more firms, which also contributes to firm surplus. Note that the social incentive to invest does not have a business stealing effect. This is because the social planner internalizes the competition between the two platforms. Moreover, since the platforms are symmetric, there is no distribution effect either.

Now we compare the social incentive and the equilibrium incentive to invest in targeting ability. For that purpose, we assume that \( C'(\alpha) \) increases fast enough such that both \( \alpha^* \) and \( \alpha^o \) are unique. Define the LHS of (10) as \( H^*(\alpha) \) and the LHS of (10) as \( H^o(\alpha) \). Specifically,
\[ H^o(\alpha) - H^*(\alpha) = \frac{1}{2} \left[ 2tk\mu^* - y(\alpha) \frac{d\mu^*}{d\alpha} \right] + \frac{1}{2} q \left[ \int_{\tilde{v}^*}^{\mu^*} vf(v)dv - \mu^*\tilde{v}^* \right] + \alpha q\tilde{v}^* \left[ \frac{k}{4x^2 \frac{dZ^*}{d\mu^*}} - k\mu^* \frac{Z^* + \mu^*\frac{dZ^*}{d\mu^*}}{2 \frac{dZ^*}{d\mu^*}} + (\ldots) \right]. \]

According to (11), given \( \alpha \) the difference between the social incentive and equilibrium incentive can be attributed to three terms. The first term is the consumer surplus effect, which is positive and only present under the social incentive. This term implies that in equilibrium platforms tend to underinvest in targeting ability. Intuitively, this is because platforms do not care about consumer surplus per se. The second term is the difference between the direct firm surplus effect under the social incentive and the profit margin effect under the individual incentive. It can be readily seen that this term is positive. Intuitively, as the targeting ability increases, platforms cannot fully appropriate the increase in firm surplus, instead they can only get the increase in the marginal firm’s surplus. The third term is the difference between the indirect firm surplus effect under the social incentive and the business stealing effect under the individual incentive. The sign of this term is indeterminate. The following proposition describes situations under which platforms underinvest or overinvest.

**Proposition 5.** (i) If \( q\tilde{v} \leq 1/2 \) and \( t \geq 1/2 \), then \( \alpha^* < \alpha^o \). (ii) Suppose \( v \) is uniformly
distributed on $[0,1]$. There is an $\tilde{x}$ (defined in the proof) such that if $x(\cdot)$, then $\alpha^* < \alpha^o$. 

(iii) Suppose $v$ is uniformly distributed on $[0,1]$. Overinvestment ($\alpha^* > \alpha^o$) occurs only if $t < 1/2$ and $x(\alpha^*) < \tilde{x}$.

To understand part (i) of Proposition 5, note that when $t$ is big the competition between the two platforms is weak. This means that the business stealing effect under the individual incentive is weak. On the other hand, the consumer surplus effect under the social incentive becomes stronger as $t$ increases (the equilibrium $\mu^*$ increases and becomes less sensitive to changes in $\alpha$ as the competition between the two platforms becomes weaker). As a result, platforms underinvest in targeting ability in equilibrium if $t$ is big enough.

When firms’ profits are uniformly distributed, part (ii) of Proposition 5 provides a condition under which the indirect firm surplus effect under the social incentive dominates the business stealing effect under the individual incentive. Recall that a bigger $x$ implies stronger competition between two platforms. Thus a bigger $x$ implies a lower equilibrium volume of ads, $\mu^*$. This tends to reduce the business stealing effect, while the indirect firm surplus effect is more or less the same as the ad volume changes.\footnote{In particular, as the ad volume decreases, a given increase in $\alpha_i$ will only allow platform $i$ to steal fewer consumers from platform $j$.} Therefore, when $x$ is high enough the indirect firm surplus effect under the social incentive dominates the business stealing effect under the individual incentive, leading to underinvestment in targeting ability in equilibrium.

Overinvestment will occur only if the business stealing effect is stronger than other effects. Part (iii) of Proposition 5, which is directly implied by parts (i) and (ii), specifies the necessary conditions for overinvestment to occur: the transportation cost $t$ should be small, and the nuisance cost of ads $\gamma$ should be relatively small (and the marginal cost of investing in targeting ability is low enough) so that the the equilibrium $\mu^*$ is big enough.

In the following example overinvestment occurs. Suppose $q = 0.1$, $\gamma = 0.001$, $\lambda = 0.001$, $t = 0.00003$, and $C'(\alpha) = \max\{0, e^{5(\frac{\alpha}{10})} - 1\}$. Then $\alpha^* = 0.9774 > 0.9702 = \alpha^o$. In the numerical examples that we run, it is verified that overinvestment occurs in a very restrictive parameter space, while underinvestment occurs for most of the parameter space.\footnote{In the previous version, we also consider an alternative setting in which only platform $B$ chooses targeting ability in the investment stage while platform $A$ has a low and fixed targeting ability (more relevant for an offline platform competing with an online one). In this alternative setting we obtain similar qualitative results.}

The result that quantitatively it is hard for overinvestment to occur is somewhat surprising. One would have thought that, as long as the transportation cost $t$ is small enough, the business stealing effect under the individual incentive will be strong enough to dominate and overinvestment will occur. However, as mentioned earlier, if the nuisance cost $\gamma$ does not decrease correspondingly with $t$, stronger competition between two platforms will lead
to a small equilibrium $\mu^*$ and reduce the profits of platforms. This will dampen platforms' incentive to invest in targeting ability as they will get a lower profit any way.

6 Conclusion and Discussion

This paper studies targeted advertising in two-sided markets. Two platforms, with different targeting abilities, compete for single-homing consumers, while advertising firms are multi-homing. Our first set of main results illustrate how asymmetry in targeting ability translates into asymmetric equilibrium outcome, including ad price, total number of ads, consumer welfare, and advertising firms’ profits. In particular, our results generate the following potentially testable implications. Compared to the platform with the lower targeting ability, in equilibrium the platform with the higher targeting ability always has more advertising firms, attracts more consumers, and is more profitable. However, the advantaged platform could charge a higher or a lower ad price, and has a higher or lower total number of ads, than the other platform.

Our second set of main results illustrate how an increase in the targeting ability of the advantaged platform affects the equilibrium outcome. First, the disadvantaged platform increases its ad price, and its total number of ads decreases; the ad price charged by the advantaged platform and its total number of ads could increase or decrease. Again, these predictions are potentially testable. Second, all consumers, on either platform, are unambiguously better off. Moreover, low-profit advertising firms are better off, and all firms are better off if the profitability of firms follows a uniform or an exponential distribution. When the targeting ability of the disadvantaged platform increases, all consumers are again unambiguously better off, but low-profit firms are worse off.

Finally, we compare social incentives and private incentives to invest in targeting abilities. It turns out that in equilibrium platforms could underinvest as well as overinvest in targeting ability. However, quantitatively underinvestment is much more likely to occur, while overinvestment occurs only under very special conditions.

In an earlier version of the paper, we also study the situation in which two platforms are owned by a single monopoly. Like Anderson and Coate (2005), we found that, compared to the equilibrium of competing platforms, under monopoly ownership each platform has more participating firms, and charges lower ad prices. This is because the monopoly owner internalizes the competition between the two platforms. We also found the following allocation results that are absent in Anderson and Coate, as they focus on symmetric platforms. Compared to the equilibrium of competing platforms, under monopoly ownership the advantaged (disadvantaged) platform has a bigger (smaller) consumer base. Moreover, under monopoly ownership, the disadvantaged platform accommodates more firms, and charges a
lower price than the advantaged platform does, a pattern opposite to that under competing platforms. The underlying reason for the above results is that, with competition internalized, the monopoly owner tries to steer more consumers to the advantaged platform, as each consumer on that platform generates more revenue. In order to achieve that, it intentionally increases the ads per consumer on the disadvantaged platform by much more. Since under monopoly ownership consumer allocation is more skewed toward the more efficient platform, it implies that monopoly ownership could lead to a higher social welfare than competition.

In an early version we also compare the advertising levels and consumer allocation under social optimum with those under equilibrium. Similar to Anderson and Coate (2005), in equilibrium the platforms could under-provide or over-provide ads. We focus on the allocation results that are absent in Anderson and Coate’s model of symmetric platforms. While usually both platforms under-provide or over-provide ads at the same time, interestingly it could sometimes be the case that one platform over-provides while the other one under-provides ads. Moreover, in equilibrium the advantaged platform could have more or fewer consumers relative to the social optimum. One surprising result is that, under some conditions, the social planner could even allocate more consumers to the disadvantaged platform than the advantaged one. The intuition hinges on the analysis of two effects under social optimum: a distribution effect and a surplus effect. Roughly speaking, the distribution effect tends to make the socially optimal consumer base of platform $A$ smaller relative to the equilibrium level. This is because platform $B$ is more efficient, and thus steering some marginal consumers from platform $A$ to platform $B$ will increase social welfare. However, the only way to steer consumers from platform $A$ to platform $B$ is to increase the number of ads on platform $A$, or decrease the number of ads on platform $B$, which tends to reduce the surplus generated on both platforms (the surplus effect). Thus, the surplus effect tends to make the optimal consumer base of platform $A$ bigger relative to the equilibrium level. Overall, whether the equilibrium consumer base of platform $A$ is bigger or smaller than the social optimum depends on which effect dominates.
1 Introduction

As one of the most widely discussed policies in education, school choice programs aim to assign students to public schools with consideration for families’ true preferences. The centralized allocation procedure is usually based on students’ submitted rankings of schools and schools’ priority orderings over students. In the existing literature, a standard assumption is that students have complete knowledge about their own preferences, or in other words, such knowledge is completely free to them.

In practice, however, it can be rather costly for students to learn about their preferences over sometimes hundreds of candidate schools. For example, before submitting their preferences to the program, students or their parents often spend a lot of time and money visiting campuses, reading school profiles, attending school choice fairs, to mention just a few. Behavioral evidence shows that compared to the rational behaviors predicted by economic theory, people tend to acquire too much costly information or pay too much for information.\footnote{The evidence of excess information acquisition is found in various environments including public school choice (Chen and He; 2015), voting (Bhattacharya et al.; 2015), and cognitive operations (Gabaix et al.; 2006).} In the scenario of school choice, such excess investment can negatively affect student welfare and the efficiency of a society. Therefore, this study aims to investigate whether certain designs of the school allocation procedure can help to reduce wasteful information acquisition and to improve efficiency.

One common practice of many school choice programs in the field is to constrain the length of students’ submitted preference lists. For instance, out of approximately 500 schools
in New York City, students cannot submit more than 12 choices. In New Orleans and Chicago, students are not allowed to include more than 8 schools and 4 schools in their preference lists, respectively. Similar policies also exist in Ghana, England, etc. It has been argued in the literature that compared to an unconstrained program, limiting the number of choices may force students to strategically misrepresent their preferences and lead to less efficient market outcomes.\(^2\) However, from a behavioral and an informational perspective, the constraint may also reduce wasteful information acquisition by encouraging students and their parents to focus their research on a smaller set of schools: the schools that are more relevant to them in terms of location, availability, etc.\(^3\) Hence, the overall effect of such a policy on student welfare and efficiency can be ambiguous once we take costly information acquisition into account. This paper tries to shed some light on the above issue.

An experimental approach is used because students’ true preferences are rather difficult to observe or elicit in the field, while in the lab, they can be easily induced with monetary incentives.

In an environment where students have incomplete information about others’ preferences, I theoretically study the effect of such a constraint under both a Deferred Acceptance mechanism (DA) and a Boston mechanism (BOS). The result shows that ex-ante stability can only be ensured under an unconstrained DA, but not under a constrained DA, an unconstrained BOS, or a constrained BOS. In a lab experiment, I find that the constraint also affects students’ information acquisition behavior. Specifically, when faced with a constraint, students tend to acquire less wasteful information and distribute more efforts to acquire relevant information under DA; such an effect is not significant under BOS. Overall, the constraint has a negative effect on efficiency and stability under both mechanisms.

## 2 The Model

### 2.1 A School Choice Problem

A school choice problem is defined as follows. There are a finite set of students \( I = \{i_1, \ldots, i_n\} \) and a finite set of schools \( S = \{s_1, \ldots, s_m\} \). A student cannot be assigned to more than one schools; schools’ capacity vector is denoted as \( q = (q_1, \ldots, q_m) \).

At each school \( s \in S \), there is a strict priority ordering of students \( \succ_s \), where \( \{i\} \succ_s \{i'\} \) means that student \( i \) has a higher priority than student \( i' \) at school \( s \). Denote the profile of

---

\(^2\)See Haeringer and Klijn (2009), Calsamiglia et al. (2010), and Klijn et al. (2013).

\(^3\)Students and their parents usually have some prior knowledge about schools’ reputation and priority orderings. Here I focus on the acquisition of refined knowledge (see the tiered structure in the experimental design).
schools’ priority orderings as \( \succeq = (\succeq_1, \ldots, \succeq_m) \). In the environment of school choice, schools’ priorities are often set by law and are common knowledge to the entire market.

For each student \( i \in I \), the preference relation \( P_i \) is a strict linear order over \( S \cup i \), where \( i \) denotes remaining unmatched. Thus, \( \{s\} P_i \{s'\} \) means that student \( i \) strictly prefers school \( s \) to school \( s' \) and \( \{s\} P_i \{i\} \) means that school \( s \) is acceptable to \( i \). Associate preferences with vNM utilities: when student \( i \) is assigned to school \( s \), she yields a utility value of \( v_i^s \).

Let \( P = (P_1, \ldots, P_n) \), \( V_i = (v_i^1, \ldots, v_i^n) \) and \( V = (V_1, \ldots, V_n) \).

To specify the information environment of the model, below I first define the extension of an irreflexive binary relation.

**Definition 1.** Let \( B \) be an irreflexive binary relation on a set \( X \). Formally, \( B \) a subset of \( X \times X \); we write \( (x, y) \in B \) as \( x By \). An irreflexive binary relation \( B' \) is an extension of \( B \) if \( B \subseteq B' \); it is a strict extension of \( B \) if it is an extension, and in addition there is a pair that is unordered by \( B \) but ordered by \( B' \).

For each student \( i \in I \), \( R_i \) and \( T_i \) are two strict partial orders over \( S \cup i \).\(^4\) Let \( i \)'s preference relation \( P_i \) be an extension of \( T_i \) and \( T_i \) be an extension of \( R_i \), that is, \( R_i \subseteq T_i \subseteq P_i \). The vector \( R = (R_1, \ldots, R_n) \) is common knowledge to the entire market. Let \( f_i \) be a probability distribution over the set of all linear orders that extends \( R_i \): \( E_i = \{B_i : R_i \subseteq B_i \text{ and } B_i \text{ is a linear order}\} \), which is also common knowledge. \( T_i \) is the private information of student \( i \); denote the vector of private information as \( T = (T_1, \ldots, T_n) \).

The outcome of a matching market is known as a matching. Formally, a matching is a function \( \mu : I \rightarrow S \cup I \) such that (i) for all \( i \in I \), if \( \mu(i) \notin S \), then \( \mu(i) = i \), and (ii) for all \( s \in S \), it must be that \( |\{i : \mu(i) = s\}| \leq q_s \). Thus, \( \mu(i) \) denotes the assignment of student \( i \) under matching \( \mu \).

In the matching literature, stability is used as a central criterion to evaluate a matching outcome. Under the current structure, we can either set such a criterion according to students’ true preferences, or according to their private information about their preferences. They are called ex-post stability (or stability with respect to \( P \)) and ex-ante stability (or stability with respect to \( T \)), respectively. The definitions are given below.

**Definition 2.** A student-school pair \( (i, s) \in I \times S \) blocks matching \( \mu \) with respect to \( P \) if \( sP_i \mu(i) \) and one or more of the following is true: (i) there is some \( i' \) with \( \mu(i') = s \) and \( i \succeq_s i' \), or (ii) \( |\{i : \mu(i) = s\}| < q_s \).

A matching \( \mu \) is ex-post stable (or stable with respect to \( P \)) if there is no blocking pair.

**Definition 3.** Similarly, a pair \( (i, s) \in I \times S \) blocks matching \( \mu \) with respect to \( T \) if \( sT_i \mu(i) \) and one or more of the following is true: (i) there is some \( i' \) with \( \mu(i') = s \) and \( i \succeq_s i' \), or (ii) \( |\{i : \mu(i) = s\}| < q_s \).

\(^4\)A strict partial order is a binary relation that is irreflexive, transitive, and antisymmetric.
A matching $\mu$ is *ex-ante stable* (or *stable with respect to $T$*) if there is no pair that blocks the matching with respect to $T$.

Below I use a simple example to illustrate the environment.

**Example 1.** There are four schools $S = \{s_1, s_2, s_3, s_4\}$ and two students $I = \{i_1, i_2\}$ on the market. For each student $i \in I$, it is common knowledge that $\{s_1, s_2\} R_i \{s_3, s_4\} R_i \{i\}$, and that $f_i$ assigns probability $\frac{1}{4}$ to each of the following strict linear orders that extend $R_i$:

\[
\begin{align*}
\{s_1\} P_i \{s_2\} P_i \{s_3\} P_i \{s_4\} P_i \{i\}; \\
\{s_2\} P_i \{s_1\} P_i \{s_3\} P_i \{s_4\} P_i \{i\}; \\
\{s_1\} P_i \{s_2\} P_i \{s_4\} P_i \{s_3\} P_i \{i\};
\end{align*}
\]

and

\[
\{s_2\} P_i \{s_1\} P_i \{s_4\} P_i \{s_3\} P_i \{i\}.
\]

The private information of student $i_1$ about her own preferences is given by

\[
\{s_1, s_2\} T_1 \{s_3\} T_1 \{s_4\} T_1 \{i_1\},
\]

and that of student $i_2$ is

\[
\{s_1\} T_2 \{s_2\} T_2 \{s_3, s_4\} T_1 \{i_2\}.
\]

Their preference relations are given by

\[
\begin{align*}
\{s_1\} P_1 \{s_2\} P_1 \{s_3\} P_1 \{s_4\} P_1 \{i_1\}
\end{align*}
\]

and

\[
\begin{align*}
\{s_1\} P_2 \{s_2\} P_2 \{s_4\} P_2 \{s_3\} P_2 \{i_2\},
\end{align*}
\]

respectively.

The above information setting can be interpreted as a “tiered” structure, where schools $s_1$ and $s_2$ belong to the first tier while $s_3$ and $s_4$ belong to the second tier. It is common knowledge that each student prefers any school to remaining unmatched and prefers a first-tier school to a second-tier school; but between the two schools in the same tier, she prefers either one with equal probability. Each student has some private information, but not full knowledge on her own within-tier preferences. For instance, student $i_1$ only knows how to compare the two second-tier schools but not those from the first tier.
2.2 Two Matching Algorithms

To select a matching for the school choice problem defined above, a systematic procedure, called a “matching algorithm,” allocates students to schools depending on students’ submitted preferences and schools’ priority ordering. Denote the profile of students’ submitted preferences as $\hat{P} = (\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n)$.

2.2.1 The Boston Algorithm (BOS)

The Boston mechanism was first described in the literature by Alcalde who called it the “Now-or-Never” mechanism. The term “Boston mechanism” was coined by Abdulkadiroğlu and Sönmez because the mechanism was used in the Boston school district until recently. Given students’ submitted preferences and schools’ priority orderings, the Boston algorithm finds a matching through the following steps.

**Step 1:** Only the 1st choices of all students are considered. For each school, consider the students who have listed it as their 1st choice; assign seats of the school to these students one at a time following their priority ordering until either there are no seats left or there are no students left who have listed it as their 1st choice.

In general, Step $k \ (k \geq 1)$ can be described as follows.

**Step $k$:** Only the $k$th choices of the remaining students (who have not been assigned a seat previously) are considered. For each school with still available seats, consider the students who have listed it as their $k$th choice; assign the remaining seats to these students one at a time following their priority ordering until either there are no seats left or there are no students left who have listed it as their $k$th choice.

The procedure terminates after any step $k$ when every student is assigned a seat at some school, or if the only students who remain unassigned listed no more than $k$ choices.

2.2.2 The Deferred Acceptance Algorithm (DA)

The DA algorithm was introduced by Gale and Shapley. Given students’ submitted preferences and schools’ priority orderings, the Deferred Acceptance algorithm finds a matching through the following steps.

**Step 1:** Each student $i$ proposes to the school that is ranked first in $Q_i$ (if there is no such school then $i$ remains unassigned). Each school $s$ tentatively assigns up to $q_s$ seats to its proposers one at a time following the priority order $f_s$. Remaining students are rejected.
In general, Step $k$ ($k \geq 1$) can be described as follows.

**Step $k$:** Each student $i$ that is rejected in Step $l - 1$ proposes to the next school in the ordered list $Q_i$ (if there is no such school then $i$ remains unassigned). Each school $s$ considers the new proposers and the students that have a (tentative) seat at $s$. School $s$ tentatively assigns up to $q_s$ seats to these students one at a time following the priority order $f_s$. Remaining students are rejected.

The algorithm stops when no student is rejected. Each student is assigned to his final tentative school.

### 2.3 Constrained Preference Submission

### 2.4 Theoretical Analysis

A student-school pair $(i, s) \in I \times S$ blocks matching $\mu$ if $sP_i \mu(i)$ and one or more of the following is true: (i) there is some $i'$ with $\mu(i') = s$ and $i \succ_s i'$, or (ii) $|\{i : \mu(i) = s\}| < q_s$.

A matching $\mu$ is **stable** if there is no blocking pair.

Similarly, a pair $(i, s) \in I \times S$ blocks matching $\mu$ with respect to $T$ if $sT_i \mu(i)$ and one or more of the following is true: (i) there is some $i'$ with $\mu(i') = s$ and $i \succ_s i'$, or (ii) $|\{i : \mu(i) = s\}| < q_s$. A matching $\mu$ is **stable with respect to $T$** if there is no pair that blocks the matching with respect to $T$.

**Proposition 1.** Under DA with unconstrained preference submission, a student $i$ can do no better than submitting a strict linear order that extends $T_i$.

*Proof.* Consider a student who has the information that she prefers school $s$ to $s'$, that is, $sT_i s'$. If the strategy of ranking $s'$ higher than $s$ is not dominated, then it does better for at least some realization of her preferences. This contradicts the fact that in every realization, $s$ is preferred to $s'$ and that truth-telling is the dominant strategy (Roth; 1989).

**Proposition 2.** Under DA with unconstrained preference submission, the outcomes of Bayesian Nash Equilibrium in weakly undominated strategies are stable with respect to $T$.

*Proof.* Without loss of generality, let $\hat{P} = \left(\hat{P}_1, \hat{P}_2, ..., \hat{P}_n\right)$ be the submitted preferences in equilibrium, where $\hat{P}_i$ is a strict linear order that extends $T_i$, $\forall i \in I$. Let $\mu$ be the matching produced by DA given $\hat{P}$. Suppose, contrary to the proposition, $\mu$ is unstable with respect to $T$. Then there exists a blocking pair $(i, s)$, in which $sT_i \mu(i)$ and one or more of the following is true: (i) there is some $i'$ with $\mu(i') = s$ such that $i \succ_s i'$, or (ii) $|\{i : \mu(i) = s\}| < q_s$.

But $\hat{P}_i$, as an undominated strategy, lists $s$ before $\mu(i)$, so $i$ proposed to $s$ at some step in the DA algorithm, was rejected, and went on to propose to $\mu(i)$. This contradicts the fact
that \((i, s)\) is a blocking pair, because for case (i) \(s\) would not reject \(i\) and subsequently accept a student with a lower priority \(i'\), and for case (ii), \(s\) would not reject \(i\) and subsequently remain unmatched.

\[ \text{Proposition 3.} \quad \text{Under DA with preference submission constrained by a quota } k, \]
\[ \quad (i) \text{ if a student } i \text{ finds at most } k \text{ schools acceptable, then she can do no better than submitting a strict linear order that extends } T_i. \]
\[ \quad (ii) \text{ if a student } i \text{ finds more than } k \text{ schools acceptable, then he can do no better than selecting } k \text{ schools and ranking them according to a strict linear order that extends } T_i. \]

\[ \text{Proof.} \quad \text{Consider a student who has the information that she prefers school } s \text{ to } s'. \text{ If the strategy of ranking } s' \text{ higher than } s \text{ is not dominated, then it does better for at least some realization of her preferences or others' preferences. This contradicts the fact that in every realization, } s \text{ is preferred to } s' \text{ and misrepresentation is a dominated strategy (Haeringer and Klijn; 2009).} \]

\[ \text{Proposition 4.} \quad \text{Under DA with preference submission constrained by a quota } k, \text{ the outcomes of Bayesian Nash Equilibrium in weakly undominated strategies are not necessarily stable with respect to } T. \]

\[ \text{Proof.} \quad \text{Counter Examples.} \]

\[ \text{Proposition 5.} \quad \text{Under BOS with unconstrained preference submission or with preference submission constrained by a quota } k, \text{ the outcomes of Bayesian Nash Equilibrium in weakly undominated strategies are not necessarily stable with respect to } T. \]

\[ \text{Proof.} \quad \text{Counter Examples.} \]

3 Experimental Design

3.1 An Experimental Market

The experiment is designed to compare subjects' strategies in information acquisition and preference submission under different mechanisms. An experimental market consists of six students and six schools; each school has only one seat. Schools have priority orderings over students, and students incur costs when learning about their preferences over schools. The primary variable of interest is students' average payoff, which is a measure of market efficiency. Fairness is another important concern for policymakers. In the school choice environment, it is usually evaluated by the elimination of justified envy. A student is said to have justified envy if she prefers a school to her assignment, and the school gives her higher priority than one of the students assigned to it.
The four treatments differ in two dimensions: (i) which allocation algorithm is adopted, and (ii) whether a constraint is imposed on the length of students’ submitted preference lists (Table 1). According to Haeringer and Klijn (2009), Calsamiglia et al. (2010), and Klijn et al. (2013), the constraint may have different effects on the performances of two algorithms named “Deferred Acceptance (DA henceforth)” and “Boston (BOS henceforth).” Following the existing literature, DA and BOS also serve as the candidate algorithms in this paper, where costly information acquisition is added into the discussion.5 Students can submit all six schools in an unconstrained treatment, but no more than two schools in a constrained treatment. All treatments are between subjects, that is, each subject only participates in one treatment.

<table>
<thead>
<tr>
<th></th>
<th>Deferred Acceptance</th>
<th>Boston</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconstrained</td>
<td>Unconstrained-DA</td>
<td>Unconstrained-BOS</td>
</tr>
<tr>
<td>Constrained</td>
<td>Constrained-DA</td>
<td>Constrained-BOS</td>
</tr>
</tbody>
</table>

Table 1: Treatments

The six students are numbered from 1 to 6, and the six schools are labeled as A1, A2, B1, B2, C1, and C2. Table 2 gives schools’ priority orderings over students, which are provided to students as common knowledge. Schools are divided into three tiers: A1 and A2, B1 and B2, and C1 and C2 belong to tiers named A, B, and C, respectively. It is common knowledge that every student prefers schools in Tier A to those in Tier B to those in Tier C, but that a student’s preferences within each tier follows a uniform distribution. For example, between the two schools in Tier A, a student prefers A1 or A2 with equal probability. Preferences are independently distributed for every tier and every student. A student receives a payoff $18, $15, $12, $9, $6, $3 when assigned a seat at the school ranked 1st, 2nd, ..., 5th, and 6th according to her true preferences, respectively. An unmatched student receives zero payoff.

5As two most widely discussed algorithms in the matching literature, DA and BOS have rather different properties in terms of efficiency and the elimination of justified envy. In a standard environment, DA is usually considered more favorable since it produces an allocation that eliminates justified envy and offers the highest student welfare. However, BOS may outperform DA under some settings with uncertainty or private information. See Abdulkadiroglu and Sonmez (2003), Abdulkadiroglu et al. (2009), etc.
A BDM mechanism is used to elicit a student’s willingness to pay for the information on her preference between two schools within each tier.

### 3.2 Theoretical Predictions

#### 3.2.1 Unconstrained-BOS

The WTP for the information on each tier is given below

<table>
<thead>
<tr>
<th>Student</th>
<th>WTP&lt;sub&gt;A&lt;/sub&gt;</th>
<th>WTP&lt;sub&gt;B&lt;/sub&gt;</th>
<th>WTP&lt;sub&gt;C&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>8.2</td>
<td>53.7</td>
</tr>
<tr>
<td>3</td>
<td>8.7</td>
<td>147.1</td>
<td>8.5</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>12.4</td>
<td>3.8</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>78.3</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### 3.2.2 Constrained-BOS

The WTP for the information on each tier is given below

<table>
<thead>
<tr>
<th>Student</th>
<th>WTP&lt;sub&gt;A&lt;/sub&gt;</th>
<th>WTP&lt;sub&gt;B&lt;/sub&gt;</th>
<th>WTP&lt;sub&gt;C&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>8.2</td>
<td>53.7</td>
</tr>
<tr>
<td>3</td>
<td>8.7</td>
<td>147.1</td>
<td>8.5</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>12.4</td>
<td>3.8</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>78.3</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
3.2.3 Unconstrained-DA

The WTP for the information on each tier is given below

<table>
<thead>
<tr>
<th>Student</th>
<th>WTP\textsubscript{A}</th>
<th>WTP\textsubscript{B}</th>
<th>WTP\textsubscript{C}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

3.2.4 Constrained-DA

The WTP for the information on each tier is given below

<table>
<thead>
<tr>
<th>Student</th>
<th>WTP\textsubscript{A}</th>
<th>WTP\textsubscript{B}</th>
<th>WTP\textsubscript{C}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>37.5</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

4 Experimental Procedure

Subjects are divided into groups of six. In a group, every member plays the role of a student; schools are simulated in the environment since they are not strategic. Each round of the experiment has three stages. First, for each tier of schools, a subject can choose to privately observe her true preferences within that tier at a cost of $1; she can invest in multiple tiers or no tier at all. Second, after information acquisition, each student submits an unconstrained (in Treatments Unconstrained-DA and -BOS) or a constrained preference list (in Treatments Constrained-DA and -BOS). Third, based on students’ submitted preferences and schools’ priorities, either the DA (in Treatments Unconstrained- and Constrained-DA) or the BOS algorithm (in Treatments Unconstrained- and Constrained-BOS) gives the final allocation result.

The experiment consists of 15 rounds; each follows the above procedure. At the beginning of every round, subjects are randomly divided into new groups and their identities are randomly re-assigned. At the end of the experiment, one round is randomly chosen for payment. A student receives a payoff of $30, $25, $20, $15, $10, and $5 when assigned a
seat at the school ranked 1st, 2nd, 3rd, 4th, 5th, and 6th according to her true preferences, respectively. She receives zero payoff if she is not assigned to any seat. The final payment also includes a show-up fee of $5 and the payoff from a lottery game designed to elicit risk attitudes.

This research will be conducted at the Experimental Economics Laboratory of the Ohio State University. Average hourly expense per subject is expected to be $15/hour. Expected time in the laboratory for each subject is approximately 1.5 hours.

5 Conclusion

The paper investigates a common practice of many school choice programs in the field, where the length of students’ submitted preference lists are constrained. In an environment where students have incomplete information about others’ preferences, I theoretically study the effect of such a constraint under both a Deferred Acceptance mechanism (DA) and a Boston mechanism (BOS). The result shows that ex-ante stability can only be ensured under an unconstrained DA, but not under a constrained DA, an unconstrained BOS, or a constrained BOS. In a lab experiment, I find that the constraint also affects students’ information acquisition behavior. Specifically, when faced with a constraint, students tend to acquire less wasteful information and distribute more efforts to acquire relevant information under DA; such an effect is not significant under BOS. Overall, the constraint has a negative effect on efficiency and stability under both mechanisms.
Chapter 4
The Instability of Matching with Overconfident Agents: Laboratory and Field Investigations

1 Introduction

In college admissions, colleges usually prefer students with certain qualities or aptitudes that are not readily observable. Students are thus evaluated by noisy signals such as SAT scores and high-school transcripts. Some countries—including China, South Korea, Turkey, Russia, and Greece—simply use a standardized exam as an evaluation system, which means every year around the world, more than 13 million students’ college entrance is determined by their performance on a single exam.¹

Since the exam is held only once a year, it greatly simplifies the admissions process and is thus especially favorable to large markets. However, a single exam always entails measurement error: a student who underperforms on the exam may lose her placement at a preferred college to someone with a lower aptitude. The previous literature (discussed in more detail below) proposes a hopeful hypothesis that such an error in the evaluation system could be corrected with proper market design. Under the proposed mechanism, students apply to colleges or submit preferences before taking the exam. If we assume that students have perfect knowledge of their relative aptitudes, this mechanism triggers a self-sorting process, with students of higher (lower) aptitudes targeting more (less) preferred colleges. However, given the overwhelming evidence on self-evaluation biases such as overconfidence, I argue that in practice students may not be able to sort themselves perfectly. As a result, the proposed solution may introduce a behavioral error into the admissions procedure, larger than the measurement error from the exam itself. This paper presents a tradeoff between

¹Table 8 in Appendix C provides the statistic for each representative country.
these two errors, which may emerge in real markets with the presence of overconfident agents.

Formally, the above issue can be described as a college admissions problem, which is a matching problem that involves pairing members of one group of agents (students) with members of another group of agents (colleges). A centralized procedure, called a matching mechanism, is adopted to solve the problem based on students’ submitted preferences over colleges, as well as their priority ordering at each college. The priority ordering, solely determined by the score ranking in a single standardized exam, serves as a noisy proxy for colleges’ true preferences over students.

Suppose all colleges prefer students with higher aptitudes, and students have correlated preferences over colleges. Then a matching mechanism is considered socially desirable or “fair” if it matches more preferred colleges with students of higher aptitudes, not just with those of higher realized scores. As a notion of such fairness, a market outcome is said to be stable with regard to aptitude (“aptitude-stable”) if it eliminates the case where a student with a higher aptitude is not assigned to a preferred college but instead some other student with a lower aptitude is.\(^2\)

Previous studies such as Lien, Zheng, and Zhong (2015, 2016), Jiang (2014), and Wu and Zhong (2014) proposes a hopeful hypothesis that aptitude-stability is more likely to be achieved by a “pre-exam Boston mechanism (PreExam-BOS henceforth),” which combines a Boston matching algorithm with pre-exam preference submission; that is, students are asked to submit their preferences before taking the exam.\(^3\) The argument goes as follows. A Boston algorithm is not strategy-proof: instead of truthful revelation, a student should fill her first choice with a “safer” college, at which she is more likely to win a seat. With pre-exam preference submission, each student should employ a self-sorting strategy based on her ex-ante expected exam performance, which perfectly reflects her aptitude. Then the seats at more desirable colleges will be “reserved” for students of higher aptitudes, since those of lower aptitudes will sort themselves into lower-ranked colleges.

The above argument hinges on the assumption that students are able to correctly sort themselves before taking the exam, which requires that every student have perfect knowledge of the relative standing of her aptitude among all students in the market. Such knowledge is then “reported” to the market designer through the way she misrepresents her preferences under PreExam-BOS. However, overwhelming evidence has established the existence and

\(^2\)A formal definition that considers the college’s unoccupied seats is given in Section 2.1. Aptitude-stability, defined according to the true preferences of the market participants, is simply called “stability” in the standard matching literature (Gale and Shapley, 1962). Here I emphasize the word “aptitude” because in the current setting, we can also define stability according to the exam-based priorities, which may deviate from colleges’ aptitude-based preferences due to a single exam’s measurement error (see Section 2.1 for more details).

\(^3\)The procedure of a Boston algorithm is described in Section 2.2.1.
heterogeneity of biases in self-evaluation, such as overconfidence. In other words, while the proposed solution theoretically diminishes the effect of a single exam’s measurement error, it might introduce a behavioral error due to self-evaluation biases. As I show in this paper, these biases have important consequences for aptitude-stability.

Using a college admissions model, I first give theoretical predictions regarding the market outcome under different matching mechanisms. Following the previous literature, PreExam-BOS is mainly compared to a “post-score Serial Dictatorship mechanism” (PostScore-SD henceforth), which combines a Serial Dictatorship algorithm with post-score preference submission; that is, students submit their preferences after seeing the exam results.\(^4\) In contrast to PreExam-BOS, PostScore-SD is strategy-proof. With every student truthfully revealing her preferences, the matching outcome under PostScore-SD only depends on exam-based priorities and is thus distorted from aptitude-stability by the exam’s measurement error. On the other hand, as a dominant strategy under PostScore-SD, truth-telling is not affected by a student’s self-evaluation biases. Hence, compared to PreExam-BOS, PostScore-SD is more vulnerable to the noise from the single-exam system but less influenced by students’ over- or under-confidence. Which mechanism will create smaller distortions from aptitude-stability depends on the relative magnitudes of these two effects.

Since we cannot exogenously vary the choice of mechanisms in the field, I conduct a lab experiment to investigate students’ strategic behaviors and the market outcomes under different mechanisms. In an experimental market, each subject, playing the role of a student, is asked to take an exam, guess the exam results, and submit her preferences over simulated colleges to a matching algorithm. The exam is designed as a real-effort task in the lab, and a subject’s aptitude is evaluated as her average performance in multiple exams. As treatments, different matching algorithms and timings of preference submission are adopted.

The experimental data confirm that a majority of students report their preferences truthfully under PostScore-SD, while under PreExam-BOS, their strategies are significantly skewed by over- or under-confidence. Thus, neither mechanism fully achieves aptitude-stability on a market level. To measure how much a student’s welfare is distorted from an aptitude-stable matching, I compare the desirability of her aptitude-stably matched college to that of her assignment under each treatment. The result shows PreExam-BOS creates more severe and more noisy distortions from aptitude-stability than PostScore-SD, because not only are fewer students assigned to their aptitude-stably matched colleges but also the magnitudes of such welfare distortions are more spread out among students. In other words, under PreExam-BOS some students receive a large advantage while some others are

\(^{4}\)I also discuss a third timing of preference submission named “halfway,” under which students submit preferences after the exam but before the revelation of exam results. Thus, a total of six mechanisms are considered: two algorithms (BOS and SD) combined with three timings of preference submission (pre-exam, halfway, and post-score). The procedure of a Serial Dictatorship algorithm is described in Section 2.2.2.
dramatically hurt; neither these gains nor losses can be justified by the students’ aptitudes.

The aforementioned results can be explained by three observations: (i) on average, students exhibit overconfidence; (ii) there is significant heterogeneity in their levels of overconfidence; (iii) students make more heterogeneous strategic choices under PreExam-BOS, as opposed to highly aligned truth-telling behavior under PostScore-SD. In particular, under PreExam-BOS subjects tend to choose more aggressive or optimistic strategies than self-sorting based on their guessed exam performances. Thus, PreExam-BOS introduces more noise into the admissions procedure through both self-evaluation biases and strategic behaviors. As for the welfare effect of overconfidence, I find PreExam-BOS tends to reward those who are overconfident and punish those who are underconfident. Since women exhibit less overconfidence than men, PreExam-BOS serves as a gender penalty for women.

On the other hand, previous experimental studies, without taking overconfidence into account, find evidence that PreExam-BOS can outperform PostScore-SD in terms of aptitude-stability.\(^5\) The key difference lies in how subjects obtain information regarding their aptitudes before the exam. In Lien et al. (2015) and Jiang (2014), the exam component is abstracted away from the experiment; instead, each student is simply provided with her score distribution (that is, the distribution from which her score will be drawn), together with the score distributions of all the other students in the market.\(^6\) In contrast, the design in this paper comes closer to the field setting: subjects collect information from the feedback of multiple practice exams, or “mock tests.” As a result, overconfidence severely skews pre-exam preference submission and prevents PreExam-BOS from achieving aptitude-stability.

In an effort to minimize self-evaluation biases under the current setting, as additional treatments, I help subjects with their information collection by showing them all the past scores and the average score of every student in the market. However, subjects’ overconfidence stays on the same level, and PreExam-BOS continues to be inferior to PostScore-SD. Such a result indicates that a subject’s overconfidence is mainly driven by the overoptimistic belief about how much she can improve in the upcoming exam. In other words, almost all of the observed biases stem from a source that cannot be muted even with the maximum amount of information.

For a matching market to function well in practice, the choice of a matching mechanism should be tailored to the specific market environment, in this case to the students’

\(^5\)In Lien et al. (2015), the advantage of PreExam-BOS in terms of aptitude-stability mainly appears in the additional 10-round learning treatment. Such an advantage is not significant in the original treatment due to subjects’ deviation from equilibrium strategies, which is consistent with the aggressive strategic choices observed in this paper.

\(^6\)Under an additional treatment (“Quiz” treatment) in Lien et al. (2015), subjects are asked to take a short quiz; a subject’s relative performance on the quiz determines which role she will play in a group. However, the score distribution and the relative aptitude of each role is pre-determined and is directly provided to subjects.
self-evaluation biases and the exam’s measurement error. Therefore, to investigate Chinese college admissions, I collected field data on students’ guessed and realized exam results, multiple mock test results (used to measure academic aptitudes), and demographic information. A simple analysis shows that students exhibit biases in self-evaluation, and the variance of such biases can be larger than the variance of the exam’s measurement error. Since a similar conclusion is drawn from the lab data, this suggests that the welfare consequences observed in the lab (that is, compared to PostScore-SD, PreExam-BOS tends to create more severe and more varied distortions from aptitude-stability) could actually occur in the field.\footnote{Using the data collected in the field and the strategic patterns observed in the lab, a simulation is conducted to compare the performance of different matching mechanisms in the specific market of interest. The results show that compared to PostScore-SD, PreExam-BOS tends to create more severe and more varied distortions from aptitude-stability.} This provides a potential explanation for the recent reforms in China’s college admissions policy: despite what is recommended by the previous research, most districts are currently in transition from a mechanism that resembles PreExam-BOS to a mechanism more similar to PostScore-SD.\footnote{In Chinese college admissions, when a student’s welfare is significantly and negatively distorted from aptitude-stability, a typical consequence is that she rejects the assignment and re-enters the market after a year. A newspaper article in Beijing Youth Daily reports that Beijing’s policy reform in 2015 yields a higher admission rate, and the number of such students is reduced by about 20%. This provides an anecdotal evidence that the new mechanism may create less severe distortions from aptitude-stability.} Moreover, this result is interesting due to the strikingly high level of competition and the high stakes involved in students’ self-evaluation and preference submission process. The fact that over- or under-confidence survives even with extreme incentives to have correct self-evaluations indicates the prevalence of these biases.

This study is part of the recent literature on school choice and college admissions problems with a single-exam evaluation system. Wu and Zhong (2014), Jiang (2014), and Lien et al. (2016) theoretically compare PreExam-BOS and PostScore-SD and show that PreExam-BOS can outperform PostScore-SD in terms of aptitude-stability.\footnote{Lien et al. (2016) identify the conditions under which PreExam-BOS can or cannot achieve complete aptitude-stability.} However, as mentioned above, it is assumed that students have perfect knowledge of their relative aptitudes before taking the exam. Such information is directly provided to subjects in the lab experiments conducted by Lien et al. (2015) and Jiang (2014), where they find support for the above theoretical results. In contrast, the current paper tries to relax this assumption by allowing biases in self-evaluation. Wu and Zhong (2014) conduct empirical research using data from a top university in China. They show that students admitted under PreExam-BOS exhibit similar or better college academic performance than those admitted through other mechanisms. This result provides some evidence that PreExam-BOS could match the top college with students of better aptitudes, yet it is silent about middle- or lower-ranked colleges or the overall matching outcome.
The paper also contributes to the studies addressing the merits and flaws of the Boston algorithm compared to Serial Dictatorship and other strategy-proof algorithms. In a standard setting without uncertainty or imperfect information, the Boston algorithm is often considered inferior in terms of stability and efficiency (Abdulkadiroğlu and Sönmez, 2003; Chen and Sönmez, 2006; and Ergin and Sönmez, 2006). On the other hand, the manipulability of Boston can sometimes improve ex-ante efficiency, because it reflects certain information that is otherwise unobservable to a market designer (see, for example, Abdulkadiroğlu, Che, and Yasuda, 2011 and Featherstone and Niederle, 2008).

Lastly, the paper is related to the enormous literature on overconfidence across economics, psychology, and finance. As in the current study, overconfidence has been used to explain market failures in various environments, such as excessive business entry and trading volume, corporate investment distortions, and stock market bubbles to name just a few (see Camerer and Lovallo, 1999; Glaser and Weber, 2007; Malmendier and Tate, 2005; and Odean, 1999). Moreover, it is well established that people exhibit heterogeneous levels of over- or under-confidence, which can be predicted by certain factors including personality, gender, and cognitive abilities (see Schaefer, Williams, Goodie, and Campbell, 2004; Niederle and Vesterlund, 2007; Barber and Odean, 2001; Coffman, 2014; Kleitman and Stankov, 2007 and Stankov and Crawford, 1996). Similar evidence is also found in both lab and field data of this research.

The rest of the paper is organized as follows. In Section 2, I lay out the college admissions model and make theoretical predictions. Section 3 describes the lab experiment and presents experimental results. In Section 4, I introduce the application of Chinese college admissions, describe the field data collection process, and summarize the results of data analysis and simulation. Section 5 concludes the paper.

2 The Model

2.1 A College Admissions Problem

The centralized matching market considered here is a variation of the college admissions problem (Gale and Shapley, 1962): there are a number of students; each of them is to be assigned a seat at one of a number of colleges. Each student has strict preferences over all colleges and each college has strict preferences over all students. There is a maximum capacity at each college, but the total number of seats exceeds the total number of students. The distinguishing feature of this environment is that every college has a priority ordering.

---

10 See also Pais and Pintér (2008) and Klijn, Pais, and Vorsatz (2013).
11 Moore and Healy (2008) provide a detailed overview of different ways in which the literature has defined overconfidence: overestimation, overplacement, and overprecision. They then offer a theory that reconciles these concepts and explains several inconsistencies in the existing evidence.
of all students, which is not necessarily in accord with its preference relation over students: the former is determined by students’ performance on a single exam, while the latter is determined by their intrinsic aptitudes. Formally, the college admissions problem consists of:

1. A set of students $I = \{i_1, \ldots, i_n\}$, $n \geq 2$.
2. A set of colleges $C = \{c_1, \ldots, c_m\}$, $m \geq 2$.
3. A capacity vector $q = (q_{c_1}, \ldots, q_{c_m})$.
4. A list of strict student preferences $P_I = (P_{i_1}, \ldots, P_{i_n})$. The preference relation $P_i$ of student $i$ is a linear order over $C \cup \{i\}$, where $cP_i c'$ means that student $i$ strictly prefers college $c$ to college $c'$ and $i$ denotes remaining unmatched. Students prefer any college to remaining unmatched.
5. A vector of students’ aptitudes $a = (a_1, \ldots, a_n)$ and a corresponding vector of aptitude ranks $r_a = (r_{a_1}, \ldots, r_{a_n})$, where $a_i$ denotes student $i$’s aptitude and $r_{a_i}$ denotes the rank of her aptitude among all students (with 1 being the highest rank). Ties in aptitudes are randomly broken.
6. A vector of students’ exam scores $s = (s_1, \ldots, s_n)$ and a corresponding vector of exam score ranks $r_s = (r_{s_1}, \ldots, r_{s_n})$, where $s_i$ denotes student $i$’s exam score and $r_{s_i}$ denotes the rank of her exam score among all students (with 1 being the highest rank). Ties in scores are randomly broken.
7. A list of strict college preferences $P_C = (P_{c_1}, \ldots, P_{c_m})$. The preference relation $P_c$ of college $c$ is a linear order over $I \cup \{c\}$, where $iP_c i'$ means that college $c$ strictly prefers student $i$ to student $i'$ and $c$ denotes leaving a seat empty. Colleges prefer any student to leaving a seat empty. Colleges have identical preferences over students, which are determined by students’ aptitude ranking; i.e., $iP_c i' \iff r_{a_i} < r_{a_{i'}}$, $\forall c \in C$.
8. A strict priority ordering of students at every college that is determined by students’ exam score ranking: student $i$ has a higher priority than student $i'$ at every college if and only if $r_{s_i} < r_{s_{i'}}$. All colleges have the same priority ordering.

Student $i$’s exam score is given by $s_i = a_i + \xi_i$, where $\xi_i$ is the measurement error of an exam. I assume that a student’s aptitude is the mean and the mode of her exam score distribution; that is, one exam score is an unbiased but noisy measure of aptitude.
**Assumption 1.** For student $i$, an exam’s measurement error $\xi_i$ follows a distribution on the real line with mean 0 and non-zero standard deviation.$^{12}$

Similarly, student $i$’s exam score rank is given by $r_{si} = r_{ai} - \epsilon_i$, where $\epsilon_i$ is the measurement error of the exam in terms of rank.$^{13}$

Below I make a simplifying assumption following the previous literature.

**Assumption 2.** Students have identical preferences over colleges.$^{14}$

Without loss of generality, assume in addition that a college with a smaller index is more desirable; i.e., $c_j P_i c_{j'} \iff j < j', \forall i \in I$.

The outcome of a matching market is known as a matching. Formally, a matching is a function $\mu : I \cup C \rightarrow 2^I \cup C$ such that for any $i \in I$ and any $c \in C$, (i) $\mu(i) \in C \cup i$, (ii) $\mu(c) \in 2^I$, (iii) $\mu(i) = c$ if and only if $i \in \mu(c)$, and (iv) $|\mu(c)| \leq q_c$. Thus, $\mu(i)$ denotes the assignment of student $i$ under matching $\mu$, and $\mu(c)$ denotes the set of students who are matched to college $c$ under matching $\mu$.

In the matching literature, stability is used as a central criterion to evaluate a matching outcome. Under the current structure, we can either set such a criterion according to colleges’ aptitude-based preferences, or according to their exam-based priorities. They are called stability with regard to aptitude and stability with regard to exam score, respectively.$^{15}$ The definitions are given below.

**Definition 1.** A matching $\mu$ is **stable with regard to aptitude** ("aptitude-stable") if and only if there is no student–college pair $(i, c)$ such that student $i$ prefers college $c$ to her assignment $\mu(i)$ and either (1) college $c$ has empty seats under $\mu$, or (2) at least one of the students in $\mu(c)$ has a lower aptitude than student $i$.

**Definition 2.** A matching $\mu$ is **stable with regard to exam score** ("score-stable") if and only if there is no student–college pair $(i, c)$ such that student $i$ prefers college $c$ to her assignment $\mu(i)$ and either (1) college $c$ has empty seats under $\mu$, or (2) at least one of the students in $\mu(c)$ has a lower exam score than student $i$.

Clearly, aptitude-stability is more socially desirable than score-stability since it respects colleges’ true preferences, which are assumed to be based on students’ aptitudes instead of their scores in one exam. Below I use a simple example to illustrate the environment.

---

$^{12}$The distribution of $\xi_i$ can be continuous or discrete. Assumption 1 is essentially an assumption on $s_i$ and $a_i$ because $\xi_i$ is derived from $\xi_i = s_i - a_i$. Empirically, it can be easily satisfied with normalization. See Section 4 for more details.

$^{13}$Here I use $\epsilon_i = r_{ai} - r_{si}$ instead of $\epsilon_i = r_{si} - r_{ai}$ to be consistent with the definition of overplacement in Section 2.3.

$^{14}$It reflects the reality of many college admissions markets that students’ preferences over colleges are correlated. Following the previous literature, here I simplify the environment by assuming homogeneity. Although relaxing such an assumption is a well-motivated extension, it is not the focus of this paper.

$^{15}$Similar concepts are named ex-post and ex-ante fairness in a school choice setting.
Example 1. Suppose there are three students \( I = \{i_1, i_2, i_3\} \) and three colleges \( C = \{c_1, c_2, c_3\} \) in the market. Each college has only one slot to fill \( q = \{1, 1, 1\} \). Students have homogeneous preferences over colleges: \( c_1 P_i c_2 \) and \( c_2 P_i c_3 \), \( i = 1, 2, 3 \). On a single exam, a student’s performance is consistent with her aptitude with probability 0.5; she overperforms with probability 0.25, and underperforms with probability 0.25. Table 1 shows the score distributions, which are independent across students.

Table 1: Score Distributions and Aptitudes (Example 1)

<table>
<thead>
<tr>
<th>Student</th>
<th>Aptitude</th>
<th>Score (Prob.)</th>
<th>Overperform (0.25)</th>
<th>Consistent (0.50)</th>
<th>Underperform (0.25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i_1 )</td>
<td>( a_1 = 12 )</td>
<td>( s_1 = 16 )</td>
<td>16</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>( i_2 )</td>
<td>( a_2 = 15 )</td>
<td>( s_2 = 19 )</td>
<td>19</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>( i_3 )</td>
<td>( a_3 = 9 )</td>
<td>( s_3 = 13 )</td>
<td>13</td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>

A student’s aptitude is given by the mean and mode of her score distribution. The second column of Table 1 implies students’ aptitude ranks \( r_a = (2, 1, 3) \), which determine every college’s aptitude-based preferences. So the unique aptitude-stable matching is

\[
\begin{align*}
&i_1 \quad i_2 \quad i_3 \\
&c_2 \quad c_1 \quad c_3
\end{align*}
\]

However, students may exhibit any ranking in their exam scores. With probability \( \frac{17}{64} \), the exam’s measurement error leads to the realized score ranks \( r_s = (1, 2, 3) \). In this case, the unique score-stable matching is

\[
\begin{align*}
&i_1 \quad i_2 \quad i_3 \\
&c_1 \quad c_2 \quad c_3
\end{align*}
\]

which is not aptitude-stable because both \( i_2 \) and \( c_1 \) prefer each other to their current assignments.

Example 1 shows how an exam’s measurement error could distort students’ realized score ranking from their aptitude ranking and thus prevent the score-stable matching from achieving aptitude-stability. Below I compare different mechanisms (combinations of a matching algorithm and a timing of preference submission) and examine which one is less likely to be effected by such noise from a single exam and is more likely to yield an aptitude-stable matching.

\[\text{Table 9 in Appendix C shows the probability of every possible score ranking.}\]
2.2 Two Matching Algorithms

To select a matching for the college admissions problem defined above, a systematic procedure, called a “matching algorithm,” allocates students to colleges depending on students’ submitted preferences and colleges’ priority ordering. In terms of timing, preference submission could occur before or after the exam, but the actual matching algorithm is always conducted after preference submission and the revelation of exam results.

In the literature of college admissions and school choice problems, three matching algorithms are most widely discussed: the Boston algorithm (BOS), the Gale-Sharply Deferred Acceptance algorithm (DA), and the Top Trading Cycles algorithm (TTC). In the current setting where all colleges share the same priority ordering, TTC reduces to a Serial Dictatorship algorithm (SD) and is equivalent to DA (Kesten, 2006). Therefore, BOS and SD are the two competing algorithms considered in this paper.\textsuperscript{17}

2.2.1 The Boston Algorithm (BOS)

Each student submits her preferences by ranking all colleges. Every college has the same strict priority ordering of students, which is determined by students’ exam score ranking.

**Step 1:** Only the 1st choices of all students are considered. For each college, consider the students who have listed it as their 1st choice; assign seats of the college to these students one at a time following their priority ordering until either there are no seats left or there are no students left who have listed it as their 1st choice.

In general, Step $k$ ($k \geq 1$) can be described as follows.

**Step $k$:** Only the $k$th choices of the remaining students (who have not been assigned a seat previously) are considered. For each college with still available seats, consider the students who have listed it as their $k$th choice; assign the remaining seats to these students one at a time following their priority ordering until either there are no seats left or there are no students left who have listed it as their $k$th choice.

The procedure terminates after any step $k$ when every student is assigned a seat at some college, or if the only students who remain unassigned listed no more than $k$ choices.

2.2.2 The Serial Dictatorship Algorithm (SD)

Each student submits her preferences by ranking all colleges. Every college has the same strict priority ordering of students, which is determined by students’ exam score ranking.

\textsuperscript{17}I choose SD instead of DA to be more consistent with the field environment of Chinese college admissions. See Section 4.1 for more details.
**Step 1:** The student with the highest priority is considered. She is assigned a seat at the college of her 1st choice.

**Step 2:** The student with the second highest priority is considered. She is assigned a seat at her 1st choice if that college still has empty seats left; otherwise, she is assigned a seat at her 2nd choice.

In general, Step $k$ ($k \geq 2$) can be described as follows.

**Step $k$:** The student with the $k$th highest priority is considered. She is assigned a seat at her most preferred college that has an empty seat.

The procedure terminates when every student has been considered, or when no college seats remain.

### 2.3 Three Timings of Preference Submission

Apart from matching algorithms, the timing of preference submission can also largely affect the strategic behaviors of market participants, thus influencing the market outcome. Inspired by college admissions in China, I focus on three different timings, under which students are asked to submit their preferences at different stages, or different information statuses. The timings and stages are named as follows.

The “ex-ante,” “interim,” and “ex-post” stages refer to: before the exam, after the exam but before the revelation of exam results, and after the revelation of exam results.\(^{18}\) Under the timings named “pre-exam,” “halfway,” and “post-score,” students submit their preferences at the ex-ante, interim, and ex-post stages, respectively. The following assumption specifies the information status at the ex-post stage.

**Assumption 3.** At the ex-post stage (after the revelation of exam results), it is common knowledge that every student knows her own exam score rank.

When submitting preferences under the pre-exam and halfway timings, students do not observe the exam results, which means they do not know their priority ordering at each college. In some situations (discussed in Section 2.4), their strategies may depend on their guessed exam results. Therefore, under the pre-exam timing, a component is added to the college admissions problem defined in Section 2.1:

1. A vector of students’ guessed exam scores $\hat{s}_{EA} = (\hat{s}^{EA}_1, \ldots, \hat{s}^{EA}_n)$ and a corresponding vector of students’ guessed exam score ranks $\hat{r}_{EA} = (\hat{r}^{EA}_{s_1}, \ldots, \hat{r}^{EA}_{s_n})$ at the ex-ante stage, where $\hat{s}^{EA}_i$ and $\hat{r}^{EA}_{s_i}$ denote student $i$’s guessed score and guessed rank.

\(^{18}\)The interim stage is defined as a distinct information status, because a student may obtain additional information during the exam.
Under the halfway timing, the following component is added instead:

9'. A vector of students’ guessed exam scores \( \hat{s}^{IN} = (\hat{s}^{IN}_1, \ldots, \hat{s}^{IN}_n) \) and a corresponding vector of students’ guessed exam score ranks \( \hat{r}^{IN} = (\hat{r}^{IN}_1, \ldots, \hat{r}^{IN}_n) \) at the interim stage, where \( \hat{s}^{IN}_i \) and \( \hat{r}^{IN}_i \) denote student \( i \)'s guessed score and guessed rank.

In the current setting, overconfidence, defined as a bias in self-evaluation, can be measured in score or rank. Following Moore and Healy (2008), I refer to score overconfidence as “overestimation,” and rank overconfidence as “overplacement.”

**Definition 3.** Student \( i \)'s overestimation at the ex-ante stage is given by \( \delta^{EA}_i \equiv \hat{s}^{EA}_i - E[s_i] = \hat{s}^{EA}_i - a_i \) and at the interim stage by \( \delta^{IN}_i \equiv \hat{s}^{IN}_i - E[s_i] = \hat{s}^{IN}_i - a_i \).

**Definition 4.** Student \( i \)'s overplacement at the ex-ante stage is given by \( \theta^{EA}_i \equiv r_{a_i} - \hat{r}^{EA}_s i \) and at the interim stage by \( \theta^{IN}_i \equiv r_{a_i} - \hat{r}^{IN}_s i \).

Naturally, a student is said to exhibit underconfidence when the above measures take negative values.

### 2.4 Theoretical Predictions

This section gives the theoretical predictions for strategies of market participants and the stability of matching outcomes under different combinations of matching algorithms and timings of preference submission. I refer to the SD (BOS) algorithm under pre-exam, halfway, or post-score timing as “PreExam-SD,” “Halfway-SD,” or “PostScore-SD” mechanism (“PreExam-BOS,” “Halfway-BOS,” or “PostScore-BOS” mechanism).

I first state the results for PreExam-, Halfway-, and PostScore-SD, which are largely drawn from the previous literature.

**Proposition 1.** (1) PreExam-SD, Halfway-SD, and PostScore-SD are strategy-proof. (2) PostScore-SD always yields the score-stable matching. (3) PreExam-SD and Halfway-SD yield the score-stable matching in the truth-telling equilibrium.

It is well-established in the literature that SD is strategy-proof for any realized priority ordering over students, which means truth-telling is a weakly dominant strategy for every student. First, note that overplacement is defined as \( E[r_{s_i}] - \hat{r}^{EA}_s i \) (or \( E[r_{s_i}] - \hat{r}^{IN}_s i \)) instead of \( \hat{r}^{IN}_s i - E[r_{s_i}] \) (or \( \hat{r}^{IN}_s i - E[r_{s_i}] \)), because a smaller value of rank means being better in aptitude or exam score. Second, there is a slight abuse of terminology in the definitions of \( \delta^{IN}_i \) and \( \theta^{IN}_i \). At the interim stage, a student has obtained some additional information, say a signal \( t \), about her performance on the exam. Therefore, strictly speaking, overestimation and overplacement should be measured as \( \hat{s}^{IN}_i - E[s_i|t] \) and \( E[r_{s_i}|t] - \hat{r}^{IN}_s i \), respectively. The current definitions are adopted since it is more relevant in this environment to discuss how the aptitude ranking is distorted by students’ guessed exam results. Third, because the measurement error of an exam has zero mean, overconfidence evaluated relative to posterior beliefs has the same average level as that evaluated relative to prior beliefs.
student, regardless of her knowledge about the priority ordering at the time of preference submission. Hence, no matter which timing of preference submission is adopted, SD always implements the score-stable matching outcome in the truth-telling equilibrium (see Appendix A for a more detailed proof).

In contrast, students have strong incentives to misrepresent their true preferences under BOS. The following definition specifies a strategy in preference submission at the ex-post stage.

**Definition 5.** A student $i$ is said to adopt a *score-based sorting strategy* if she lists college $c_j$ as her first choice in preference submission such that $\sum_{k=1}^{j-1} q_k < r_{s_i} \leq \sum_{k=1}^{j} q_k$.\(^{20}\)

In the current setting, score-based sorting means listing one’s score-stably matched college as the first choice. Recall the environment in Example 1, where every college has only one seat, and students’ realized score ranks are $r_s = (1, 2, 3)$. Then we say all students exhibit score-based sorting if the submitted first choices of $i_1$, $i_2$, and $i_3$ are given by $c_1$, $c_2$, and $c_3$. Hence, BOS will have every student accepted in Step 1 of the procedure and achieve score-stability. Below, Proposition 2 shows that score-based sorting is an equilibrium strategy under PostScore-BOS, and the score-stable matching is implemented in equilibrium (the formal proof is given in Appendix A).

**Proposition 2.** (1) Under PostScore-BOS, there is a Nash equilibrium where every student exhibits score-based sorting. (2) PostScore-BOS always implements the score-stable matching.

Under PreExam- and Halfway-BOS, students do not observe their exam score ranks at the time of preference submission; their strategies are thus affected by their guessed exam results at the ex-ante and interim stages, respectively. As a counterpart of score-based sorting, guess-based sorting is defined below.

**Definition 6.** Under the pre-exam (or halfway) timing, a student $i$ is said to adopt a *guess-based sorting strategy* if she lists college $c_j$ as her first choice in preference submission such that $\sum_{k=1}^{j-1} q_k < \hat{r}_{s_i}^{EA} \leq \sum_{k=1}^{j} q_k$ (or $\sum_{k=1}^{j-1} q_k < \hat{r}_{s_i}^{IN} \leq \sum_{k=1}^{j} q_k$).

For PreExam- or Halfway-BOS to implement an aptitude-stable matching, it is crucial that every student’s guessed rank perfectly reflects her aptitude rank, which should be commonly known to the market. Therefore, the previous literature makes the following assumption and gives the prediction stated in Proposition 3.

**Assumption 4.** *It is common knowledge that no student exhibits any over- or under-placement.*

\(^{20}\)The concept is also named rank bias in the literature.
Proposition 3. If Assumption 4 (common knowledge of no over- or under-placement) holds for preference submission at the ex-ante stage (or at the interim stage) and every student exhibits guess-based sorting, PreExam-BOS (or Halfway-BOS) yields the aptitude-stable matching.

The proof of the above proposition is straightforward. Under Assumption 4, every student who exhibits guess-based sorting lists her aptitude-stably matched college as the first choice. Under BOS, everyone is accepted in Step 1 and aptitude-stability is achieved.

However, if Assumption 4 fails, that is if students exhibit over- or under-placement, the conclusion in Proposition 3 will change significantly. Based on the fact that one is not aware of her own bias, and evidence on the false-consensus effect, I make the following assumption instead.\footnote{Under the false-consensus effect, people tend to believe that others are similar to them; see Ross, Greene, and House (1977) for a seminal contribution and Marks and Miller (1987) for a survey. Evidence on such an effect is also found in the lab experiment for this study (Section 3.3; Result 4).}

Assumption 4’. Every student believes that she exhibits no over- or under-placement and that other students exhibit no over- or under-placement.

The beliefs specified in Assumption 4’ will be false with the presence of over- or under-placement. Since students’ strategies under PreExam- and Halfway-BOS hinge on these beliefs, the matching outcome will be affected as well. This provides the intuition for Proposition 4.\footnote{Note that neither Propositions 3 nor 4 can give equilibrium predictions for a general environment, because students’ strategic choices under PreExam- and Halfway-BOS depend on their cardinal utilities and risk attitudes. Here they only serve as a guideline for the subsequent experimental analysis, where these claims are examined using subjects’ strategic behaviors and market outcomes in the lab.}

Proposition 4. If Assumption 4 is replaced by 4’ for preference submission at the ex-ante stage (or at the interim stage) and every student exhibits guess-based sorting, PreExam-BOS (or Halfway-BOS) may fail to achieve aptitude-stability.

On the other hand, truth-telling under PreExam-, Halfway-, and PostScore-SD and score-based sorting under PostScore-BOS do not depend on students’ guessed exam results. Therefore, the market outcomes under these four mechanisms are less easily affected by over- or under-confidence. Below I illustrate the theoretical predictions in the setting of Example 1.

Example 1 (Cont.) (i) Recall that students’ aptitude ranks are \( r_a = (2, 1, 3) \) and their exam score ranks are \( r_s = (1, 2, 3) \). Suppose their guessed ranks at the ex-ante stage are given by \( \hat{r}^E_{EA} = (1, 1, 2) \). Then both PreExam- and PostScore-SD yield the score-stable matching in the truth-telling equilibrium; under PostScore-BOS, if \( i_1, i_2, \) and \( i_3 \) all exhibit...
score-based sorting by submitting $c_1$, $c_2$, and $c_3$ as their first choices respectively, the score-stable matching is again implemented. However, under PreExam-BOS, if $i_1$, $i_2$, and $i_3$ all exhibit guess-based sorting by submitting $c_1$, $c_1$, and $c_2$ as their first choices, the following matching is implemented:

\[
\begin{array}{ccc}
  i_1 & i_2 & i_3 \\
  c_1 & c_3 & c_2 \\
\end{array}
\]

(ii) Now suppose the exam’s measurement error is given by $\epsilon = (0, 0, 0)$ and thus $r_s = r_a = (2, 1, 3)$; all else stays the same. Then PreExam-BOS yields the following matching with everyone exhibiting guess-based sorting:

\[
\begin{array}{ccc}
  i_1 & i_2 & i_3 \\
  c_3 & c_1 & c_2 \\
\end{array}
\]

Part (i) of the example indicates that under PreExam- and Halfway-BOS, overconfidence has two effects on the matching procedure.\(^{23}\) First, it directly skews the sorting in preference submission: under the influence of overplacement, $i_3$ submits $c_2$ as her first choice, while $i_1$ submits $c_1$ and ends up competing with $i_2$ in Step 1 of BOS. Second, it brings back the noise from the exam’s measurement error: due to the first effect, BOS needs to resolve the competition between $i_1$ and $i_2$ according to their exam scores and as a result of the exam’s measurement error, $i_1$ is matched with $c_1$ although $i_2$ has a higher aptitude. Therefore, PreExam- and Halfway-BOS can be directly affected by the presence of self-evaluation biases and meanwhile, indirectly by the noise from a single-exam evaluation system.

As for individual welfare, under the setting of Part (i), $i_2$ is unbiased but is punished since she is allocated to $c_3$ instead of her aptitude-stable match $c_1$, while both $i_1$ and $i_3$ are rewarded for being overconfident ($i_1$ is matched to $c_1$ instead of $c_2$; $i_3$ is matched to $c_2$ instead of $c_3$). On the other hand, from Part (ii) of the example, we can see the same level of overplacement hurts $i_1$ but benefits $i_3$.

Hence, regarding the effects of overconfidence on the market outcome and on individual welfare, the prediction from the model is ambiguous since it depends on the distribution of overconfidence and the realization of the exam’s measurement error. To further explore these issues, I conduct a lab experiment where subjects’ preferences are induced by monetary incentives. Such a controlled setting allows me to closely observe their strategic choices, examine market stability, and analyze individual welfare.

\(^{23}\)Given the definitions of overconfidence at the interim stage (see Definitions 3 and 4; Footnote 19), in this model the halfway timing is theoretically equivalent to the pre-exam timing. Therefore, Example 1 also has implications for Halfway-BOS and Halfway-SD.
3 A Lab Experiment

To investigate strategic behaviors and market outcomes under different mechanisms, I design an experiment with various combinations of matching algorithms and timings of preference submission. Compared to other experimental studies in the literature, the distinguishing feature of this design lies in how subjects obtain information regarding their aptitudes. In Lien et al. (2015) and Jiang (2014), the exam component is abstracted away from the experiment; instead, each student is provided with her score distribution (that is, the distribution her score will be drawn from), together with the score distributions of all the other students in the market. There is thus much less scope for over- or under-confidence since subjects are provided with perfect information of their aptitude ranking. In my design, the exam component is introduced as a real-effort task; subjects evaluate themselves at the ex-ante and interim stages using feedback from multiple practice exams, or “mock tests.” Such a setting resembles the field environment and allows us to examine the existence of self-evaluation biases. In addition, to control for factors that can potentially affect subjects’ strategic choices, at the end of the experiment I elicit their beliefs about other participants’ overconfidence level as well as their risk attitudes.

3.1 Experimental Design

Each experimental market consists of five students and five colleges. Each subject plays the role of a student; colleges are simulated in the environment since they are not strategic. Colleges are labeled as $c_1$, $c_2$, $c_3$, $c_4$, and $c_5$; each has only one slot to fill. All students have the same induced preferences over colleges: when matched with $c_1$, $c_2$, $c_3$, $c_4$, or $c_5$, a student receives a payoff of $20$, $15$, $10$, $5$, or $0$, respectively.

Students’ priority ordering at each college is determined by their score ranking in an exam. The exam consists of 20 IQ multiple choice questions, and students have 3 minutes to work. One’s score equals the number of correct answers; there is no penalty for wrong answers.\footnote{Such a design aims to reduce the gender gap. Baldiga (2013) shows that when there is a penalty for wrong answers, women answer significantly fewer questions than men and thus do significantly worse conditional on their knowledge.} In order to obtain a strict score ranking and thus a strict priority ordering, ties are broken randomly. When exam results are revealed, each subject can observe the scores and ranks of all five students.

At the ex-ante stage (before the exam), each student is asked to guess her exam score and the rank of her score in the market. Similarly, a guess of score and a guess of rank are again elicited at the interim stage (after the exam but before the revelation of exam results). These guesses are not observable to other students.\footnote{Admittedly, the elicitation procedure may influence subjects’ subsequent decisions. For example, sub-}
Table 2: Treatment Design

<table>
<thead>
<tr>
<th>Timing</th>
<th>The BOS Algorithm</th>
<th>The SD Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-exam</td>
<td>PreExam-BOS</td>
<td>PreExam-SD</td>
</tr>
<tr>
<td>Halfway</td>
<td>Halfway-BOS</td>
<td>Halfway-SD</td>
</tr>
<tr>
<td>Post-score</td>
<td>PostScore-BOS</td>
<td>PostScore-SD</td>
</tr>
</tbody>
</table>

The experiment has a three-by-two treatment design (Table 2); varying the matching algorithm (BOS or SD) and the timing of preference submission (pre-exam, halfway, or post-score). Under the pre-exam timing, preference submission follows the guess at the ex-ante stage and precedes the exam; under the halfway timing, it follows the guess at the interim stage and precedes the revelation of exam results; and under the post-score timing, it comes after the revelation of results. At the end of every treatment, an algorithm is used to match students with colleges, based on students’ submitted preferences and their exam score ranking.

Pre-exam:

![Figure 1: Timings of Preference Submission](image)

Halfway:

Post-score:

Treatments using the same algorithm (in the same column of Table 2) are implemented within-subject. Every subject makes three sequences of decision making. As illustrated in Figure 1, all sequences include the same six components (an exam, a guess at the ex-ante stage, a guess at the interim stage, the revelation of exam results, preference submission, and a matching procedure), but differ in the timing of preference submission. To ensure subjects may be more likely to base their strategies in preference submission on the elicited beliefs. However, since the beliefs are elicited at both ex-ante and interim stages under all treatments, such an effect, if exists, should not interfere with any treatment effects.
a relatively clean treatment effect, the three timings appear in a random order, and no feedback is given in between regarding other students’ submitted preferences or the final matching outcomes.

Before the three treatments, subjects are given three “mock tests;” each takes the same form as the exam (20 questions over 3 minutes). The results, including the scores and ranks of all five students, are revealed at the end of every mock test. This process is for subjects to learn about their aptitudes, as well as the relative standing of their aptitudes, in taking the exam. Such a design provides three mock tests and three exams (one in each treatment) for every subject. I use the average of these six performances as the measure of aptitude.  

3.2 Experimental Procedure

Each session of the experiment consists of three parts. In the first part, either BOS or SD is described and illustrated with an example, followed by five practice rounds of preference submission with randomly assigned ranks (designed to familiarize subjects with the matching algorithm). The second part is the main experiment, three mock tests followed by three treatments. At the end of the second part but before giving any feedback on matching outcomes, I elicit beliefs about other participants’ overconfidence level using a question like: “The computer will now randomly choose one of the other participants in the room. During the experiment, this participant has given a total of 6 guesses about his/her rank in the exam. Please give your guesses regarding the correctness of his/her responses by guessing the value of (his/her actual rank - his/her guessed rank) for each guess.” The third part elicits risk attitudes using a variation of the lottery game from Holt and Laury (2002).  

Subjects are randomly divided into groups of five and are re-grouped for every practice round of preference submission, every mock test, and every treatment. At the end of the experiment, one mock test, one guess (either a guess of one’s own score or rank or a guess of another participant’s overconfidence level), and the matching outcome in one treatment are randomly chosen for payment. A subject receives $0.25 for each correct answer in the chosen mock test, plus $2 if the chosen guess is correct, together with a payoff of $20, $15, $10, $5, or $0, if she is matched with \(c_1, c_2, c_3, c_4, \) or \(c_5\) in the chosen treatment. The final payment also includes the payoff from the lottery game, a show-up fee of $3, and a $1-payment for completing a questionnaire.

---

26 There is some evidence of learning during the three mock tests. The main results remain unchanged if I exclude all or some of the mock tests from the measure of aptitude.

27 Subjects are asked to make 20 choices between paired lotteries; each pair consists of a “safe” option and a “risky” option. Following Holt and Laury (2002), the total number of safe choices (ranging from 0 to 20) is used as an indicator of risk aversion. A majority of subjects chose the safe option when the probability of the higher payoff was small, and then crossed over to the risky option without ever going back to the safe option. Only 7 out of 95 subjects exhibited back-and-forth behavior.
The experiment was conducted in February 2015 at the Experimental Economics Laboratory of The Ohio State University. There were seven sessions in total. One session had 10 subjects; one had 20; and the other five sessions were conducted with 15 subjects. Out of 95 subjects (41 females and 64 males), there were 60 participants for treatments using BOS, and 45 participants for treatments using SD. Each session lasted approximately 75 minutes. The average payment, including a show-up fee, was about $18.28.

3.3 Experimental Results

Below I first evaluate the matching mechanisms by examining how frequently an aptitude-stable outcome is produced under different treatments. The results in Section 3.3.1 show that compared to PostScore-SD, PreExam- and Halfway-BOS create an even more severe distortion from aptitude-stability, because a smaller proportion of students are allocated to their aptitude-stably matched colleges. Such market failures under PreExam- and Halfway-BOS could stem from students’ self-evaluation biases or their deviations from the guess-based sorting strategy. Sections 3.3.2 and 3.3.3 analyze the aspects of belief and strategy, respectively. Section 3.3.4 investigates how a student’s individual welfare is affected by her beliefs and strategic choices under different mechanisms.

3.3.1 Market Outcomes

I start by examining the hypothesis on market outcomes according to the theoretical predictions in Propositions 1 to 3. Result 1 shows how frequently an aptitude-stable or a score-stable outcome is produced on a market level. For more detail, I also analyze the proportion of aptitude-stably and score-stably matched student-college pairs (Result 2).

Hypothesis 1. (i) Score-stability is achieved under PostScore-SD, Halfway-SD, PostScore-SD, and PostScore-BOS. (ii) Aptitude-stability is achieved under PreExam-BOS and Halfway-BOS.

Result 1. (i) Score-stability is achieved in all markets under PostScore-SD, most markets under PreExam-SD, Halfway-SD, and PostScore-BOS, but no markets under PreExam-BOS or Halfway-BOS. (ii) Aptitude-stability is rarely achieved under any mechanism.

Figure 2 summarizes, for each treatment, the fraction of markets that yield the score-stable or aptitude-stable matching. As shown in Figure 2b, aptitude-stability is only

---

28 For treatments using SD, the pilot data exhibit less variation since there exists a dominant strategy (truth-telling). Therefore, the power calculation prior to the experiment requires fewer data points.

29 There are a total of 12 markets under PreExam-, Halfway-, and PostScore-BOS, and a total of 9 markets under PreExam-, Halfway-, and PostScore-SD.
observed in 2 out of 9 markets under Halfway-SD, and 1 out of 9 markets under PreExam- and PostScore-SD, which could be considered as coincidences because the aptitude-stable matching also happens to be score-stable in these four markets.

To measure how severely the market outcome is distorted from score-stability (aptitude-stability) under each mechanism, I calculate the proportion of students who are allocated to their score-stably (aptitude-stably) matched colleges, that is, the proportion of score-stably (aptitude-stably) matched pairs.

**Result 2.** (i) PreExam-BOS and Halfway-BOS yield a smaller proportion of score-stably matched pairs than the other four mechanisms. (ii) PreExam-BOS and Halfway-BOS yield a smaller proportion of aptitude-stably matched pairs than Halfway-SD and PostScore-SD.
According to Figure 3a, there is a smaller proportion of score-stably matched pairs under PreExam-BOS or Halfway-BOS compared to the other four mechanisms \((p < 0.001)\).\(^{30}\) As for the aptitude-stably matched pairs presented in Figure 3b, PreExam-BOS yields a smaller proportion of such pairs than Halfway-SD \((p = 0.05)\) and PostScore-SD \((p = 0.016)\). Similarly, there is also a smaller proportion under Halfway-BOS compared to Halfway-SD \((p = 0.076)\) and PostScore-SD \((p = 0.026)\).\(^{31}\)

To sum up, the results regarding market outcomes confirm the theoretical predictions on score-stability but largely contradict the predictions on aptitude-stability under Assumption 4 (common knowledge of no over- or under-placement). In particular, not only do PreExam-BOS and Halfway-BOS fail to achieve aptitude-stability but they also create more severe distortions from aptitude-stability compared to Halfway-SD and PostScore-SD.

### 3.3.2 Overconfidence

To trace the reason behind the above-mentioned market failures under PreExam-BOS and Halfway-BOS, I first examine the existence and heterogeneity of subjects’ self-evaluation biases. Since rank is a much more relevant notion than score in the current setting, below I use overplacement as the primary measure of overconfidence. Recall \(r_{ai} (“AptitudeRank”)\) refers to a student’s rank of aptitude; \(\epsilon_i (“ExamError”)\) refers to an exam’s measurement error in terms of rank; \(\theta_{iEA} (“OverconfidenceEA”)\) and \(\theta_{iIN} (“OverconfidenceIN”)\) are defined

\(^{30}\) All the proportion tests comparing PreExam-BOS (or Halfway-BOS) to PreExam-SD, to Halfway-SD, and to PostScore-SD yield a \(p\)-value smaller than 0.001. The McNemar’s test comparing PreExam-BOS (or Halfway-BOS) to PostScore-BOS yields a \(p\)-value smaller than 0.001.

\(^{31}\) The \(p\)-values are from proportion tests.
as a subject’s level of overplacement at the ex-ante stage and the interim stage.\textsuperscript{32}

**Hypothesis 2.** (i) No student exhibits any over- or under-confidence. (ii) Every student believes that other students exhibit no over- or under-confidence.

**Result 3.** At both ex-ante and interim stages, (i) students exhibit overconfidence on average; (ii) men exhibit more overconfidence than women.

The average level of overplacement is 0.50 rankings at the ex-ante stage and is 0.26 at the interim stage. Since both values are significantly greater than zero ($p < 0.001$, t tests), students exhibit overconfidence at both stages. Moreover, $\theta_{EA}^i$ is significantly larger than $\theta_{IN}^i$ on average ($p < 0.001$, paired t and sign test).\textsuperscript{33} Figure 4 compares the distributions of $\theta_{EA}^i$ and $\theta_{IN}^i$ to the distribution of the exam’s measurement error $\epsilon_i$. While all three variables exhibit similar variances, $\epsilon_i$ has a significantly larger mass on zero compared to $\theta_{EA}^i$ ($p < 0.001$, McNemar’s test) or $\theta_{IN}^i$ ($p = 0.042$, McNemar’s test). This provides us with

\textsuperscript{32}In Section 3.3.2, data from all six treatments are pooled together because in each treatment, subjects’ guessed exam results are elicited at both ex-ante and interim stages. Moreover, as shown in Table 3, there is generally no significant treatment effect on overconfidence.

\textsuperscript{33}There is no clear evidence that learning can reduce or eliminate self-evaluation biases. See a detailed analysis in Appendix B.1.
some intuition behind Result 2: compared to the exam’s measurement error, the behavioral error due to self-evaluation biases could lead to more mismatched pairs, that is, more severe distortions from aptitude-stability.\footnote{Since overconfidence directly skews the sorting in preference submission, not only the heterogeneity but also an overall tendency in self-evaluation biases will lead to distortions from aptitude-stability. On the other hand, the exam’s measurement error in terms of rank has zero mean by construction.}

To understand which factors can influence and thus predict a student’s overconfidence level, I run an OLS regression of $Overconfidence_{EA}$ and $Overconfidence_{IN}$, with the data clustered by subject. The results are displayed in Table 3 and briefly summarized as follows. First, men are more overconfident than women (the marginal effect of $Female$ at the mean of $RiskAverse$ is $-0.184$ at the ex-ante stage and is $-0.113$ at the interim stage; such a negative effect is more significant for more risk-averse subjects). Second, at the ex-ante stage, those who are less risk averse tend to be more overconfident, and the coefficient of the interaction term $Female \times RiskAverse$ indicates this effect is mainly driven by men. Third, students with lower aptitudes (larger values of $AptitudeRank$) exhibit more overconfidence.\footnote{The correlation between aptitude and overconfidence may be partially driven by a ceiling effect: the student with the highest aptitude rank cannot have a positive level of overconfidence. See a similar remark on the field result in Footnote 50.}

There are no consistent treatment effects (except the 10%-level significance of $PreExam-BOS$ and the 5%-level significance of $PostScore-BOS$ on $Overconfidence_{IN}$).

Recall at the end of the experiment, each subject is asked to estimate the levels of overplacement $\theta_{j}^{EA}$ and $\theta_{j}^{IN}$ for a randomly drawn other subject $j$; I refer to the estimates for $\theta_{j}^{EA}$ and $\theta_{j}^{IN}$ as “$GuessedOther_{EA}$” and “$GuessedOther_{IN}$”. Result 4 suggests that subjects are not aware of the general tendency of overplacement.

\textbf{Result 4.} \textit{At both ex-ante and interim stages, subjects underestimate other students’ average level of overconfidence.}

Without any significant treatment effect, the average level of $GuessedOther_{EA}$ is 0.03 and that of $GuessedOther_{IN}$ is $-0.21$. Comparing to the mean of $\theta_{i}^{EA}$ (0.50) and that of $\theta_{i}^{IN}$ (0.26), we conclude that on average, subjects underestimate others’ overconfidence level at both ex-ante and interim stages ($p < 0.001$, t tests). Such a result provides evidence for Assumption 4’, which could be explained by the unawareness of one’s own bias, together with the false-consensus effect.
Table 3: Predicting Factors of Overconfidence (OLS)

<table>
<thead>
<tr>
<th>Dep. Var.</th>
<th>OverconfidenceEA</th>
<th>OverconfidenceIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-1.008** (0.430)</td>
<td>-0.974* (0.500)</td>
</tr>
<tr>
<td>RiskAverse</td>
<td>-0.074*** (0.020)</td>
<td>-0.063*** (0.021)</td>
</tr>
<tr>
<td>Female × RiskAverse</td>
<td>0.067** (0.033)</td>
<td>0.070* (0.037)</td>
</tr>
<tr>
<td>AptitudeRank</td>
<td>0.603*** (0.044)</td>
<td>0.525*** (0.046)</td>
</tr>
<tr>
<td>PreExam-BOS</td>
<td>0.053 (0.167)</td>
<td>0.321* (0.183)</td>
</tr>
<tr>
<td>Halfway-BOS</td>
<td>-0.114 (0.167)</td>
<td>0.154 (0.188)</td>
</tr>
<tr>
<td>PostScore-BOS</td>
<td>0.053 (0.175)</td>
<td>0.388** (0.185)</td>
</tr>
<tr>
<td>Halfway-SD</td>
<td>0.022 (0.182)</td>
<td>0.067 (0.183)</td>
</tr>
<tr>
<td>PostScore-SD</td>
<td>-0.133 (0.136)</td>
<td>0.133 (0.140)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.330 (0.316)</td>
<td>-0.714** (0.326)</td>
</tr>
</tbody>
</table>

Observations 315 315

Notes: Robust standard errors are shown in parentheses, allowing for clustering by subject. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. In the regression, Female is a dummy variable that equals 1 for a female subject and 0 otherwise; RiskAverse is the total number of safe choices made by a subject during risk attitude elicitation; Female × RiskAverse is the interaction term between Female and RiskAverse; PreExam-BOS, Halfway-BOS, PostScore-BOS, Halfway-SD, and PostScore-SD are dummy variables for the corresponding treatments. The descriptive statistics of key variables are summarized in Table 10 of Appendix C.

3.3.3 Preference Submission

Besides belief biases, subjects’ strategic deviations can also affect the performance of a mechanism. Recall in Section 2.4, I discussed three strategies under different treatments: truth-telling under PreExam-, Halfway-, and PostScore-SD; score-based sorting under PostScore-BOS; and guess-based sorting under PreExam- and Halfway-BOS. From the results below, we can see all three strategies are common in the experimental data.

**Hypothesis 3.** Students report their preferences truthfully under PreExam-, Halfway-, and PostScore-SD, exhibit score-based sorting under PostScore-BOS, and exhibit guess-based sorting under PreExam-BOS and Halfway-BOS.

**Result 5.** Regardless of the timing of preference submission, more than 80% of the students report their preferences truthfully under SD, while more than 80% of the students misrepresent their preferences under BOS.
Figure 5: Truth-telling in Preference Submission

Figure 5 summarizes the proportions of truth-telling subjects under different mechanisms. As predicted by the model, truthful revelation dominates under SD, while preference misrepresentation dominates under BOS. The following result describes the general patterns of misrepresentation under BOS.

**Result 6.** 
(i) Under PostScore-BOS, about 70% of the students exhibit score-based sorting. 
(ii) Under PreExam-BOS and Halfway-BOS, students tend to exhibit guess-based sorting or adopt slightly more aggressive strategies than guess-based sorting.

To identify score-based or guess-based sorting under BOS, I focus on the variable $FirstChoice$, which is given by the index of the college listed on top of one’s submitted preferences. For example, $FirstChoice = 3$ if a subject chooses college $c_3$ as her first choice. Moreover, recall the variables $r_{s_i}$ (“Rank”), $\hat{r}_{EA}^{i}$ (“GuessedRankEA”), and $\hat{r}_{IN}^{i}$ (“GuessedRankIN”) are defined as one’s realized rank, guessed rank at the ex-ante stage, and guessed rank at the interim stage, respectively.

Under PostScore-BOS, a subject is said to exhibit score-based sorting in the experiment if $FirstChoice = r_{s_i}$, because the index of her score-stably matched college equals $r_{s_i}$. Figure 6c is a bubble chart that shows the relationship between $FirstChoice$ and $r_{s_i}$ under PostScore-BOS; the size of each bubble is determined by frequency. We can see that a majority of the data is on the 45-degree line, meaning most subjects (71.67%) exhibit score-based sorting. The bubbles under the 45-degree line represents those who adopt a more aggressive strategy since $FirstChoice < r_{s_i}$, that is, the college of one’s first choice is more desirable than her score-stable match. Most students with such a strategy are ranked 5th
in the exam.\footnote{This confirms the theoretical prediction that under the current setting, students ranked 5th in the exam are indifferent among all strategies in equilibrium (see the proof of Proposition 2 in Appendix A).}

Under PreExam-BOS, a student is said to exhibit guess-based sorting if \( \text{FirstChoice} = \hat{r}^{EA}_{s_i} \). 61.67\% of the subjects use this strategy (see bubbles on the 45-degree line of Figure 6a). We also observe a considerable mass (30\%) on \( \text{FirstChoice} = \hat{r}^{EA}_{s_i} - 1 \), indicating a slightly more aggressive strategy than guess-based sorting. Similar patterns are observed under Halfway-BOS (Figure 6b). While 46.67\% of the students adopt guess-based sorting with \( \text{FirstChoice} = \hat{r}^{IN}_{s_i} \), 35\% of them exhibit \( \text{FirstChoice} = \hat{r}^{IN}_{s_i} - 1 \).

The aforementioned aggressive strategic choices suggest that subjects tend to be overoptimistic about the extent of competition in the market. In an environment with more uncertainty, like PreExam- or Halfway-BOS, they appear to be gambling on the chance that no others will choose more desirable colleges as their first choice, leaving them the oppor-
tunity to get in. From the subsequent regression analysis (Table 4), we will be able to examine whether such behaviors are related to one’s beliefs about other students’ over- or under-confidence levels.

**Result 7.** (i) Under PreExam-BOS and Halfway-BOS, a subject’s first choice in preference submission is predicted by her aptitude rank and overconfidence level. (ii) Under PostScore-BOS, a subject’s first choice is predicted by her aptitude rank and the exam’s measurement error.

Table 4 displays the results for ordered logit regressions of FirstChoice under (1) PreExam-BOS, (2) Halfway-BOS, and (3) PostScore-BOS. Regression (3) shows significant effects of both AptitudeRank and ExamError on FirstChoice under PostScore-BOS. Since by definition, an exam outcome is composed of one’s aptitude together with a measurement error, such a result echoes the fact that a majority of subjects exhibit score-based sorting (Result 6).

Now I focus on treatments PreExam-BOS and Halfway-BOS. First, by regressions (1) and (2), a subject with a higher level of overconfidence (an increase in OverconfidenceEA or OverconfidenceIN) or a better rank of aptitude (a decrease in AptitudeRank) tends to choose a more desirable college as the first choice (a decrease in FirstChoice). Significant marginal effects for each outcome are displayed in Table 11 of Appendix C. For example, under PreExam-BOS, a subject is 30.1% more likely to choose the best college $c_1$ as her first choice if her overplacement is increased by one rank; she is 26.9% less likely to choose $c_1$ if her aptitude is placed one rank worse in the market. Hence, not only aptitudes, but also overconfidence levels enter students’ strategic choices, thus influencing the performance of these two mechanisms.

Second, by regressions (1) and (2), GuessedOtherEA, GuessedOtherIN, and RiskAverse do not have any significant influence on FirstChoice. Recall that under PreExam- and Halfway-BOS, a considerable number of subjects adopt more aggressive strategies than guess-based sorting. Apparently, such behaviors are not correlated with risk attitudes and cannot be rationalized by beliefs on others’ overconfidence levels. Therefore, optimism is displayed on two levels: not only are subjects overconfident in guessing the exam outcomes, but also they tend to “shoot for the stars” in preference submission.

\footnote{In the regressions, I exclude all variables that a subject does not observe at the time of preference submission. For example, OverconfidenceIN is excluded from regression (1), because preferences are submitted before the guess at the interim stage. The conclusions remain unchanged if these variables are included.}
### Table 4: First Choice in Preference Submission (Ordered Logit)

<table>
<thead>
<tr>
<th>Dep. Var.</th>
<th>(1) PreExam-BOS</th>
<th>(2) Halfway-BOS</th>
<th>(3) PostScore-BOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OverconfidenceEA</td>
<td>-2.687***</td>
<td>-1.320***</td>
<td>-0.507</td>
</tr>
<tr>
<td></td>
<td>(0.523)</td>
<td>(0.448)</td>
<td>(0.516)</td>
</tr>
<tr>
<td>OverconfidenceIN</td>
<td>-0.984**</td>
<td>-0.704</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.379)</td>
<td>(0.494)</td>
<td></td>
</tr>
<tr>
<td>AptitudeRank</td>
<td>2.391***</td>
<td>2.214***</td>
<td>4.691***</td>
</tr>
<tr>
<td></td>
<td>(0.482)</td>
<td>(0.429)</td>
<td>(0.958)</td>
</tr>
<tr>
<td>ExamError</td>
<td></td>
<td></td>
<td>-4.307***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.848)</td>
</tr>
<tr>
<td>GuessedOtherEA</td>
<td>0.246</td>
<td>0.120</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.207)</td>
<td>(0.277)</td>
</tr>
<tr>
<td>GuessedOtherIN</td>
<td>-0.131</td>
<td>0.595*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.240)</td>
<td>(0.360)</td>
<td></td>
</tr>
<tr>
<td>RiskAverse</td>
<td>0.054</td>
<td>-0.049</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.065)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>Observations</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

Notes: Standard errors are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

### 3.3.4 Individual Welfare

To measure how much a student’s welfare is distorted from the aptitude-stable matching, I define the variable *WelfareDistortion* as *AptitudeRank*—which equals the index of one’s aptitude-stably matched college—minus the index of one’s currently matched college. A positive value of *WelfareDistortion* indicates that a student is allocated to a college with a smaller index than her aptitude-stable match, which means the current mechanism is giving her an “unfair” advantage that cannot be justified by her aptitude.

The distributions of *WelfareDistortion* under different treatments are illustrated in Figure 7. We can clearly see a smaller mass at 0 under PreExam- and Halfway-BOS compared to Halfway- and PostScore-SD. Such a conclusion has already been drawn in Result 2, stating that PreExam- and Halfway-BOS tend to yield a smaller proportion of aptitude-stably matched pairs. In addition, the distribution under PostScore-SD exhibits a smaller variance compared to those under PreExam-BOS ($p = 0.006$, variance ratio test) and Halfway-BOS ($p = 0.040$, variance ratio test).
As shown in Section 2.4, being overconfident under PreExam- or Halfway-BOS could hurt or benefit a student’s welfare, depending on the distribution of overconfidence and the realization of the exam’s measurement error in the market. Figure 8 helps us to take a first look at the relationship between overconfidence and individual welfare in the experimental data. Both graphs exhibit a generally positive correlation, which means PreExam- and Halfway-BOS tend to reward those who are overconfident and punish those who are underconfident. Combined with the fact that men are more overconfident than women (Result 3), we can conclude males tend to receive an unfair advantage under these two mechanisms. Figure 9 clearly shows that the gender difference in overconfidence (Figure 38) This conclusion also uses the fact that gender does not impose a direct effect on WelfareDistortion (see
9a) is translated into a gender penalty for women in terms of individual welfare (Figure 9b) under PreExam- and Halfway-BOS.

![Figure 9: Gender Penalty under PreExam-BOS and Halfway-BOS](image)

To obtain specific marginal effects, I run an OLS regression of WelfareDistortion under (1) PreExam-BOS, (2) Halfway-BOS, and (3) PostScore-BOS. The main results are summarized as follows.

**Result 8.** On an individual level, PreExam-BOS and Halfway-BOS create more severe and more varied distortions from aptitude-stability than PostScore-SD. Such distortions are affected by one’s overconfidence level and strategic choice, as well as the exam’s measurement error. Specifically, a student tends to be matched to a better college if

(i) she exhibits a higher level of overconfidence,

(ii) she performs better in the exam, or

(iii) she adopts a more aggressive strategy in preference submission.

Regression (3) in Table 12 shows that ExamError significantly and positively affects WelfareDistortion. This means under PostScore-BOS, a student’s performance on one exam has a direct influence on her welfare, which is not surprising given the strong evidence for score-based sorting (Result 6).

Recall in Section 2.4, I use Example 1 to illustrate that under PreExam- and Halfway-BOS, the presence of overconfidence can cause welfare distortions both directly (it skews the sorting in preference submission) and indirectly by bringing back the noise of the exam’s measurement error (it creates conflicts in submitted preferences, thus forcing BOS to resolve regressions (1) and (2) in Table 5).

---

39Tables 12 and 13 in Appendix C present the ordered logit regressions of WelfareDistortion and the significant marginal effects.
them using exam scores). Such an intuition is well supported by regressions (1) and (2), where both OverconfidenceEA (or OverconfidenceIN) and ExamError impose a significant influence on WelfareDistortion under PreExam-BOS (or Halfway-BOS). On average, if a student’s level of overplacement is increased by one, her WelfareDistortion rises by 0.493 under PreExam-BOS and by 0.789 under Halfway-BOS; if a student’s score rank in the exam is increased by one, her WelfareDistortion increases by 0.379 under PreExam-BOS and by 0.288 under Halfway-BOS.\footnote{See Appendix B.2 for a detailed discussion on the correlation between overconfidence and the exam’s measurement error.}

<table>
<thead>
<tr>
<th>Dep. Var.</th>
<th>(1) PreExam-BOS</th>
<th>(2) Halfway-BOS</th>
<th>(3) PostScore-BOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OverconfidenceEA</td>
<td>0.493***</td>
<td>-0.217</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.252)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>OverconfidenceIN</td>
<td></td>
<td>0.789***</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.250)</td>
<td>(0.084)</td>
</tr>
<tr>
<td>ExamError</td>
<td>0.379**</td>
<td>0.288**</td>
<td>0.943***</td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.123)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>AggressiveStrategy</td>
<td>0.689***</td>
<td>0.423*</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(0.259)</td>
<td>(0.224)</td>
<td>(0.131)</td>
</tr>
<tr>
<td>GuessedOtherEA</td>
<td>0.172</td>
<td>0.166</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.120)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>GuessedOtherIN</td>
<td>-0.089</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.059)</td>
<td></td>
</tr>
<tr>
<td>RiskAverse</td>
<td>0.020</td>
<td>0.057</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.039)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.009</td>
<td>-0.304</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>(0.371)</td>
<td>(0.330)</td>
<td>(0.144)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.772</td>
<td>-0.834*</td>
<td>-0.288</td>
</tr>
<tr>
<td></td>
<td>(0.523)</td>
<td>(0.493)</td>
<td>(0.197)</td>
</tr>
</tbody>
</table>

Notes: Standard errors are shown in parentheses. *** , **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Besides biases in beliefs, we also observe a considerable proportion of students adopt strategies other than guess-based sorting under PreExam- and Halfway-BOS.\footnote{Figure 18 in Appendix C categorizes observations on two dimensions: (i) whether a subject’s belief is unbiased, underconfident, or overconfident, and (ii) whether her strategy follows guess-based sorting, is more...}

81
the extent of such a deviation, I define the variable \textit{AggressiveStrategy} as \textit{GuessedRankEA} (or \textit{GuessedRankIN}) minus \textit{FirstChoice} under PreExam-BOS (or Halfway-BOS). A positive value of \textit{AggressiveStrategy} indicates a strategy more aggressive than guess-based sorting, because the college of one’s first choice is more desirable than her score-stable match. According to regressions (1) and (2), a more aggressive strategy tends to have a significant and positive effect on individual welfare. On average, a unit increase in \textit{AggressiveStrategy} raises \textit{WelfareDistortion} by 0.689 under PreExam-BOS and by 0.423 under Halfway-BOS. Under PostScore-BOS, \textit{AggressiveStrategy} is defined as \textit{Rank} minus \textit{FirstChoice} and measures the extent of a deviation from score-based sorting. Since a majority of subjects adopt the equilibrium strategy under PostScore-BOS, \textit{AggressiveStrategy} does not exhibit a significant effect in regression (3).

3.4 No Effects of Additional Information

The above results clearly suggest that under PreExam- and Halfway-BOS, overconfidence serves as a major obstacle to the implementation of an aptitude-stable matching. In this section, I explore whether an improved information condition could help to reduce overconfidence and thus enhance the performance of these two mechanisms. While a detailed description is given in Appendix B.3, below I briefly introduce the design and summarize the main results and implications.

In the new environment, after being re-grouped at the beginning of each treatment, every subject is provided with all the past performances of her new group members, including their scores in the three mock tests and in all the exams they have taken in the previous treatments. The average score of each member is calculated and displayed as well.\footnote{Recall in each session of the original design, about 15 participants are randomly divided into groups of five and are re-grouped for every mock test and every treatment. From the results of the three mock tests, subjects should be able to obtain a relatively clear picture of their aptitude ranking. However, for each treatment they do not receive any specific information on the other four group members. Under such a setting, there exist two major sources of overconfidence: overconfidence about the group composition (“I might be grouped with less smart people in this exam”) and overconfidence about one’s own performance (“I can score higher in this exam”). The new information condition is essentially muting the former while keeping the latter.}

The data show that the provision of such additional information has very little influence over subjects’ levels of overconfidence, their strategic behaviors, and the market outcomes. This is a rather negative result since it indicates that almost all of the biases observed before stem from one’s belief about herself and thus cannot be reduced even with very detailed information regarding the rest of the market. Therefore, it posts a even bigger challenge to PreExam- and Halfway-BOS in some field environments. For example, in Chinese college admissions, since mock tests are mostly created and organized by different high schools, a
student cannot obtain direct information on the percentile of her score in the entire market (see Section 4.2). In this case, the conclusion of this section raises the possibility that even with perfect information revelation, the performance of PreExam- and Halfway-BOS still cannot be improved because self-evaluation biases stay on the same level.

4 Field Evidence from Chinese College Admissions

One of the most important features of Chinese college admissions lies in the strikingly high stakes involved in students’ self-evaluation and preference submission process. Hence, the first question we should ask is whether self-evaluation biases like overconfidence continue to exist in such a setting. Next, I analyze the field data on the distribution of overconfidence and the realization of the exam’s measurement error, which helps us to infer the welfare comparisons across different mechanisms in the specific market.

4.1 Chinese College Admissions

According to statistics released by the Ministry of Education of the People’s Republic of China, in the year 2014, about 9,390,000 applicants competed for seats at 2,246 higher education institutions. While the admission rate for these institutions was around 74%, it fell to 39% for universities that offer a bachelor’s degree, and to about 2% for the top 39 universities.

In this centralized matching procedure, every college has an identical priority ordering over students, which is fully determined by their score ranking on a single standardized exam: the college entrance exam, also known as gaokao. Each year, high school graduates take the exam held by their residential districts and submit a preference list over colleges. Each district makes its own admissions policy. Students can choose colleges outside their own districts. But because for each district, the capacity (or “quota”) of each college is predetermined and announced in advance, the college admissions market in every district is an independent market.

Since its introduction in 1952, the centralized procedure has undergone frequent reforms.\textsuperscript{43} In addition, the specific matching mechanism varies across the country, mainly in two dimensions: the matching algorithm and the timing of preference submission. There are two primary the matching algorithms – a sequential and a parallel algorithm – and three primary timings of preference submission – pre-exam, halfway, and post-score.

The sequential algorithm is very similar to BOS; it tries to accommodate as many students as possible into their reported first choices. A parallel algorithm, on the other hand, is a combination of BOS and SD. Under such an algorithm, students’ preference lists

\textsuperscript{43}See Chen and Kesten (2013) for a more detailed description.
are composed of three or four tiers. While SD is applied within each tier, BOS is applied between tiers. Chen and Kesten (2013) show that although a parallel algorithm is still not fully strategy-proof, it is more strategy-proof than a sequential algorithm.\footnote{Chen and Kesten (2013) characterize a parallel mechanism as a combination between BOS and Deferred Acceptance (DA). In the context of Chinese college admissions with a unique priority ordering of students, SD is equivalent to DA. Here I choose to use the SD specification since it is more similar to the official description of a parallel mechanism.}

In recent years, Chinese college admissions are in the transition from a sequential mechanism with pre-exam or halfway timing to a parallel mechanism with post-score timing. By 2014, the parallel algorithm had been introduced to almost all districts in China. As for the timing of preference submission, in 2014, pre-exam was only used in Shanghai and Beijing; halfway was only used in Xinjiang; all the remaining districts used post-score.\footnote{In 2015, Xinjiang and Beijing also changed to the post-score timing.} Therefore, I collected data from Shanghai and Xinjiang to investigate the pre-exam and halfway timings, respectively.\footnote{Xinjiang and Shanghai have different cultural backgrounds. Therefore, this paper does not intend to directly compare the two field studies and makes no claim that any difference in the two samples stems only from their different timings of preference submission. It only provides better understanding of these two markets separately.}

4.2 Data

Shanghai was under the pre-exam timing in 2014. The data collection mainly took place in May at Shanghai Pengpu High School. Students submitted their preferences about three weeks before the college entrance exam. In the meantime, their guessed exam results were elicited for the purpose of this study.\footnote{The official preference submission procedure was conducted online. However, students were also asked to submit written copies to the school in order to get feedback and advice from their teachers. The forms used to elicit their guessed exam results were distributed together with the empty preference lists for them to fill out. Students’ guesses are not monetary incentivized.} To evaluate students’ academic aptitudes, scores from seven mock tests were also collected. Follow-up data on exam results were obtained after the revelation in June. The sample size is 95, including 40 male and 55 female students.

In Xinjiang, the data were collected in June at No. 6 High School of Kuerle City. The procedure resembles that in Shanghai: students’ guessed exam results were elicited at the time of their preference submission. However, under the halfway timing, it took place after the college entrance exam but before the revelation of exam results. Three mock test scores and the exam results were also obtained.\footnote{Since mock tests are organized by different high schools, the number of mock tests varies across schools and districts.} The sample size is 119, including 54 male and 65 female students.

In the field environment, mock tests are mostly created and organized by different high schools. Thus, students cannot obtain direct information on the relative standing of their
aptitudes in the entire market. Since the exam score distribution stays relatively stable from year to year, most students infer such information by fitting their guessed scores into the distribution from the previous year. Therefore, below I mainly discuss the results in terms of scores rather than ranks and use overestimation as the primary measure of overconfidence.

4.3 Results

Since the difficulty and thus the score distribution can vary across mock tests and the college entrance exam, for each sample all mock test scores are normalized so that the distributions have the same mean and variance as the exam score distribution. A student’s Aptitude is then evaluated as the average of all her mock test scores and her exam score. Moreover, the variable ExamError refers to the exam’s measurement error in terms of score; OverconfidenceEA and OverconfidenceIN are given by a student’s level of overestimation at the ex-ante stage and the interim stage. Since the total score differs in Shanghai and Xinjiang, for the convenience of comparison, I report the relative values of Aptitude, ExamError, OverconfidenceIN, and OverconfidenceEA to the standard deviation of Aptitude.

Result 9. (i) Under the pre-exam timing in Shanghai, students exhibit overconfidence at the ex-ante stage; the variance of such biases is larger than the variance of the exam’s measurement error. (ii) Under the halfway timing in Xinjiang, students exhibit underconfidence at the interim stage; the variance of such biases is not significantly different from the variance of the exam’s measurement error.

Under the pre-exam timing in Shanghai, students’ average level of overestimation at the ex-ante stage is given by 0.185 times the standard deviation of Aptitude, which is significantly greater than 0 ($p = 0.004$, t test). Figure 10a compares the distribution of OverconfidenceEA with the distribution of ExamError. We can see that OverconfidenceEA exhibits a larger variance than ExamError ($p < 0.001$, variance ratio test). On the other hand, under the halfway timing in Xinjiang, the mean of OverconfidenceIN is -0.249, significantly lower than 0 ($p < 0.001$, t test). Moreover, OverconfidenceIN and ExamError do not exhibit significant different variances (Figure 10b).49

49Since Shanghai and Xinjiang have very different cultural backgrounds, we cannot simply attribute the difference in overconfidence to the difference in the timing of preference submission. In fact, this paper does not intend to make any direct comparison between the two field samples; it only provides better understanding of these two markets separately.
The following two results summarize the predicting factors of a student’s overconfidence level in Shanghai and Xinjiang.

**Result 10.** Under the pre-exam timing in Shanghai, students are heterogeneous in overconfidence at the ex-ante stage, which can be partially predicted by:

(i) gender: female students are more overconfident than male students;

(ii) aptitude level: those who have lower aptitudes exhibit more overconfidence.

**Result 11.** Under the halfway timing in Xinjiang, students are heterogeneous in overconfidence at the interim stage, which can be partially predicted by:

(i) gender: male students are more overconfident than female students;

(ii) aptitude: those who have lower aptitudes exhibit more overconfidence;

(iii) ethnic group: students of the Uyghur ethnic group exhibit more overconfidence.

I run an OLS regression of \( \text{Overconfidence}_{EA} \) using the Shanghai sample and a regres-
In Shanghai, the average overestimation for female students is higher than that for male students by 0.289 at the mean of Aptitude; the interaction term Female × Aptitude indicates such a difference is smaller for those who have higher aptitudes. Moreover, students with lower aptitudes exhibit more overconfidence: when one’s aptitude is increased by 1, a male student’s overestimation decreases by 0.241 on average, while a female student’s overestimation decreases by 0.514 on average.

Table 6: Overconfidence in Shanghai and Xinjiang (OLS)

<table>
<thead>
<tr>
<th>Dep. Var.</th>
<th>Overconfidence(\delta^EA_{i}) (\text{(Shanghai)})</th>
<th>Overconfidence(\delta^IN_{i}) (\text{(Xinjiang)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>3.521*** (1.272)</td>
<td>-1.550*** (0.518)</td>
</tr>
<tr>
<td>Uyghur</td>
<td></td>
<td>2.238*** (0.519)</td>
</tr>
<tr>
<td>Aptitude</td>
<td>-0.241*** (0.077)</td>
<td>-0.249*** (0.063)</td>
</tr>
<tr>
<td>Female × Aptitude</td>
<td>-0.273** (0.107)</td>
<td>0.235*** (0.085)</td>
</tr>
<tr>
<td>Uyghur × Aptitude</td>
<td>-0.294*** (0.086)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.872*** (0.927)</td>
<td>1.034*** (0.374)</td>
</tr>
</tbody>
</table>

Observations 95 119

Notes: Standard errors are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

In Xinjiang, the marginal effect of Female at the mean of Aptitude is −0.151; such a negative effect is less significant as Aptitude increases. Second, when one’s aptitude is improved by 1, a male student’s overestimation decreases by 0.249, while a female student’s overestimation decreases by 0.014 on average.\(^50\) Third, since Xinjiang Region has a significant population of the Uyghur minority ethnic group, about 50% of students in the sample took the exam in Uyghur language. The result shows these students exhibit a higher average overestimation by 0.483 at the mean of Aptitude, compared to the other students, mostly of the Han majority.

From Results 9 to 11, we can clearly see that distributions and predicting factors of students’ self-evaluation biases could vary dramatically across different markets. We even observe overall underconfidence in Xinjiang and women being more overconfident than men.

\(^50\) The correlation between aptitude and overconfidence in both Shanghai and Xinjiang can be partially driven by a ceiling effect: since every exam has a perfect score, a student with the highest possible aptitude cannot have a positive level of overconfidence.
in Shanghai, which contradicts most findings in the literature of overconfidence. Therefore, in studying a specific market, we may need to tailor our choice of matching mechanism to the behavioral attributes of its participating agents.

In both samples collected from Chinese college admissions, students exhibit overall biases in self-evaluation, as well as significant heterogeneity in the magnitudes of their biases. Since a similar pattern is observed in the lab data, this suggests that in a field setting, PreExam- or Halfway-BOS may also create more severe and more varied distortions from aptitude-stability compared to PostScore-SD. This provides a potential explanation for the recent reforms in China’s college admissions policy: most districts are currently in transition from a mechanism that resembles PreExam-BOS into a mechanism more similar to PostScore-SD.

5 Conclusion

Many centralized college admissions markets adopt a standardized exam to evaluate students’ aptitudes and determine their priorities in the matching procedure. Since every exam entails a measurement error, the exam-based priorities can only serve as a noisy proxy for colleges’ aptitude-based preferences. The previous literature suggests the effect of this noise can be diminished with a “PreExam-BOS” mechanism (a Boston algorithm combined with preference submission before the exam). Using a lab experiment, I conclude otherwise: (i) since pre-exam preference submission is skewed by overconfidence, PreExam-BOS creates more severe and more varied welfare distortions than the PostScore-SD mechanism (a Serial Dictatorship algorithm combined with preference submission after the revelation of exam results); (ii) PreExam-BOS introduces unfairness by rewarding overconfidence and punishing underconfidence, thus serving as a gender penalty for women.

In a field investigation on actual Chinese students, I find similar behavioral biases such as overconfidence. This suggests a potential explanation for China’s recent policy reform from a mechanism that resembles PreExam-BOS to a mechanism more similar to PostScore-SD. Admittedly, the field study only provides us with belief biases but not strategic patterns, which limit its ability to give general predictions for the complex field environment. But it is not the main purpose of this paper. Instead, this study intends to introduce a behavioral perspective and to present a tradeoff that can emerge in real markets. Further efforts should

51Because students’ true preferences are difficult to elicit in the field setting, in order to make welfare comparison across different mechanisms, I run a simulation using the field data and the strategic patterns observed in the lab. The results show that PreExam- or Halfway-BOS tend to create more severe and more varied distortions from aptitude-stability compared to PostScore-SD. Although such a simulation method has its limitations in predicting for the complex field environment, it still provides us with important intuitions regarding whether the observed biases could generate similar welfare distortions as in the lab. See details of the simulation in the working paper version at http://siqipan.weebly.com/uploads/5/1/4/8/51488789/instability_with_overconfidence_workingpaper.pdf
be made for a more thorough field investigation.

Although preferable to PreExam-BOS, PostScore-SD is still largely affected by the exam’s measurement error. This implies the challenge of obtaining a fair market outcome when students are evaluated with a single standardized exam. Thus, it raises the need for policymakers to weigh the benefits and costs when adopting such a noisy evaluation system. In practice, similar systems exist in various environments including public school choice, college admissions, as well as labor market clearinghouses. Such prevalence calls for more research in the future. In addition to the behavioral aspects considered in this paper, more issues like heterogeneous preferences, asymmetric information, and constrained choices in preference submission should be added into the discussion.
References


Appendix A: Proofs

A.1. Proofs of Chapter 1

Proof of Proposition 1

Proof. Consider a worker who has received an early open offer from firm $f_j$ ($j = 1, 2, \ldots, F - 1$). As long as a better firm $f_{j'}$ ($j' > j$) moves in a later period in equilibrium, she strictly prefers holding to accepting or rejecting the offer right away.

Firstly, holding is preferred to accepting. By choosing to hold but not to accept, the worker is strictly better off if she receives a better offer in a later period. If not, she is not worse off since she still has the open offer from $f_j$. Secondly, holding is preferred to rejecting. By choosing to hold but not to reject, the worker is not worse off if she receives a better offer in a later period; she is strictly better off otherwise.

Knowing this, the best firm $f_F$ strictly prefers to wait until the last period, so that all workers will stay in the market and the one of the best quality can be perfectly identified. Since no offer is accepted in Period 1, the other firms cannot make themselves better off by moving early; instead, they incur the risk of being rejected in the last period and remaining unmatched.

Hence, there always exist profitable deviations from an equilibrium with partial or full unraveling. Only an equilibrium without unraveling can sustain. It yields the assortative matching according to the true ranking of workers $\succ$, which is the unique stable matching in the current environment with strict rankings and aligned preferences.

Proof of Lemma 1

Proof. When making an offer in Period 2, every firm is indifferent because an open offer is equivalent to an exploding offer. Now consider an offer in Period 1. The best firm $f_F$ is still indifferent since neither an exploding offer nor an open offer will be rejected by any worker. However, for other firms, making an open offer is never the strictly best response.

First, when making an exploding offer that has to be accepted within the same period, a firm always knows whether it will be accepted. This is because there is no information
asymmetry in the current setting, and the information status remains the same within a period.

Next, if an exploding offer will be accepted, making an exploding offer yields the same or a higher payoff than making an open offer. Suppose a firm makes an open offer in Period 1, there are three possible responses: (i) it is accepted right away; (ii) it is held and accepted in Period 2; (iii) it is held and rejected in Period 2. Compared to an exploding offer, the firm yields the same payoff in cases (i) and (ii), but is strictly worse off in case (iii).

Finally, if an exploding offer will be rejected, waiting yields the same or a higher payoff than making an exploding offer or an open offer. A firm never wants to make an exploding offer knowing it will be rejected because then it cannot make an offer to the same worker again. On the other hand, if a firm makes an open offer in Period 1, there are three possible responses: (i) it is rejected right away; (ii) it is held and accepted in Period 2; (iii) it is held and rejected in Period 2. Compared to waiting, the firm is strictly worse off in cases (i) and (iii), and is weakly worse off in case (ii).

\[ \square \]

Proof of Proposition 2

*Proof.* Below I first calculate the updated beliefs of firms and workers after they observe a signal in Period 1.

After observing a signal \( \hat{\succ} \), posteriors on the true state are given by

\[
\text{Pr}(\hat{\succ} | \hat{\succ}) = \frac{[\alpha + (1 - \alpha) \frac{1}{W!}] \frac{1}{W!} + (W! - 1) [(1 - \alpha) \frac{1}{W!}] \frac{1}{W!}}{[\alpha + (1 - \alpha) \frac{1}{W!}] \frac{1}{W!} + (W! - 1) [(1 - \alpha) \frac{1}{W!}] \frac{1}{W!}} = \alpha + \frac{1 - \alpha}{W!}.
\]

For any \( \succ' \neq \hat{\succ} \),

\[
\text{Pr}(\succ' | \hat{\succ}) = \frac{1 - \frac{\alpha W! + (1 - \alpha)}{W!} + (W! - 1) [(1 - \alpha) \frac{1}{W!}] \frac{1}{W!}}{1 - \frac{\alpha}{W!}}.
\]

Posterior on types are given by

\[
\text{Pr}(\hat{r} | \hat{r}) = \text{Pr}(\hat{\succ} | \hat{\succ}) + [(W! - 1)! - 1] \text{Pr}(\succ' | \hat{\succ}) = \alpha + \frac{1 - \alpha}{W},
\]

and \( \forall r' \neq \hat{r} \)

\[
\text{Pr}(r' | \hat{r}) = \frac{1 - (\alpha + \frac{1 - \alpha}{W})}{W - 1} = 1 - \frac{\alpha}{W}.
\]
Thus, the expected quality of a type-$\tilde{r}$ worker is

$$EV(\tilde{r}) = \left(\alpha + \frac{1 - \alpha}{W}\right) \tilde{r} + \sum_{r' \neq \tilde{r}} \frac{1 - \alpha}{W} r' = \alpha \tilde{r} + \frac{(1 - \alpha)(W + 1)}{2}.$$ 

Next, consider the case where $W > F$. In an equilibrium without unraveling, no actions are taken in Period 1. In Period 2, after $\succ$ is revealed, $f_F$ makes an offer to $w_W$, $f_{F-1}$ to $w_{W-1}$, ..., and $f_1$ to $w_{W-F+1}$. All offers are accepted.

It is clear that workers do not have any incentive to deviate, nor does firm $f_F$. Given all others are playing the equilibrium strategy, a firm $f_j$ with $j = 1, 2, ..., F - 1$ will not deviate and make an offer to a different worker in Period 2, since it will not be accepted by a worker better than its current match $w_{j-F+W}$. I now focus on checking the deviation of $f_j$ in Period 1. Such a deviation involves both sides of the market: a firm should want to make an early exploding offer to a worker who wants to accept it. So a sufficient condition for the existence of an equilibrium without unraveling is that, for each worker type $\hat{r}$, the firms whose offer would be accepted are not willing to offer.

Suppose $f_j$ deviates by making an early offer to a type-$\hat{r}$ worker in Period 1, and $\hat{r} \neq \hat{r}(j)$. The offer is accepted if

$$j \geq \left(\alpha + \frac{1 - \alpha}{W}\right) \times \tilde{r}(\hat{r}) + \frac{1 - \alpha}{W} \times \left(\sum_{r'=W-F+1}^{W} j(r') - j(\hat{r})\right) - \frac{1 - \alpha}{W} \times 1,$$

or equivalently,

$$j \geq j_{A1}^j(\hat{r}) \equiv \alpha \times (\hat{r} - W + F) + \frac{(1 - \alpha)(F^2 + F - 2)}{2W}. \tag{2}$$

The function $j_{A1}^j(\hat{r})$ is defined as the lowest ranked firm that is accepted by type $\hat{r}$. On the other hand, if accepted, the firm is willing to make such an offer if

$$j + W - F \leq EV(\hat{r}), \tag{3}$$

or equivalently,

$$j \leq j_{O1}^j(\hat{r}) \equiv \alpha \hat{r} + \frac{(1 - \alpha)(W + 1)}{2} - W + F. \tag{4}$$

The function $j_{O1}^j(\hat{r})$ is defined as the highest ranked firm that wants to make an early offer to type $\hat{r}$. The sufficient condition for no deviation in this case is that for each type, there does not exist a firm that is willing to offer, and is accepted. That is, $\forall \hat{r}$, we need to have

$$j_{A1}^j(\hat{r}) \geq j_{O1}^j(\hat{r})$$

and $j_{O1}^j(\hat{r}) \geq 1$. \tag{5}
or
\[
j_{\hat{r}}^I(\hat{r}) < 1,
\]
which solves
\[
W \geq 2 + F.
\]
(6) ensures that there does not exist a \( j \) such that \( j^A(\hat{r}) \leq j \leq j_{\hat{r}}^I(\hat{r}) \). (6) is a boundary condition where no firms are willing to make an offer to a type-\( \hat{r} \) worker.\(^{52}\)

Now we consider the deviation of a firm \( f_j \) to its signal-suggested type \( \hat{r}(j) \). The offer is accepted if
\[
j \geq \left( \alpha + \frac{1 - \alpha}{W} \right) \times j(\hat{r}(j)) + \frac{1 - \alpha}{W} \times \left( \sum_{r'=W-F+1}^{W} j(r') - j(\hat{r}(j)) \right) - \left( \alpha + \frac{1 - \alpha}{W} \right),
\]
(8)
or equivalently,
\[
j \geq j_2^A \equiv \frac{F^2 + F - 2}{2W} - \frac{\alpha}{1 - \alpha}.
\]
If accepted, the firm is willing to make such an offer if
\[
j + W - F \leq EV(\hat{r}(j)),
\]
(10)
or equivalently,
\[
j \leq j_2^O \equiv \frac{1}{2} + F - \frac{W}{2}.
\]
The sufficient condition for no deviation in this case is that there does not exist a firm willing to make an offer to its signal-suggested type, and is accepted. That is,
\[
j_2^A \geq j_2^O \text{ and } j_2^O \geq 1,
\]
or
\[
j_2^O < 1.
\]
Combining with (7), the two sufficient conditions when \( W > F \) are given by (i) \( W \geq 2F \), or (ii) \( F + 2 < W \leq 2F - 1 \) and \( \alpha \leq \frac{(W-F)^2 - W + F - 2}{(W-F)^2 + W + F - 2} \).

Now I move on to the case where \( W \leq F \). In an equilibrium without unraveling, no actions are taken in Period 1. In Period 2, after \( \succ \) is revealed, \( f_F \) makes an offer to \( w_W \), \( f_{F-1} \) to \( w_{W-1} \), \ldots, and \( f_{F-W+1} \) to \( w_1 \). All these offers are accepted.

Suppose a firm \( f_j \) deviates by making an early offer to its signal-suggested type \( \hat{r}(j) = \hat{r} \).
\( j - F + W \). The offer is accepted if
\[
j \geq \left( \alpha + \frac{1 - \alpha}{W} \right) \times j(\hat{r}(j)) + \frac{1 - \alpha}{W} \times \left( \sum_{r'=1}^{W} j(r') - j(\hat{r}(j)) \right) - \left( \alpha + \frac{1 - \alpha}{W} \right),
\]
(12)
or equivalently,
\[
j \geq j_A^4 \equiv \frac{2F - W + 1}{2} - \frac{1}{W} - \frac{\alpha}{1 - \alpha}.
\]
(13)
The firm wants to make such an offer if
\[
j + W - F \leq EV(\hat{r}(j)),
\]
or equivalently,
\[
j \leq j_O^3 \equiv \frac{1}{2} + F - \frac{W}{2}.
\]
(15)

It is easy to show that \( j_A^3 < j_O^3 \), that is, the sufficient condition for no deviation never holds for \( W \leq F \).

Therefore, the equilibrium without unraveling always exists if (i) \( W \geq 2F \), or (ii) \( F + 2 < W \leq 2F - 1 \) and \( \alpha \leq \frac{(W-F)^2-W+F-2}{(W-F)^2+W+F-2} \).

**Proof of Proposition 3**

*Proof.* For a deviation from the equilibrium without unraveling to occur, there must exist a type of worker and a firm in Period 1 such that, the firm is willing to offer and the worker is willing to accept.

When \( W > F \), for a deviation between a non-signal-suggested pair to exist, we need \( \exists \hat{r} \) such that
\[
\bar{j}_1^O(\hat{r}) - \underline{j}_1^A(\hat{r}) > 1 \text{ if } \underline{j}_1^A(\hat{r}) \geq 0,
\]
(16)
or
\[
\bar{j}_1^O(\hat{r}) > 1 \text{ if } \underline{j}_1^A(\hat{r}) < 0.
\]
(17)
(16) ensures that the range between \( \bar{j}_1^O(\hat{r}) \) and \( \underline{j}_1^A(\hat{r}) \) is larger than 1, so that there always exists an integer in between. (17) is a boundary case where \( \underline{j}_1^A(\hat{r}) < 0 \). Then we need the range to be even larger so that at least \( f_1 \) is willing to offer.\(^{53}\) It is easy to confirm that the two conditions above never hold.

Similarly, for a deviation between a signal-suggested pair to exist, we need
\[
\bar{j}_2^O - \underline{j}_2^A > 1 \text{ if } \underline{j}_2^A \geq 0,
\]
(18)
\(^{53}\)There is another boundary where \( \bar{j}_1^O(\hat{r}) > F \), which never holds since \( f_F \) never wants to deviate.
or
\[
\bar{j}_2^O > 1 \text{ if } j_2^A < 0,
\] (19)

which solves \( F < W < 2F - 1 \) and \( \alpha > \frac{(W-F)^2+W+F-2}{(W-F)^2+3W+F-2}. \)

When \( W \leq F \), for a deviation between a signal-suggested pair to exist, we need
\[
\bar{j}_3^O - j_3^A > 1 \text{ if } j_3^A \geq 0,
\] (20)

or
\[
\bar{j}_3^O > 1 \text{ if } j_3^A < 0,
\] (21)

which solves \( W \leq F \) and \( \alpha > \frac{W-1}{2W-1} \).

Therefore, the equilibrium without unraveling never exists if (i) \( W \leq F \) and \( \alpha > \frac{W-1}{2W-1} \); or (ii) \( F < W < 2F - 1 \) and \( \alpha > \frac{(W-F)^2+W+F-2}{(W-F)^2+3W+F-2}. \)

\[\square\]

**Proof of Proposition 4**

*Proof.* Consider the case where \( W > F \).

In an equilibrium with full unraveling, after \( \hat{\succ} \) is revealed in Period 1, \( f_F \) makes an offer to type \( \hat{r} = W, f_{F-1} \) to \( \hat{r} = W - 1, \ldots \), and \( f_1 \) to \( \hat{r} = W - F + 1 \). All offers are accepted. Such an equilibrium can never sustain because after Period 1, there are \( W - F \) workers left in the market. Given all the other firms move early, each firm has an incentive to deviate to Period 2, in which case it becomes the only firm left in the market and can choose the best remaining worker.

Consider the case where \( W \leq F \).

In an equilibrium with full unraveling, after \( \hat{\succ} \) is revealed in Period 1, \( f_F \) makes an offer to type \( \hat{r} = W, f_{F-1} \) to \( \hat{r} = W - 1, \ldots \), and \( f_{F-W+1} \) to \( \hat{r} = 1 \). All these offers are accepted. No workers are left in the market after the first period. Therefore, a firm has no incentive to deviate as long as in its deviation, no worker would reject her current offer and become available in Period 2. That is, in the subgame after any firm’s deviation, all workers still accept their offers in Period 1.

Suppose a firm \( f_{j'} \) deviates and waits until Period 2, \( j' = F - W + 1, F - W + 2, \ldots, F \). A worker of type \( \hat{r} \) would still accept her current offer if \( j' \leq j(\hat{r}) \), that is, a worker would never deviate for a firm that is worse than her offer in equilibrium, which is from her signal-suggested firm \( j(\hat{r}) \). The binding condition for the existence of an equilibrium with full unraveling then requires type \( \hat{r} = 1 \) not to unilaterally reject her offer in Period 1 in the deviation of \( f_F \), that is,
\[
F - W + 1 \geq \frac{1 - \alpha}{2} F + \left( \alpha + \frac{1 - \alpha}{2} \right) (F - W).
\] (22)
The RHS of (22) is the worker’s payoff if she accepts her offer in equilibrium. The LHS is the worker’s expected payoff if she rejects. In this case, after the first period, there are two workers (\( \hat{r} = 1 \) and \( \hat{r} = W \)) and \( F - W + 2 \) firms \((f_F, f_{F-W+1}, f_{F-W}, f_{F-W-1}, \ldots, \text{and } f_1)\) left in the market. Type \( \hat{r} = 1 \) is matched with \( f_F \) if she turns out to have a higher quality than \( \hat{r} = W \), and is matched with \( f_{F-W} \) otherwise since \( f_{F-W+1} \) is no longer available to her after the rejection. Compared to a higher type, the probability of a worker having a higher quality in the true state is given by \( \frac{1-\alpha}{W} \times \frac{W+1}{2} = \frac{1-\alpha}{2} \). Equation (22) solves

\[
\alpha \geq \frac{W - 2}{W}.
\]

Together with the constraint \( W \leq F \), an equilibrium with full unraveling exists if \( W \leq F \) and \( \alpha \geq \frac{W-2}{W} \).

On the other hand, when \( W \leq F \) and \( \alpha < \frac{W-2}{W} \), an equilibrium with full unraveling never exists since type \( \hat{r} = 1 \) has an incentive to reject her offer in Period 1 in the deviation of \( f_F \), which makes \( f_F \) strictly prefer to deviate. Combined with the fact that such an equilibrium never exists when \( W > F \), we obtain the sufficient and necessary condition for the existence of an equilibrium with full unraveling: \( W \leq F \) and \( \alpha \geq \frac{W-2}{W} \).

### A.2. Proofs of Chapter 4

**Proof of Proposition 1**

Proof. (1) In the current setting, where all colleges have the same strict priority ordering over students, the SD algorithm is a special case of the TTC algorithm. By Abdulkadiroğlu and Sönmez (2003), TTC is strategy-proof for any realized priority ordering over students. Therefore, truth-telling is a weakly dominant strategy for every student, regardless of her knowledge about the priority ordering at the time of preference submission. This proves the strategy-proofness of the PreExam-SD, Halfway-SD, and PostScore-SD mechanisms.

(2) By Kesten (2006), since the priority structure here satisfies the acyclic condition, the matching outcome of TTC is stable according to priorities. In addition, the uniqueness of such an outcome is proved by Haeringer and Klijn (2009). Translating into terms under the current setting, PostScore-SD always yield the score-stable matching outcome.

(3) From (1), we know in the truth-telling equilibrium, students’ submitted preferences stay the same under PreExam-SD and Halfway-SD as those under PostScore-SD. And colleges’ priority ordering depends only on students’ exam score ranking. Therefore, given the fact that a matching algorithm only considers students’ submitted preferences and colleges’ priority ordering, the PreExam-SD and Halfway-SD also implement the score-stable matching outcome in the truth-telling equilibrium.
Proof of Proposition 2

Proof. Define the total number of seats at colleges $c_1$, $c_2$, ..., and $c_k$ as $Q_k = \sum_{j=1}^{k} q_j$.

First, for a student with score rank $r_{s_i} = 1, 2, ..., Q_1$, it is a dominant strategy to list $c_1$ as her first choice. This is because she will be accepted by the best college $c_1$ regardless of other students’ submitted preferences. Any other first choice will make her strictly worse off, because she will always be accepted by her first choice.

Given this, a student with score rank $r_{s_i} = Q_1 + 1, ..., Q_2$ best responds by listing $c_2$ as the first choice. Deviating to $c_1$ will get her rejected in the first step and thus cannot make her better off. Any other first choice will make her strictly worse off.

Similarly, it follows that a student with score rank $r_{s_i} = Q_2 + 1, ..., Q_3$ best responds by listing $c_3$ as the first choice, a student with score rank $r_{s_i} = Q_3 + 1, ..., Q_4$ best responds by listing $c_4$ as the first choice, and so on.

Finally, consider students with lowest score ranks $r_{s_i} = Q_M + 1, ..., n$, where $Q_M < n$ and $Q_{M+1} \geq n$. Given other students’ equilibrium strategies, these students are indifferent among all strategies that list $c_{M+1}$ above $c_{M+j}$ ($j \geq 2$).

(1) From the reasoning above, in any Nash equilibrium, students with $r_{s_i} = 1, ..., Q_1$ list $c_1$ as the first choice; students with $r_{s_i} = Q_1 + 1, ..., Q_2$ list $c_2$ as the first choice; ... and students with $r_{s_i} = Q_{M-1} + 1, ..., Q_M$ list $c_M$ as the first choice. Any remaining choices of these students and for students with $r_{s_i} = Q_M + 1, ..., n$, any strategies that list $c_{M+1}$ above $c_{M+j}$ ($j \geq 2$) can exist in a Nash equilibrium. Therefore, those Nash equilibria where students with $r_{s_i} = Q_M + 1, ..., n$ list $c_{M+1}$ as the first choice are the ones where every student exhibit score-based sorting.

(2) Given the characterization of students’ equilibrium strategies, it is easy to see that all equilibria yield the same matching, where a student with $r_{s_i} = 1, ..., Q_1$ is assigned a seat at her first choice, and a student with $r_{s_i} = Q_M + 1, ..., n$ is assigned a seat at $c_{M+1}$. Such a matching is stable according to exam-based priorities or is score-stable. □
Appendix B: Additional Analysis

B.1. Learning and Overconfidence

This section investigates whether students exhibit less self-evaluation biases as they become more experienced. In the experiment, treatments using the same algorithm are implemented within-subject, and under every treatment, subjects’ beliefs about their performance on the exam are elicited at both the ex-ante and the interim stage. Therefore, I have three observations on each subject’s ex-ante belief ($Overconfidence_{EA,1}$, $Overconfidence_{EA,2}$, $Overconfidence_{EA,3}$) and three on interim belief ($Overconfidence_{IN,1}$, $Overconfidence_{IN,2}$, $Overconfidence_{IN,3}$), from which I can investigate learning behavior.\footnote{The order of treatments is randomized to ensure that learning does not interfere with any treatment effects.}

![Figure 11: Three Observations of $Overconfidence_{EA}$ and $Overconfidence_{IN}$ (Mean)](image-url)
As shown in Figure 11, there is not a clear downward trend in the average level of overconfidence. At the ex-ante stage, \( \text{Overconfidence}_{EA} \_1 \) is greater than \( \text{Overconfidence}_{EA} \_2 \) \((p = 0.01, t \text{ test})\) but not than \( \text{Overconfidence}_{EA} \_3 \) \((p = 0.22, t \text{ test})\) on average. At the interim stage, \( \text{Overconfidence}_{IN} \_1 \) is greater than \( \text{Overconfidence}_{IN} \_2 \) \((p < 0.01, t \text{ test})\) but \( \text{Overconfidence}_{IN} \_2 \) is smaller than \( \text{Overconfidence}_{IN} \_3 \) \((p = 0.07, t \text{ test})\) on average. All six values are significantly greater than zero \((p < 0.001, t \text{ tests})\). In addition, these beliefs are from more experienced subjects since they are elicited after three mock tests and feedback. Hence, the lab data do not provide clear evidence that learning can reduce or eliminate self-evaluation biases.

**B.2. Overconfidence and The Exam’s Measurement Error**

This section discusses the correlation between the exam’s measurement error and the behavioral error due to self-evaluation biases. Such a correlation is not assumed away in the paper, which is one of the reasons why the model gives an ambiguous prediction on the performance of PreExam- and Halfway-BOS, depending on the distribution of these two errors.

![Figure 12: The Mean of Overconfidence EA, Overconfidence IN, and ExamError by Aptitude Rank](image)

Figure 12: The Mean of \( \text{Overconfidence}_{EA} \), \( \text{Overconfidence}_{IN} \), and \( \text{ExamError} \) by \( \text{Aptitude \_Rank} \)
In the experimental data, OverconfidenceEA and ExamError exhibit a positive correlation of 0.30, and OverconfidenceIN and ExamError exhibit a positive correlation of 0.34. This means the two errors reinforce each other in terms of welfare distortion: a student who underperforms in the exam is worse off on average; meanwhile, she tends to be less overconfident, which hurts her even more. Such a positive correlation can be largely explained by the fact that students with higher aptitudes are less overconfident and more likely to underperform in an exam (see the mean of OverconfidenceEA, OverconfidenceIN, and ExamError by the rank of aptitude in Figure 3). It can be overestimated in the lab environment because small markets impose a strong ceiling effect: the student with the highest aptitude rank cannot have a positive level of overconfidence or overperform in an exam. Such a ceiling effect may also exist in other environments but could be much weaker in a market with more students.

B.3. Treatments with Additional Information

The provision of additional information adds a third dimension to the original treatment design. From the overwhelming truth-telling behaviors under SD regardless the timing of preference submission, we can conclude that subjects’ decision-making and the market outcomes are not affected by different information conditions. Therefore, I focus on the three mechanisms using BOS; the three new treatments with additional information are named “PreExam-BOS-INFO,” “Halfway-BOS-INFO,” and “PostScore-BOS-INFO,” respectively. All the other details of the experimental design and procedure are similar to those described in Sections 3.1 and 3.2.\footnote{For the new treatments, four additional sessions (one with 20 subjects, three with 15 subjects) were conducted in September 2015 at the Experimental Economics Laboratory of The Ohio State University. There were a total of 65 participants (22 females and 43 males).}

**Result 12.** The provision of additional information fails to reduce the magnitude of overconfidence at either the ex-ante or interim stage.

Under the new treatments, subjects’ average level of overplacement is 0.51 at the ex-ante stage and is 0.24 at the interim stage.\footnote{Again, data from all three treatments are pooled together since there is no significant treatment effect on overconfidence. Figure 13 shows the distributions of $\theta_{i}^{EA}$ and $\theta_{i}^{IN}$.} Comparing to the average without additional information (0.50 at the ex-ante stage and 0.26 at the interim stage), the magnitudes of biases are clearly not reduced ($p > 0.44$).
Figure 13: Levels of Overconfidence with Additional Information

Given the above result, together with the fact that subjects’ strategic choices are not
significantly affected by the additional information (Figure 14), we can expect that our conclusion regarding aptitude-stability remain unchanged.

**Result 13.** *The provision of additional information fails to improve the performance of PreExam-BOS and Halfway-BOS: compared to PostScore-SD, they still create more severe and more varied distortions from aptitude-stability.***

On a market level, aptitude-stability is rarely achieved under any of the three new treatments (except one market under PreExam-BOS-INFO). From Figure 15, we can see in terms of the proportion of aptitude-stably matched pairs, PreExam-BOS-INFO does not perform significantly better than PreExam-BOS ($p = 0.36$), while Halfway-BOS-INFO performs slightly worse than Halfway-BOS ($p = 0.07$).

![Figure 15: Aptitude-Stably Matched Pairs](image)

Next I compare the three new mechanisms with PostScore-SD. PostScore-SD is the best-performing mechanism in Section 3.3, and as mentioned above, it is natural to assume that its performance is not affected by the change of information environment. Figure 15 shows that PostScore-SD yields a larger proportion of aptitude-stably matched pairs than PreExam-BOS-INFO ($p = 0.031$), Halfway-BOS-INFO ($p < 0.001$), and PostScore-BOS-INFO ($p = 0.014$). As for the variable *WelfareDistortion* (Figure 16), the distribution under PostScore-SD still exhibits a smaller variance compared to PreExam-BOS-INFO ($p = 0.008$) and Halfway-BOS-INFO ($p = 0.016$). Moreover, the positive relationship between overconfidence and individual welfare remain unchanged under PreExam-BOS-INFO and Halfway-BOS-INFO (see Figure 17 and Table 7).
Figure 16: Distributions of Individual Welfare Distortion from Aptitude-Stability

Figure 17: Overconfidence and Individual Welfare Distortion
Table 7: Individual Welfare Distortion (OLS)

<table>
<thead>
<tr>
<th></th>
<th>Dep. Var.</th>
<th>WelfareDistortion</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PreExam-BOS-INFO</td>
<td>Halfway-BOS-INFO</td>
<td>PostScore-BOS-INFO</td>
<td></td>
</tr>
<tr>
<td>OverconfidenceEA</td>
<td></td>
<td>0.495***</td>
<td>0.161</td>
<td>-0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.177)</td>
<td>(0.193)</td>
<td>(0.060)</td>
<td></td>
</tr>
<tr>
<td>OverconfidenceIN</td>
<td></td>
<td>0.553**</td>
<td>0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.210)</td>
<td>(0.059)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ExamError</td>
<td></td>
<td>0.565***</td>
<td>0.631***</td>
<td>0.989***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.150)</td>
<td>(0.150)</td>
<td>(0.038)</td>
<td></td>
</tr>
<tr>
<td>AggressiveStrategy</td>
<td></td>
<td>0.726***</td>
<td>0.255</td>
<td>-0.018</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.189)</td>
<td>(0.215)</td>
<td>(0.054)</td>
<td></td>
</tr>
<tr>
<td>GuessedOtherEA</td>
<td></td>
<td>0.230**</td>
<td>-0.119</td>
<td>-0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.102)</td>
<td>(0.110)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>GuessedOtherIN</td>
<td></td>
<td>0.037</td>
<td>-0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.121)</td>
<td>(0.040)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RiskAverse</td>
<td></td>
<td>0.037</td>
<td>0.017</td>
<td>-0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.037)</td>
<td>(0.035)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>0.091</td>
<td>0.514*</td>
<td>0.090</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.319)</td>
<td>(0.297)</td>
<td>(0.090)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>-0.882*</td>
<td>-0.711</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.484)</td>
<td>(0.436)</td>
<td>(0.138)</td>
<td></td>
</tr>
</tbody>
</table>

Observations: 65  65  65

Notes: Standard errors are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.
## Appendix C: Supplementary Tables and Figures

### Table 8: Standardized Exams in Representative Countries (Year: 2014)

<table>
<thead>
<tr>
<th>Country</th>
<th>Standardized Exam</th>
<th>Number of Participants or Applicants</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>National College Entrance Exam (Gaokao)</td>
<td>9,390,000</td>
<td><a href="http://gaokao.eol.cn/">http://gaokao.eol.cn/</a></td>
</tr>
<tr>
<td>Greece</td>
<td>Panhellenic Exams</td>
<td>104,616</td>
<td><a href="http://edu.klimaka.gr/">http://edu.klimaka.gr/</a></td>
</tr>
<tr>
<td>Russia</td>
<td>Unified State Exam</td>
<td>757,303</td>
<td><a href="http://vestnikkavkaza.net/articles/society/57810.html">http://vestnikkavkaza.net/articles/society/57810.html</a></td>
</tr>
<tr>
<td>South Korea</td>
<td>College Scholastic Ability Test</td>
<td>640,619</td>
<td><a href="http://www.kice.re.kr/main.do?s=suneung">http://www.kice.re.kr/main.do?s=suneung</a></td>
</tr>
<tr>
<td>Turkey</td>
<td>Higher Education Exam-Undergraduate Placement Exam</td>
<td>2,086,115</td>
<td><a href="http://www.osym.gov.tr/">http://www.osym.gov.tr/</a></td>
</tr>
</tbody>
</table>

Notes: The statistic for Greece is from the year 2015.

### Table 9: Distribution of Score Rankings (Example 1)

<table>
<thead>
<tr>
<th>Score Ranking $r_s$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2,1,3)</td>
<td>26/64</td>
</tr>
<tr>
<td>(2,3,1)</td>
<td>17/64</td>
</tr>
<tr>
<td>(1,2,3)</td>
<td>17/64</td>
</tr>
<tr>
<td>(1,3,2)</td>
<td>1/64</td>
</tr>
<tr>
<td>(3,2,1)</td>
<td>1/64</td>
</tr>
<tr>
<td>(3,1,2)</td>
<td>2/64</td>
</tr>
</tbody>
</table>
Table 10: Descriptive Statistics of Key Variables (Lab Experiment)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>AptitudeRank</td>
<td>3</td>
<td>1.416</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>ExamError</td>
<td>0</td>
<td>1.205</td>
<td>-3</td>
<td>3</td>
</tr>
<tr>
<td>OverconfidenceEA</td>
<td>0.505</td>
<td>1.211</td>
<td>-3</td>
<td>4</td>
</tr>
<tr>
<td>OverconfidenceIN</td>
<td>0.257</td>
<td>1.192</td>
<td>-4</td>
<td>4</td>
</tr>
<tr>
<td>GuessedOtherEA</td>
<td>0.029</td>
<td>1.515</td>
<td>-4</td>
<td>3</td>
</tr>
<tr>
<td>GuessedOtherIN</td>
<td>-0.213</td>
<td>1.401</td>
<td>-4</td>
<td>3</td>
</tr>
<tr>
<td>RiskAverse</td>
<td>12.305</td>
<td>4.232</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Female</td>
<td>0.390</td>
<td>0.489</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11: Ordered Logit Marginal Effects for First Choice

<table>
<thead>
<tr>
<th>PreExam-BOS</th>
<th>Marginal Effects for FirstChoice=</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>OverconfidenceEA</td>
<td>0.301***</td>
</tr>
<tr>
<td>AptitudeRank</td>
<td>-0.269***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Halfway-BOS</th>
<th>Marginal Effects for FirstChoice=</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>OverconfidenceEA</td>
<td>0.079**</td>
</tr>
<tr>
<td>OverconfidenceIN</td>
<td>0.063**</td>
</tr>
<tr>
<td>AptitudeRank</td>
<td>-0.127***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PostScore-BOS</th>
<th>Marginal Effects for FirstChoice=</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>AptitudeRank</td>
<td>-0.012</td>
</tr>
<tr>
<td>ExamError</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Figure 18: Biased Beliefs and Non-guess-based Sorting Strategies
Table 12: Individual Welfare Distortion (Ordered Logit)

<table>
<thead>
<tr>
<th>Dep. Var.</th>
<th>(1) PreExam-BOS</th>
<th>(2) Halfway-BOS</th>
<th>(3) PostScore-BOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OverconfidenceEA</td>
<td>0.733***</td>
<td>-0.628</td>
<td>-0.150</td>
</tr>
<tr>
<td></td>
<td>(0.232)</td>
<td>(0.434)</td>
<td>(0.433)</td>
</tr>
<tr>
<td>OverconfidenceIN</td>
<td></td>
<td>1.632***</td>
<td>0.512</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.485)</td>
<td>(0.532)</td>
</tr>
<tr>
<td>ExamError</td>
<td>0.571**</td>
<td>0.547***</td>
<td>5.369***</td>
</tr>
<tr>
<td></td>
<td>(0.238)</td>
<td>(0.209)</td>
<td>(0.925)</td>
</tr>
<tr>
<td>OverconfidenceIN</td>
<td>1.010***</td>
<td>0.851**</td>
<td>-0.280</td>
</tr>
<tr>
<td></td>
<td>(0.389)</td>
<td>(0.388)</td>
<td>(0.776)</td>
</tr>
<tr>
<td>GuessedOtherEA</td>
<td>0.251</td>
<td>0.261</td>
<td>-0.079</td>
</tr>
<tr>
<td></td>
<td>(0.160)</td>
<td>(0.205)</td>
<td>(0.283)</td>
</tr>
<tr>
<td>GuessedOtherIN</td>
<td></td>
<td>-0.135</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.249)</td>
<td>(0.352)</td>
</tr>
<tr>
<td>RiskAverse</td>
<td>0.039</td>
<td>0.118**</td>
<td>0.142**</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.067)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>Female</td>
<td>0.053</td>
<td>-0.623</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>(0.560)</td>
<td>(0.562)</td>
<td>(0.851)</td>
</tr>
<tr>
<td>Observations</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

Notes: Standard errors are shown in parentheses. *** , ** , and * indicate significance at the 1%, 5%, and 10% levels, respectively.
Table 13: Ordered Logit Marginal Effects for Individual Welfare Distortion

(1) PreExam-BOS

<table>
<thead>
<tr>
<th>WelfareDistortion=</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>OverconfidenceEA</td>
<td>-0.046**</td>
<td>-0.062**</td>
<td>-0.030</td>
<td>0.109***</td>
<td>0.032*</td>
</tr>
<tr>
<td>ExamError</td>
<td>-0.036*</td>
<td>-0.048*</td>
<td>-0.023</td>
<td>0.085**</td>
<td>0.025*</td>
</tr>
<tr>
<td>AggressiveStrategy</td>
<td>-0.064*</td>
<td>-0.086*</td>
<td>-0.041</td>
<td>0.150**</td>
<td>0.044</td>
</tr>
</tbody>
</table>

(2) Halfway-BOS

<table>
<thead>
<tr>
<th>WelfareDistortion=</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>OverconfidenceIN</td>
<td>-0.043</td>
<td>-0.260***</td>
<td>0.062</td>
<td>0.125**</td>
<td>0.138**</td>
</tr>
<tr>
<td>ExamError</td>
<td>-0.014</td>
<td>-0.087**</td>
<td>0.021</td>
<td>0.042**</td>
<td>0.046**</td>
</tr>
<tr>
<td>AggressiveStrategy</td>
<td>-0.022</td>
<td>-0.136**</td>
<td>0.032</td>
<td>0.065*</td>
<td>0.072*</td>
</tr>
<tr>
<td>RiskAverse</td>
<td>-0.003</td>
<td>-0.019*</td>
<td>0.004</td>
<td>0.009</td>
<td>0.010*</td>
</tr>
</tbody>
</table>

(3) PostScore-BOS

<table>
<thead>
<tr>
<th>WelfareDistortion=</th>
<th>-1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ExamError</td>
<td>-0.435**</td>
<td>-0.115</td>
<td>0.551**</td>
</tr>
</tbody>
</table>