Behavioral Market Design for Social Good

by

Yingzhi Liang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2021

Doctoral Committee:

Professor Yan Chen, Chair
Professor Stephen Leider
Professor David A. Miller
Professor Tanya Rosenblat

Yingzhi Liang
yingzhi@umich.edu
ORCID iD: 0000-0003-3366-1289

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Prof. Yan Chen. Prof. Chen offered her meticulous and generous guidance at every stage of my doctoral study. Her consistent, selfless, insightful mentorship encouraged me to not give up during difficult times. As a mentor, Prof. Chen is always available and hands-on. Since my first year of being a doctoral student, we have always had weekly one-on-one meetings. The weekly meetings continued, uninterrupted, when I was working on my research internship in China, and during COVID-19. This has helped give a good momentum to my research. When I first started to conduct economic experiments, she would read every word of my instructions and correct even the smallest mistake. While I was conducting lab sessions, she sat in the back of the room, and kept notes on what I could improve. She is also the first reader of every one of my papers. I can always expect to receive line-by-line editing from her. I am very fortunate to have Prof. Chen as my advisor, who always leads by example. She has shown me female economists can produce superb work, and gain respect from peers. I am now hired as an Assistant Professor at the Chinese University of Hong Kong, and will soon have my own students. I hope I can be an inspiring and helpful mentor to my students as Prof. Chen has been to me. Her dedication will keep motivating me to advance my own research and teaching.

I would also like to extend my deepest gratitude to my committee members. Prof. Tanya Rosenblat has been my biggest cheerleader. She sees the potential of my research when I have self-doubt. Her knowledge and wisdom not only helped me in developing research agenda, but also guided me in working with field partners. Prof. David A. Miller is always the first one I turn to when I have theoretical problems. He helped me in modeling and proving the main theoretical properties in Chapter 2. He also helped me to develop the theoretical framework of the field experiment with DiDi. Prof. Stephen Leider has opened the door to business school for me. While I was on the job market, he provided meticulous feedback to my application packet. Without his help, I wouldn't be able to find my current job.

Besides my committee members, I am also very grateful for my undergraduate advisor and co-author, Prof. Binglin Gong. She introduced me to the field of experimental economics and helped me apply for doctoral programs. Prof. Gong also funded the college admission experiment and run all the sessions in the behavioral lab at Fudan University. Without her help, the completion of this chapter will be much delayed.

While working on Chapter 4, I received enormous help from colleagues in DiDi. I would like to thank Prof. Zhixi Wan, Prof. Jieping Ye for hiring me as a research intern, connecting me with different teams, and supporting my research agenda. I would like to thank colleague Lihua Xu and Wenjie Shang for helping me understand and extract the data. I thank Tracy (Qun) Li for managing this project. Lastly, I would like to thank DiDi for the generous financial support (through MIDAS), with award ID AWD010555.

Throughout the seven years, I am fortunate to make many lifetime friends. We carried each others through tough times, and shared joy and sorrow. They witnessed my every bit of struggle, progress and achievement. Among these friends, I especially like to thank Wei Ai, Yuyan Han, Chanda Phelan, Fangyang Wang, Shiyan Yan, Hongjiu Zhang, and Zhengting Zou. I wish them all the best for their personal and professional life.

I must also thank The University, The Rackham Graduate School, and The School of Information for providing a nourishing, inclusive, and open-minded environment which helped my intellectual growth tremendously. I was very fortunate to receive The Barbour Scholarship, a prestigious award given to women doctoral students from eastern countries. With this generous funding, I was able to focus on dissertation completion and job searching without worrying about financial situation.

Last but not least, I would like to express my deepest gratitude to my parents, Lanping Liang and Hongzhi Ding, who always support my decisions unconditionally.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

APPENDIX

ABSTRACT

Market design studies allocation rules of marketplaces, institutions, and economic environments. In addition to solving the equilibrium solution for rational agents in strategic environment, real-world market design problems also have to take into account that individuals may fail to coordinate on equilibrium behavior due to various human limitations. With the help of lab and field experiments, I study individual behavior in markets in three different contexts. Firstly, I examine a dynamic feedback mechanism in the context of college admissions. I find that although the dynamic mechanism is theoretically inferior to the deferred acceptance mechanism (DA), its performance in the lab can be as good as the DA due to its simplicity. Secondly, I investigate the optimal group size in the context of public goods contribution. I find that the optimal group size depends on the complementarity of group efforts. When efforts are highly complementary, players will coordinate on the high contribution equilibrium if the group size is smaller than an upper bound; whereas when efforts are less complementary, players will coordinate on the high contribution equilibrium if the group size is larger than a lower bound. Lastly, I study the effect of incorporating private information into centralized matching in the context of the gig economy. Through a field experiment on the largest ride-sharing platform in China, I examine the effect of incorporating location preferences into centralized matching. I find that allowing drivers to self-define working regions increases their working hours and income by more than 4%, without lowering their productivity.

CHAPTER 1

Introduction

Market design studies allocation rules of marketplaces, institutions, and economic environments. It is widely used to allocate scarce resources in both private and public sectors, such as the gig economy (Castillo et al., 2017; Liu et al., 2019), school choice (Braun et al., 2010; Kojima and Ünver, 2014; Abdulkadiroğlu et al., 2011; Chen and Kesten, 2017; Calsamiglia and Güell, 2018), college admissions (Kübler, 2012; Chen and Kesten, 2017), entry-level labor markets (Roth and Peranson, 1997), housing assignment (Chen and Sönmez, 2002), organs exchange and donation (Roth et al., 2004, 2005).

Neoclassical economics assuming individuals in markets have unbounded rationality, however, market design problems in the real world has to take into account for human limitations. As a result of these limitations, individuals may fail to coordinate on equilibrium behavior predicted by theoretical models. One example is the comparison between the second price auction and the English auction. Both auction formats are strategy-proof, but participants constantly fail to play the dominant strategy under the second price auction, but having no problems playing the dominant strategy under the English auction. It turns out that the dominant strategy under the second price auction is less "obvious" than under the English auction: A cognitively limited agent without the ability of contingent reasoning can figure out that truth-telling is the dominant strategy under the English auction, but not under the second price auction (Li, 2017). The second example is the auction on eBay. eBay uses an auction format that is similar to the second price auction, "a bidder can submit his reservation price (called a proxy bid) early in the auction and have the resulting bid register as the minimum increment above the previous high bid" (Roth and Ockenfels, 2002). The dominant strategy for this auction is to submit the true value as the reservation price, however, a considerable share of buyers revise their bids in the last hour, causing system congestion (Roth and Ockenfels, 2002; Ariely et al., 2005; Ockenfels and Roth, 2006). It turns out this sniping can be the best response to several types of naive strategies, such as the incremental bidding strategy.

Experiments play an important role in exploring the gap between theoretical predictions of rational agents and the real behavior of ordinary humans. In this dissertation, I use lab and field experiments to study individual behavior in markets in the absence of unbounded rationality, unbounded selfishness, and complete information. In the following chapters, I solve incentive problems in three different markets. The first one is the Chinese college admissions, which involves 10 million students and 7 million college seats each year. The assumptions violated in this market are unbounded rationality and complete information. The second market is an artificial market in the lab, where players contribute to public goods. The assumption violated in this market is unbounded selfishness. The third market is the largest ride-sharing platform in China, DiDi, which involves 21 million drivers. The assumptions violated in this market is complete information. These markets impose distinct market design challenges on matching individuals with resources and matching individuals into groups. Matching decisions in these markets have a significant impact on individual education and labor outcomes, as well as social welfare.

In the Chinese college admissions market, I study the dynamic mechanism implemented in Inner Mongolia, China. Under this mechanism, students are given real-time allocation feedback and are permitted to revise their choices before an ending time. I first model the dynamic mechanism using a stochastic process and find that the dynamic mechanism is stable and efficient only when the revision opportunities are frequent enough. I then compare its performance with DA in the lab. Even though the dynamic mechanism is theoretically inferior to DA, students are more truthful under the dynamic mechanism, and the matching outcome is as stable as DA in an environment where student preference correlation is low. This suggests that when student preference correlation is low, the dynamic mechanism can be a good substitute for DA.

In the artificial market in the lab, I study the optimal group size for coordination. Previous research shows mixed evidence on the effect of group size. I reconcile these findings by varying the complementarity of efforts. I find that the optimal group size depends on the level of complementarity. When the complementarity is high, there is an upper bound for the optimal team size, but when the complementarity is low, there is a lower bound for the optimal team size. These theoretical predictions are validated via a lab experiment.

In the ride-sharing market, I investigate how to incorporate workers' preferences into centralized matching. Like other ride-sharing platforms, DiDi drivers do not have control over their assigned trips. Drivers could end up in an unfamiliar area far away from home, or get stuck in bad traffic for hours. One way to incorporate driver location preference is by asking drivers to submit a destination. However, setting a specific destination increases driver and passenger waiting time, therefore reducing matching efficiency. Another way to incorporate driver location preference is by asking drivers to submit a coarse location preference. This compromise between having no information and having overly precise information might lead to driver welfare gain without reducing matching efficiency significantly. I evaluate a field experiment where 15% of drivers in a city were randomly assigned to receive a flexibility feature, where drivers could choose working regions. I find that treatment drivers increase working hours and income by more than 4% while maintaining productivity, measured by hourly earnings. This shows that incorporating workers' preferences into task assignments can be a win-win situation for both the workers and the company.

CHAPTER 2

A Dynamic Matching Mechanism for College Admissions: Theory and Experiment

2.1 Introduction

Market design has provided many managerial insights into why certain market institutions fail and proposed alternative mechanisms to solve known problems. For instance, soft ending rules are proposed to solve the sniping problem in online auctions (Roth and Ockenfels, 2002; Ariely et al., 2005; Ockenfels and Roth, 2006); frequent batch auctions are proposed to solve the high-frequency trading arms race in financial markets (Budish et al., 2015); team contests are proposed to solve the low productivity problem in the ride-sharing economy (Ai et al., 2019a). Matching theories have guided the allocation of scarce resources in many real-world markets, such as labor market clearinghouses (Roth, 1984; Roth and Peranson, 1999; Echenique et al., 2016), allocation of organs (Roth et al., 2004, 2005), and ride-sharing platforms (Liu et al., 2019). In this paper, we analyze a new form of dynamic matching mechanism emerged from the Chinese college admissions, which could potentially be applied to other matching markets, including the placement of medical residents and employee recruitment.

In the past years, researchers have proposed several static centralized matching mechanisms. Strategy-proof mechanisms, such as the Deferred Acceptance (DA) and the Top Trading Cycles (TTC), are theoretically appealing (Gale and Shapley, 1962; Dubins and Freedman, 1981; Roth, 1982; Abdulkadiroglu and Sönmez, 2003). However, their complexity makes the dominant strategy (i.e., truth-telling) less obvious (Li, 2017). As a result, the realized stability of DA and the realized efficiency of TTC fall well short of full stability and efficiency (Chen and Sönmez, 2006; Pais and Pintér, 2008; Calsamiglia et al., 2010; Klijn et al., 2013; Chen and Kesten, 2019; Hassidim et al., 2016, 2017; Shorrer and Sóvágó, 2018; Rees-Jones and Skowronek, 2018; Rees-Jones, 2018). To reduce the cognitive complexity, market designers might consider breaking down the matching game into several stage games, each of which has a reduced strategy space and strategic uncertainty. Market designers could also provide opportunities for participants to observe others' choices and revise their own. However, until recently, these mechanisms have been infeasible due to computational limitations.

In comparison, non-strategy-proof mechanisms, such as the Boston mechanism, require students to acquire information on others' preferences to best respond. When implemented among a large number of students, this information acquisition cost could be substantial. As a result, students often rely on luck to get into their desired schools. Though having information on others' preferences could lead to welfare gains (Chen and He, 2017), information provision might be difficult to implement under the static version of the classic matching mechanisms.

With innovations in information technology, there has been a recent flurry of mechanisms emerging from the field, which address both the issue of complexity and costly information acquisition. For example, the public school assignments in Wake County, North Carolina, uses a dynamic truncated Boston mechanism, where students could see the number of first-choice applicants and revise their choices within two weeks (Dur et al., 2018). Similarly, college admissions in Brazil allow students to revise choices in four consecutive days, after knowing the cutoff score of each university (Bo and Hakimov, 2018).

In this paper, we focus on a new dynamic mechanism implemented in college admissions in Inner Mongolia (IM), China. We refer to it as the IM Dynamic mechanism for the rest of the paper. Under this mechanism, students submit one choice at a time and are free to change their choices at any time before the pre-announced ending time. In real time, the system shows unfilled quotas, the number of applicants for each school, and these applicants' standardized test scores (i.e., students' priorities). At the ending time, each school accepts students up to its quota and rejects the rest. In this paper, we first formally model the IM Dynamic mechanism. We then compare it with two well-studied mechanisms: the Boston and DA mechanisms. We use DA as a benchmark because DA is the theoretically optimal mechanism in the Chinese college admissions problem. We use Boston as a benchmark because Inner Mongolia used to use the truncated Boston mechanism before switching to the dynamic mechanism in 2011. It is a general practice to compare the old and the new mechanisms side by side when studying a new mechanism to inform policy.

In Chinese college admissions, colleges have an identical preference over students[1]. In this setting, a stable and efficient outcome can be reached by performing the Serial Dictatorship mechanism (SD) or DA over students' true preferences. Theoretically, we show that the IM Dynamic mechanism is similar to DA, in that the stable and efficient outcome can be obtained under every rationalizable strategy profile. However, unlike DA, the IM Dynamic mechanism does not have a dominant strategy. Compared with the Boston mechanism, which is efficient but not stable in the incomplete information setting, the IM Dynamic mechanism has theoretical advantages in stability.

[1] Chinese college admissions is a centralized matching via a standardized test. See the online appendix of Chen and Kesten (2017).

To test our theoretical predictions, we conduct a lab experiment. Lab experiments are essential to understand how people respond to the incentives provided by each mechanism, because preferences are generally unknown in the field. We specifically choose two environments with different preference correlation levels so that the strategic complexity of best responding varies. Our low correlation environment has higher strategic uncertainty and complexity than our high correlation environment, formally measured by the number of Bayesian Nash Equilibrium and the number of schools in the “ minimum strategy” (See Section 2.5 for details). Our experimental results show that in the less complex environment, consistent with theoretical predictions, the IM Dynamic mechanism is less stable and less efficient than DA, more stable, and less efficient than Boston. In the more complex environment, however, the IM Dynamic mechanism outperforms theoretical predictions. It is as stable as DA, and as efficient as Boston, with a higher truth-telling level than both of them. These results indicate that the IM Dynamic mechanism indeed has behavioral advantages over static mechanisms in complex settings. We propose two explanations for these behavioral advantages. The first explanation is that when enough time has elapsed, the truth-telling strategy is the unique rationalizable strategy under the IM Dynamic mechanism. Even though the truth-telling strategy is a weakly dominant strategy under DA, it is neither the unique dominant strategy, nor the unique rationalizable strategy. The other explanation is that under the IM Dynamic, players only need to rank one school to best respond, whereas, under DA and Boston, players need to rank several schools to best respond. The strategy space at each strategy point of IM Dynamic is much smaller than the strategy space of DA and Boston.

Our work has broad policy implications. In many real-world matching markets, the strategic complexity and uncertainty are enormous due to incomplete information, diversity of preferences, and a large number of participants. Strategy-proof mechanisms, like DA, are not always feasible due to a prohibitively large number of choices. Even when DA is feasible, participants may still manipulate their preferences and fail to reach equilibrium. Our findings suggest that the IM Dynamic mechanism may be an attractive alternative.

Our paper is organized as follows. Section 2.2 provides background information about the Chinese college admissions process and the Inner Mongolia dynamic mechanism. Section 2.3 presents a literature review. Section 2.4 offers our formal definitions and proves the basic theoretical properties under the IM Dynamic mechanism. Section 2.5 presents our experimental design. Section 2.6 reports the experimental results. Section 2.7 concludes and discusses extensions to the private sectors.

2.2 Chinese College Admissions System and Inner Mongolia Dynamic Mechanism

Each year in China, roughly 10 million high school seniors compete for 7 million college seats[2]. College admissions in China involves centralized matching via a standardized test. The college admissions process is centralized at the provincial level, and each university determines a quota for each province. Each province may have its own college entrance exam, the score of which determine student priorities. Using students' reported preferences and their entrance exam scores, each province then matches students to colleges using its own centralized matching mechanism. The matching mechanisms used in China fall into three classes: sequential, parallel, and dynamic (Chen and Kesten, 2017). By the year 2016, 28 out of 31 provinces had adopted some version of a parallel mechanism. Among the remaining three provinces, Qinghai and Jilin use a hybrid of parallel and sequential mechanisms, while Inner Mongolia has adopted the IM Dynamic mechanism[3].

According to Rongfei Han, Director of the Center for College Admissions in Inner Mongolia: "Our mechanism makes it easy to figure out which college to apply to based on students' scores. There is no ambiguity regarding the mapping from scores to colleges. The application and admission processes are completely transparent. Students know which college they have been admitted to by the time the system closes. Fairness is assured under this mechanism. Compared to 2007 [when the truncated Boston mechanism was used][4], after the IM Dynamic mechanism was implemented in 2011, the percentage of students who repeated their last year of high school and retook the college entrance exam dropped from 23.4% to 7%, and the percentage of students who accepted their matches increased from 91% to 99.03%."[5]

China is not the only country where student priority in college admissions is determined by a centralized standard test. University assignments in Turkey uses a similar scheme, though scores in various subjects are weighted differently for different majors (Balinski and Sönmez, 1999). Hungary (Biró, 2012) and South Korea (Avery et al., 2019) use a hybrid which considers standard test scores along with high school grades.

[2]http://gaokao.chsi.com.cn/

[3]Table A.11 in Appendix A.5

[4]From the year 2008 to 2010, Inner Mongolia experimented with the IM Dynamic mechanism in early admissions, which are only for art and military colleges. The year 2007 was the last year that the truncated Boston mechanism was used in all admissions. The year 2011 was the first year that the IM Dynamic mechanism was used in general admissions.

[5]Inner Mongolia Daily, June 20th, 2013

2.3 Literature Review

One of the most widely-used and well-studied mechanisms is the Boston mechanism, also known as the Immediate Acceptance Mechanism. In the Boston mechanism, each student submits a ranking of all schools based on her preferences. Each school first considers students who list it as their first choice and admits these students to up its quota and reject the rest. If there are still slots left, the school considers students who list it as their second choice if other schools have not yet admitted them, and so on. Note that every admission decision is final in each round. While the Boston mechanism is Pareto efficient if all students submit their true preferences, it has long been criticized because it is not strategy-proof. That is, students have an incentive to misreport their preferences because they may lose their priorities if they do not rank a school highly enough. Because of misreporting of preferences, the realized efficiency under the Boston mechanism may be low. In addition, the Boston mechanism is not stable, which leads to justified envy and thus is considered unfair.

Given the limitations of the Boston mechanism, the DA and the TTC mechanisms have been proposed as alternatives (Abdulkadiroglu and Sönmez, 2003). DA, also called the Gale-Shapley mechanism or student-optimal mechanism, was first proposed by Gale and Shapley (1962). Unlike the Boston mechanism, allocations under DA are temporary: in each round, schools consider new applicants together with all the applicants kept from the previous rounds and reject students with low priorities. The allocation is finalized in the last round when no one gets rejected anymore. DA is strategy-proof and stable (Dubins and Freedman, 1981; Roth, 1982; Gale and Shapley, 1962). Although it is not Pareto efficient when both sides have heterogeneous preferences, When the priority structure is acyclical, as defined in Ergin (2002), DA is efficient. In the context of Chinese college admissions, where colleges have an identical preference over students, DA is both stable and efficient (Balinski and Sönmez, 1999; Ergin, 2002).

Exhaustive research has compared these two mechanisms theoretically, empirically, and experimentally. For example, Ergin and Sönmez (2006) prove that, with complete information, the set of Nash equilibrium outcomes of a game under the Boston mechanism equals the set of stable matchings. DA, on the other hand, yields the optimal outcome among all stable matchings, and therefore its outcome Pareto dominates Boston's Nash equilibrium outcome. In another study, Abdulkadiroglu et al. (2006) examined data before and after changing from the Boston to the DA mechanism and found that unsophisticated players were exploited by sophisticated players under the Boston mechanism. Experimental studies confirm that DA is more stable and less manipulable than the Boston mechanism, and that the efficiency comparison is environment-dependent (Chen and Sönmez, 2006; Pais and Pintér, 2008; Calsamiglia et al., 2010; Klijn et al., 2013; Chen et al., 2018a; Chen and Kesten, 2019). However, under incomplete information, others have found that in specific environments, truth-telling can be the equilibrium strategy under the Boston mechanism and *ex-ante*, every student may strictly prefer the Boston mechanism to DA (Featherstone and Niederle, 2016). Abdulkadiroğlu et al. (2011) further prove that under incomplete information, when students all have the same ordinal preference, each student is weakly better off under the Boston mechanism using any symmetric tie-breaking.

Both Boston and DA become unwieldy when there are too many schools to rank. In China, where there are more than 2,000 colleges, it is thus common to use a truncated version (also called the Parallel and Sequential mechanism in Chen and Kesten (2017)), where the entire choice list is divided into several choice bands, and students only submit a limited number of choices in each of the choice bands.

However, this truncation can affect both the stability and efficiency of the final allocation (Calsamiglia et al., 2010). Indeed, Pathak and Sönmez (2013) found that, in a truncated DA, the mechanism becomes more manipulable as the number of permitted choices decreases. Chen and Kesten (2017) further analyze the whole family of parallel mechanisms and prove that DA is more stable and less manipulable than any other version of the parallel mechanism.

Recently, the matching literature has investigated several different dynamic mechanisms, including four that are closely related to our mechanism: sequential DA mechanism (Echenique et al., 2016), the dynamic DA mechanism (Klijn et al., 2018), the iterative DA mechanism (Bó and Hakimov, 2019), and the continuous feedback mechanism (Stephenson, 2016). In the sequential DA mechanism, unmatched proposers make simultaneous proposals and are informed of temporary allocation results in each round. This mechanism is theoretically equivalent to the static DA (Echenique et al., 2016). The dynamic DA mechanism of Klijn et al. (2018) is equivalent to the iterative DA (IDAM-NC) in Bó and Hakimov (2019). This mechanism differs from the IM Dynamic mechanism in two ways. First, only unmatched students are allowed to make adjustments under the dynamic DA, whereas every student is free to make adjustments under IM Dynamic. Second, students have less information under dynamic DA as they are only informed of their temporary allocation results. Under IM Dynamic, on the other hand, students see a complete list of other students' scores. In another variation of the iterative DA, students are informed of the tentative cutoff scores for acceptance at each school at each step (IDAM) (Bó and Hakimov, 2019). Though similar to our mechanism, it differs in that only unmatched students are allowed to make new proposals. Finally, the IM Dynamic mechanism differs from Stephenson (2016)'s continuous feedback mechanism, which allows students to select more than one school at a time. The IM Dynamic mechanism only allows students to choose one school at a time. These differences make the IM Dynamic mechanism theoretically unique. For a detailed discussion, the reader is referred to Hakimov and Kübler (2021) for a survey of various dynamic mechanisms. Overall, while the IM Dynamic mechanism shares similarities with several other dynamic mechanisms, it is theoretically distinct.

Our paper is also broadly related to papers that investigate Chinese college admission mechanisms. Several of them are related to preference submission timing. Under pre-exam preference submission, the Boston mechanism can be ext-ante fair and efficient (Wu and Zhong, 2014; Lien et al., 2016, 2017), and can also outperform the DA in stability with respect to true ability rather than test scores (Jiang, 2014). The downside of pre-exam preference submission is that students might be over-confident and thus cause market instability (Pan, 2019). Chen and Kesten (2017) formally models the whole family of parallel and sequential mechanisms using a unified framework. Nevertheless, none of them examined the dynamic mechanism in Inner Mongolia.

2.4 Theoretical Analysis

In this section, we formally define the three mechanisms before focusing on the theoretical properties of the IM Dynamic mechanism and stating our two main theorems.

2.4.1 Definitions and Assumptions

We begin by defining the college admission problem. The college admission problem (Balinski and Sönmez, 1999) is a set (I, C, q, P_I, P_C), consisting of (1) a set of students $I = \{1, 2, \cdots, n\}$; (2) a set of colleges $C = \{1, 2, \cdots, m\} \bigcup \{\emptyset\}$, where $\emptyset$ denotes a student's outside option; (3) a capacity vector $q = (q_1, q_2, \cdots, q_m)$ where q_k is the capacity of college k; (4) a list of student preferences $P^I = (P_1^I, P_2^I, \cdots, P_n^I)$ where P_i^I is the strict preference of student i over colleges including the no-college option; (5) a list of college preferences $P^C = (P_1^C, P_2^C, \cdots, P_m^C)$ where P_k^C is the strict preference of college k over a set of students. In other words, P_k^C determines student priorities in college k.

As in Chinese college admissions, we assume that college preferences P_k^C is determined by student scores on a centralized college entrance exam, and a student has the same priority across different colleges, therefore $P_k^C = P_l^C$, $\forall k, l \in C$. Student priorities are public information and therefore cannot be manipulated. We assume that student preferences $P^I = (P_1^I, P_2^I, \cdots, P_n^I)$ are private information, and students have cardinal preferences over colleges. The utility of student i getting into college k is u_i^k. Furthermore, we assume that students are expected utility maximizers and their beliefs about others' preferences and strategies have full support. We only consider the welfare of students. College seats are considered as public resources to be allocated.

A matching μ is a many-to-one mapping from I to C, $\mu : I \to C$, such that $|\mu^{-1}(k)| \leq q_k, \forall k$, where $\mu^{-1}(k)$ is the set of students admitted to school k.

- A matching contains *justified envy* if there exist two students i and j, where i prefers j's allocation to his current allocation and i has higher priority than j in j's assigned school.

- A matching is *non-wasteful* if no student prefers a school with an unfilled quota compared to his assignment.

- A matching μ is *stable* if there does not exist any justified envy, and the matching is non-wasteful.

- A matching is *Pareto efficient* if there is no other matching which makes all students at least as well off and at least one student better off. Note that we use "efficient" and "Pareto efficient" interchangeably in this paper.

- A *static mechanism* selects a matching for each reported preference profile.

- A static mechanism is *strategy-proof* if truth-telling is a weakly dominant strategy.

- A *dynamic mechanism* creates an extensive form game for participants to play. Each information set is a decision point at which a participant observes information and takes an action. The mechanism selects a matching as a function of the accumulated actions.

- A mechanism, regardless static or dynamic, is always Pareto efficient (stable) if the selected matching is Pareto efficient (stable) with respect to a reported preference profile.

2.4.1.1 Boston Mechanism

The Boston mechanism, formally introduced by Abdulkadiroglu and Sönmez (2003), has been used to allocate students to public schools in Boston, Cambridge, Charlotte, Denver, Minnesota, Seattle, and St. Petersburg-Tampa (Ergin and Sönmez, 2006). The Boston mechanism is implemented using the following algorithm for a given problem:

- Step 1: Each college considers students who have listed it as their first choice and assigns seats to these students based on their priority orders until either there are no seats left or there are no students remaining who have listed it as their first choice. Students whose priorities are lower than the college's quota are rejected.

- Step k ($k > 1$): For the students who have been rejected after step $k-1$, only their kth choices are considered. For each college with available seats, they consider those students who have listed it as their kth choice and assigns the remaining seats to these students one at a time following their priority orders.

- The allocation is finalized when either there are no seats left or there are no students left who have listed it as their *k*th choice.

Under complete information and heterogeneous priorities, the Boston mechanism is efficient if everyone plays the truth-telling strategy. However, since it is not strategy-proof, it may not lead to efficient allocations. In addition, it is not stable.

When we assume incomplete information, full support of beliefs and identical priorities, according to Featherstone and Niederle (2016)'s Proposition B.1., truth-telling is an ordinal Bayesian Nash equilibrium (Ehlers and Massó, 2007) under the Boston mechanism. Furthermore, it is the unique Bayesian Nash equilibrium in anonymous strategies[6]. If players play the truth-telling strategy under Boston, the allocation is always efficient ex-post. However, depending on the realization of preferences, the allocation may not be stable.

2.4.1.2 Deferred-Acceptance Mechanism

The DA was introduced by Gale and Shapley in 1962. DA is implemented using the following algorithm for a given problem:

- Step 1: Students apply their first-choice colleges. Colleges temporarily keep students based on students' priorities up to their enrollment capacities on hold and reject the rest.

- Step k ($k > 1$): Each rejected student is sent to the next college on her reported preference list. Each college considers both students who have been temporarily accepted and new applicants. It then temporarily accepts students based on their priorities, up to its enrollment capacity and rejects the rest.

- The allocation is finalized when no one gets rejected. Each student is assigned to the college in which they temporarily hold a seat.

Under complete information and heterogeneous priorities, the DA mechanism is strategy-proof and stable, but it is not efficient.

When we assume incomplete information, full support of beliefs and identical priorities, the DA mechanism is stable, efficient, and strategy-proof. Furthermore, we can characterize all rationalizable strategies under DA: truthfully report to the point where the sum of reported schools' quotas exceeds one's ranking. This set of strategies strictly dominates other strategies.

[6]For the definition of anonymous strategies, see Featherstone and Niederle (2016).

2.4.1.3 IM Dynamic Mechanism in Real Life

The IM Dynamic mechanism was introduced by the Inner Mongolia Center of College Admissions in 2008. To our knowledge, Inner Mongolia is the only place where the mechanism is used. The IM Dynamic mechanism is implemented using the following descriptive algorithm:

- Students are divided into N groups based on a sequence of cutoff scores $s_1, s_2, \cdots, s_{N-1}$, where $s_i < s_{i-1}$, for $i = 2, \cdots, N-1$. Students with exam scores higher than s_1 comprise the first group. Students with scores between s_{i-1} and s_i comprise the ith group. Students with scores below s_{N-1} comprise the Nth group.

- All students, regardless of their group, can enter an online admission system at the same time to submit their college choices. Students can submit only one choice at a time, but they can change their choices as many times as they want within their allocated time period.

- In the system, students can see each college's quota, the number of students applying to that college, and those students' scores in real-time.

- Each group i has an ending time T_i, where $T_{i-1} < T_i$, for $i = 2, \cdots, N-1$. At time T_i, the system closes for the ith group.

- For each group, by the time the system closes, each school k admit up to q_k students who are currently applying to school k. If more students are applying to school k than its quota q_k, the school admits the q_k students with the highest priorities and rejects the rest.

2.4.1.4 A Model of IM Dynamic Mechanism

In this subsection, we model the IM Dynamic Mechanism when all students belong to one group. The grouping of students is not essential to the mechanism and will not change its properties. We will discuss the effect of grouping in Section 2.4.2.

We model the IM Dynamic mechanism as a Poisson process with parameter λ. Each student i has her Poisson recognition process R_i, which is independent of others'. At each arrival, the student is recognized to have a chance to revise. We use z to represent the number of arrivals, and $R_i(z)$ represents the corresponding time of that arrival.

We choose to model the timing of this game using the Poisson process for several reasons. Modeling students' recognition process as Poisson processes restricts each student to act at no more than a countable number of instances, making their history of actions tractable. Compared with deterministic arrivals, stochastic arrivals can capture uncertainty in real life. For example, students may not be able to update their choices due to network connection failure or network congestion; therefore, their total number of revision opportunities is non-deterministic. Furthermore, the probability that any two students are recognized at the same instant is zero. The Poisson process is commonly used in modeling asynchronous revision games (Kamada and Kandori, 2011).

We now formally define the game created by IM Dynamic mechanism.

- At time 0: Each student i knows her own preference over colleges P_i^I, each college's quota q_k and its preference over students P_k^C, $\forall k \in C$, as well as the ending time T.
- When student i is recognized for the zth time by the Poisson process R_i, she has a decision point at which:
  - Student i observes every other student's most recent choice.
  - Student i makes a choice c_i^z

    If $z = 1$, meaning this arrival is student i's first arrival, student i can choose to apply to any college or not apply (empty choice), $c_i^z \in C \cup \{\emptyset\}$.

    If $z > 1$, meaning student i previously had an arrival, student i can choose not to make a revision, in which case $c_i^z = c_i^{z-1}$, or switch to a different college $c_i^z \neq c_i^{z-1}$, where $c_i^z \in C$.
- At time T: The system closes. If a student is ranked within the quota of her applied college, she gets admitted. Otherwise, she gets rejected.

One important note is that even though the arrivals of this game are exogenous, the revision decisions are endogenous. At each arrival, a student can always choose between acting or not. Arrivals merely provide students chances of acting. When the arrivals are frequent enough, students can almost act whenever they want.

We define student i's *best school available* $\widehat{c}_i^z$ at arrival z as the highest-ranked school on player i's preference list with seats not occupied by students ranked ahead of student i. In other words, student i's *best school available* is her favorite school among all schools in which she can temporarily have a seat at the decision point $R_i(z)$. We now define the myopic best responding strategy and the truth-telling strategy under the IM Dynamic mechanism.

- A player is playing *myopic best responding strategy* if $c_i^z = \widehat{c}_i^z$ for all z.
- A player is playing *truth-telling strategy* if there does not exist an arrival z, such that $c_i^z \neq c_i^{z-1}$, and $c_i^z \neq \widehat{c}_i^z$.

The difference between the myopic best response and truth-telling is that myopic best response requires best responding at each arrival. In contrast, truth-telling only requires best responding when players decide to revise. The reason we want to define truth-telling this way is to tolerate inattention. While it is a natural concept to best respond at each arrival in the theoretical model, it is almost impossible to keep track of all the information and make an action in every arrival in real life when the arrivals are frequent enough. In the lab experiment, when we see a subject not acting at a certain moment, we simply cannot infer whether this lack of responsiveness is due to strategic misrepresentation of preference or inattention. Therefore, when evaluating human subject behavior, we need a less strict concept than the myopic best response but still captures the essential aspect of being truthful. The truth-telling defined here is to exclude all intentional misrepresentation of preferences.

2.4.2 Theoretical Analysis of the IM Dynamic Mechanism

In this section, we first state our main theorems under the assumption that all students belong to one group. We then generalize the results to having a different deadline for different student groups and discuss its implication.

To analyze the performance of the IM Dynamic mechanism, we define player i as the player with the ith highest ranking. That is, player 1 has the highest-ranking score, player 2 has the second-highest ranking, etc. We assume that players' beliefs regarding others' preferences and strategies have full support. Furthermore, we assume that players have cardinal preferences over colleges and are expected utility maximizers. In the college admissions problem, when students' priorities are the same across different colleges, there is a unique stable and efficient matching (Balinski and Sönmez, 1999). The stable and efficient matching can easily be achieved by performing DA or serial dictatorship over students' true preferences.

In the IM Dynamic mechanism, we show that the stable and efficient matching can be achieved almost certainly when students have sufficient revision opportunities. This is our Theorem 1.

Theorem 1. *When the Poisson parameter λ is large enough, the stable and efficient outcome arises with arbitrarily high probability under any rationalizable strategy profile.*

Proof. We first characterize the rationalizable strategy profiles. Define an exogenous time sequence $t_0, t_1, t_2, \cdots, t_n$, where $t_0 = 0$, $t_k = \varepsilon k$, $\varepsilon > 0$. Also define ρ to be the largest utility ratio between a less preferred school and a more preferred school among all students.

$$\rho = \max_i \{\max_{k \neq l} \{\frac{u_i^k}{u_i^l}\}\}; \forall k, l \in C, \text{such that } u_i^k < u_i^l; \forall i \in I.$$

We prove that when $\lambda > -\frac{ln[1-\rho^{\frac{1}{n-1}}]}{\varepsilon}$, player i's undominated strategies satisfy: at player i's first arrival within $(t_{i-1}, t_i]$, conditional on seeing players ranked ahead of her have seats in their current choices, she chooses the best school available; for all subsequent arrivals, if players ranked ahead of her do not change, she does not make any changes. This is obviously true for player 1, since player 1 has a strictly dominant strategy to choose her most-preferred school at her first arrival within time $(0, \varepsilon]$, and does not make any changes for all subsequent arrivals. A strategy that chooses any other school at some arrival is dominated because there is a chance that player 1 will not have another arrival after choosing a less-preferred school. For player 2, at his first arrival within $(t_1, t_2]$, conditional on seeing player 1 made a choice, player 2 knows that player 1 has already picked her favorite school and will not make a change in subsequent arrivals, therefore player 2 should choose his best school available. Since player 1 will not make a change in future arrivals, there is no need for player 2 to make a change in future arrivals. Player 3's decision making process is more complicated than player 2, because seeing player 1 and player 2 have seats in their current choices does not guarantee that player 1 and player 2 have had arrivals in sequence. Player 2's choice could have been made before player's choice. Therefore, player 3 needs to consider the possibility of being outranked by player 2 in future arrivals. At player 3's first arrival within $(t_2, t_3]$, player 3 compares the expected utility between choosing the best school available and choosing another school. Player 3 faces the following cases:

1. Player 3 has future arrivals. This happens with greater than $1 - e^{-\lambda(T-3\varepsilon)}$ probability.
2. Player 3 does not have future arrivals. This happens with less than $e^{-\lambda(T-3\varepsilon)}$ probability.
   - 2A. Player 1 had at least one arrival in $(t_0, t_1]$ and player 2 had at least one arrival in $(t_1, t_2]$. This happens with $(1 - e^{-\lambda\varepsilon})^2$ probability.
   - 2B. Either player 1 did not have an arrival in $(t_0, t_1]$, or player 2 did not have an arrival in $(t_1, t_2]$. This happens with $1 - (1 - e^{-\lambda\varepsilon})^2$ probability.

In Case 1, where player 3 has future arrivals, player 3's choice at this arrival does not matter. Whatever she does at a future arrival will supersede her current choice. Therefore, the expected utility of choosing the best school available and choosing another school is the same under Case 1. Hence, we only need to focus on Case 2, where player 3 does not have future arrivals. Assume the utility of getting into the best school available is u_3^*, and the maximum utility of getting into any other school is u_3'.

If player 3 chooses the best school available, she will get $u_3 **$ under Case 2A, and the worst scenario is getting unmatched under Case 2B. Hence, the expected utility of choosing the best school available under Case 2 is:

$$U(\text{best school available}) \geq (1 - e^{-\lambda\varepsilon})^2 \times u_3^* + [1 - (1 - e^{-\lambda\varepsilon})^2] \times 0 \tag{2.1}$$

If player 3 chooses another school, she will get u_3' under Case 2A, and best scenario is not being outranked and get matched to that school under Case 2B. Hence, the expected utility of choosing another school under Case 2 is:

$$U(\text{another school}) \leq u'_3 \tag{2.2}$$

When $\lambda > -\frac{ln[1-\rho^{\frac{1}{2}}]}{\varepsilon}$, $(1-e^{-\lambda\varepsilon})^2 \times u^*_3 + [1-(1-e^{-\lambda\varepsilon})^2] \times 0 > u'_3$. Choosing the best school available dominates choosing another school.

If player 1 and player 2 have not made any changes, player 3 has no incentive to make changes in future arrivals.

For any player i ranked below player 3, she makes similar decisions as player 3. Player i faces the following cases:

1. Player i has future arrivals. This happens with greater than $1-e^{-\lambda(T-i\varepsilon)}$ probability.
2. Player i does not have future arrivals. This happens with less than $e^{-\lambda(T-i\varepsilon)}$ probability.
   A. All players ranked ahead of player i have had at least one arrival in their designated time interval. This happens with $(1-e^{-\lambda\varepsilon})^{(i-1)}$ probability.
   B. At least one player ranked ahead of player i haven't had an arrival in their designated time interval. This happens with $1-(1-e^{-\lambda\varepsilon})^{(i-1)}$ probability.

Assume player i's utility of getting into the best school available is u^*_i, and the maximum utility of getting into any other school is u'_i. The expected utility of choosing the best school available under Case 2 is:

$$U(\text{best school available}) > (1-e^{-\lambda\varepsilon})^{(i-1)} \times u^*_i + [1-(1-e^{-\lambda\varepsilon})^{(i-1)}] \times 0 \tag{2.3}$$

The expected utility of choosing another school under case 2 is:

$$U(\text{another school}) \leq u'_i \tag{2.4}$$

When $\lambda > -\frac{ln[1-\rho^{\frac{1}{i-1}}]}{\varepsilon}$, $(1-e^{-\lambda\varepsilon})^{i-1} \times u^*_i + [1-(1-e^{-\lambda\varepsilon})^{i-1}] \times 0 > u'_i$. Choosing the best school available dominates choosing another school. As long as players ranked ahead have not made any changes, player i does not have incentive to change her choice for all subsequent arrivals.

When $\lambda > -\frac{ln[1-\rho^{\frac{1}{n-1}}]}{\varepsilon}$, the above argument is true for all players.

Under any rationalizable strategy profile, players reach the stable and efficient outcome if each player i has at least one arrivals in their designated time period $(t_{i-1}, t_i]$, which happens with probability $(1-e^{-\lambda\varepsilon})^n$. When $\lambda \to \infty$, $(1-e^{-\lambda\varepsilon})^n \to 1$. The stable and efficient outcome can be reached with arbitrarily high probability. □

An even weaker assumption than common knowledge of rationality is sufficient for obtaining such an outcome. To illustrate, let $K_i()$ be an indicator function of whether player i knows the argument inside the parentheses. The assumption that for all i, $K_i(K_{i-1}(\cdots K_{j+1}(\text{j is rational}))) = 1$, for all $j < i$, is sufficient to prove Theorem 1. Suppose player i is player 3. The above condition means that this player knows that player 2 is rational, and further knows that player 2 knows that player 1 is rational. A similar argument can be made for all players in the game.

While DA and the IM Dynamic mechanism yield similar outcomes in every rationalizable strategy profile, there are several differences. First, the DA mechanism requires only individual rationality, as every player has a dominant strategy. By contrast, the IM Dynamic mechanism requires player i to know that all players ahead of him are rational, that player $i-1$ knows that all players ahead of her are rational, and so on. Second, under DA, if every player is rational, the stable and efficient outcome is a certainty. However, under the IM Dynamic mechanism, there always exists a slight possibility that the allocation will not be stable and efficient, since some players may not have a chance to respond to changes made by players ranked ahead of them. Unlike DA, there does not exist a dominant strategy for every player under the IM Dynamic mechanism. Our Theorem 2 explains why.

Theorem 2. *Any player whose ranking number is larger than the quota of her most preferred school does not have a weakly dominant strategy.*

Proof. We have characterized the rationalizable strategy profiles in Theorem 1. If a weakly dominant strategy exists, it should belongs to the set of rationalizable strategies. Now we show that no strategy in rationalizable strategies is weakly dominant. For a player i whose rank is larger than the quota of her most preferred school, suppose the 1st ranked player's strategy is to always choose whichever school that player i has chosen, then player i's best response is to choose her most preferred school among schools with at least two empty seats. If there is no school with at least two empty seats, then player i's best response is to choose her most preferred school available at her last arrival before the system closes, hoping that the 1st ranked player does not have another arrival after her. None of these best responses belong to the rationalizable strategies we characterized in Theorem 1. Thus a weakly dominant strategy does not exist. □

We now discuss the implication of dividing students into groups and assigning earlier ending times to higher score groups, like described in Section 4.1.3. When arrivals are frequent enough, dividing students into N groups and assigning earlier ending times to higher score groups is equivalent to having N sequential markets where students only participate in their corresponding market. The first market consists of students whose scores are higher than s_1 and all college seats. The second market consists of students whose scores are higher than s_2 but lower than s_1 and the remaining college seats left in the first market. The ith market consists of students whose scores are higher than s_i but lower then s_{i-1} and remaining college seats left in the previous $i-1$ markets. Our Theorem 1 – 2 still hold in each sequential market. When the outcome in each sequential market is stable and efficient, since students in a later market can never have justified envy against students in a previous market, the outcome in the pooled market is also stable and efficient. The benefit of dividing students into different score groups is that each market contains less students. Therefore the stable and efficient outcome is more likely to arise under the same arrival frequency.

Our Theorem 1 and 2 shows that the IM Dynamic is close to, but not as good as DA in stability and efficiency. Therefore we expect the IM Dynamic to be less efficient and stable than DA.

When comparing IM Dynamic and Boston mechanisms, we expect IM Dynamic to be more stable, but less efficient than Boston, as it is almost as stable and efficient as DA.

The comparison between the Boston and DA mechanisms under incomplete information is more complicated. Our setting of full support of beliefs and identical preferences among schools satisfy the *school-symmetric* condition in Featherstone and Niederle (2016). According to their Proposition B.1., truth-telling is an ordinal Bayesian Nash equilibrium (Ehlers and Massó, 2007) under the Boston mechanism. Furthermore, it is the unique Bayesian Nash equilibrium in anonymous strategies[7]. If players play the truth-telling strategy under Boston, the allocation is always efficient ex-post. However, depending on the realization of preferences, the allocation may not be stable. Since the Bayesian Nash equilibrium allocation under DA is always stable and efficient, we expect Boston to be less stable than, while as efficient as DA.

These theoretical predictions may not necessarily hold in the field due to various assumptions made. For example, we assume common knowledge of rationality in our theoretical analysis, but it is well documented that humans have bounded rationality (Simon, 1972). In the next section, we test these theoretical predictions in a controlled lab setting, as direct testing in the field would be infeasible.

2.5 Experimental Design

We start with the simplest setting, which captures the essential strategic components of Chinese college admissions. In China, students know their test scores and rankings, but not others' preferences. To represent this environment, we use an incomplete information setting with correlated preferences. In each group, there are four colleges $\{A,B,C,D\}$, each of which has one seat. Correspondingly, there are four students $\{1,2,3,4\}$. Students' preferences have two different types. Each group has two type I students and two type II students. While individual preferences are private information, the composition of the group is common knowledge. Similar to reality, students also know their score ranks among other students in the same group.

[7] For the definition of anonymous strategies, see Featherstone and Niederle (2016).

We deliberately select two environments with different levels of preference correlations, so that the difficulty of reaching a Bayesian Nash equilibrium is different. We measure the difficulty of reaching a Bayesian Nash equilibrium in two ways. The first way is how easily players can reach BNE by playing a random strategy. The larger the set of BNE strategy profiles is, the easier it is to reach a BNE by chance. The second way is how many schools a player needs to rank in order to best respond if players are not required to rank all schools. The more schools needed to rank, the more cognitively demanding for players to best respond. Consider the most extreme case, where student preferences are identical $A \succ B \succ C \succ D$. Under all three mechanisms, when assuming rationality, students can best respond by only reporting one school. The first-ranked student chooses school A as her first choice; the second-ranked student chooses school B as her first choice; the third-ranked student chooses school C as her first choice, and the fourth-ranked student chooses school D as her first choice. In this setting, three mechanisms are likely to perform the same, but this performance does not reflect the fundamental properties of the three mechanisms. Consider a slightly different environment, where two students have the preference $A \succ B \succ C \succ D$, and two students have the preference $A \succ C \succ B \succ D$. Under DA, the top two students can still best respond by reporting one school. The third-ranked student needs to rank school B and C to best respond, and the fourth-ranked student can best respond by any strategy. Under Boston, the top two students can also best respond by reporting one school. The third-ranked student has to report school B,c, and D; and the fourth-ranked student needs report school C and D. Under IM Dynamic, students can always best respond by ranking one school. Lower ranked students can wait until higher ranked students made their choices and then best respond by choosing the best school available. Therefore, in the latter environment, the difficulty of best responding is greatest under Boston, followed by DA, then by IM Dynamic. The above examples show that when we decrease the preference correlation, the complexity of strategically best responding increases disproportionally across the three mechanisms. This difference in complexity is central to test the performance of mechanisms.

To vary complexity, we have a “high correlation environment” and a “low preference environment”. In both environments, we fix the first preference type to be $A \succ B \succ C \succ D$, and vary the second preference type. In the high correlation environment, preference type II is $B \succ A \succ D \succ C$. In the low environment, preference type II is $B \succ D \succ C \succ A$. The two environments differ in the number of BNE strategy profiles and the number of schools needed to rank to best respond. For simplicity, We define the best responding strategy which requires the least number of reported schools the “minimum strategy”. Column (4) in Table 2.2 reports the number of BNE strategy profiles under each environment and mechanism. Column (5) in Table 2.2 characterizes the minimum strategy. Under both measures, the complexity of the low correlation environment is always higher than the complexity of the high correlation environment.

The two environments are also directly comparable. The set of Bayesian Nash equilibrium strategies under DA and Boston in our low correlation environment is a subset of that in our high correlation environment, respectively. A detailed examination of these two environments can be found in Section 5.1. Table 2.2 column (2) reports the BNE strategies under each environment.

Subjects receive 16 Yuan for getting admitted to the most preferred school, 11 Yuan for the second-most preferred school, 7 for the third-most preferred school, and 5 for the least preferred school (Table 2.1). This setting ensures a sufficient distinction between the monetary incentives for getting into different schools. The same payoff differentiation was used in Chen and Kesten (2019).

Table 2.1: Payoffs Table

High Correlation	A	B	C	D
Payoff of Type I	16	11	7	5
Payoff of Type II	11	16	5	7
Low Correlation	**A**	**B**	**C**	**D**
Payoff of Type I	16	11	7	5
Payoff of Type II	5	16	7	11

Subjects are asked to report a complete list of preferences under the DA and Boston mechanisms, taking as much time as needed. Under the IM Dynamic mechanism, the decision time is 30 seconds. To mimic the real IM Dynamic mechanism, we divide the four students into two subgroups, with the top two students in the first subgroup, and the bottom two students in the second subgroup. Students in both subgroups enter the decision stage, but the stage ends 15 seconds earlier for the first subgroup[8].

The experiment lasts for 20 periods. At the start of each period, subjects are randomly rematched to form new groups. Subjects' preference types are fixed during the experiment, but their rankings rotate every five periods. We use a between-subjects design among different treatments.

2.5.1 Theoretical Analysis of the Experimental Environments

In this section, we characterize all Bayesian Nash equilibrium strategies under the DA and Boston mechanisms in each environment. A detailed proof is provided in Appendix A.1. To simplify the

[8]We also tried 60 seconds of decision time for the second subgroup and 30 seconds for the first subgroup in our pilot. We find that subjects usually finish adjusting choices within the first 20 seconds, and they get bored waiting for the stage to end. Therefore we decide to shorten the decision period to 30 seconds for the second subgroup and 15 seconds for the first subgroup, and we find no significant difference in subject reported preferences and allocation outcomes between our 60-second sessions and 30-second sessions in our pilot.

notation, we use numbers and asterisks to represent strategies. Number *i* denotes a student's *i*th most preferred school on her preference list. Asterisks represent arbitrary schools. For example, strategy $(1,2,*,*)$ means that the player reports her most preferred school first, followed by her second-most preferred school, while her third and fourth choices can be either her third- or fourth-most preferred school.

In our high correlation environment, under DA, for the highest-ranked student, any strategy which lists her favorite school as the first choice is a Bayesian Nash equilibrium strategy. For a student who has the second-highest ranking, any strategy which lists the student's two most preferred schools ahead of the two least preferred schools is a Bayesian Nash equilibrium strategy. Given that students who ranked the first and second were rational, a student with the third-highest ranking would not be able to make into the top two schools, so any strategy which ranks her third ahead of her fourth most preferred school is a Bayesian Nash equilibrium strategy. A student with the lowest ranking has no choice but to accept the remaining school. Therefore, any strategy is a Bayesian Nash equilibrium strategy for her.

Under Boston, the highest-ranked student has the same dominant strategy, which is to report her first choice truthfully. Given this, the second and the third highest-ranked students can best respond by truthfully reporting their first choice and report their third most preferred school ahead of their fourth most preferred school. Since the lowest-ranked student could not get into her top two schools in equilibrium, her best response is to report her third most preferred school ahead of her fourth most preferred school.

In our low correlation environment, under DA, the Bayesian Nash equilibrium strategies for the highest- and lowest-ranked students are the same as in the high correlation environment. For a student with the second-highest ranking, however, her Bayesian Nash equilibrium strategy is to report her first two choices truthfully. For a student with the third-highest ranking, her Bayesian Nash equilibrium strategy is to report her last two choices truthfully.

Under Boston, again, the highest-ranked student has a dominant strategy to report her first choice truthfully. For the second- and third-ranked students, their best response is complete truth-telling. For the lowest-ranked student, since she could not get into her most preferred school in equilibrium, she can best respond by truthfully report the ranking of her second to fourth most preferred school.

Table 2.2 column (2) outlines the respective Bayesian Nash equilibrium strategies in each environment. Any combination of Bayesian Nash equilibrium strategies among the four players is a Bayesian Nash equilibrium in our game. Note that although there are different Bayesian Nash equilibria, the equilibrium outcome is the same under each environment in each realized preference type.

We expect Boston to perform worse than DA in stability in equilibrium because when the

Table 2.2: Comparison of The Two Environments

Ranking (1)	BNE Strategies (2)	Num of Strategies (3)	Num of Profiles (4)	Minimum Strategy (5)
		DA – High Correlation		
1st	$(1,*,*,*)$	6	6912	(1, $*$, $*$, $*$)
2nd	$(1/2 > 3/4)$	4		(1, 2, $*$, $*$) or (2, 1, $*$, $*$)
3rd	$(3 > 4)$	12		(3, $*$, $*$, $*$)
4th	any strategy	24		irrelevant
		DA – Low Correlation		
1st	$(1,*,*,*)$	6	576	(1, $*$, $*$, $*$)
2nd	$(1,2,*,*)$	2		(1, 2, $*$, $*$)
3rd	$(*,*,3,4)$	2		(1, 2, 3, 4) or (2, 1, 3, 4)
4th	any strategy	24		irrelevant
		Boston – High Correlation		
1st	$(1,*,*,*)$	6	648	(1, $*$, $*$, $*$)
2nd	$(1,3 > 4)$	3		(1, 3, $*$, $*$)
3rd	$(1,3 > 4)$	3		(1, 3, $*$, $*$)
4th	$(3 > 4)$	12		(3, $*$, $*$, $*$)
		Boston – Low Correlation		
1st	$(1,*,*,*)$	6	24	(1, $*$, $*$, $*$)
2nd	$(1,2,3,4)$	1		(1, 2, 3, $*$)
3rd	$(1,2,3,4)$	1		(1, 2, 3, $*$)
4th	$(2 > 3 > 4)$	4		(2, 3, $*$, $*$)

top two students share the same preference, the BNE allocations under Boston are not stable. In particular, in the high correlation environment, the second-ranked student is matched with her third most preferred school, where she has justified envy over the third-ranked student; in the low correlation environment, when the top two students both have preference type I, the second-ranked student has justified envy over the third-ranked student as well. On the other hand, because both mechanisms lead to efficient allocations in equilibrium, we expect DA and Boston perform similarly in efficiency. A detailed characterization of the Bayesian Nash equilibrium allocation under each realized preference type is in Table A.1 in Appendix A.1.

2.5.2 Experimental Procedure

In each session, we randomly assign subjects a seat in front of a terminal in the laboratory. The experimenter reads the instructions aloud at the front of the lab. After the instructions are read aloud, subjects are given 10 minutes to read the instructions at their own speed and ask questions. All questions are repeated and answered publicly. Subjects are then asked to complete a quiz consisting of 17 questions; for each correct answer, subjects are paid 0.25 Yuan. The quiz is designed to help subjects understand the experimental setting and the corresponding mechanism. The first question of the quiz is to solve an allocation problem under the corresponding mechanism. After subjects finish the quiz, the correct answers and corresponding explanations for each question are presented on their computer screens. After this, the experimenter solves the allocation problem on the blackboard to make sure subjects understand the allocation rules, and explains all the other quiz questions. Any other questions about the quiz are also answered publicly. After the quiz feedback, subjects take part in 20 periods of the college admissions game. One of every five periods with the same priority is randomly drawn for payment. At the end of 20 periods, subjects fill out a demographics survey. Subjects are paid in private at the end of the experiment. The quiz, matching game, and survey are programmed using z-Tree (Fischbacher, 2007).

The experiment was conducted from October 2015 to December 2015 at Fudan University in Shanghai. Subjects are students at Fudan University. For each mechanism and each environment, we conduct five independent sessions, with the exception of the IM Dynamic mechanism in the low correlation environment, for which we have an unexpectedly large number of subjects showed up. Each session has 12 subjects who are randomly re-matched at the beginning of each period. No subject participates in more than one session. In total, we conduct 31 sessions, with 372 subjects. Detailed information is presented in Table 2.3. Each session lasts approximately 1.5 hours. The average payment, including payment for the quiz, is 46.12 Yuan (about 7.5 US dollars, which is above the average hourly pay at Fudan). The experimental instructions are included in Appendix A.4.

Table 2.3: Number of subjects in each treatment

Environment	DA	IM Dynamic	Boston
High Correlation	5 sessions × 12 subjects	5 sessions × 12 subjects	5 sessions × 12 subjects
Low Correlation	5 sessions × 12 subjects	6 sessions × 12 subjects	5 sessions × 12 subjects

2.6 Experimental Results

We first discuss individual level results, such as truth-telling behavior and choice revision time. We then discuss group level results, such as stability and efficiency.

In the following discussion, we use $A > B$ to indicate that a measure under mechanism A is greater than under mechanism B at the 5% significance level, $A \geq B$ to denote that a measure under mechanism A is greater than under mechanism B at the 10% significance level, and $A \sim B$ to denote that the measured difference is not significant at the 10% level.

2.6.1 Individual Behavior

We first compare preference manipulation across the three mechanisms among all players, then compare the truth-telling behavior among players for each score rank, using truth-telling and first choice truth-telling as measures.

2.6.1.1 Overall Truth-telling

In the DA and Boston mechanisms, the truth-telling strategy is defined as the reported order of preference being the same as the true preference. However, since we could only observe one choice at a time in the IM Dynamic mechanism, the definition of the truth-telling strategy had to be based on other players' choices at each stage game.

Specifically, we define the truth-telling strategy in the IM Dynamic mechanism as selecting the best school available at stage game when a player revises. It is worth mentioning that the truth-telling strategy here differs from the myopic best response. The myopic best response requires a player best-responding at every stage game, whereas the truth-telling strategy only look at stage games when a player revises. The formal definition of the truth-telling strategy and the myopic best response strategy can be found in Section 4.1.4.

We expect to see a higher level of truth-telling under DA than that under both Boston and IM Dynamic, since truth-telling is a dominant strategy under DA.

Hypothesis 1. *The proportion of truth-telling under the DA mechanism is higher than that under the Boston and IM Dynamic mechanism in any environment.*

Our results show that the Boston mechanism indeed is more manipulable than DA, regardless of the environment. While the IM Dynamic mechanism performs as good as DA in the high correlation environment, it even performs better than DA in the low correlation environment.

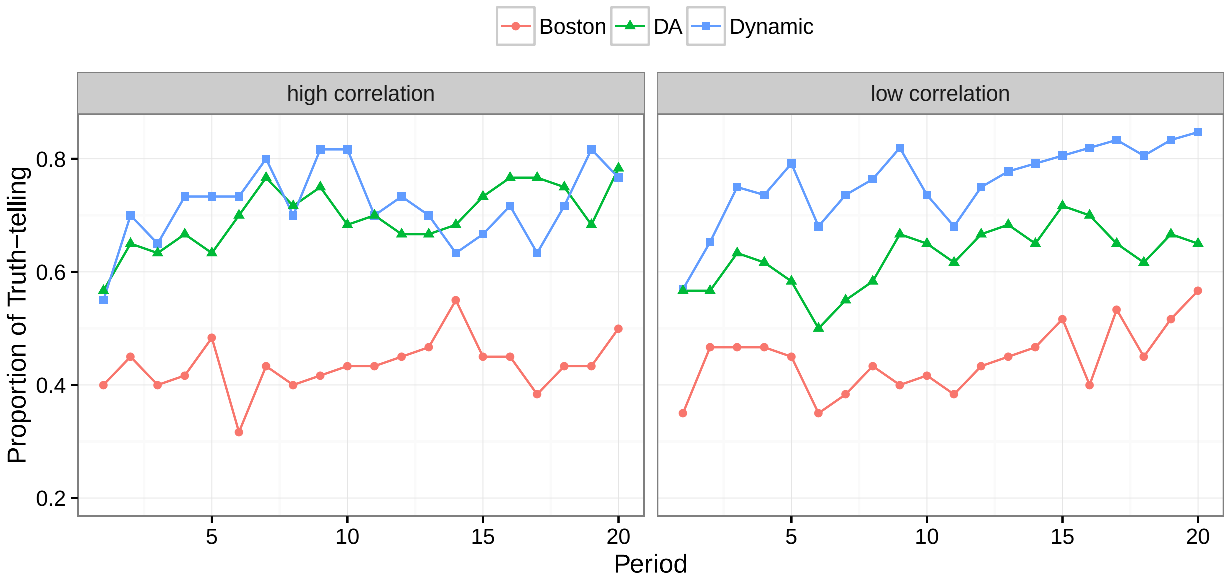


Figure 2.1: Truthtelling Comparison

Table 2.4: Truth-telling

Mechanism (1)	Prop (High) (2)	Prop (Low) (3)	Ha (4)	p-value (5)
IM DY	0.716	0.759	High = Low	0.221
DA	0.698	0.627	High = Low	0.119
BOS	0.435	0.445	High = Low	0.786
Hypothesis (6)	p-value (High) (7)	p-value (Low) (8)		
IM DY >DA	0.278	0.011		
DA>BOS	0.004	0.004		
IM DY = BOS	0.008	0.004		

Result 1. *(Truth-telling) In the environment where preferences are highly correlated, the proportion of truth-telling has the following order: IM Dynamic $\sim$ DA $>$ Boston. In the environment with low preference correlation, the order is IM Dynamic $>$ DA $>$ Boston. All three mechanisms are largely unchanged across environments.*

Support. *Table 2.4 reports the proportions of truth-telling in each treatment and the p-values of permutation tests, using each session as an independent observation. The proportions of truth-telling in the high and low correlation environments are in column (2) and (3), respectively. Columns (4) and (5) compare the performance of the same mechanism across environments. Columns (6) (7) and (8) present pair-wise comparisons of the three mechanisms under the same environment. Figure 2.1 depicts the trend of truth-telling under the three mechanisms in both environments.*

Our predicted truth-telling comparison between DA and Boston in Hypothesis 1 is supported in both environments, whereas the comparison between DA and the IM Dynamic is rejected. The IM Dynamic performs as good as DA in the low correlation environment and better than DA in the high correlation environment.

Our proportions of truth-telling under DA and Boston are consistent with those in previous studies, which report that the proportion of truth-telling is between 40% – 50% under the Boston and between 65% – 75% under DA (Pais and Pintér, 2008; Calsamiglia et al., 2010; Chen and Kesten, 2019). The proportion of truth-telling in our experiment is higher than results from Chen and Sönmez (2006) and Chen et al. (2015), possibly because our group size is smaller.

Note that students report only one school at a time under the IM Dynamic mechanism. Thus, it may be unfair to compare a sequence of reported preferences with a complete list. For example, under both DA and the Boston mechanism, students who have the highest ranking have a dominant strategy to report their first choice truthfully. In this case, while manipulation of preferences from the second to the fourth choices may exist, it does not impact allocation results. By contrast, we will not observe this type of manipulation under IM Dynamic, simply because the second to fourth choices are not reported. Therefore, using complete truth-telling as our measure favors the IM Dynamic mechanism. To solve this problem, we re-run our analysis using the first choice truth-telling as our measure. Doing so gives us the same results. Table A.2 in Appendix A.3 reports the results from permutation tests using first choice truth-telling as the measure. We also repeat our analysis using truth-telling up to one's rank under the DA mechanism as the measure. All comparisons stay the same. Table A.3 in Appendix A.3 reports the results using truth-telling up to one's rank as the measure under the DA mechanism.

2.6.1.2 Rank Effect in Truth-telling

In this subsection, we break down truth-telling behavior into the behavior of each score rank. We are more interested to see how low ranked students are affected differently under the three mechanisms since the highest-ranked students have no incentive to misreport their preferences. We find that going down one rank increases preference manipulation under all three mechanisms.

However, the rank effect is most severe under the Boston mechanism, followed by DA, then the IM Dynamic. Not surprisingly, all students, besides the highest-ranked student, are more likely to manipulate their preferences under the Boston mechanism. The comparison between DA and the IM Dynamic is more interesting, with the two lowest-ranked students less likely to manipulate their preference under the IM Dynamic. Table A.4 in Appendix A.3 reports the proportion of truth-telling for each ranking, mechanism, and environment, together with p-values from permutation tests, using each session as an independent observation.

Table 2.5: The effect of ranking, period and quiz scores on truth-telling, stability and efficiency

	(1)	(2)	(3)	(4)	(5)
Dependent Variables	Truth-telling	Truth-telling(DA)	Justified envy	Stability	Efficiency
DA	0.104		-0.0597	0.450***	0.0457
	(0.0737)		(0.0640)	(0.0795)	(0.0541)
IM Dynamic	-0.103		-0.00377	0.156***	-0.108***
	(0.0681)		(0.0851)	(0.0522)	(0.0373)
Period	0.00400	0.00622***	-0.00236**	0.00555**	0.00233
	(0.00273)	(0.00154)	(0.00116)	(0.00235)	(0.00207)
Period×DA	0.00206		-0.00116	0.00227	0.00485
	(0.00312)		(0.00162)	(0.00386)	(0.00391)
Period×IM Dynamic	0.00199		-0.00636***	0.0123**	0.00423
	(0.00311)		(0.00225)	(0.00480)	(0.00349)
Ranking	-0.223***	-0.172***	0.0800***		
	(0.0141)	(0.00954)	(0.00534)		
Ranking×IM Dynamic	0.131***		-0.0384***		
	(0.0181)		(0.00906)		
Ranking×DA	0.0506**		-0.0298*		
	(0.0199)		(0.0170)		
Low Corr Eniv	0.0135	-0.0447*	-0.00168	-0.0391	-0.0778***
	(0.0314)	(0.0267)	(0.0125)	(0.0270)	(0.0275)
Low Corr Eniv×DA	-0.0827*		0.135***	-0.284***	-0.0260
	(0.0473)		(0.0324)	(0.0735)	(0.0421)
Low Corr Eniv×IM Dynamic	0.0270		0.0552**	-0.0887*	0.0792**
	(0.0405)		(0.0241)	(0.0526)	(0.0358)
Quiz Score		0.139***	-0.0335**		
		(0.0361)	(0.0143)		
Quiz Score×DA			-0.0234		
			(0.0213)		
Quiz Score×IM Dynamic			0.00133		
			(0.0210)		
Observations	7,440	2,400	5,580	1,860	1,860

Notes: (1) (2) (3) (4) (5) are Probit regressions reporting marginal effect; standard errors in parentheses are clustered at session level. The high correlation environment is coded as 0, and the low correlation environment is coded as 1.
*** p<0.01, ** p<0.05, * p<0.1

Result 2. *(Rank Effect in Truth-telling) Lower ranked students are more likely to manipulate their preferences. The magnitude of manipulation on average has the following order: BOS > DA > IM Dynamic.*

Support. *Table 2.5 reports the marginal effects from probit regressions. Regression (1) reports the effects of mechanism, period, ranking (a higher ranking means a lower priority), and environment on one's truth-telling behavior. The marginal effects of going down one rank are 22.3%, 17.2%, and 9.2%, respectively, under the Boston, DA, and IM Dynamic mechanisms, with any pair-wise difference significant.*

Finally, the results show that subjects who have higher quiz scores are more likely to report their preferences truthfully under the DA mechanism. The marginal effect of correctly answering four questions is 13.9%. Since quiz scores measure how well subjects understand the mechanism, our result suggests that further illustration and explanation might help subjects figure out the dominant strategy and can thus improve allocation outcomes under the DA mechanism. Quiz scores do not play a role in either the IM Dynamic or Boston mechanism, where the truth-telling strategy is not the dominant strategy.

2.6.1.3 Decision Time in the IM Dynamic

In this subsection, we look at the choice revision dynamics under the IM Dynamic mechanism. In particular, we examine whether there are last-second revisions. Last-second revisions are referred to as sniping in previous literature, and are prevalent in online auctions with hard timing rules. Although bid sniping may cause lower expected revenues, it can also be a best response to incremental bidding strategies (Roth and Ockenfels, 2002; Ariely et al., 2005; Ockenfels and Roth, 2006).

Similar to their setting, under IM Dynamic, sniping is also a best response to the reverse truth-telling strategy and random strategy[9]. The following analysis clarifies how prevalent this behavior is in our experiment.

In our experiment, subjects with different rankings have different ending time. Subjects with the top two priorities have 15 seconds to make revisions, while subjects with the last two priorities have 30 seconds. We define the sniping behavior as revising in the last two seconds (i.e., the 14th and 15th second for the top two players, and the 29the and 30th second for the bottom two players). Figure 2.2 shows the number of last revisions made at each second for both the top two students and the bottom two students.

In general, sniping is not as frequent as we expected. Only 6% of last revisions are classified as sniping. We further investigate if the sniping behavior prevents other group members from best responding. For every player's last revision, we compute whether that revision is a best response

[9]We define reverse truth-telling strategy as selecting the least preferred school among all available schools and adjusting to a more preferred school on following arrivals based on the reverse order of the truth preference list. We define random strategy as randomly choosing a school among all available schools at each arrival.

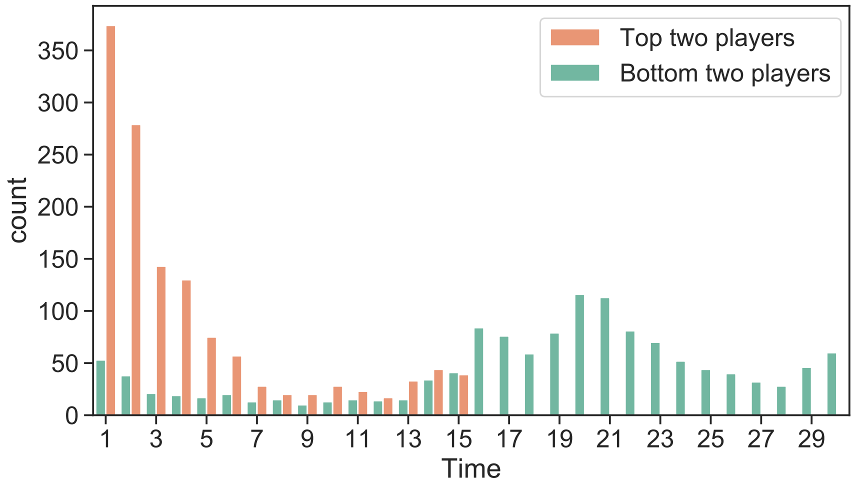


Figure 2.2: Last Revision Time

based on other group members' last revisions. We find that 6.135% of last revisions are sub-optimal, and only 11.111% of these sub-optimal revisions are caused by the sniping behavior of other players.

In sum, though sniping behavior exists in our experiment, it is not the main contributor to instability and inefficiency under the IM Dynamic mechanism.

To further alleviate the sniping problem of the IM Dynamic mechanism, policymakers may consider introducing a soft ending rule, where the system closes when no one has made any change in a defined amount of time (e.g., 10 minutes), or the number of people making revisions drops below a certain threshold. Another alternative would be to allow students to list more than one school at a time in a preferred order, similar to the continuous feedback mechanism in Stephenson (2016). A third option would be to restrict temporarily-allocated students from changing their choices, similar to the dynamic DA mechanism in Klijn et al. (2018) and Bó and Hakimov (2019).

2.6.2 Aggregate Performance

In this section, we investigate the aggregate performance of the three mechanisms. We first compare stability using the proportion of justified envy and the proportion of stable allocations. We then compare efficiency using the proportion of efficient allocation and normalized efficiency.

2.6.2.1 Stability

We consider an allocation to be stable when no one in the group has justified envy. We say that player i's envy is justified if player i prefers player j's allocation to his current allocation, and player i has a higher priority than j in j's assigned school. We consider a matching to be unstable when there is at least one player who has justified envy. In many cases, there could be more than one player who has justified envy. To better measure stability, We define proportion of justified envy as the number of players who has justified envy divided by the total number of players. The proportion of justified envy is a more precise measure than the proportion of stable outcomes.We also repeat our analysis using the proportion of stable allocations as the measure and find similar results.

Based on Theorem 1, the IM Dynamic is almost as good as DA in stability when the revision chances are frequent enough. Therefore, we expect DA to overperform the IM Dynamic in stability. We expect the IM Dynamic to overperform Boston, as we have shown in Subsection 5.1 that the BNE allocations under Boston are unstable when the top two students share the same preference. Accordingly, we have the following hypothesis.

Hypothesis 2. *The proportion of justified envy under the Boston mechanism is higher than that under the IM Dynamic mechanism, which in turn is higher than that under the DA mechanism in any environment.*

Table 2.6: Justified Envy

Mechanism (1)	Prop (High) (2)	Prop (Low) (3)	Ha.2 (4)	p-value (5)
IM DY	0.042	0.072	High = Low	0.069
DA	0.011	0.073	High = Low	0.008
BOS	0.154	0.152	High = Low	0.960
Hypothesis (6)	**p-value (High) (7)**	**p-value (Low) (8)**		
IM DY>DA	0.012	0.498		
DA<BOS	0.004	0.004		
IM DY<BOS	0.004	0.002		

Result 3. *(Stability) In the environment where preferences are highly correlated, the proportions of justified envy have the following order: Boston > IM Dynamic > DA. In the environment with low preference correlation, the proportions of justified envy have the following order: Boston > IM Dynamic ∼ DA. DA and the IM Dynamic mechanisms have higher proportions of justified envy in the low correlation environment, while the proportion of justified envy is largely unchanged under the Boston mechanism across environments.*

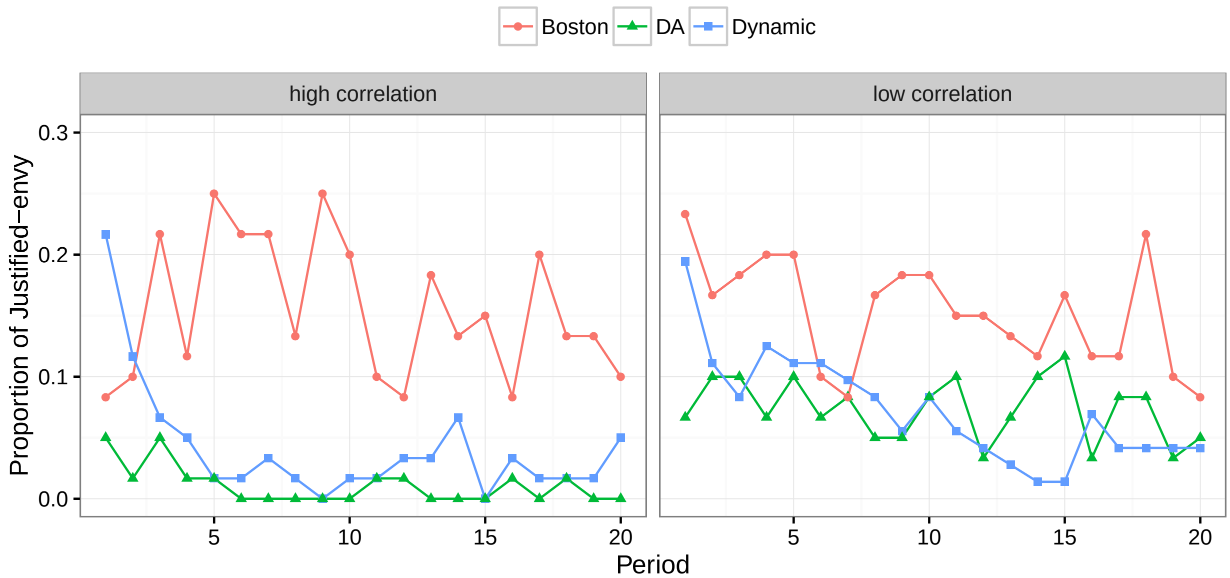


Figure 2.3: Justified envy comparison

Support. *Table 2.6 presents the average proportion of justified envy in each treatment and the p-values of permutation tests, using each session as an independent observation. The proportions of justified envy in the high and low correlation environments are in column (2) and (3), respectively. Columns (4) and (5) compare the performance of the same mechanism across environments. Columns (6) (7) and (8) present pair-wise comparisons of the three mechanisms under the same environment. Figure 2.3 depicts the trend of justified envy over the 20 periods in both environments. For robustness check, we use the proportion of stable allocations as an alternative measure and find similar results. Table A.6 in Appendix A.3 reports the proportion of stable allocations for each mechanism in each environment.*

Based on our results, we conclude that Hypothesis 2 is supported in the high correlation environment, but not in the low correlation environment, where the IM Dynamic mechanism performs as well as DA in stability. This result shows that the IM Dynamic mechanism has behavioral advantages when preference correlation is low, as the IM Dynamic mechanism reveals more information about student preferences and allows revisions.

In terms of cross-environment comparison, for the DA mechanism, the higher proportion of justified envy in the low correlation environment can be explained by the difference in the proportion of Bayesian Nash equilibrium strategies played in each environment. By design, the set of Bayesian Nash equilibrium strategies in the low correlation environment is smaller than that in the high correlation environment, making it more difficult for subjects to reach an equilibrium. For the IM Dynamic mechanism, the higher justified envy level in the low correlation environment comes from a lack of revisions. As we will point out later in Section 2.6.2.3, subjects fail to best respond even when there is plenty of time. The higher response failure rate in the low correlation environment might be due to its complexity. Previous studies have shown that subjects who are cognitively overloaded are more likely to make mistakes and perform worse in lab experiments (Sweller, 1988; Fayol et al., 1994).

2.6.2.2 Rank Effect in Justified Envy

Result 4. *(Rank Effect in justified envy) As student rankings increase, the proportion of justified envy increases under all three mechanisms. The order of magnitude is Boston $\geq$ DA, Boston $>$ IM Dynamic, DA $\sim$ IM Dynamic.*

Support. *Column (3) in Table 2.5 presents the marginal effects of mechanism, period, ranking, environment, and quiz score on justified envy from the Probit regressions. These results show that decreasing rank by one increases the likelihood of having justified envy by 8.00%, 5.01%, and 4.16%, respectively, under the Boston, DA, and IM Dynamic mechanisms. Pair-wise F tests reveal a significant difference between the Boston and IM Dynamic mechanisms, but not between DA and the IM Dynamic mechanisms.*

Our results suggest that, compared to the Boston mechanism, the DA and IM Dynamic mechanisms are more friendly to students with the second or third rankings. Note that the regression does not account for students with the lowest ranking, as these students do not have justified envy. The Boston mechanism offers students with the lowest rankings better schools than they can obtain under the DA or the IM Dynamic mechanism, to the detriment of the second- and third-ranked students.

Finally, we find a small but significant learning effect under all three mechanisms, with different magnitudes. Specifically, the marginal effects of having one more period are 0.24% (the effect of period in Table 2.5 column (3)) and 0.35% (the combination of period and period$\times$ da in Table 2.5 column (3)) under Boston and DA, respectively, with no significant difference between the two. The marginal effect under the IM Dynamic is 0.87% (the combination of period and period$\times$ IM Dynamic in Table 2.5 column (3)), which is the highest among the three.

2.6.2.3 Efficiency

We consider a matching to be efficient when there does not exist a subset of students in which at least one is better off without making any other student worse off when allocations are switched. In this section, we first compare efficiency using the proportion of efficient allocation as the measure. For the robustness check, we repeat our analysis using normalized efficiency as the measure and find similar results. We define normalized efficiency as:

$$\text{Normalized Efficiency} = \frac{\text{maximum group rank} - \text{actual group rank}}{\text{maximum group rank} - \text{minimal group rank}},$$

where group rank is the sum of ranks of each assigned school in everyone's preference list.

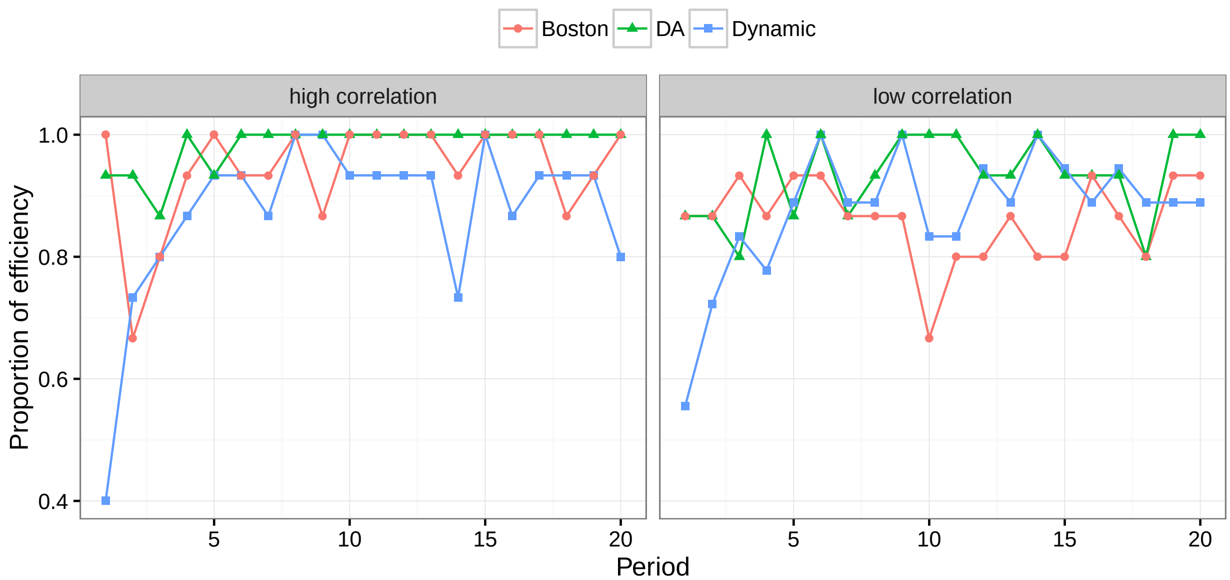


Figure 2.4: Efficiency comparison

Based on Theorem 1, the IM Dynamic is almost as good as DA in efficiency when the revision chances are frequent enough; therefore, we expect DA to overperform IM Dynamic in efficiency in both environments. Since both DA and Boston yield efficient allocation in BNE, we expect they perform similarly in efficiency. Accordingly, we have the following hypothesis:

Hypothesis 3. *The proportion of efficient allocations under the DA and Boston mechanisms are the same. Both of them have higher proportions of efficient allocations than the IM Dynamic mechanism.*

Result 5. *(Efficiency) In the environment where preferences are highly correlated, the proportion of efficient allocations has the following order: DA $\sim$ Boston $\geq$ IM Dynamic. In the environment with low preference correlation, the proportion of efficient allocations has the following order: DA $>$ IM Dynamic $\sim$ Boston. Both DA and the Boston mechanisms have higher proportions of efficient allocations in the high correlation environment. The proportion of efficient allocations under the IM Dynamic mechanism is largely unchanged across environments.*

Table 2.7: Efficiency

Mechanism (1)	Prop (High) (2)	Prop (Low) (3)	Ha (4)	p-value (5)
IM DY	0.873	0.875	High = Low	1.000
DA	0.983	0.933	High = Low	0.024
BOS	0.943	0.860	High = Low	0.032
Hypothesis (6)	**p-value (High) (7)**	**p-value (Low) (8)**		
IM DY<DA	0.004	0.045		
DA = BOS	0.119	0.008		
IM DY <BOS	0.036	0.355		

Support. *Table 2.7 presents the average proportion of efficient allocation in each treatment and the p-values of permutation tests, using each session as an independent observation. The proportions of efficient allocations in the high and low correlation environments are in column (2) and (3), respectively. Columns (4) and (5) compare the performance of the same mechanism across environments. Columns (6) (7) and (8) present pair-wise comparisons of the three mechanisms under the same environment. Figure 2.4 depicts the trend of efficient allocations over the 20 periods of our experiment in both environments. To test the robustness of our results, we use normalized efficiency as an alternative measure and find similar results. A comparison of the three mechanisms using the normalized efficiency measure is provided in Appendix A.3, Table A.7.*

Overall, we conclude that our Hypothesis 3 is supported in the high correlation environment, but not in the low correlation environment. In the high correlation environment, the IM Dynamic mechanism is the least efficient; in the low correlation environment, it is as efficient as the Boston, and less efficient than DA. One drawback of the IM Dynamic mechanism is that students may not be accepted by any school at all. In fact, unallocated students account for 60.39% of inefficient allocations in the high correlation environment, and 68.90% of inefficient allocations in the low correlation environment. In real life, when colleges have unfilled quotas, the central planner will call for another round of applications to have these quotas filled. To see how this would improve the efficiency of the IM Dynamic mechanism, we run a random re-matching among unallocated students and unfilled seats. We find that, with random re-matching, the proportion of efficient allocation goes up to 0.960 in the high correlation environment, and 0.958 in the low correlation environment, which makes the IM Dynamic mechanism as efficient as the Boston mechanism in the high correlation environment (p-value = 0.242 from one-sided permutation test), and as efficient as the DA in the low correlation environment (p-value = 0.877 from one-sided permutation test).

When it comes to cross-environment comparison, both DA and Boston are less efficient in the low correlation environment, due to less Bayesian Nash equilibria being reached, whereas the performance of the IM Dynamic mechanism is robust across environments. One may wonder if the low efficiency of the IM Dynamic mechanism is caused by subjects not having enough time to best respond (i.e., that the "arrivals are frequent enough" condition in our Theorem 1 is not met); however, this is not the case in our experiment. As we have pointed out in Subsection 6.1.3, only 11.111% of sub-optimal last revisions are caused by not having enough time to respond, which accounts for only 0.682% of all last revisions. The percentage of efficiency losses caused by not having enough time to respond is 1.333% in the high correlation environment, and 2.500% in the low correlation environment, respectively. Even without this efficiency loss, the IM Dynamic still performs worse than DA.

We observe that subjects choose not to revise, consciously or not, when they should have revised, which could have been caused by confusion or inattention. We further look at whether subjects learn to revise more actively when they are outranked in later periods. We find that they did not. The inefficiency caused by not revising is not significantly different between earlier periods and later periods. Hence, we attribute most of the low efficiency in IM Dynamic mechanism not to confusion, but to inattention.

We are uncertain about how much of this inattention is due to the low stake of the environmental setting and how much of it would carry out in the field when the stakes are much higher. To address the issue of inattention, the system could be modified to send out an automatic notice when someone is being outranked.

In sum, we find that DA is the most efficient, regardless of the environment. Although theory predicts that the IM Dynamic mechanism should be the least efficient in both environments, it is as efficient as Boston in the low correlation environment. Finally, the efficiency of the IM Dynamic mechanism can be improved by performing after-market matching.

2.7 Conclusion

This study examines a new market design in the context of college admissions, where students submit one choice at a time and are free to change their choices before the pre-announced ending time. We find that theoretically, this dynamic mechanism is close to, but not as good as DA, in that it has no dominant strategy, and the matching may not be efficient and stable even if everyone plays rationally. Compared with the efficient but not stable Boston mechanism, the IM Dynamic is less efficient but more stable. Our lab experiment shows that these theoretical predictions hold when student preference correlation is high. In contrast, when student preference correlation is low, the IM Dynamic outperforms theoretical predictions in truth-telling, stability, and efficiency.

It is as stable as DA, and as efficient as Boston, with a higher truth-telling level than both of them.

Our experimental results show that the IM Dynamic mechanism has behavioral advantages over static mechanisms, such as DA and Boston. The performance of DA and Boston become worse when the strategic complexity of the environment increases. The performance of IM Dynamic, on the other hand, is largely unchanged across environments. By having the truth-telling strategy as the unique rationalizable strategy, the IM Dynamic mechanism reduces the strategic complexity. Moreover, players only need to rank one school to best respond, so the strategy space at each strategy point of IM Dynamic is much smaller than DA and Boston.

In the Chinese college admissions, even though students might have highly correlated preferences over the most prestigious colleges, preferences across the other colleges are much less correlated, as students might prefer different majors or locations. With hundreds of thousands of students[10], and thousands of colleges on each side of the market, the strategic complexity is tremendous under DA or Boston; therefore, the IM Dynamic mechanism would be an attractive alternative.

We further learned that certain practices can improve the efficiency and stability of the IM Dynamic. One practice is to have practice rounds before the actual matching process, as there is a strong learning effect under the IM Dynamic in our experimental data. Another practice is to perform re-matching among unfilled seats and unallocated students after the market to improve the efficiency.

[10]See Table A.11 for the number of college applicants in each province.

CHAPTER 3

Optimal Team Size in Public Goods Provision: Theory and Experiments

3.1 Introduction

Teams are ubiquitous in organizations and online communities. They often emerge for the production of public goods. For example, the English Wikipedia currently has over 2,000 WikiProjects of various sizes. A WikiProject is comprised of a group of Wikipedia editors who joined voluntarily to coordinate their efforts to improve the quality of Wikipedia articles within its scope. WikiProject Economics consists of more than 200 editors to edit Wikipedia articles in Economics. The first peer-to-peer prosocial lending online community, Kiva.org, initiated lending teams to coordinate loan efforts to help small entrepreneurs in developing countries. Research shows that membership in active teams causally increases team member activities and contributions on Kiva using both a field experiment (Ai et al., 2016) and naturally occurring data (Chen et al., 2017). Using a Wikipedia data dump, researchers find that WikiProject membership causally increases both the intensity and diversity of an editor's contributions (Chen et al., 2018b). Therefore, leaders in organizations and online community designers should actively encourage and help organize teams. An open question emerged from the field is the optimal size of teams.

Prior research on team size provides mixed evidence. Early research found little effect of group size on public goods contributions (Ledyard, 1995), whereas recent research shows more complex results depending on a group's ability to monitor defectors and to control migration between groups (Chaudhuri, 2011). Interestingly, social psychology research on social loafing, which focuses on contributions of effort, suggests that people work less hard toward collective goals in larger groups (Karau and Williams, 1993). Further complicating the issue, research in the context of online communities suggests that group size has both positive and negative effects on community sustainability, in the sense that larger groups attract more new members but also drive away more existing ones (Butler, 2001). Large groups are also associated with a stronger sense of social presence, which in turn can have a positive effect on members' attachment and commitment (Dabbish et al., 2012). To reconcile mixed findings regarding group size and composition, we use a game-theoretic approach to test the idea that the optimal team size depends crucially on the degree of complementarity in the production of public goods.

When a platform designer organizes teams and assigns social roles within a team, the degree of complementarity of the public goods production function affects team members' incentives to contribute and the optimal team size. We use two extreme cases as examples.

Consider the case where contributions are perfect complements, such as identifying worthy borrowers versus making a loan, captured by the following minimum-effort utility function: $\pi_i(x_1,\ldots,x_n) = \beta \min\{x_1,\ldots,x_n\} - cx_i + w$ where β and c captures the marginal benefit and marginal cost of contribution, respectively, w is a fixed utility level and $x_i \geq 0$ is the contribution provided by user i. This potential game (Monderer and Shapley, 1996) has multiple Pareto-ranked pure-strategy Nash equilibria. To achieve efficient production, group size cannot be too large, $n < \beta/c$. This is a crisp prediction, validated by many lab experiments (Chen and Chen, 2011).

In contrast, when everyone's contributions are perfect substitutes, such monetary contributions towards a loan in Kiva, which are voluntary contributions mechanism with a linear production function: $\pi_i(x_1,\ldots,x_n) = \beta \sum_{i=1}^{n} x_i - cx_i + w$. Again, we obtain a potential game, which yields the crisp prediction that group size has no effect on contributions.

From the two extreme cases, we now turn to the general form of the constant elasticity of substitution production function, $F = \beta(\sum_i^n x_i^\gamma)^{\frac{1}{\gamma}}$, which nests the perfect complements ($\gamma \to -\infty$) and perfect substitutes ($\gamma = 1$) as special cases, but also covers various degrees of production complementarity. We prove that when the complementarity is low ($0 < \gamma < 1$), there is a lower bound for optimal group size, $\underline{n} = (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$; whereas when the complementarity is high ($\gamma < 0$), there is an upper bound for optimal group size, $\overline{n} = (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$. We conduct a lab experiment to test these theoretical predictions. In the low complementarity condition, the larger group (size 10) performs significantly better than the smaller group (size 4) in providing the public goods, whereas in the high complementarity condition, the smaller group performs significantly better than the larger group. We also find that initial contribution levels are biased towards the middle in all treatments.

Our findings have clear policy implications for team formation in real life. Our study shows that to form effective teams, practitioners need to consider the complementarity of the tasks. When a task relies on complementary efforts from team members, a smaller team can perform better. Whereas when a task relies on substitutable efforts from team members, a large team can perform better. When team size is not changeable, platform designers can consider changing the complementarity of the task. For a large team, reducing the complementarity might improve team performance. In contrast, for a small team, increasing complementarity by introducing different roles might improve team performance.

3.2 Literature Review

Understanding the effect of the group size on cooperative behavior has long been a major topic in economics (Guttman, 1986), psychology (Sweeney, 1974; Karau and Williams, 1993), political science (Olson, 1965), and sociology (Marwell and Ames, 1979).

Theoretically, the effect of group size on the contribution level might be ambiguous. On the one hand, a large group makes non-cooperative behavior difficult to detect, leading to a lower contribution level from self-interested players. On the other hand, a large group magnifies the benefit of cooperation, leading to more contribution from altruistic players (Ledyard, 1995).

Previous laboratory experiments examining the effect of group size employed different forms of production functions. These production functions include step-wise functions, where different levels of group contribution lead to different levels of group returns (Chamberlin, 1978; Marwell and Ames, 1979; Pereda et al., 2019); linear functions, where the group return is proportional to group contribution (Isaac and Walker, 1988; Isaac et al., 1994; Goeree et al., 2002; Nosenzo et al., 2015; Shank et al., 2015; Diederich et al., 2016; Weimann et al., 2019); concave functions, such as quadratic functions (Laury et al., 1999; Laury and Holt, 2008), power functions (Guttman, 1986), and Cobb-Douglas functions (Andreoni, 1993). As different functional forms affect the incentive to contribute and generate different predictions on the effect of group size, we structure our discussion based on the classes of production functions, in the order of linear, quadratic and power functions.

In the linear public goods setting, a robust positive effect of group size has been found with low to medium marginal per-capita returns (MPCR). The positive effect of group size is robust across relatively small groups (Isaac and Walker, 1988; Goeree et al., 2002; Nosenzo et al., 2015; Shank et al., 2015) with less than twelve subjects per group and large groups with up to one hundred subjects per group (Isaac et al., 1994; Diederich et al., 2016; Weimann et al., 2019). The effect is also confirmed in the general population (Diederich et al., 2016) and when monitoring and punishment are available (Carpenter, 2007). With MPCRs of 0.75 or above, however, researchers find either no group size effect (Isaac and Walker, 1988) (MPCR = 0.75) and Shank et al. (2015) (MPCR = 0.8), or a negative group size effect Nosenzo et al. (2015) (MPCR = 0.75).

When the production function is concave, contributing zero is no longer a Nash equilibrium. Instead, there is an interior solution for the optimal total group contribution. Among experiments where the production function is a concave function of the sum of players' effort, only two studies vary the group size. Guttman (1986) used a power function as the production function and Laury et al. (1999) used a quadratic function as the production function. Both find evidence that increasing the group size leads to an increase in aggregate contributions to the group but a decrease in average contribution.

Despite using different production functions, group members' efforts are perfect substitutes in both the linear and the concave public goods game because the group return is a function of the sum of group efforts in both cases.

In addition to evidence from the lab, studies using field data also provide insight into the effect of group size on cooperative behavior. Using data from the Wolong Nature Reserve, Yang et al. (2013) found that medium-size groups perform better than small or large groups in collective actions (forest monitoring) and resource outcomes (changes in forest cover), indicating a non-linear effect of group size. Using data from Wikipedia, Zhang and Zhu (2011) found that contributors decreased their contributions by 42.8% after an exogenous reduction of group size due to the blocking of Chinese Wikipedia in mainland China, indicating a positive group size effect. Using data from a large MOOC class, Baek and Shore (2020) found that larger forums elicit more contribution per person, indicating a positive group size effect. Regardless, neither of the studies identified the form of the production function, nor the complementary in group members' efforts.

The previous literature on the effect of group size assumes that group members' efforts are perfect substitutes or fail to identify the level of complementarity. Our paper introduces complementarity into the theoretical model and proves that the optimal group size depends on the level of complementarity. When the complementarity is high, there is an upper bound for the optimal group size; whereas when the complementarity is low, there is a lower bound for the optimal group size.

One paper that also incorporates complementarity of efforts is Fenig et al. (2018). However, their focus is different from our paper, and their group size is fixed at four. Their experimental results can still validate our theory's predictions. Table B.1 in Appendix B.2 compares our theoretical prediction with their experiment results.

Our theoretical predictions are also validated by previous literature on the minimum effort game. Summarized by Chen and Chen (2011), previous experiments on minimum efforts game showed that when the group size was smaller than the ratio of the marginal benefit of the public good (β) divided by the individual marginal cost of effort (c), participants coordinated on the high contribution equilibrium; whereas when the group size was larger than the ratio, participants coordinated on the low contribution equilibrium. This ratio β/c is the predicted upper bound for the high contribution equilibrium in our theoretical model.

3.3 A Theoretical Framework and Hypotheses

In this section, we set up a theoretical framework to derive our hypotheses. Let $i = 1, 2, \cdots, n$ be a member of a group of size n. Each member has a private good endowment, $\omega \in [\underline{\omega}, \bar{\omega}]$, where $\bar{\omega} > \underline{\omega} > 0$. An individual's endowment can be kept for private consumption or contributed for public good production.

To model the complementarity of inputs, we use the Constant Elasticity of Substitution (CES) function as the public good production function. While the CES production function is discussed in standard microeconomics textbooks, e.g., Varian (1984, page 30), we include its basic properties here for completeness. Formally, we define the production function as:

$$F(x_1,\ldots,x_n) = \beta(\sum_i^n x_i^\gamma)^{\frac{1}{\gamma}}, \tag{3.1}$$

where $x_i > 0$ is group member i's contribution, $\beta > 0$ is a constant, and $\gamma \in (-\infty, 1]$ is a complementarity parameter, with a smaller γ indicating more complementarity.[1]

Using the CES function as the production, the utility function for player i in the voluntary contribution mechanism is assumed to be:

$$\pi_i = \omega - cx_i + \beta\left(\sum_i^n x_i^\gamma\right)^{\frac{1}{\gamma}}, \tag{3.2}$$

where $c \geq 0$ is the marginal cost of effort. Equation (3.2) incorporates a wide range of games. We discuss three special cases below:

1. When $\gamma = 1$, individual inputs are perfect substitutes. The CES production function becomes $F = \beta\sum_i^n x_i$, and Equation (3.2) defines the class of linear public goods games (Ledyard, 1995).

2. When $\gamma = -\infty$, inputs are perfect complements. The CES production function becomes $F = \beta\min\{x_1,\ldots,x_n\}$, and Equation (3.2) defines the minimal-effort game (Van Huyck et al., 1990).

3. When $\gamma = 0$, the CES production function becomes the Cobb-Douglas function $F = \prod_i^n x_i^{\frac{1}{n}}$, under the condition that $\beta = (1/n)^{\frac{1}{\gamma}}$, as the CES function is not well-defined for other values of β.

We now derive the Nash equilibrium predictions in this class of utility functions. We first derive the interior equilibrium, with individual contribution, $x_i \in (\underline{\omega}, \bar{\omega})$, for all i. The first order condition for player i is as follows:

$$\beta x_i^{\gamma-1}\left(\sum_{i=1}^n x_i^\gamma\right)^{\frac{1-\gamma}{\gamma}} - c = 0, \quad x_i \in (0,\omega), \forall i. \tag{3.3}$$

[1] As the CES function is not well-defined at $x_i = 0$ if $\gamma < 0$, one could define the function arbitrarily in this case. It is not relevant for our experimental setup.

Define $S \equiv \sum_{i=1}^{n} x_i^{\gamma}$. We can rewrite Equation (3.3) as

$$\beta x_i^{\gamma-1} (S)^{\frac{1-\gamma}{\gamma}} = c, \quad x_i \in (0, \omega), \forall i. \tag{3.4}$$

It follows that, for any player $i \neq j$, we can take the ratio of their first-order conditions and obtain

$$\frac{x_i^{\gamma-1}}{x_j^{\gamma-1}} = 1, \quad \text{or} \quad x_i = x_j.$$

That is, any interior Nash equilibrium must be symmetric, which implies that Equation (3.3) becomes

$$\beta n^{\frac{1-\gamma}{\gamma}} - c = 0, \tag{3.5}$$

which is independent of x_i. We now check the second-order condition for the interior solution and obtain:

$$\frac{\partial^2 \pi_i}{\partial x_i^2} = \beta(1-\gamma) x_i^{\gamma-2} S^{\frac{1-\gamma}{\gamma}} (\frac{x_i^{\gamma}}{S} - 1) \tag{3.6}$$

Because $x_i^{\gamma} < S$, $(\frac{x_i^{\gamma}}{S} - 1) < 0$. When $\gamma < 1$, $\frac{\partial^2 \pi_i}{\partial x_i^2} < 0$, hence, the payoff function is concave. When $\gamma = 1$, $\frac{\partial^2 \pi_i}{\partial x_i^2} = 0$, hence, the payoff function is a linear function.

We now summarize three classes of solutions:

- **Case 1**: Interior solutions: When $\beta n^{\frac{1-\gamma}{\gamma}} - c = 0$, for any $x_i \in (\underline{\omega}, \bar{\omega})$, the first-order condition holds. This means that we have an infinite number of symmetric interior Nash equilibria, i.e., for any $x \in (\underline{\omega}, \bar{\omega})$, $x_i = x$, $\forall i$, is a Nash equilibrium.

- **Case 2**: Boundary solutions: When $\beta n^{\frac{1-\gamma}{\gamma}} - c > 0$, for any value $x_i \in [\underline{\omega}, \bar{\omega})$, a player always want to contribute more, which means that $x_i = \bar{\omega}$ is a Nash equilibrium.

- **Case3**: Boundary solutions: When $\beta n^{\frac{1-\gamma}{\gamma}} - c < 0$, for any value $x_i \in (\underline{\omega}, \bar{\omega}]$, a player always want to contribute less, which means that $x_i = \underline{\omega}$ is a Nash equilibrium. .

We now derive the implication on group size under each case:

- **Case 1**: When $n = (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$, group size does not affect the contribution level.

- **Case 2 & Case 3**: When $0 < \gamma < 1$, if $n > (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$, the high contribution equilibrium exists; if $n < (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$, the low contribution equilibrium exists. When $\gamma < 0$, if $n > (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$, the low contribution equilibrium exists; if $n < (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$, the high contribution equilibrium exists.

Case 2 and Case 3 lead to the same prediction: When $0 < \gamma < 1$, there is a lower bound $\underline{n}$, such that any group size larger than $\underline{n}$ would lead to the high contribution equilibrium, whereas any group size smaller $\underline{n}$ would lead to the low contribution equilibrium. When $\gamma < 0$, there is an upper bound $\bar{n}$, such that any group size larger than $\bar{n}$ would lead to low contribution equilibrium and any group size smaller than $\bar{n}$ would lead to high contribution equilibrium. This gives us Theorem 1.

Theorem 3. *When the complementarity is relatively low ($0 < \gamma < 1$), there is a unique high contribution equilibrium when the group size is larger than $(\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$ and a unique low contribution equilibrium when the group size is smaller than $(\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$. When the complementarity is relatively high ($\gamma < 0$), there is a unique low contribution equilibrium when the group size is larger than $(\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$ and a unique high contribution equilibrium when the group size is smaller than $(\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$.*

Theorem 1 implies when the complementarity is low, players will coordinate on the high contribution equilibrium in large groups. In contrast, when the complementarity is high, players will coordinate on the high contribution equilibrium in small groups. Accordingly, we have the following hypotheses:

Hypothesis 4. *When the complementarity is low ($0 < \gamma < 1$), there is a lower bound for optimal group size, $\underline{n} = (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$*

Hypothesis 5. *When the complementarity is high ($\gamma < 0$), there is an upper bound for optimal group size, $\overline{n} = (\frac{c}{\beta})^{\frac{\gamma}{1-\gamma}}$*

Both theoretical and experimental studies of public goods provision point to the importance of learning dynamics. When incorporating dynamic learning models, it is useful to examine the potential function of the game, as described by Monderer and Shapley (1996) and defined below. One interesting property of potential games is that several learning algorithms converge to the argmax set of the potential, including a log-linear strategy revision process Blume (1993), myopic learning based on a one-sided better reply dynamic and fictitious play Monderer and Shapley (1996). Under these learning dynamics, the potential-maximizing equilibrium has the largest basin of attraction. It is for this reason that we study the potential function of the public goods game with CES production functions.

Monderer and Shapley (1996) formally define *potential games* as games that admit a potential function P such that:

$$\pi_i(x_i, x_{-i}) \geq \pi_i(x_i^{'}, x_{-i}) \quad \Leftrightarrow \quad P(x_i, x_{-i}) \geq P(x_i^{'}, x_{-i}), \quad \forall i, x_i, x_i^{'}, x_{-i}. \tag{3.7}$$

A potential function is a global function defined on the space of pure strategy profiles such that the change in any player's payoffs from a unilateral deviation is exactly matched by the change in the potential P. To determine whether a game has a potential function, Ui (2000) notes that every potential game has a symmetric structure. The Cournot oligopoly game with a linear inverse demand function is a well-known example of a potential game, where each player's payoff depends on a symmetric market aggregate of all players' outputs (the inverse demand function), and also on her own output (the cost of production). Similarly, the minimum-effort game is also a potential game.

When the payoff functions are twice continuously differentiable, Monderer and Shapley (1996) present a convenient characterization of potential games. That is, a game is a potential game if and only if the cross partial derivatives of the utility functions for any two players are the same, i.e.,

$$\frac{\partial^2 \pi_i(x_i, x_{-i})}{\partial x_i \partial x_j} = \frac{\partial^2 \pi_j(x_j, x_{-j})}{\partial x_i \partial x_j} = \frac{\partial^2 P(x_i, x_{-i})}{\partial x_i \partial x_j}, \quad \forall i, j \in N. \tag{3.8}$$

Equation (3.8) can be used to identify potential games. If (3.8) holds, the potential function P can be calculated by integrating (3.8). Similar conditions hold for non-differentiable payoff functions by replacing "differentials" with "differences" Monderer and Shapley (1996).

For public goods games with a CES production function with a payoff function defined by Equation (3.2) is a potential game with the potential function:

$$P(x_1, \ldots, x_n) = \beta \left(\sum_i^n x_i^\gamma \right)^{\frac{1}{\gamma}} - c \sum_{i=1}^n x_i. \tag{3.9}$$

3.4 Experimental Design

To test the hypotheses, we use two production functions, one with low complementarity ($\beta = 0.4$, and $\gamma = 0.7$), the other with high complementarity ($\beta = 15$, $\gamma = -2$). In addition, we set the marginal cost of effort $c = 1$ for all conditions. These two production functions are chosen to have different equilibrium predictions for the group size of four and ten. For the low complementarity production function, the lower bound is $0.4^{\frac{0.7}{0.7-1}} = 8.48$. Therefore, we expect the group of size four will coordinate on the low contribution equilibrium, whereas the group of size ten will coordinate on the high contribution equilibrium. For the high complementarity production function, the upper bound is $15^{\frac{-2}{-2-1}} = 6.08$. Therefore, we expect the group of size four will coordinate on the high contribution equilibrium, whereas the group of size ten will coordinate on the low contribution equilibrium. This setting is summarized in Table 3.1. We refer to the high complementarity, 4-person group treatment as the High_Small treatment. The other three treatments are High_Large, Low_Small, and Low_Large, correspondingly.

We conducted our experiment at the School of Information Behavioral Lab, University of

Table 3.1: Parameters and Predictions

Complementarity	γ	β	Threshold	Group of 4 Prediction	Group of 10 Prediction
Low	0.7	0.4	8.48	Low Contribution	High Contribution
High	-2	15	6.08	High Contribution	Low Contribution

Michigan, between December 2016 and March 2017. To keep the same session size across different treatments, we have 20 subjects in each session, so that in the treatments with the 4-person group, there are five groups, while in the treatments with the 10-person group, there are two groups. The game is played for 20 periods. To avoid the formation of reputation, at the beginning of each period, subjects are randomly re-matched. Table 3.2 summarizes feature of experimental sessions.

Table 3.2: Features of expeirmental sessions

Complementarity	Group Size	Subjects per session	# of sessions	Total # of subjects
Low	4	20	3	60
	10	20	3	60
High	4	20	3	60
	10	20	3	60

At the beginning of each session, after subjects were seated in front of a terminal, the experimenter read the instructions aloud. Subjects were then given time to read the instructions on their own and ask questions. This process lasted 10 minutes. We then asked subjects to answer 12 review questions with $0.2 for each correct answer. After subjects went over the correct answer for review questions, they played 20 rounds of the game. After the game, subjects were asked to play the Holt and Laury lottery (Holt and Laury, 2002) and to fill out a demographic survey. At the end of the experiment, subjects' payment is the sum of their quiz earnings, their earnings from 1 round of the public goods game, and their earnings from the lottery. On average, each session lasted 50 minutes, and subjects earned $13.5. We did a balanceness check over characteristics collected in the survey. Table B.2 reports the summary of statistics on gender, age, risk attitude, ethnicity, student status, major, and household income. We did not find statistical significant difference in any of these characteristics, except in risk attitude. We further used pairwise t-test with Holm–Bonferroni correction to find which two treatments are different in risk attitude. It turns out that the High_Large treatment is 0.852 more risk averse than the Low_Four treatment using the Holt and Laury risk measure (p-value= 0.042, after multiple hypotheses testing correction). We later showed that risk attitude has no effect on contributions (Table B.4). Results from any other pairwise comparisons are not significant.

To help subjects make decisions, we provided a calculator with build-in production functions. Subjects can calculate their return from the group account, individual account, and total return by simply putting in others' hypothetical contributions and their own hypothetical contribution.

3.5 Results

In this section, we first examine the level of contributions in each treatment. We then explain the contribution dynamics using behavioral models.

3.5.1 Contribution Level

We first look at the first period contribution level. Since subjects have not interacted with each other in the first period, their beliefs are not affected by other subjects' contribution level, and their decisions are not affected by social norms formed in this experiment. We then look at how contribution levels evolve over the twenty periods.

3.5.1.1 First Period

Since subjects have not interacted with each other in the first period, we treat each individual as an independent observation. Therefore we have 60 observations in each treatment.

Figure 3.1 presents the empirical distributions of first-period contributions in each of the four treatments. In treatments (High_Small and Low_Large) where we have the high contribution equilibrium, subjects are more likely to contribute 10 points and 5 points. In contrast, in treatments (High_Large and Low_Small) where we have the low contribution equilibrium, subjects are more likely to contribute 1 point and 5 points. We run two non-parametric tests on the empirical distributions. One is the one-sided permutation test using the Monte Carlo method. The other is the one-sided Wilcoxon rank-sum test. Both tests give us the same results: the contribution levels in predicted high contribution treatments are significantly higher than in predicted low contribution treatments. Table 3.3 reports the first-period treatment average, the null hypotheses, and p-values.

This shows that contribution levels in different treatments already differ in the first period. Without forming beliefs on others' contribution levels, subjects already respond to the incentive structure of different treatments.

Figure 3.1: Distributions of First-Period Contributions by Treatments

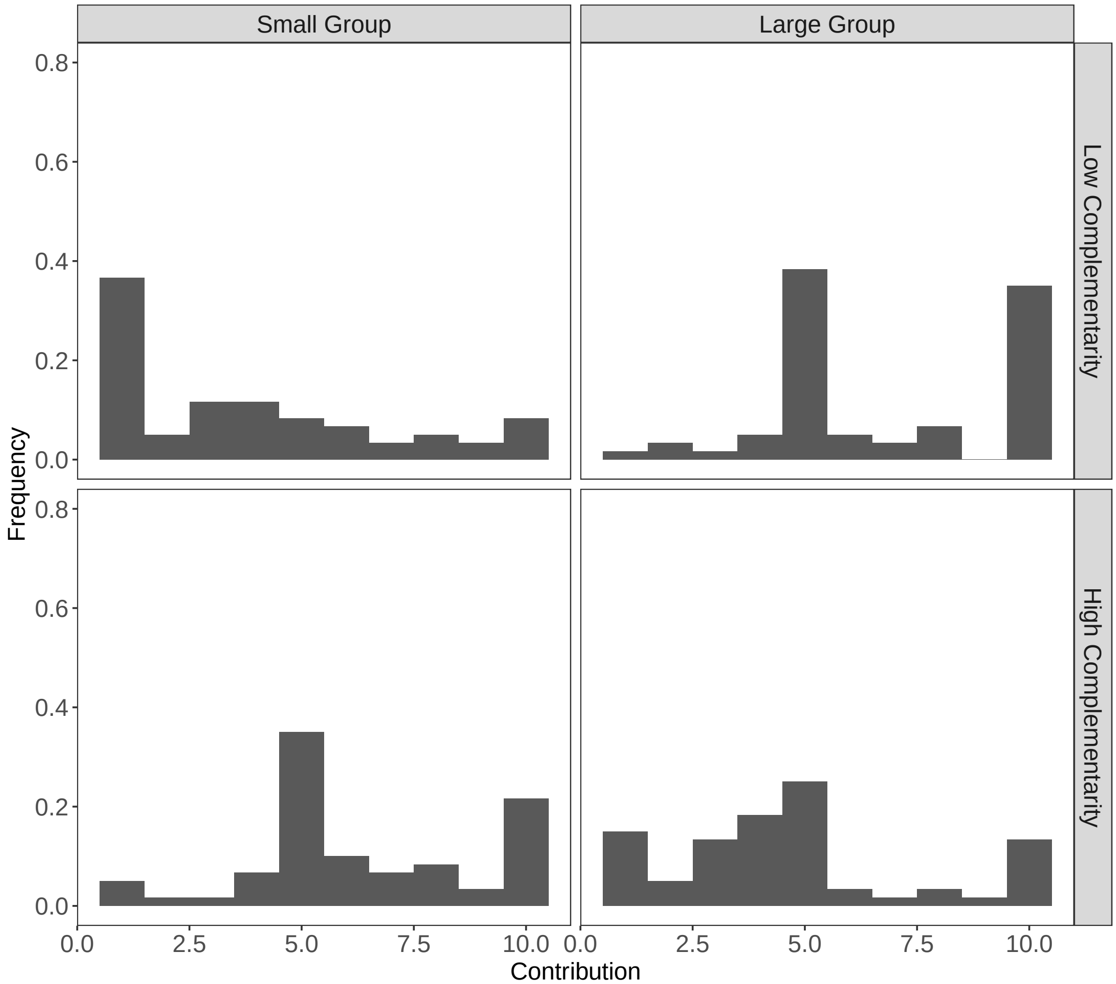


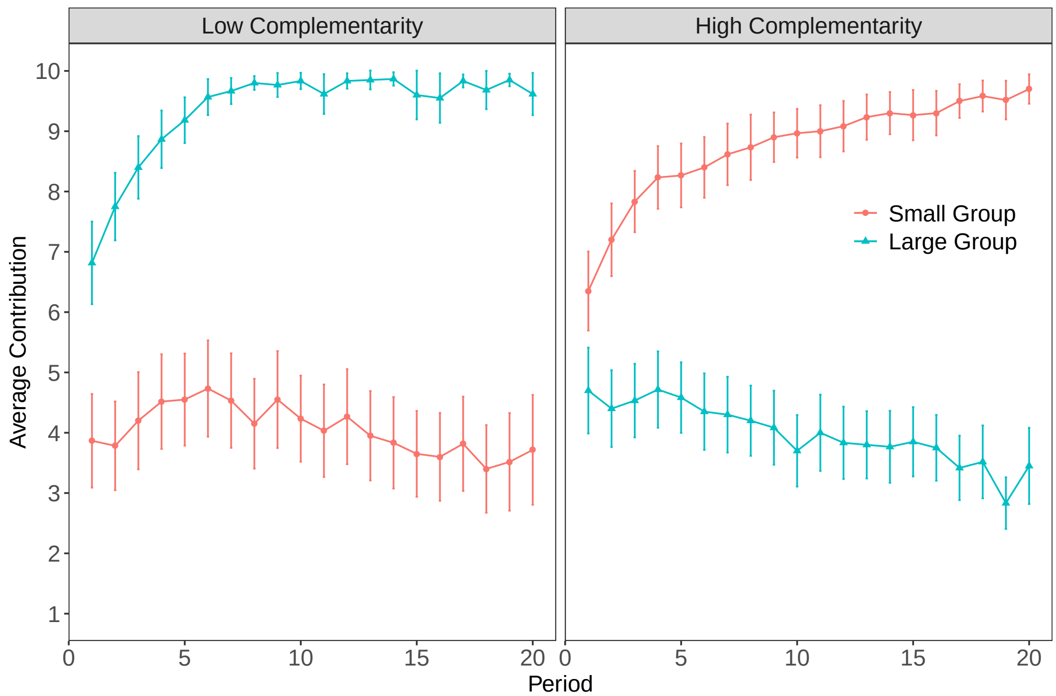
Table 3.3: Average Contribution in The First Period

Complementarity	Group Size	Average Contribution	Ha	P-value (Permutation)	P-value (Rank-Sum)
Low	Small Large	3.867 6.817	Small < Large	0.001	< 0.001
High	Small Large	6.350 4.700	Small > Large	0.001	< 0.001

3.5.1.2 All Periods

We then look at subjects' contribution levels over time. Figure 3.2 shows the average contribution in each period. Under low complementarity, subjects contribute more in the large group, whereas under high complementarity, subjects contribute more in the small group. The difference between small and large groups expands over time. The average contribution levels in the two high contribution treatments converge to full contribution in the last period, whereas the average contribution levels in the two low contribution treatments decrease to 3 points.

Figure 3.2: Average Contributions Over Time: Error bars indicate 95% confidence intervals.



We treat each session as one independent observation. Hence we have three independent observations under each treatment. To test our hypotheses, we run the one-sided permutation test. Under both the high complementarity and low complementarity treatments, we are able to reject the null hypotheses in favor of the alternative hypotheses at the 0.05 significance level. Table 3.4 summarizes average contributions, null hypotheses, and p-values.

Table 3.4: All Periods Average Contribution

Complementarity	Group Size	Average Contribution	Ha	P-value (Permutation)
Low	Small Large	4.045 9.348	Small < Large	0.05
High	Small Large	8.749 3.989	Small > Large	0.05

3.5.2 Contribution Dynamics

We try to explain the contribution dynamics using three different behavioral models. The first model is the conditional cooperation model, where a player's contribution is affected by other players' contributions in the same group. The second model is the best-response model, where a player best responds to her belief about other players' contribution. The third model is the fictitious play model, where a player exponentially discounts her learning experience in previous periods and best responds to her updated belief.

3.5.2.1 Conditional Cooperation Model

Because our production function is quite complicated and solving best responses is cognitively costly, subjects might simply follow what others do in previous periods. To see how players' contribution is affected by other players' contribution in the same group, we regress players' contribution on others' average contribution in previous periods. We find that other players' average contribution level in the previous period is predictive of players' contribution level in the current period in all treatments. However, others' contribution levels two periods before are only predictive in the High_Small treatment and High_Large treatments. Adding more periods into the regression does not improve the prediction. Table 3.5 and Table 3.6 show the detailed regressions. In sum, others' contribution in the previous period is predictive of subjects' contribution, but only explains a small proportion of subjects' contribution.

Table 3.5: Effect of previous four periods others' average contributions (Low Complementarity)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contribution	Small Group				Large Group			
Other_Ave lag1	0.359*	0.308**	0.300**	0.299**	0.613***	0.585***	0.512***	0.420***
	(0.113)	(0.070)	(0.054)	(0.031)	(0.013)	(0.016)	(0.048)	(0.039)
Other_Ave lag2		0.201	0.193*	0.168		0.020	-0.054	-0.016
		(0.098)	(0.066)	(0.078)		(0.032)	(0.041)	(0.049)
Other_Ave lag3			0.060	0.049			0.085	0.086
			(0.100)	(0.077)			(0.046)	(0.049)
Other_Ave lag4				0.083				-0.005
				(0.069)				(0.065)
Constant	2.598*	1.991	1.793	1.571	3.755***	3.850**	4.469**	5.034***
	(0.734)	(0.824)	(1.001)	(1.084)	(0.161)	(0.485)	(0.499)	(0.144)
Observations	1,140	1,080	1,020	960	1,140	1,080	1,020	960
R-squared	0.050	0.068	0.073	0.078	0.177	0.127	0.085	0.048

Robust standard errors in parentheses
We exclude the first period for the regressions on the previous period.
We exclude the first two periods for the regressions on the previous two periods, etc.
*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Effect of previous four periods others' average contributions (High Complementarity)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contribution	Small Group				Large Group			
Other_Ave lag1	0.636***	0.465***	0.392***	0.379***	0.749**	0.584***	0.553***	0.520***
	(0.061)	(0.024)	(0.027)	(0.029)	(0.096)	(0.052)	(0.036)	(0.003)
Other_Ave lag2		0.253***	0.203**	0.178***		0.226*	0.138	0.155
		(0.010)	(0.024)	(0.014)		(0.073)	(0.083)	(0.072)
Other_Ave lag3			0.167***	0.147**			0.148***	0.045
			(0.016)	(0.026)			(0.005)	(0.075)
Other_Ave lag4				0.076				0.133
				(0.030)				(0.055)
Constant	3.344**	2.668***	2.312**	2.167**	0.943	0.682	0.534	0.441
	(0.337)	(0.217)	(0.334)	(0.408)	(0.542)	(0.550)	(0.494)	(0.468)
Observations	1,140	1,080	1,020	960	1,140	1,080	1,020	960
R-squared	0.270	0.285	0.289	0.287	0.141	0.144	0.150	0.155

Robust standard errors in parentheses
We exclude the first period for the regressions on the previous period.
We exclude the first two periods for the regressions on the previous two periods, etc.
*** p<0.01, ** p<0.05, * p<0.1

3.5.2.2 Best Response Model

In this section, we explore whether subjects best respond to their beliefs. We use subjects' inputs in the calculator as their beliefs. The usage of the calculator is rare on average. Only 42.3% participants ever use a calculator to make contribution decisions, and the calculator usage only appears in 9.4% of the decisions in the experiment. The usage of calculators decreases over rounds. On average, a subject uses the calculator 0.545 times in the first round, and 0.109 times in the last round. Furthermore, the usage of calculators are not predicted by observable characteristics, such as gender, age, risk attitude, ethnicity, student status, and household income. The only characteristic that is predictive of calculator usage is major. "Math/Science/Engineering/Computer Science" students usage the calculator significantly more than students from other majors (p-value = 0.030).

We do not find evidence that the usage of the calculator has positive effects on payoff at the individual level and at the choice level. At the individual level, there is a small negative correlation (-0.149) between the number of time an individual use the calculator and her total earning in twenty rounds. Only in the High_Large treatment, the usage of calculator leads to a small increase (0.244 points) in the payoff. The effect of calculator usage for other treatments is not significant. Detailed statistics are reported in Table 3.7, with standard errors clustered at the session level.

Among the 9.4% of decisions that involves calculator usage, we examine the following questions:

1. Do subjects have correct beliefs over other subjects' contribution?

2. Do subjects best respond to their beliefs?

Table 3.7: Effects of calculator usage on payoff

	All	Low Complementarity		High Complementarity	
		Small	Large	Small	Large
Payoff	(1)	(2)	(3)	(4)	(5)
Period	0.481**	-0.0938	0.952**	1.143**	-0.133*
	(0.195)	(0.0400)	(0.137)	(0.218)	(0.0398)
# of Calculator Usage	-1.933*	0.153	-0.567*	-0.698	0.244***
	(1.035)	(0.0901)	(0.193)	(0.243)	(0.0108)
Constant	48.81***	19.04**	90.65***	51.91**	20.24***
	(9.648)	(0.521)	(2.526)	(7.498)	(1.677)
Observations	4,400	1,200	1,200	800	1,200
R-squared	0.013	0.026	0.367	0.262	0.040

Standard errors clustered at session level

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

For the Low_Small treatment, we only have data on two sessions.

To answer the first question, we define "others' complementary sum" s_i as the sum of other player's contribution to the power of the complementarity parameter:

$$s_i = \sum_{j \neq i} x_j^{\gamma} \tag{3.10}$$

We define this terminology because knowing others' complementary sum is sufficient to compute the best response:

$$x_i^* = \left(\frac{\sum_{j \neq i} x_j^{\gamma}}{\beta^{\frac{\gamma}{\gamma-1}} - 1}\right)^{\frac{1}{\gamma}} = \left(\frac{s_i}{\beta^{\frac{\gamma}{\gamma-1}} - 1}\right)^{\frac{1}{\gamma}} \tag{3.11}$$

Furthermore, we define the hypothetical best response as:

$$\hat{x}_i^* = \left(\frac{\sum_{j \neq i} \hat{x}_j^{\gamma}}{\beta^{\frac{\gamma}{\gamma-1}} - 1}\right)^{\frac{1}{\gamma}} = \left(\frac{\hat{s}_i}{\beta^{\frac{\gamma}{\gamma-1}} - 1}\right)^{\frac{1}{\gamma}}, \tag{3.12}$$

where $\hat{x}_j$ is the calculator input of others' hypothetical contribution, and $\hat{s}_i$ is the hypothetical complementary sum.

To measure the accuracy of belief, we compare the difference between x_i^* and $\hat{x}_i^*$. The difference between the two captures the difference between s_i and $\hat{s}_i$. For each decision, we first compute the best response x_i^* using other players' real contributions. We then compute the hypothetical best response $\hat{x}_i^*$ for every set of calculator inputs associated with this decision. Note that, a player may have multiple sets of calculator inputs for one decision, therefore, there exists multiple $\hat{x}_i^*$ for the same x_i^*. We compute the difference between each $\hat{x}_i^*$ and its corresponding x_i^* and plot Figure 3.3. As we can see from Figure 3.3, subjects' beliefs are highly concentrated around others' real contribution in all treatments, except the High_Small treatment. Table 3.8 shows the mean and standard deviation of the difference between the two best responses, together with p-value from two-sided t tests. In general, subjects' beliefs are accurate. In the High_Large, and Low_Small treatments, hypothetical best responses are not statistically significantly different from real best responses. In the High_Small and Low_Large treatments, even though the hypothetical best responses are downward biased, meaning that the belief about others' contribution is also downward biased, the biased is small in the Low_Large treatments. Only in the High_Small treatment, the biased belief leads to 1.177 units lower contribution if subjects were best responding to their beliefs.

Table 3.8: The Accuracy of Beliefs

Complementarity	Group Size	$\hat{x}_i^* - x_i^*$	STD	Ha	P-Value
High	Small	-1.177	2.791	$\hat{x}_i^* - x_i^* \neq 0$	0.000
	Large	-0.022	0.764	$\hat{x}_i^* - x_i^* \neq 0$	0.508
Low	Small	-0.022	0.786	$\hat{x}_i^* - x_i^* \neq 0$	0.752
	Large	-0.434	0.979	$\hat{x}_i^* - x_i^* \neq 0$	0.000

The second question is: Do subjects best respond to their beliefs? To answer this question, for each set of calculator inputs, we first compare subjects' calculator input for their own contribution (represented by x_i') with the hypothetical best responses ($\hat{x}_i^*$). If $x_i' = \hat{x}_i^*$, it means that subjects are able to find out the best response. We then examine if the hypothetical contribution ($\hat{x}_i^*$) in the calculator relates with the real contribution (x_i). For each real contribution (x_i), there may exist multiple hypothetical contributions ($\hat{x}_i^*$).

We find that subjects' calculator input is highly concentrated around the hypothetical best response in all treatments. Figure 3.4 plots the difference between between each pair of x_i' and $\hat{x}_i^*$. As we can see from Figure 3.4, in all treatments, the histogram is heavily concentrated around zero. Nevertheless, the calculator inputs are biased in a certain way. The inputs are downward biased for the two treatments with the high contribution equilibrium (High_Small and the Low_Large treatments). In contrast, in the two treatments with low contribution equilibrium (High_Large and the Low_Small treatments), the inputs are upward biased. Given that subject beliefs are largely unbiased, this shows that subjects' calculator inputs are biased towards the middle. Table 3.9 presents the statistics summary of the difference between the calculator input contribution and the hypothetical best response.

We then examine how the hypothetical contribution in the calculator relates to the real contribution? To answer this question, we first check whether the real contribution belongs to the set of

Figure 3.3: Difference between the hypothetical best response and the real best response by treatments

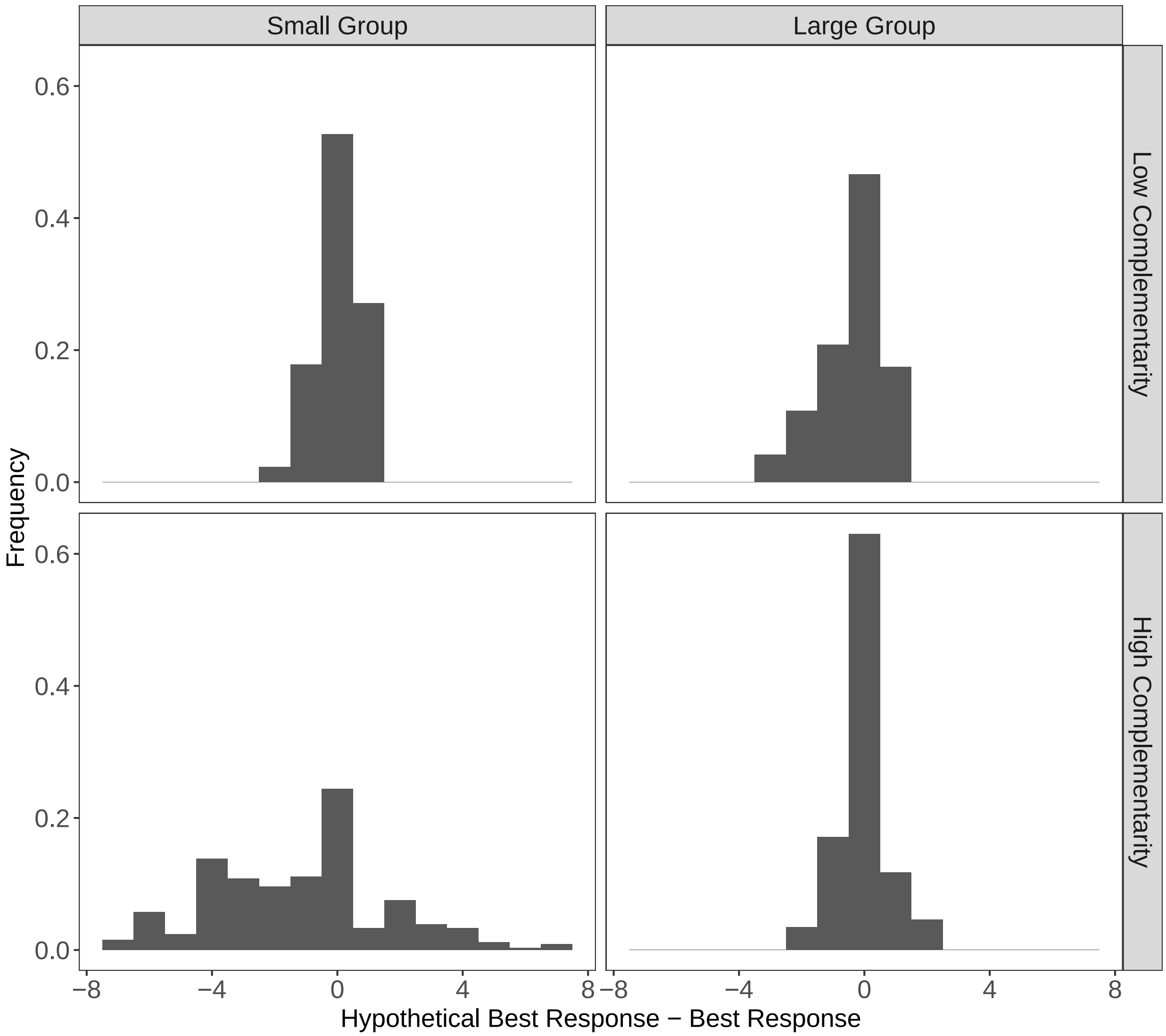
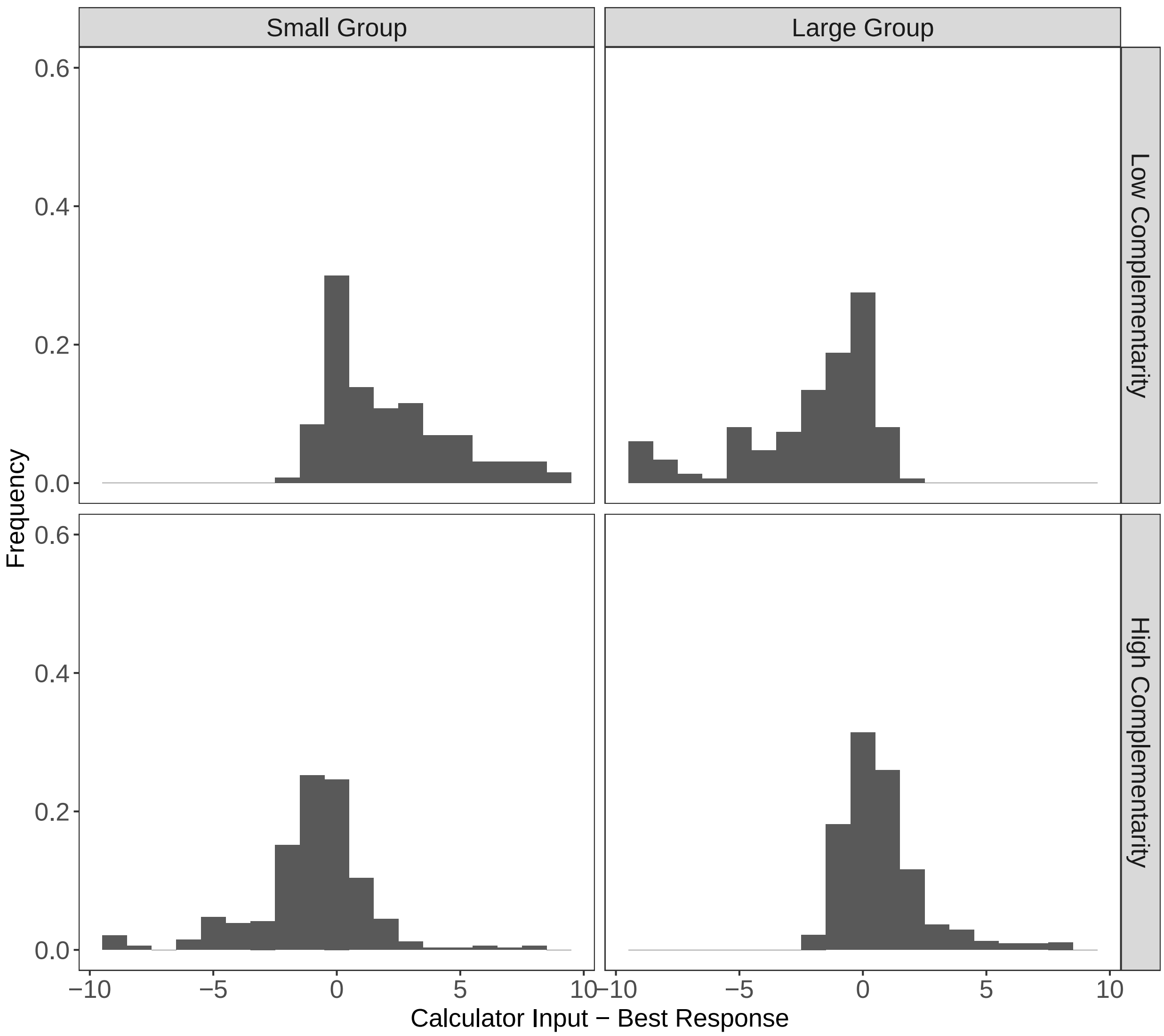


Table 3.9: Difference Between The Calculator Input and The Best Response

Complementarity	Group Size	$x_i' - \hat{x}_i^*$	STD	H_0	P-Value
High	Small	-1.036	2.414	$x_i' = \hat{x}_i^*$	0.000
	Large	0.760	1.731	$x_i' = \hat{x}_i^*$	0.000
Low	Small	2.008	2.532	$x_i' = \hat{x}_i^*$	0.000
	Large	-2.121	2.785	$x_i' = \hat{x}_i^*$	0.000

Figure 3.4: Difference between the calculator contribution input and the hypothetical best response



hypothetical contributions experimented in the calculator. If it coincides with one of the hypothetical contributions, we further decompose it into three categories: the hypothetical best response, the first input, and the last input. It turns out on average, for all contributions that involve calculator usage, the proportion of contributions that belongs to hypothetical contributions is 62.4%. Among these contributions that belong to the hypothetical contributions, 47.5% of them are the hypothetical best response, while 25.5% of them are the first input in the calculator, and 20.1% of them are the last input in the calculator. The percentage under each treatment varies. These statistics are reported in Table 3.11. This shows that for the contributions that involve calculator usage, the contribution level may not be directly related to calculator inputs. For those directly related to calculator inputs, only less than half of them are the hypothetical best responses.

Contributions that does not belong to hypothetical contributions are biased in the same way as how hypothetical contributions are biased from hypothetical best responses: in the two high contribution treatments, the contributions are biased downward; whereas in the two low contribution treatments, the contributions are biased upward. Table 3.10 reports the magnitudes of this bias and the corresponding t-tests.

Table 3.10: Difference Between Contributions Not Related to Calculator Inputs and The Best Response

Complementarity	Group Size	Contribution - Best Response	P-value (t-test)
Low	Small	2.571	0.000
	Large	-2.615	0.013
High	Small	-1.236	0.000
	Large	1.158	0.000

Table 3.11: Relationship between real contribution and hypothetical contribution

Complementarity	Group Size	% hyp contribution	% hyp best responses	% first input	% last input
High	Small	0.452	0.383	0.426	0.277
	Large	0.794	0.518	0.134	0.143
Low	Small	0.477	0.333	0.333	0.333
	Large	0.622	0.609	0.435	0.217
Total		0.624	0.475	0.255	0.201

We first check if a contribution is a hypothetical best response.
If not, we check if it is the first input. If not, we check if it is the last input.

In sum, among choices that involve calculator usage, only 30% of them can be explained as best responding to beliefs. This shows that among subjects who use calculators, the best response model cannot characterize their behavior well.

Since only a small amount of choices involves calculator usage, we lose most of our data by only looking at this dataset. To make the most use of our data, we use others' contribution in the previous period as subjects' belief and investigate whether subjects best respond to others' contribution in previous periods. To do this, we first compute others' complementary sum in all previous periods using Equation 3.10. We then compute the best response for each complementary sum and regress a player's current period contribution to the best responses in previous periods. The detailed regressions are in Table 3.13 and Table 3.12. We find that only the best response to the previous period is predictive of the current period contribution under all four treatments. The best response to two periods before the current period is only predictive in treatment High_Small, but not constantly predictive in the other three treatments. In the two treatments where the equilibrium prediction is the low contribution (Low_Small and High_Large), subjects over-contribute than the best responses (coefficient 1.348 and 1.059). In the two treatments where the equilibrium prediction is the high contribution (Low_Large and High_Small), subjects under-contribute than the best responses (coefficient 0.444 and 0.372). This is consistent with what we see using calculator inputs, that contributions are biased towards the middle.

Table 3.12: Effect of previous four periods hypothetical best responses (Low Complementarity)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contribution	Small Group				Large Group			
Best_Res lag1	1.348*	1.145*	1.120**	1.108**	0.444***	0.405***	0.348***	0.284**
	(0.440)	(0.279)	(0.215)	(0.136)	(0.008)	(0.015)	(0.033)	(0.034)
Best_Res lag2		0.764	0.745*	0.651		0.031	-0.027	-0.004
		(0.366)	(0.241)	(0.285)		(0.019)	(0.024)	(0.035)
Best_Res lag3			0.193	0.152			0.067	0.063
			(0.367)	(0.278)			(0.029)	(0.033)
Best_Res lag4				0.318				0.002
				(0.243)				(0.044)
Constant	2.641*	2.057	1.883	1.666	4.102***	4.229**	4.845***	5.406***
	(0.747)	(0.821)	(0.979)	(1.043)	(0.151)	(0.441)	(0.471)	(0.178)
Observations	1,140	1,080	1,020	960	1,140	1,080	1,020	960
R-squared	0.050	0.068	0.073	0.078	0.176	0.126	0.085	0.048

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.5.2.3 Fictitious Play Model

By fitting our data using the best response model, we learn that the contribution level is biased towards the middle in all treatments. This indicates that subjects learn slower than best responding to the previous period. In other words, subjects' belief about others' contribution may not simply

Table 3.13: Effect of previous four periods hypothetical best responses (High Complementarity)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contribution	Small Group				Large Group			
Best_Res lag1	0.372***	0.260***	0.222***	0.215***	1.059*	0.851**	0.842***	0.716***
	(0.032)	(0.011)	(0.015)	(0.019)	(0.251)	(0.133)	(0.062)	(0.016)
Best_Res lag2		0.159***	0.109***	0.100***		0.403	0.268*	0.340**
		(0.010)	(0.010)	(0.009)		(0.160)	(0.073)	(0.063)
Best_Res lag3			0.118***	0.095**			0.299*	0.178
			(0.005)	(0.011)			(0.095)	(0.081)
Best_Res lag4				0.052**				0.265***
				(0.010)				(0.025)
Constant	4.841***	4.375***	4.086***	3.946***	1.741	1.298	0.924	0.689
	(0.159)	(0.156)	(0.279)	(0.387)	(0.766)	(0.721)	(0.556)	(0.409)
Observations	1,140	1,080	1,020	960	1,140	1,080	1,020	960
R-squared	0.276	0.291	0.301	0.305	0.126	0.138	0.160	0.171

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

be others' contribution in the previous period, rather can be a function of others' contribution in all previous periods.

To explore this possibility, we fit our data into a fictitious play model where a player exponentially discounts her learning experience in previous periods. Suppose players have belief over the distribution of others' complementary sum s_i and the support of their belief is the set of s_i that ever occurred in their experience.

The weight function in period $t+1$ over others' complementarity sum s_i is defined as:

$$w_i^{t+1}(s_i) = \delta w_i^t(s_i) + \begin{cases} 1 & \text{if} \quad s_i = s^t \\ 0 & \text{otherwise} \end{cases}, \tag{3.13}$$

where δ is the discount factor. If the others' complementarity sum s_i is played in period t ($s_i = s^t$), the weight over s_i is $\delta w_i^t(s_i) + 1$; otherwise, the weight over s_i is $\delta w_i^t(s_i)$.

In each period, a player believes that s_i occurs with the probability $\mu_i^t(s_i)$:

$$\mu_i^t(s_i) = \frac{w_i^t(s_i)}{\sum_{s_i \in S} w_i^t(s_i)}, \tag{3.14}$$

where S is the set of s_i that ever occurred in the game.

We assume that a player maximize her expected payoff over her belief $\mu_i^t(s_i)$ in each period. To account for decision error in players' choices, we the logit equilibrium model. Under the logit equilibrium, the strategy with the highest expected payoff is not chosen with certainty, rather higher expected payoffs are chosen more frequently based on the logit distribution:

$$f_i^t(x_i) = \frac{\exp[\lambda \bar{u}_i^t(x_i)]}{\sum_{x_k} \exp[\lambda \bar{u}_i^t(x_k)]}, \tag{3.15}$$

where λ is the noise parameter, and $\bar{u}_i^t(x_i)$ is the expected payoff of player i in period t by choosing strategy x_i.

We estimate the noise parameter λ and the discount parameter δ in the following way: For each session, we did a grid search for λ and δ in the space $[0,10] \times [0,1]$, with increments of 0.1 for both parameters. For each pair of (λ, δ), we take subjects' contribution in the first round and compute all possible combinations of others' complementary sum. We use this empirical distribution of others' complementary sum as subjects' belief distribution for the first period. Starting from the second period, for each player, the previous period's belief distribution is discounted by δ, and updated by realized others' complementary sum using Equation 3.13. For each player, based on the belief distribution, we compute the expected payoff for each strategy. To accommodate decision error, each strategy is chosen based on the logit distribution (Equation 3.15). Therefore, for each player, each period, we have the predicted choice distribution over the discrete strategy space $[1,2,3,4,5,6,7,8,9,10]$. We then compute the difference between the predicted choice distribution with the actual choice using the mean squared error. This gives us one mean squared error for each player in each period. We sum up all the mean squared errors within the same session. The (λ, δ) pair that minimizes the sum of mean squared error in each session is picked as the estimated λ and δ.

Table 3.14 reports the estimated λ and δ in each session. We find that across treatments, the estimated λ varies from 0.2 to 7.2, indicating different levels of decision error. The estimated δ is small in most sessions except session 1 in the High_Large treatment. This indicates that subjects discount information in past periods heavily and mostly rely on information in the most recent period. Nevertheless, because estimated λ and δ vary substantially across different treatments, the fictitious play model does not fit well with the overall behavior of subjects in this experiment.

3.5.3 Comparison Between Different Models

We compare the three models on their prediction accuracy. We first estimate model-specific parameters using data in the first ten periods. We then use the estimated parameters to simulate contribution levels from period ten to period twenty. Lastly, we compare the distributions of predicted contributions and real contributions to see which model has the most accurate prediction.

In the conditional cooperation model, we assume that the current contribution is a linear function of others' average contribution in the previous period. This assumption is based on the analysis in Section 3.5.2.1, which shows that others' contribution in two periods before are not predictive of the current period contribution in all treatments. In particular, we regress contribution (y) on

Table 3.14: Logit equilibrium under the fictitious play model

Complementarity	Group Size	Session	λ (noise)	δ (discount)
Low	Small	1	0.7	0.0
		2	0.2	0.1
		3	3.2	0.3
	Large	1	7.2	0.0
		2	7.2	0.0
		3	7.2	0.0
High	Small	1	0.4	0.3
		2	1.3	0.3
		3	2.5	0.0
	Large	1	0.7	0.6
		2	0.2	0.0
		3	0.4	0.0

others' average contribution in the previous periods (x): $y = ax + b$. To estimate parameter a, and b, we use the first ten periods' data in all treatments. As a result, we get the following estimation:

$$y = 0.897x + 0.798. \tag{3.16}$$

We then use equation 3.16 to predict the contribution in the final period, starting from period eleven.

In the best response model, we assume that players only best respond to others' contribution level in the previous period. This assumption is made based on Table 3.12 and 3.13, which show that only the best response to the previous period is predictive of the current period contribution under all treatments. In particular, we regress contribution (y) on the best response in the previous periods (x): $y = ax + b$. To estimate parameter a, and b, we use the first ten periods' data in all treatments. As a result, we get the following estimation:

$$y = 0.485x + 3.574. \tag{3.17}$$

We then use Equation 3.17 to predict the contribution in the final period, starting from period eleven.

The last model is the fictitious play model. Using a similar procedure as described in Section 3.5.2.3, we estimate λ and δ. Instead of estimating a separate (λ, δ) for each session, we estimate a universal (λ, δ) using the first ten periods' data in all treatments. It turns out that $\lambda = 0.7$ and $\delta = 0$, so the data can be best explained by a logit equilibrium model best responding the previous period.

Besides the three models, we have two benchmarks. The first one is the random choice, where players are equally likely to choose any strategy between 1 and 10. The other one is the Nash equilibrium prediction. Suppose players are perfectly rational, then their contribution level in the last period will be the Nash equilibrium prediction.

To compare the predicted contribution with the real contribution, we calculate the cosine similarity between the predicted contribution distribution and the real contribution distribution. The cosine similarity is defined as

$$cos(\theta) = \frac{\mathbf{p} \cdot \hat{\mathbf{p}}}{\|\mathbf{p}\| \|\hat{\mathbf{p}}\|} = \frac{\sum_{i=1}^{10} p_i \hat{p}_i}{\sqrt{\sum_i^{10} p_i^2} \sqrt{\sum_i^{10} \hat{p}_i^2}}, \tag{3.18}$$

where $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_1, \cdots; \hat{p}_{10})$ is the predicted contribution distribution, $\hat{p}_i$ is the predicted probability of contributing i. $\mathbf{p} = (p_1, p_1, \cdots; p_{10})$ is the real contribution distribution; p_i is the real probability of contributing i. The cosine similarity $cos(\theta) = 1$ when $\hat{\mathbf{p}} = \mathbf{p}$, and $cos(\theta) = 0$ when $\hat{\mathbf{p}} \perp \mathbf{p}$.

Table 3.15 compares the the prediction accuracy among different models. Among the three models, the conditional cooperation model strictly dominate the best response model in all treatments. Between the conditional cooperation model and the fictitious play model, the fictitious player model has superior performance in the Low_Small treatment, but the conditional cooperation model is better in all the other treatments. Compared the two benchmarks, the fictitious play model dominates the random choice model, and is the only model that dominates the random choice model. The Nash equilibrium prediction weakly dominates the best response model. In sum, the conditional cooperation model and the fictitious play model perform better than the best response model. Only the fictitious play model performs better than the random choice model and none of the three models perform better than the Nash equilibrium prediction.

Table 3.15: Cosine Similarity Between Predicted and Real Contribution Distribution

Complementarity (1)	Group Size (2)	Conditional Cooperation (3)	Best Response (4)	Fictitious Play (5)	Random Choice (6)	Nash Equilibrium (7)
Low	Small	0.282	0.031	0.710	0.588	0.899
	Large	0.998	0.990	0.417	0.376	0.990
High	Small	0.962	0.923	0.910	0.374	0.987
	Large	0.825	0.183	0.797	0.693	0.475

3.6 Conclusion

We study the optimal group size in coordination. To account for complementarity in group efforts, we use the CES function as the production function in a voluntary contribution mechanism. We find that the optimal group size depends on the complementarity of group efforts. When efforts are highly complementary, players will coordinate on the high contribution equilibrium if the group size is smaller than an upper bound. Whereas when efforts are less complementary, players will coordinate on the high contribution equilibrium if the group size is larger than a lower bound. We conduct a lab experiment to test these predictions. In the low complementarity condition, the group of size 10 performs significantly better than the group of size 4 in providing the public goods. Whereas in the high complementarity condition, the group size of 4 performs significantly better than the group of size 10 in providing the public goods. We also find that contribution levels are biased towards the middle in all treatments. Because subjects' beliefs are largely unbiased, this bias in contribution level cannot be explained by subjects best responding to their belief, nor can it be explained by other-regarding preferences, because other-regarding preferences will lead to an upward bias in all treatments. We fit our experimental data into three behavioral models, the social norm model, the best response model, and the fictitious play model. None of them provides an exceptional fit. Our findings have clear policy implications in team formation. For teamwork that requires high complementary efforts, smaller teams would perform better than large teams, whereas, for teamwork that requires low complementary efforts, larger teams would perform better than smaller teams.

CHAPTER 4

Adding Private Information Into Centralized Algorithms: A Field Experiment at a Ride-Sharing Platform

4.1 Introduction

The rise of the gig economy has been providing many new working opportunities. Being a gig worker has been increasingly popular among people who work part-time, and those who are between jobs. Services through online intermediaries accounted for 0.5% of all workers in 2015 (Katz and Krueger, 2019), and grew at a 10-fold speed (Farrell and Greig, 2016) from 2012 to 2015. The Pew Research Center estimated that about 24% of American adults earned money from the gig economy in 2016 (Smith, 2016). These gig economy platforms include labor platforms such as Uber, Lyft, TaskRabbit, Upwork, etc., as well as capital platforms such as eBay and Airbnb.

The gig economy provides more flexible working hours compared with traditional workplaces. By turning on and off the app, gig workers decide when to work and how long to work. In fact, 87% of Uber drivers in a 2014 survey listed "to be my own boss and set my own schedule" as their major reason to partner with Uber (Hall and Krueger, 2018). However, gig workers have no control over the tasks they are assigned to. Platforms assign workers tasks through a centralized algorithm, and punish declining tasks by lowering worker's scores. As noted in Ravenelle (2019):

"Concerned that she would slip below the 85 percent acceptance rate, Sarah felt a lot of pressure to accept any work she was offered. "I had no control over what I would be getting and when. So I just took pretty much everything that I could," she said . . . " (Ravenelle, 2019, p. 2)

Similar constraints happen on ride-sharing platforms. A centralized algorithm matches drivers with nearby passengers to minimize the sum of pick-up distances. Drivers have no control over where they are going, and how long it will take. Platforms punish declining rides by lowering drivers' priority in future matches. Drivers could end up in an unfamiliar area far away from home, or get stuck in bad traffic for hours. As a result, there exists misalignment between drivers' working location preferences and rides.

Because driver working location preferences are private information, in order to incorporate this information into centralized matching, the platform needs to first elicit driver preferences. One way to elicit location preference is to let drivers choose a destination they would like to go. This has been implemented in major ride-sharing platforms when drivers need to go home. However, setting a specific destination greatly reduces the number of qualified rides, thus increasing drivers' waiting time to be matched, or even prevents drivers from being matched. When many drivers set destinations at the same time, passengers' waiting time will also increase, because the number of unconstrained drivers decreases. To prevent drastic increase in drivers and passengers' waiting time, platforms restrict the number of times a driver can set a destination in a day. Uber and Lyft drivers can only set destinations twice a day, and DiDi drivers can only set destinations three times a day. Consequently, rides in the middle of the day might still be misaligned with driver location preferences.

Another way to elicit driver location preference is to let drivers choose a region instead of a location. Because choosing a region as the destination is less restrictive than choosing a location, it is possible that this compromise between having no information and having overly precise information could lead to driver welfare gain while not reducing matching efficiency significantly.

To understand the effect of incorporating a coarse location preference on driver welfare and matching efficiency, I utilize a randomized field experiment on the largest ride-sharing platform in China, DiDi. In this experiment, the platform first divide a city into ten regions based on neighborhoods and administrative districts. The platform then randomly assigns 14% of drivers in the treatment group to receive a flexibility tool. Drivers in the treatment group are allowed to specify one or several regions as their destination regions, and the platform will only assign them rides with destinations within the pre-specified regions. The control group drivers are not offered this tool during the experimental period. After the experiment, the platform enabled this tool for every driver.

I find that treatment drivers increase working hours and income by approximately 4%, while maintaining productivity, measured by hourly earnings. I further examine the usage pattern. I find that only 14.565% of destination region trips are end-of-session trips, while the rest are mid-session trips. This shows that driver mainly use this feature to exert location preferences in the middle of the day rather than to go home at the end of the day. For mid-session destination region trips, the majority of them are consecutive trips within a certain area (81.321%).

This shows that drivers have strong location preferences either due to traffic conditions, familiarity, or gas costs. This strategy, though, lowers drivers' hourly income by 2.412%.

Our results have broad policy implications to the gig economy and traditional workplaces. Many other gig platforms also use a centralized algorithm to assign gig workers tasks. Our results show that incorporating gig workers' preferences into the matching can increase both gig workers' welfare and the platform's revenue. Traditional workplaces, such as the retailing and service industry, often use a shift work schedule. Incorporating workers' preferences in assigning shifts can also leads to workers' welfare improvement without sacrificing the revenue.

Our paper is organized as follows. Section 2 presents a literature review. Section 3 provides background information about the platform and the experimental design. Section 4 analyzes the treatment effects and usage patterns. Section 5 concludes.

4.2 Literature Review

The study of labor supply and demand has been at the center of economic studies. The critical assumption of labor supply theory is that a representative agent compares her rate of substitution between income and leisure to the market wage, and choose to work the amount of time that equates her rate of substitution to the market wage. This framework assumes that an individual is free to adjust her working time continuously, which is problematic because most of the jobs have a pre-specified working contract designed by employers, such as forty hours a week, and employees have little space of adjustment. Several explanations can be given as to why employers might prefer a rigid working schedule, as discussed in Gielen (2008). These reasons include fixed employment and production cost, coordination cost, long-term contract to smooth labor supply fluctuation, imperfect competition, and costly labor mobility. Earlier work examined the discontinuity of labor supply includes King (1978); Altonji and Paxson (1988); Presser (1995).

Theoretically, Rebitzer and Taylor (1995) proved that adverse selection problems might lead to a less than optimal equilibrium where the labor market fails to provide enough short-hour jobs. Empirical evidence shows that the misalignment of working arrangements and employees' preferences exists universally in different countries and industries among different population groups. Using the British Household Panel Survey, Stewart and Swaffield (1997) showed approximate 40% of male workers prefer to work less, and on average, the preferred working hours are 4.3 less than actual hours per week; Böheim and Taylor (2004) showed that hour constraints are highly predictive of job mobility and early retirement; Gielen (2008) showed that over employed senior women choose to retire early, hence adopting a more flexible working hours may increase labor supply among older workers. Using the Dutch Socio-Economic Panel, Euwals (2001) showed that flexibility of working hours within jobs is low, about 34% of the desired change for workers who want to work more, and 21% of the desired change for workers who want to work less. Using panel data from the Household, Income and Labour Dynamics in Australia, Drago et al. (2009) showed that there are mismatches between employees' desired working hours and actual working hours for employees who go motherhood, widowhood of men, and job loss. Using The Panel Study of

Income Dynamics data in the US, several researchers have confirmed that hours constraints also exist in the US (Senesky, 2005; Bryan, 2007; Altonji and Usui, 2007).

The discrepancy between actual and preferred work arrangements is not only of interest of economists but also sociologists and psychologists. Research in sociology showed that mismatch between preferred and actual working hours generally exists in the four countries in the study: Japanese, Swedish, Ist German, and the US, however, the form of mismatch is different, in that the US has the largest worker population who what to work more, and Swedish has the largest worker population who want to work less (Reynolds, 2004). Research in psychology showed that allowing employees to adopt flexible working arrangements increases job satisfaction, and mitigate work-life conflict (Hill et al., 2001, 2010).

The traditional workplace constraints fall apart in the gig economy, as workers can work whenever they want and as much or as little as they want. A survey conducted by Benenson Strategy Group shows that the majority of Uber drivers are drawn to the platform because of flexibility[1] (Hall et al., 2017). Empirical analysis using Uber data validates a highly flexible labor supply on the platform. Among drivers who drive at least 16 hours during a 36-week study, correlations between labor supply of contiguous weeks are generally low, and there is only a roughly 47% of probability that a driver works in a particular hour block will continue work in the same hour block the following week (Chen et al., 2019). Based on this highly elastic nature of the labor supply on Uber, the authors further estimated Uber drivers' surplus from providing ride-sharing service compared with less flexible arrangements. They first estimate a mean reservation wage for each driver in a certain hour block during a week, then compare the mean reservation wage with the real average wage. If the real average wage is higher than the mean reservation wage, but the driver is not working, it indicates that the driver has a positive shock to her reservation wage. Under a fully flexible schedule, a driver can avoid working in hours when there are positive shocks in her reservation wage. Compared with an arrangement where drivers are not able to make hourly working time adjustment, but only daily adjustment, the drivers' surplus reduce by approximately 68%. Their work demonstrated that the value of flexibility is huge in the ride-sharing economy. Nevertheless, their identification strategy requires various assumptions on drivers' behavior, such as drivers have a relatively stable mean reservation wage across weeks, and drivers have the knowledge of the real average wage at any given time in the city. These assumptions seem to be strong, and if they do not hold, the estimation could be biased.

The centralized matching system, flexible working arrangement, and dynamic pricing scheme make the ride-sharing platform an ideal market to study many economic research questions that

[1]More than 85% of the responses point out the flexible schedule as one of the main reasons to be an Uber driver. Flexibility is even more valued for female drivers. 42% of female drivers explicitly say that they work for Uber because they can only work for part-time or flexible jobs, whereas the number is 29% for men.

cannot be answered in the traditional marketplace. Utilizing the centralized matching system and dynamic pricing scheme, Castillo et al. (2017) reveals a mechanism where surge pricing would reduce inefficiency caused by drivers picking up distant passengers. Using data from a unique decentralized ride-sharing market in China, Liu et al. (2019) showed that centralized matching could improve match quality and number of matches compared with decentralized matching. Utilizing the dynamic pricing scheme, research on labor supply has found that surge pricing leads to longer working sessions, and that Uber drivers' labor supply is highly elastic (Chen and Sheldon, 2016; Hall et al., 2017). Utilizing the detailed records of working time and location, Cook et al. (2018) identifies a 7% of gender earning gap and decompose the gender earning gap into driver experience, location preference, and driving speed preference. Utilizing the flexible working arrangement, Ai et al. (2019b) shows that team formation and contest could increase drivers' engagement and labor supply. Ride-sharing platforms have also been used to study racial and gender discrimination (Ge et al., 2016), the moral hazard problem of taxi drivers (Liu et al., 2018), and effective ways of repeated apologies (Halperin et al., 2019).

Our paper is also broadly related to the literature on wage elasticity among New York City taxi drivers. Camerer et al. (1997) found a negative wage elasticity and proposed that drivers have a daily income target. Later research on richer data sets disputes the daily income target model in favor of either a daily hour target model (Farber, 2005), or a double-target model of daily income-hour target (Crawford and Meng, 2011). Using a comprehensive data set (all trips taken during five years), Farber (2015) decomposes wage variations into unanticipated and anticipated variations, and shows little evidence of reference-dependence preference. A recent investigation using Uber data finds little support for reference-dependence preference and negative wage elasticity (Chen and Sheldon, 2016).

However, none of the above research studies the effect of increasing flexibility on labor supply and outcome, except for Hill et al. (2010), where they showed through a survey that employees at IBM work one or two days more per week when they have workplace flexibility. Our work is the first one to identify the causal effect of flexibility on labor supply and outcome through an experiment.

4.3 Field Experiment

4.3.1 Setting

DiDi is the largest ride sharing company in China with more than 450 million users and 21 million active drivers who earned income from DiDi in 2016[2]. It operates in over 400 cities in China and

[2]These statistics are from *DiDi Chuxing Corporate Citizenship Report 2017*

in major cities in Australia, Brazil, Japan and Mexico.

Unlike popular ride-sharing platforms in the US, such as Uber and Lyft, DiDi does not use a dynamic pricing mechanism due to low passenger satisfaction. Instead, it divides each day into different hour blocks, such as the morning peak hours, the evening peak hours, late-night hours, and non-rush hours, and have a fix per-mile and per-minute rate for each hour block. This fixed pricing mechanism significantly eliminates the price uncertainty for both the drivers and passengers.

DiDi matches drivers and passengers using a centralized dispatching mechanism[3]. The matching is centralized at the city level. Every few seconds, the server runs a bipartite matching between available drivers and riders, and finds a match which minimizes the weighted total pick up distance subject to constraints. To avoid wild goose chase, the platform only matches drivers and riders within a certain distance.

Because of the use of the dispatching mechanism, drivers can end up in any area of the city arbitrarily, which can be far away from their homes. To make it easier for drivers to go home at the end of the day, DiDi allows drivers to set specific destinations three times a day, in which the platform will assign drivers trips within three kilometers of their pre-specified destinations. This feature is called "destination location" feature. Although this feature provides drivers with more flexibility, it also causes inefficiency. Naturally, a driver can only be assigned a trip to her pre-specified destination if there is a passenger who wants to go to that location. Often, a driver needs to wait much longer to get a trip that goes to a certain destination location, or even fails be matched with a trip at all.

To increase the number of matches and reduce waiting time, the platform introduces a feature which allows drivers express a coarse location preference. Instead of entering a specific destination location, drivers select one or several regions as their destination. Once destination regions are selected, the platform will only assign drivers' rides to destinations that are within the selected destination regions. This feature is called "destination region" feature.

Destination regions are defined by the platform mainly based on natural administrative districts and neighborhoods. Each city usually is divided into six to ten destination regions. Unlike setting a location as a destination, which a driver can only use up to three times in a day, a driver can set destination regions as many times as they want.

As part of a product evaluation, the platform did a randomized experiment to test the effect of offering the destination region feature in the city Dalian, between the period 2018-05-25 and 2018-07-16, in total 53 days. Dalian is a coastal city in Liaoning Province of northeastern China, with a population of over 5.96 million. The GDP per capita of Dalian is 97,470 Yuan, approximately

[3]We are aware of the existence of other matching mechanism on DiDi, but the dispatching mechanism accounts for most of the driver-passenger matches, and is the default option for DiDi Express, which is the service line that I are focusing on.

13,846 US dollars, which is 1.65 times the national average. The monthly average wage for urban non-private sector employees is 6,147 Yuan (873 US Dollars)[4]. Dalian is one of the major cities that DiDi operates

4.3.2 Definitions

In this section, I define some concepts that will be useful in the following analysis.

- Working hours: I define drivers' working hours as the amount of time they are active on the app. This includes the waiting time before they receive a ride, the time they pick up passengers, the time they drive passengers to their destinations. Our definition is consistent with other papers that study labor supply on ride-sharing platforms, such as Cook et al. (2018); Chen et al. (2019).

- Work session: I define a set of trips as a work session if the starting trip is at least one hour after the previous trip, and the ending trip is at least one hour ahead of the next trip. This is because waiting for more than one hour without being assigned a ride is extremely rare on this platform, given the high demand. A driver might have several work sessions in one day.

4.3.3 Experimental Design

To test the effect of offering drivers the destination region feature, the platform did a randomized experiment Dalian. The randomization is done in the following procedure. The platform first extracted all drivers who opened their app during the time period 2018-04-23 to 2018-05-22 in this city. Then, they randomized about 14% of drivers into the treatment group, while the rest are in the control group. Drivers in the treatment group are provided with the destination region feature, while drivers in the control group are not. The only difference in the application for drivers in the treatment and the control group was slight difference in user interface. For drivers in the treatment group, they are able to see "set destination regions" on their setting page. Once they click on it, they will see the region names together with a map, which shows where a region is. By clicking on the region names, they can select or unselect destination regions. For drivers in the control group, this function did not exist on their setting page. Treatment group drivers are only informed about this function by in-app pushes, while there are no in-app pushes for the control group drivers. Because drivers are not contacted by text messages or emails, there was no contacting effect. If a driver did not turn on her app during the experimental period, she would not be aware of the experiment. Both treatment and control groups drivers still have access to the destination location feature.

[4]Source: Dalian Statistical Yearbook, 2017

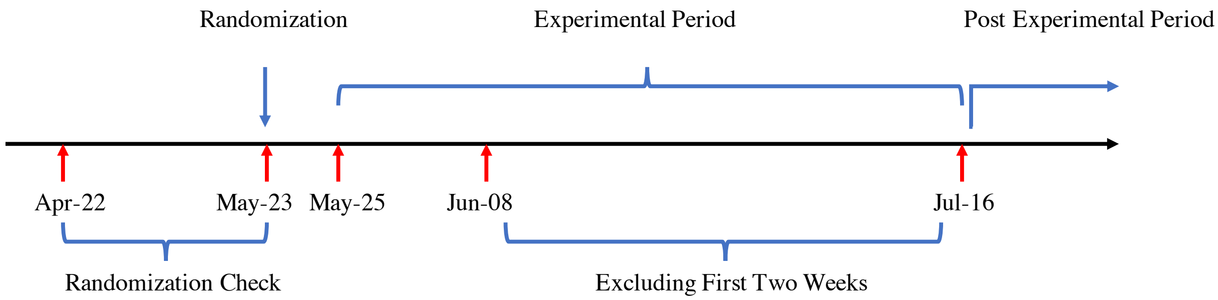


Figure 4.1: Timeline of the experiment

The experiment started at 2018-05-25 and ended on 2018-07-16. In total, the experiment lasted for 53 days. Drivers in the control group began to receive the destination region function on 2018-07-17. A graphic representation of the experimental timeline is in Figure 4.1.

4.4 Analysis

4.4.1 Summary Statistics and Randomization Check

We first do a randomization test to see if driver characteristics in the treatment and control groups are balanced.

The randomization test is conducted on the period between 2018-04-23 and 2018-05-22 on three samples. The first sample is the whole population of drivers who turned on their apps between 2018-04-23 to 2018-05-22. I call this sample the "full sample". The second sample is the subgroup of drivers who turned on their apps during the experimental period 2018-05-25 and 2018-07-16. I call this the "online sample". The third sample is the subgroup of drivers who completed at least one trip during the experimental period. I call this sample, the "trip-completed sample". The relationship between the three samples are shown in Figure 4.2. Because of the high attrition rate, only 78.220% drivers in the full sample are in the online sample, and 72.517% drivers are in the trip-completed sample.

We further explore whether treatment and control group drivers differ in attrition. It turns out the proportion of treatment and control drivers appear in the online sample, and the trip-completed sample is almost identical to the proportion in the full sample. The exact numbers on attrition are in Table 4.1. The treatment group and the control group are not significantly different in the proportion of drivers online during the experimental period (p-value = 0.642). I further look at whether they differ in the proportion of drivers who complete at least one trip during the experimental period and find no difference (p-value= 0.967). There is still a chance that the treatment drivers in the trip-completed sample are not aware of this new function, because the in-app pushes may not get

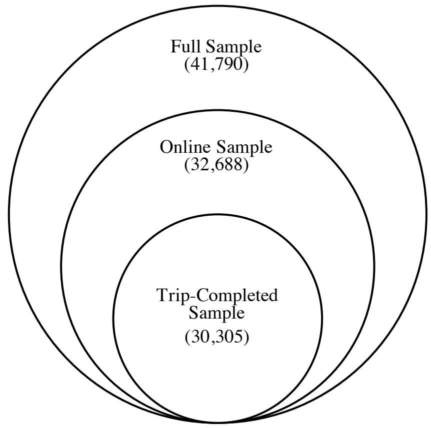


Figure 4.2: Relationship of the three samples

their attention, therefore, even for the treatment effect estimated using the trip-completed sample, the effect size is still diluted.

Table 4.1: Experimental Sample Size

	Full Sample	Online Sample	Online/Full	Trip-completed Sample	Trip-completed/Full
Treatment	5,874	4,581	0.780	4,258	0.725
Control	35,916	28,107	0.783	26,044	0.725
Total	41,790	32,688	0.782	30,302	0.725

We compare drivers' daily working hours, daily income, daily number of trips completed, gender, experience, hometown location, as well as whether they are classified as full-time drivers by DiDi. Drivers' experience is measured by the total number of trips they have completed on the platform prior to the experiment. I consider a driver as a local driver if her self-reported hometown is in Liaoning province. Table 4.2 shows the detailed statistics for the randomization checks. We do not find any statistically significant difference for any features in all three samples except for gender and experience. The treatment group has slightly higher proportion of female drivers and more experienced drivers. The difference is only significant under t-tests at the 10% significant level.

A summary of statistics is in Table 4.2

4.4.2 Take-up Rate

Before analyzing the treatment effect, I first check the take-up rate. Among treatment drivers in the trip-completed sample, 26.144% of them completed at least one destination region trip. The

Table 4.2: Randomization checks

Daily	Working Hours	Income (Yuan)	Trips completed	Male	Local	Full-Time	Age	Trips in history
			Full Sample					
Treatment Average	2.468	61.148	5.124	0.843	0.387	0.276	38.044	2139.738
Control Average	2.431	60.248	5.058	0.853	0.383	0.268	38.004	2060.457
T-Test	0.381	0.428	0.484	0.034	0.514	0.196	0.783	0.090
KS-Test	0.320	0.161	0.361	0.619	1.000	0.898	0.771	0.553
			Online Sample					
Daily	Working Hours	Income (Yuan)	Trips completed	Male	Local	Full-Time	Age	Trips in history
Treatment Average	2.974	73.996	6.189	0.851	0.402	0.349	38.540	2380.873
Control Average	2.923	72.803	6.104	0.862	0.398	0.338	38.435	2284.511
T-Test	0.316	0.381	0.447	0.059	0.631	0.120	0.518	0.070
KS-Test	0.168	0.115	0.196	0.784	1.000	0.649	0.777	0.248
			Order-Completed Sample					
Daily	Working Hours	Income (Yuan)	Trips completed	Male	Local	Full-Time	Age	Trips in history
Treatment Average	3.161	78.869	6.596	0.851	0.399	0.375	38.678	2514.525
Control Average	3.116	77.785	6.522	0.862	0.398	0.364	38.542	2416.806
T-Test	0.384	0.449	0.532	0.057	0.824	0.154	0.421	0.081
KS-Test	0.195	0.131	0.181	0.776	1.000	0.734	0.670	0.324

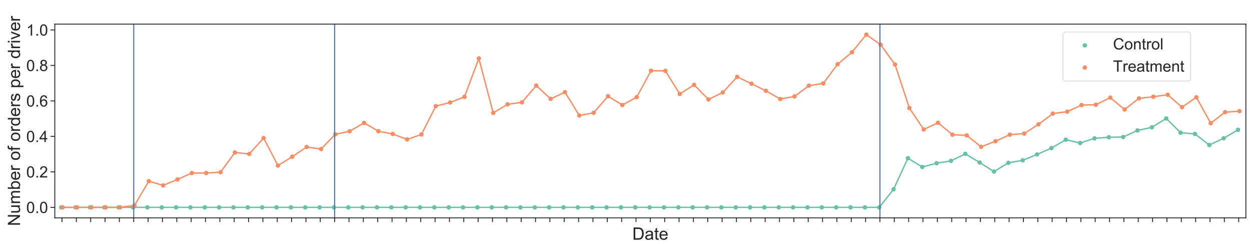


Figure 4.3: Daily average destination region trips per driver

take-up rate was low at the beginning and gradually Int up during the experimental period. For the first week of the experimental period, treatment drivers in the trip-completed sample completed 0.144 destination region trips per day, while the number Int up to 0.796 in the last week of the experimental period. The usage of the destination region feature was unevenly distributed among treatment drivers, in the way that some drivers used this feature intensively, while others nearly used it. Among these 26.144% of treatment drivers who used this function, 50% of them completed less than nine destination region trips, while the top 10% of drivers each completed more than 144 destination region trips.

4.4.3 Average Treatment Effect

We evaluate our main treatment effect on all three samples.

We look at two different periods because treatment drivers did not pick up the intervention immediately, as I have shown in figure 4.3. The interpretation of the treatment effects below is based on the online sample two weeks after the start of the experiment. On average, the treatment drivers increase their daily working hours by 0.116 hours, which is 4.460% of their average daily working time. Their daily income increases by 2.965 Yuan, which is 4.116% of their daily income. The increase in revenue highly resembles the increase in income, because the platform takes approximately 20% commission from the revenue generated by drivers. The treatment drivers take, on average, 0.242 more trips per day, which is 4.197% of their daily trips on average. The detailed statistics are presented in Table 4.3. The upper left panel presents the average treatment effect on the full sample for the whole experimental period. The upper right panel presents the results for the full sample two weeks after the start of the experiment. The middle two panels present the results for the online sample, and the bottom two panels present the results for the trip-completed sample. As I can see, the average treatment effects do not differ much for the four specifications, but the standard deviation in the full sample whole period is too large to detect statistical significance. I use the t-test to test for difference in mean, and KS-test to test for difference in distribution.

We also examined whether the treatment has any effect on drivers' productivity, and do not find any difference in drivers' hourly income in any of the four specifications. This indicates that, on average, the treatment does not make drivers more productive.

To get an intuitive idea of the treatment effects, I plotted drivers' daily average working hours, income, and trips completed by treatments. Figure 4.4 depicts the trend of the daily treatment effects, with the error bar representing the 95% confidence interval. As I can see from the graphs, the treatment and control drivers have almost identical averages before the start of the experiment. After the beginning of the experiment, as treatment drivers start to use the destination region feature, the difference gradually appears. After the experiment, the control drivers begin to catch up a

Table 4.3: Treatment Effects

	Full Sample					
Period	Whole Experimental Period			Excluding First Two weeks		
Outcome (Daily)	Working Hours	Income	Trips completed	Working Hours	Income	Trips completed
Control	2.084	54.867	4.586	2.036	54.047	4.512
Treatment	2.151	56.666	4.734	2.119	56.191	4.687
T-Test	0.110	0.129	0.133	0.050	0.075	0.081
KS-Test	0.307	0.162	0.158	0.221	0.129	0.110

	Online Sample					
Period	Whole Experimental Period			Excluding First Two weeks		
Outcome (Daily)	Working Hours	Income	Trips completed	Working Hours	Income	Trips completed
Control	2.663	70.111	5.86	2.601	69.064	5.766
Treatment	2.758	72.645	6.069	2.717	72.029	6.008
T-Test	0.058	0.075	0.076	0.024	0.041	0.044
KS-Test	0.126	0.054	0.049	0.069	0.041	0.026

	Complete Sample					
Period	Whole Experimental Period			Excluding First Two weeks		
Outcome (Daily)	Working Hours	Income	Trips completed	Working Hours	Income	Trips completed
Control	2.871	75.665	6.324	2.805	74.534	6.223
Treatment	2.964	78.105	6.525	2.92	77.442	6.46
T-Test	0.151	0.074	0.062	0.105	0.052	0.043
KS-Test	0.077	0.102	0.105	0.032	0.057	0.061

bit, but the difference still exists because of the slow adaption of the new feature.

The general declining trend in Figure 4.4 is due to driver attrition. I are gradually losing drivers who have previously been assigned to our experiment.

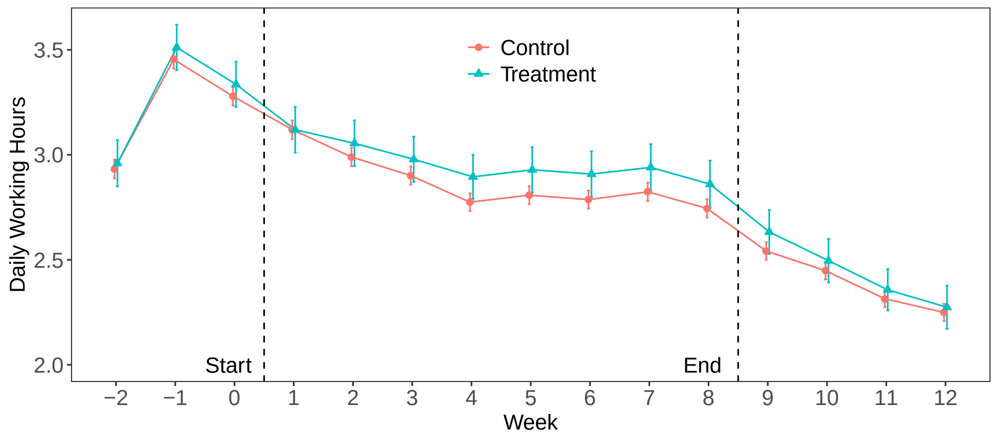
4.4.4 Heterogeneous Treatment Effects

To see how drivers' characteristics affect the treatment effects, I run panel regressions with day of the week fixed effect. I choose to run a fixed effect model on the day of the week because I see a clear day of the week pattern in the daily treatment effect in Figure 4.4. I include two controlling variable: drivers working hours in the past 30 days (Before Working Hours), and trip completion rate in the past 30 days (Before Completion Rate). The characteristics that I are interested in are gender (Male), whether a driver is a local driver in the same province (Local), and whether a driver is a part-time driver (Local). Because I only have data on these three characteristics on half of the drivers, I lose half of our observations when including these variables. Between the drivers with or without data on these three variables, there is no systematic difference in the features that I can observe, such as working hours, income, trip completion during the experimental period. As seen in Table 4.4, columns (1), (3), and (5), Before Working Hours and Before Completion Rate, are significantly correlated with longer working hours, higher income and more completed rides in the experimental period. After controlling Before Working Hours and Before Completion Rate, the treatment effect is still there. The effect of treatment on working hours (0.081 hours), income (2.085 yuan), and trips completed (0.165 trips) are slightly smaller then in our non-parametric analysis in Table 4.3, but the effect sizes are still comparable. Most of the treatment effects are driven by full-time drivers, as part-time treatment drivers work 0.247 hours less, earn 3.071 Yuan less compared with full-time treatment drivers. However, when I break the driver population into part-time and full-time drivers, the treatment effects are not significant for either group because I don't have enough observations. It is not surprising that part-time drivers are affected by the treatment less because they work less and therefore complete much fewer destination region trips. When it comes to gender and local drivers, I face the same problem of lacking observations in either category. I see positive directional effects in favor of locals (positive Treatment $\times$ Local) and females (negative Treatment $\times$ Male), but I do not have enough data to claim statistical significance.

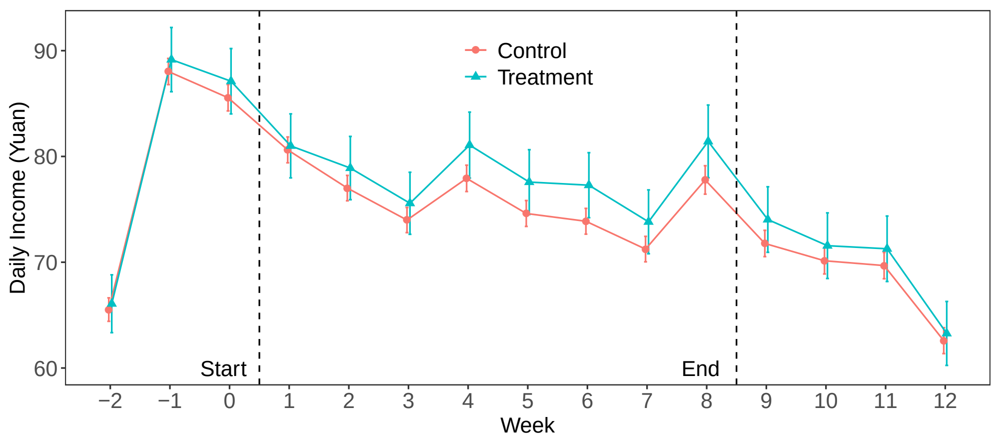
4.4.5 Usage Patterns

In this subsection, I investigate when drivers set destination regions and which regions they select. To look at the timing, I classify destination region trips into two categories: end-of-session trips and mid-session trips. End-of-session trips are more likely to suffer from necessity constraints.

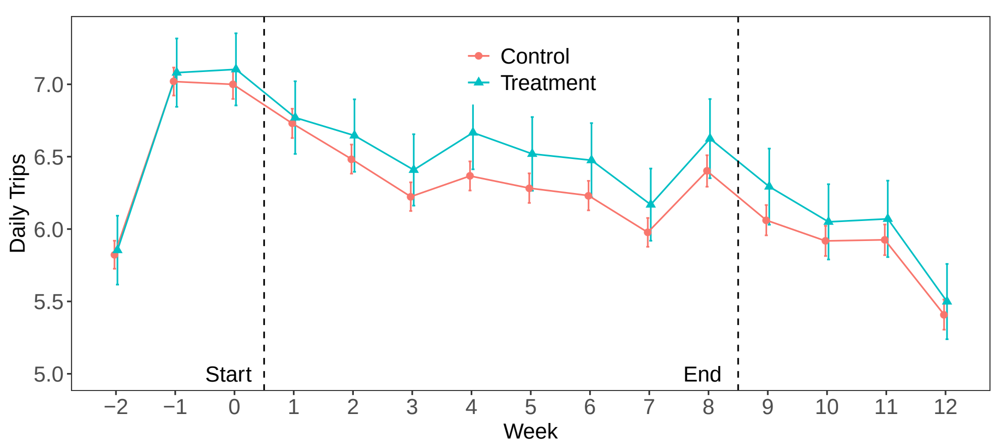
Figure 4.4: Treatment Effects by week on the Trip-completed Sample



(a) The weekly average daily working hours



(b) The weekly average daily income



(c) The weekly average daily completed trips

Table 4.4: Heterogeneous Treatment Effect (Panel Regressions)

VARIABLES	(1) Working Hours	(2) Working Hours	(3) Income	(4) Income	(5) Trips Completed	(6) Trips Completed
Treatment	0.081**	0.291	2.085*	2.637	0.165*	0.460
	(0.037)	(0.227)	(1.067)	(6.730)	(0.089)	(0.559)
Before Working Hours	0.024***	0.023***	0.659***	0.577***	0.054***	0.047***
	(0.000)	(0.000)	(0.005)	(0.012)	(0.000)	(0.001)
Before Completion Rate	0.703***	0.520***	17.154***	12.155***	3.767***	3.169***
	(0.079)	(0.091)	(2.132)	(2.325)	(0.200)	(0.226)
Part-time		-0.554***		-30.562***		-2.912***
		(0.070)		(2.292)		(0.190)
Treatment × Part-time		-0.246**		-3.071		-0.240
		(0.104)		(3.159)		(0.266)
Local		-0.104**		-5.335***		-0.053
		(0.043)		(1.304)		(0.106)
Treatment × Local		0.143		3.122		0.248
		(0.110)		(3.238)		(0.266)
Male		0.092		2.812		0.143
		(0.087)		(2.410)		(0.207)
Treatment × Male		-0.123		-0.410		-0.290
		(0.197)		(5.821)		(0.482)
Constant	-0.203***	0.435***	-6.666***	27.863***	-2.597***	0.596*
	(0.068)	(0.137)	(1.847)	(3.922)	(0.175)	(0.348)
Observations	1,159,665	614,211	1,159,665	614,211	1,159,665	614,211
R-squared	0.334	0.380	0.296	0.338	0.311	0.355

Robust standard errors clustered on driver level
Sample: Trip-completed sample two weeks after
*** p<0.01, ** p<0.05, * p<0.1

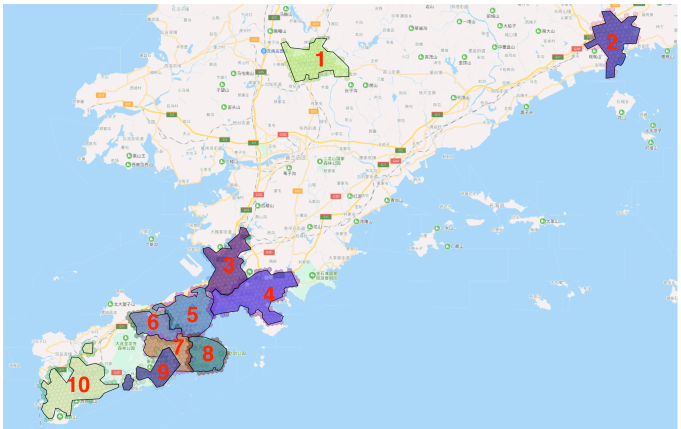


Figure 4.5: Region id

Region id	Expected hourly earnings (Yuan)
8	32.808
9	31.378
6	30.830
7	30.130
5	29.664
4	27.315
3	25.469
10	22.312
1	20.741
2	17.568

Table 4.5: Expected Hourly Earnings by Regions

These trips could be trips to go home, trips to go to an area to pick up children, or trips to have meals. Therefore, allowing drivers to set destination regions may enable them to work at the time when they previously wouldn't be able to. Mid-session trips are less likely to suffer from necessity constraints compared with end-of-session trips, as drivers continue working after finishing these trips. Therefore, these trips exhibit location preferences or strategic concerns. For example, drivers may prefer to drive in a familiar region because they can easily find routes, parking or food. Drivers may also prefer to drive in less congested areas. or they may strategically choose more lucrative regions to increase their earnings.

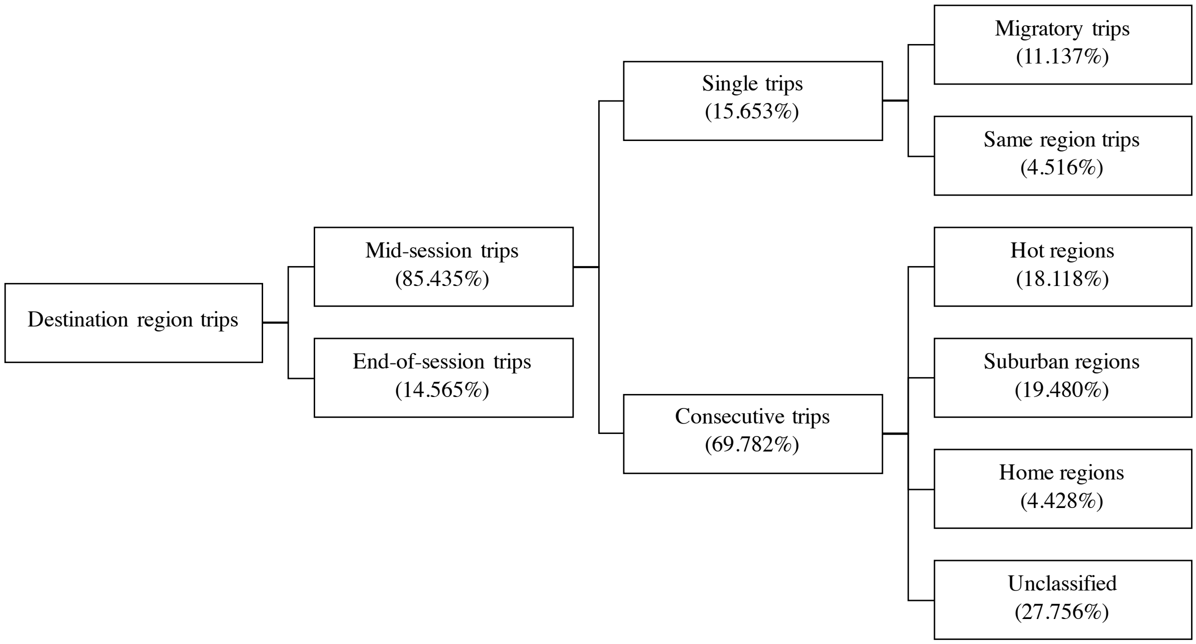
Only 14.565% of destination region trips are classified end-of-session trips, while the majority of destination trips are classified as mid-session trips. The proportion of end-of-session trips stay stable as the take-up rate goes up later in the experimental period, indicating that the way drivers use the destination region feature does not change over time.

We then look at patterns among mid-session trips. There are two main patterns. One is selecting one or several destination regions and complete consecutive trips within the same area. The other is completing only one trip under a destination region setting. The first pattern shows a strong location preference, whereas the second pattern demonstrates working location migration within a city.

Among all mid-session trips, 81.679% of them are consecutive trips within an area, while the rest are stand-alone trips. Among the 18.321% stand-alone trips, 71.144% of them are migrating to a different part of the city, while the rest are staying within the same region as the previous trip. This shows that the primary usage of the destination region feature is to avoid going to certain areas, rather than migrating across different areas.

Among all mid-session consecutive trips, only 6.345% contains their home region as one of the destination regions, showing that familiarity is not a major preference for DiDi driver in Dalian. I

Figure 4.6: Usage Patterns Break Down

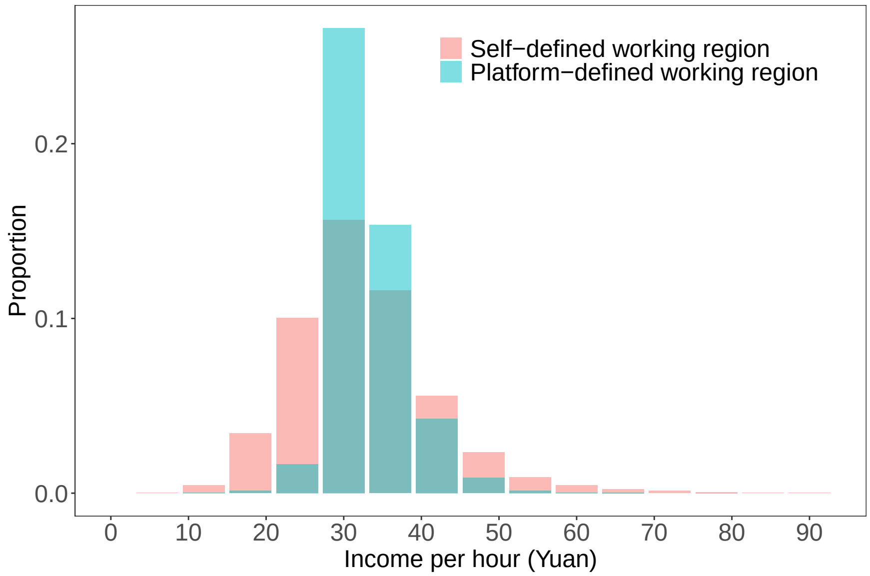


then look at whether drivers prefer to go to lucrative regions (hot regions). In order to do this, I compute the expected hourly earnings in each region using the data one month before the experiment. As shown in Figure 4.5, region 8, 9, 6, 7 and 5 are the central five regions. Their expected earnings are significantly higher (p-value ¡0.01) than region 4 and 3 (Table 4.5), which are suburban areas. The expected earnings in region 4 and 3 are significantly higher (p-value ¡0.01) than distance regions 10, 1 and 2 (Table 4.5). Surprisingly, only a small proportion (25.964%) of settings have one or more regions among the central five regions. A slighter higher proportion of settings (27.915%) are region 4 and 3, while the rest (39.775%) cannot be classified into any categories. This shows that while some drivers try to game the system by going to more lucrative regions, the majority of drivers prefer less congested suburban area or have their own location preferences that cannot be easily classified. The detailed usage pattern is in Figure 4.6.

I show how consecutive destination region trips affects drivers' income. I compose a comparison group in the following way: for each driver who self-defines a working region, I find a comparable group of drivers who start at the same hour and the same location but do not set destination regions. The group of drivers also have to complete as least as many trips as the driver with self-defined working regions. I then compare the hourly income of the driver with the self-defined working region with the average hourly income of the group of matched drivers for the same number of trips they complete.

In total, I have 8662 groups of consecutive destination region trips completed by 752 drivers during the 53 experimental period. I are able to find a matching group for 8343 groups out of 8662

Figure 4.7: Hourly Earnings Comparison



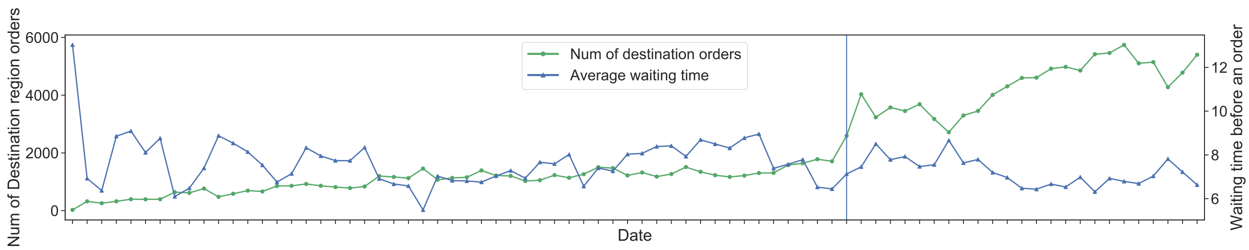
groups. I drop the unmatched 319 groups in the following analysis. For 291 groups of consecutive destination region trips, I find only one match. For all the other groups, I find at least two matches.

I identify a small but significant productivity difference between drivers with self-defined working region and drivers with no control of working regions. On average, drivers with a self-defined working region earn 0.728 Yuan (p-value < 0.01) less per hour than drivers with no control of working regions. The detailed distribution of hourly earnings for the two groups of drivers is presented in Figure 4.7.

4.4.6 Post Experiment

After the experiment, the platform decided to offer this flexibility feature to the whole population of drivers. One concern is whether the positive effect of flexibility on labor supply and income persists once flexibility is offered to everyone. A higher demand for destination region trips may lead to longer waiting times and therefore lower the productivity of drivers. I compare the number of destination trips with the average waiting time of destination trips and find out that this is not the case. The total amount of destination region trips are still too small to affect waiting time given the large volume of demand from the passenger side. The daily trend of the number of destination region trips and the average waiting time is presented in Figure 4.8. As shown in the graph, the number of destination region trips increases sharply after the end of the experiment, but this does not lead to an increase in the waiting time. A t-test between the waiting time 25 days before and after the end of the experiment shows that, on average, the waiting time is 0.502 minutes shorter

Figure 4.8: Number of Destination Region Trips vs. Waiting time



after the end of the experiment compared with before (p-value < 0.01).

4.5 Conclusion

Despite the flexible working schedule that the gig economy provides, many gig workers feel that they are tethered to work because they have no control over the algorithm that assigns them tasks. Utilizing a field experiment at the largest ride-sharing platform in China, I examine the effect of incorporating drivers' private information into centralized matching. In particular, treatment drivers are allowed to self define a working region by specifying one or several destination regions. I find that this autonomy feature increases drivers' labor supply and income by 4.460% and 4.116%, respectively, while not lowering their productivity. A close investigation of drivers' usage patterns shows that the majority of destination region trips are consecutive trips completed within the same area, rather than single migratory trips across the city or end-of-session trips, which suffers from necessity constraints. This indicates that drivers have a strong location preference.

Our findings indicate that giving drivers more autonomy is beneficial for both drivers and the platform without sacrificing efficiency. Even though this is only one piece of evidence from one platform, it has broad implications for other algorithm-based centralized matching marketplaces. Incorporating more personalized matching criteria can lead to higher participation and better match quality.

CHAPTER 5

Conclusion and Discussion

The three chapters solve different market design problems in three distinct markets. In the college admissions market, the goal is to achieve stability and efficiency. Using theory and a lab experiment, I studied a dynamic feedback mechanism emerged from the field. I found that the dynamic mechanism has behavioral advantages over static mechanisms due to its simplicity and straightforwardness. In the artificial market in the lab, the goal is to facilitate cooperation in groups. Using theory and a lab experiment, I found that the contribution level depends on the interplay of group size and complementarity. When the complementarity is high, there is an upper bound for the optimal team size, but when the complementarity is low, there is a lower bound for the optimal team size. In the ride-sharing market, the goal is to provide location flexibility to ride-sharing riders. Using a field experiment, I studied the effect of incorporating workers' location preference in task assignments in the context of the ride-sharing economy. I found that drivers who can self-define their working regions increase their labor supply and income without lowering the matching efficiency.

Even though the three studies have different goals and design challenges, they all show the importance of relaxing the rational agent assumption in neoclassical economics. Limited cognitive ability, other-regarding preferences, and private information all play a huge part in determining the performance of a mechanism.

Although the three studies shed light on human behavior in markets, they also lead to more open questions. For example: What is the optimal ending rule for a dynamic mechanism? A hard ending rule is the simplest to implement, but it can cause congestion. Soft ending rules may be theoretically superior, but are hard to be implemented in practice. Another question is how does one characterize different dynamic mechanisms? Seemingly different dynamic mechanisms may have the same theoretical properties. Lastly, how does one combine dynamic mechanisms with static mechanisms? Adding a dynamic matching period before DA or IA can significantly increase their performance while avoiding the congestion under the dynamic mechanism. In the context of team formation, previous research has found that groups made up of like-minded people become "echo chambers" that lead to group polarization. However, there is mixed evidence on groups made up of people with diverse opinions. What is the optimal mixture of members who hold different opinions? How divergent or similar should opinions be in order to have any meaningful conversation? Does the size of a group affect the convergence of opinions? In the context of the gig economy, what is the optimal time window to evaluate worker performance? A longer period may allow workers to procrastinate in achieving their goals, whereas, in a shorter period, performance may be heavily influenced by uncertainty than reflecting actual productivity. Another question is the timing of the reward, i.e., immediate reward vs. delayed reward. An immediate reward gives workers a piece rate bonus today, whereas a delayed reward gives workers a piece rate bonus tomorrow. If workers are naive about their present-bias, they may overestimate their future productivity and find the delayed reward more attractive. If workers are sophisticated about their present-bias, they may use the delayed reward as a commitment device to overcome self-control problems. These questions will be explored in my future research.

APPENDICES

APPENDIX A

Appendix for Chapter 2

A.1 Proof

A.1.1 Bayesian Nash equilibrium under DA

Here, we prove that any combination of strategies listed in Table 2.2 is indeed a Bayesian Nash equilibrium under the DA mechanism. Furthermore, these are the only Bayesian Nash equilibria. We know that in the college admissions problem, the allocation outcome under the DA mechanism equals the allocation outcome under the serial dictatorship mechanism. In other words, students need to care about only those other students who have a higher score rank than they do. We use this property to prove the strategies listed in Table 2.2 are Bayesian Nash equilibrium strategies and are the only Bayesian Nash equilibrium strategies. There are six different preference type realizations in our experiment; each of them happens with probability $\frac{1}{6}$. The six realizations are listed below:

Case1 The 1st student and 2nd student have preference type I, and the 3rd student and 4th student have preference type II

Case2 The 1st student and 2nd student have preference type II, and the 3rd student and 4th student have preference type I

Case3 The 1st student and 3rd student have preference type I, and the 2nd student and 4th student have preference type II

Case4 The 1st student and 3rd student have preference type II, and the 2nd student and 4th student have preference type I

Case5 The 1st student and 4th student have preference type I, and the 2nd student and 3rd student have preference type II

Case6 The 1st student and 4th student have preference type II, and the 2nd student and 3rd student have preference type I

First, let's look at the high correlation environment. Recall that, in the high correlation environment, preference type I is $A \succ B \succ C \succ D$, while preference type II is $B \succ A \succ D \succ C$.

- For the 1st student, her dominant strategy is to report her first choice truthfully, regardless of her own and others' preferences.

- For the 2nd student, she needs to consider the realization of preference types. When she has the same preference as the 1st student (case1, 2), the best she can do is to get into her second most preferred school. She can obtain this allocation by reporting her second most preferred school ahead of her third and fourth most preferred schools. Since her most preferred school is already occupied by the 1st student, the reported position of her most preferred school does not matter. Hence, her best response in case1,2 is $(2 > 3/4)$. When she has a different preference from the 1st student (case3, 4, 5, 6), she can get into her most preferred school by either playing $(1,2,*,*)$ or $(2,1,*,*)$. When playing the latter strategy, since her second most preferred school has been occupied by the 1st student, she will be kicked out and send to her most preferred school. Therefore, her best response in case3, 4, 5, 6 is $(1/2 > 3/4)$. Since ex-ante, the 2nd student does not know the realization of the preference type, only strategies that are best responses in all six cases will be Bayesian Nash equilibrium strategies. These strategies are $(1/2 > 3/4)$.

- For the 3rd student, regardless of the realization of preference types, in equilibrium, she cannot get into her top two choices. The best she can do is to get into her third-most preferred choice. She can do so by reporting her third-most preferred choice ahead of her fourth most preferred choice $(3 > 4)$

- For the 4th student, any strategy is a best response because she can do nothing but accept what is left.

Together, this proves that any combination of strategies in Table 2.2 in the high correlation environment under DA is a Bayesian Nash equilibrium strategy. These are the only Bayesian Nash equilibrium strategies, since any other strategies are strictly dominated.

Second, let's look at the low correlation environment. The preference type II is $B \succ D \succ C \succ A$

- For the 1st student, her dominant strategy is to report her first choice truthfully, regardless of her own and others' preferences.

- For the 2nd student, she needs to consider the realization of preference types. When she has the same preference as the 1st student (case1, 2), the best she can do is to get into her second most preferred school. She can obtain this allocation by reporting her second most preferred school ahead of her third and fourth most preferred schools. Since her most preferred school will be occupied by the 1st student, the reported position of her most preferred school does not matter. Hence, her best response in case1, 2 is $(2 > 3/4)$. When she has a different preference from the 1st student, and the 1st student has preference type I (case3, 5), she can get into her most preferred school by playing $(1 > 2/3)$. The ranking of her least preferred school does not matter because it will be occupied by the 1st student. When she has a different preference from the 1st student, and the 1st student has preference type II (case4, 6), she can get into her most preferred school by playing $(1 > 3/4)$. The ranking of her second-most preferred school does not matter because it will be occupied by the 1st student. Since ex-ante, the 2nd student does not know the realization of the preference type, only strategies that are best responses in all six cases will be Bayesian Nash equilibrium strategies. These strategies are $(1,2,*,*)$.

- For the 3rd student, in case1, 4, 5, she can get into her second most preferred school by playing $(2 > 3)$. The rankings of her most- and fourth-most preferred school do not matter because they will be occupied by the top two students. In case2, she can get into her most preferred school by playing $(1 > 3)$. In case3,6, she can get into her third-most preferred school by playing $(3 > 4)$. The intersection of the three strategy sets is $(*,*,3,4)$, which is her Bayesian Nash equilibrium strategy.
- For the 4th student, any strategy is a best response because she can do nothing but accept what is left.

Together, this proves that any combination of strategies in Table 2.2 in the low correlation environment under DA is a Bayesian Nash equilibrium strategy. These are the only Bayesian Nash equilibrium strategies, since any other strategies are strictly dominated.

A.1.2 Bayesian Nash equilibrium under Boston

In this subsection, we show that any combination of strategies listed in Table 2.2 is a Bayesian Nash equilibrium under Boston and they are the only Bayesian Nash equilibria. Truth-telling is one Bayesian Nash equilibrium strategy.

First, let's look at the high correlation environment, where a type I player has the preference $A \succ B \succ C \succ D$, and a type II player has the preference $B \succ A \succ D \succ C$. There are six different realizations of preference combination, similar as under DA, with each combination happens with probability $\frac{1}{6}$:

- For the 1st student, she has a dominant strategy, which is to truthfully report her first choice $(1,*,*,*)$.

- Assume the 1st student is rational. For the 2nd student, there is $\frac{1}{3}$ of probability that she has the same preference as the 1st student (case1, 2), and $\frac{2}{3}$ probability that she has a different preference from the 1st student (case3, 4, 5, 6). Ex-ante, she has the incentive to report her first choice truthfully, because, with probability $\frac{2}{3}$, she will get in. Any untruthful first choice will make her strictly worse off. When the 2nd student happens to have the same preference as the 1st student, she will not get into her first choice. The worse case would be that the 3rd student was truthful, and therefore the 3rd student would get into the 2nd student's second most preferred school. In this case, the 2nd player would have justified envy for the 3rd student, but there is nothing the 2nd player could do to prevent this justified envy, because ex-ante truthfully reporting her first choice dominates all the other strategies. The 2nd player also has no incentive to misreport her ranking over her third and fourth most preferred schools, because the 4th student prefers the 3rd student's fourth most preferred school over the third most preferred school. Hence, for the 2nd student, her best response is $(1, 3 > 4)$. The ranking of her second most preferred school is not important, because, in equilibrium, she will not get into it.

- Given the 1st and the 2nd students are truthful with their first choice, the 3rd student's best response is to truthfully report her first choice as well. When the 1st and the 2nd student have the same preference type, which happens with probability $\frac{1}{3}$ (case1, 2), the 3rd student will be able to get into her first choice. In all the other cases (case3, 4, 5, 6), her most and second preferred schools have already been taken by the 1st and the 2nd student, the best she can get is her third most preferred school; therefore, her best response is $(1, 3 > 4)$.

- Given the first three students' strategies, the 4th student can best respond by reporting her third most preferred school ahead of her fourth most preferred school $(3 > 4)$, because the best she can get into under each of the six cases is her third most preferred school.

Next, we look at the low correlation environment, where a type I player has the preference $A \succ B \succ C \succ D$, and a type II player has the preference $B \succ D \succ C \succ A$. Similar to the high correlation environment, there are six different realizations of preference types, and each of them happens with the probability of $\frac{1}{6}$.

- For the 1st student, her dominant strategy is to truthfully report her first choice $(1, *, *, *)$.

- For the 2nd student, truthfully report her first choice is also a dominant strategy. When the 1st and the 2nd students have different preferences (case3, 4, 5, 6), the 2nd student can get into her most preferred school by reporting her first choice truthfully. In the case where the 1st and 2nd students have preference type II (case2), the 2nd student can best respond by truthfully list her first two choices, whereas in the case where the 1st and 2nd students have preference type I (case1), the best she can get is her third-most preferred school given that she lists her first two choices truthfully. Therefore, the Bayesian Nash equilibrium strategy for the 2nd student is complete truth-telling $(1,2,3,4)$.

- Given the strategies of the top two players, the 3rd student can best respond by playing the truth-telling strategy. She will get into her most preferred school when the top two students have the same preference (case1, 2), and her second most preferred school when the 1st and 3rd students have preference type II (case4), or the 1st and the 4th students have preference type I (case6). In the rest two cases (case3, 5), the 3rd student will get into her third-most preferred school.
- Given the strategies of the top three students, the 4th student can best respond by truthfully reporting the order of her last three choices $(2 > 3 > 4)$.

Table A.1 shows the BNE allocation in each case. The capital letters represent the school that a student is assigned to. The numbers in the parentheses represent the corresponding payoffs. In the high correlation environment, under Boston, the allocation in case1 and case2 are not stable; in the low correlation environment, the allocation in case1 is not stable. BNE allocations under DA are always efficient and stable.

Table A.1: Bayesian Nash Equilibrium Outcomes in DA and Boston

	Boston											
	High Correlation						Low Correlation					
Ranking	Case1	Case2	Case3	Case4	Case5	Case6	Case1	Case2	Case3	Case4	Case5	Case6
1st	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)
2nd	C(7)	D(7)	B(16)	A(16)	B(16)	A(16)	C(7)	D(11)	B(16)	A(16)	B(16)	A(16)
3rd	B(16)	A(16)	C(7)	D(7)	D(7)	C(7)	B(16)	A(16)	C(7)	D(11)	D(11)	C(7)
4th	D(7)	C(7)	D(7)	C(7)	C(7)	D(7)	D(11)	C(7)	D(11)	C(7)	C(7)	D(11)
Efficient	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stable	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes

	DA											
	High Correlation						Low Correlation					
Ranking	Case1	Case2	Case3	Case4	Case5	Case6	Case1	Case2	Case3	Case4	Case5	Case6
1st	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)	A(16)	B(16)
2nd	B(11)	A(11)	B(16)	A(16)	B(16)	A(16)	B(11)	D(11)	B(16)	A(16)	B(16)	A(16)
3rd	D(7)	C(7)	C(7)	D(7)	D(7)	C(7)	D(11)	A(16)	C(7)	D(11)	D(11)	C(7)
4th	C(5)	D(5)	D(7)	C(7)	C(7)	D(7)	C(7)	C(7)	D(11)	C(7)	C(7)	D(11)
Efficient	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

A.2 Generalize the IM Dynamic mechanism to heterogeneous college preferences

One key assumption of this paper is that colleges have identical preferences over students. In this section, we generalize our main theoretical result to a special case under heterogeneous college preferences. Our results cannot be generalized to general heterogeneous college preferences.

Theorem 1 says that when arrivals are frequent enough, the stable and efficient outcome arises with arbitrarily high probability under the IM Dynamic mechanism. However, under heterogeneous college preferences, a stable and efficient outcome does not exist. The question becomes whether students can reach the same stable outcome as under student-proposing DA. Lemma 1 proves that when there are only two students and two colleges, such a stable outcome arises with arbitrarily high probability under any rationalizable strategy profile under IM Dynamic mechanism.

Lemma 1. *When the Poisson parameter λ is large enough, if there are only two students and two colleges, each with one seat, the stable outcome arises with arbitrarily high probability under any rationalizable strategy profile.*

Proof. Suppose we have two students i and j, and two colleges a and b. If the two colleges have the same preference, either $i \succ j$ or $j \succ i$, Theorem 1 applies. Therefore, the stable outcome arises with arbitrarily high probability. When the two colleges have different preferences, without losing generality, assume $i \succ_a j$ and $j \succ_b i$. If either student has the highest priority at her most preferred college, without losing generality, assume $a \succ_i b$, then student i has a dominant strategy which is to choose her most preferred school at her first arrival. Conditional on seeing student i has a seat at the college where she has the highest priority, the only rationalizable strategy for student j is to choose the remaining school at her next arrival. In this case, the two students will reach the stable outcome as long as student j has an arrival after student i, which happens with arbitrarily high probability when λ is large enough. If neither student has the highest priority in their most preferred school, without losing generality, assume $b \succ_i a$ and $a \succ_j b$. Assuming rationality, if student j sees student i choose college b, then student j knows that i's preference must be $b \succ_i a$, otherwise, if student i's preference is $a \succ_i b$, then student i has a dominant strategy to choose college a at her first arrival. Conditional on seeing student i choose college b, the only rationalizable strategy for student j is to choose college a at her next arrival. Knowing this, student i has no incentive to misreport her preference, so she will choose b at her first arrival. The same is also true *vice versa*. Under this case, the two students will reach the stable outcome as long as each of them have an arrival, which happens with arbitrarily high probability when λ is large enough. In both cases, the two students will reach the stable outcome as long as there exist three arrivals in the sequence $i > j > i$ (or

$j > i > j$), which happens with arbitrarily high probability when λ is large enough.

□

Definition 1 (Ergin (2002) acyclicality)**.** *A preference profile for colleges over students is said to be acyclical if there does not exist three students i, j, k and two colleges a, b such that* $i \succ_a j \succ_a k \succ_b i$.

Theorem 4. *When the Poisson parameter λ is large enough, if college preferences over students are acyclical, the stable outcome arises with arbitrarily high probability under any rationalizable strategy profile.*

Proof. By acyclicality, there are at most two students who have the highest priority at some colleges. If only one such student exists, then she has the highest priority at all colleges. In this case, this student has a dominant strategy. Otherwise, there are precisely two students who rank the 1st or the 2nd in all colleges. Denote these two students using i and j. Using Lemma 1, we know that these two students will reach a stable outcome between themselves as long as there exist three arrivals in the sequence $i > j > i$ (or $j > i > j$). For simplicity, we call arrivals in the sequence of $i > j > i$ (or $j > i > j$) "mixed arrivals". Again by acyclicality, there exist a 3rd ranked student or a student pair, such that all colleges prefer them over the remaining students. Conditional on seeing the top student or student pair have seats in their current choice, the rationalizable strategy profile of the 3rd ranked student or student pair is characterized by Lemma 1, and they will reach a stable outcome between themselves as long as there exist three mixed arrivals. The same logic applies till the last ranked student or student pair. Since a lower ranked group cannot have justified envy over a higher ranked group, the entire allocation outcome is also stable. Using the technique of proving Theorem 1, we can assign a small time interval for each rank group. As long as there are three mixed arrivals in each time interval for the group of two and one arrival in each time interval for the group of one, students will reach the stable outcome under any rationalizable strategy profile. With λ large enough, the desired arrival sequence arrives with arbitrarily high probability. □

A.3 Tables and Regressions

Table A.3 presents the proportion of truth-telling under each mechanism, where truth-telling under DA is defined as truth-telling up to rank. Table A.2 presents the proportion of first choice truth-telling under each mechanism. Table A.4 presents the proportion of truth-telling, whereas Table A.5 presents the proportion of first choice truth-telling in each score rank under each mechanism. Table A.6 compares the stability of the three mechanisms using the proportion of stable allocation as the measure. Table A.7 compares the efficiency of the three mechanisms using normalized efficiency as the measure. Table A.8 presents session level averages in proportion of truth-telling, justified envy, and efficient allocation, as well as the standard deviation at the treatment level.

Table A.2: Proportion of first choice truth-telling

High Correlation			
Mechanism	Proportion	Ha	p-value
IM DY	0.763	IM DY >DA	0.317
DA	0.746	DA>BOS	0.008
BOS	0.611	IM DY = BOS	0.008
Low Correlation			
Mechanism	Proportion	Ha	p-value
IM DY	0.801	IM DY >DA	0.002
DA	0.664	DA>BOS	0.012
BOS	0.611	IM DY= BOS	0.004

Table A.3: Proportion of truth-telling (DA truth-telling-up-to-rank)

High Correlation			
Mechanism	Proportion	Ha	p-value
IM DY	0.716	IM DY >DA	0.444
DA	0.712	DA>BOS	0.004
BOS	0.435	IM DY = BOS	0.008
Low Correlation			
Mechanism	Proportion	Ha	p-value
IM DY	0.759	IM DY >DA	0.011
DA	0.632	DA>BOS	0.004
BOS	0.445	IM DY= BOS	0.004

Table A.4: Truth-telling by rank

High Correlation									
Rank	IM DY	DA	BOS	Ha	p-value	Ha.1	p-value.1	Ha.2	p-value.2
1st	0.907	0.947	0.900	IM DY<DA	0.091	DA>BOS	0.151	IM DY>BOS	0.464
2nd	0.837	0.830	0.513	IM DY<DA	0.595	DA>BOS	0.004	IM DY>BOS	0.004
3rd	0.500	0.567	0.177	IM DY<DA	0.123	DA>BOS	0.004	IM DY>BOS	0.004
4th	0.620	0.450	0.150	IM DY>DA	0.020	DA>BOS	0.004	IM DY>BOS	0.004

Low Correlation									
Rank	IM DY	DA	BOS	Ha	p-value	Ha.1	p-value.1	Ha.2	p-value.2
1st	0.925	0.967	0.927	IM DY<DA	0.024	DA>BOS	0.103	IM DY>BOS	0.563
2nd	0.742	0.683	0.430	IM DY>DA	0.175	DA>BOS	0.004	IM DY>BOS	0.002
3rd	0.653	0.437	0.230	IM DY>DA	0.015	DA>BOS	0.016	IM DY>BOS	0.002
4th	0.717	0.420	0.193	IM DY>DA	0.002	DA>BOS	0.008	IM DY>BOS	0.002

Table A.5: First choice truth-telling by rank

High Correlation									
Rank	IM DY	DA	BOS	Ha	p-value	Ha.1	p-value.1	Ha.2	p-value.2
1st	0.943	0.983	0.980	IM DY<DA	0.016	DA>BOS	0.500	IM DY<BOS	0.036
2nd	0.850	0.847	0.637	IM DY<DA	0.563	DA>BOS	0.008	IM DY>BOS	0.004
3rd	0.607	0.637	0.460	IM DY<DA	0.313	DA>BOS	0.016	IM DY>BOS	0.012
4th	0.653	0.517	0.367	IM DY>DA	0.083	DA>BOS	0.044	IM DY>BOS	0.004

Low Correlation									
Rank	IM DY	DA	BOS	Ha	p-value	Ha.1	p-value.1	Ha.2	p-value.2
1st	0.944	0.977	0.987	IM DY<DA	0.063	DA<BOS	0.278	IM DY<BOS	0.035
2nd	0.767	0.713	0.490	IM DY>DA	0.182	DA>BOS	0.004	IM DY>BOS	0.002
3rd	0.750	0.490	0.390	IM DY>DA	0.002	DA>BOS	0.095	IM DY>BOS	0.002
4th	0.744	0.477	0.277	IM DY>DA	0.002	DA>BOS	0.020	IM DY>BOS	0.002

Table A.6: Stability comparison

High Correlation			
Mechanism	Proportion	Ha	p-value
IM DY	0.850	IM DY<DA	0.012
DA	0.957	DA>BOS	0.004
BOS	0.543	IM DY>BOS	0.004

Low Correlation			
Mechanism	Proportion	Ha	p-value
IM DY	0.731	IM DY<DA	0.461
DA	0.733	DA>BOS	0.004
BOS	0.490	IM DY>BOS	0.004

Table A.7: Normalized efficiency comparison

Mechanism	Proportion	Ha	p-value
	High Correlation		
Mechanism	Proportion	Ha	p-value
IM DY	0.776	IM DY<DA	0.020
DA	0.847	DA = BOS	0.722
BOS	0.860	IM DY <BOS	0.004
	Low Correlation		
Mechanism	Proportion	Ha	p-value
IM DY	0.913	IM DY<DA	0.130
DA	0.932	DA = BOS	0.167
BOS	0.914	IM DY <BOS	0.5t35

Table A.8: Session Average and Standard Deviation

Measure	Mechanism	Session1	Session2	Session3	Session4	Session5	Session6	SD
				High Correlation				
Measure	Mechanism	Session1	Session2	Session3	Session4	Session5	Session6	SD
Truth-telling	IM DY	0.733	0.675	0.738	0.704	0.729		0.026
	DA	0.771	0.725	0.713	0.629	0.654		0.057
	Boston	0.408	0.483	0.442	0.458	0.383		0.040
Justified envy	IM DY	0.046	0.033	0.050	0.025	0.054		0.012
	DA	0.008	0.004	0.004	0.004	0.033		0.013
	Boston	0.171	0.158	0.121	0.146	0.175		0.022
Efficiency	IM DY	0.817	0.883	0.833	0.933	0.900		0.048
	DA	0.983	0.983	1.000	1.000	0.950		0.020
	Boston	0.950	0.983	0.967	0.950	0.867		0.045
				Low Correlation				
Measure	Mechanism	Session1	Session2	Session3	Session4	Session5	Session6	SD
Truth-telling	IM DY	0.679	0.846	0.758	0.817	0.750	0.704	0.064
	DA	0.562	0.558	0.621	0.713	0.679		0.069
	Boston	0.392	0.379	0.442	0.525	0.487		0.062
Justified envy	IM DY	0.050	0.033	0.100	0.079	0.062	0.108	0.029
	DA	0.083	0.100	0.092	0.050	0.042		0.026
	Boston	0.138	0.188	0.117	0.179	0.142		0.030
Efficiency	IM DY	0.933	0.950	0.783	0.883	0.850	0.850	0.061
	DA	0.917	0.917	0.917	0.950	0.967		0.024
	Boston	0.900	0.850	0.900	0.783	0.867		0.048

Table A.9 shows the average earnings by rank under each environment each mechanism. Columns (5), (6), (7), (12), (13), (14) present corresponding p-values from two-sided permutation tests. The earnings of the 1st ranked student are not significantly different regardless of mechanism and environment. The 2nd ranked student constantly earns less, whereas the 4th ranked student constantly earns more under the Boston mechanism compared with under the other two mechanisms. Comparing IM Dynamic and DA, we find that in the high correlation environment, the 3rd ranked student earns slightly more under the IM Dynamic because these students have the highest priority in their score group, whereas in the low correlation environment, there is no difference between the two mechanisms.

Table A.9: Average Earnings by Rank

				High Correlation		
Ranking	IM DY	DA	Boston	IM DY $\neq$ DA	IM DY $\neq$ Boston	DA $\neq$ Boston
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	79.050	79.583	79.500	0.516	0.659	1.000
2	67.717	71.900	59.767	0.095	0.016	0.008
3	36.967	35.300	42.117	0.048	0.008	0.008
4	30.767	32.233	39.467	0.063	0.008	0.008

				Low Correlation		
Ranking	IM DY	DA	Boston	IM DY $\neq$ DA	IM DY $\neq$ Boston	DA $\neq$ Boston
(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	79.417	79.217	79.567	0.762	0.870	0.357
2	65.250	66.167	61.667	0.636	0.082	0.008
3	51.625	51.667	51.800	0.987	0.857	0.937
4	43.056	44.233	47.133	0.377	0.013	0.008

Table A.10 reports the the distribution of the first choice among the third-ranked students under the IM Dynamic mechanism. We are curious to see if students only choose the practical schools that they can get in, regardless their true preference. We choose to look at the behavior of the third-ranked students because their allocation outcome differs the most across environments. In the high correlation environment, the third-ranked students always prefer school A or B, but the best they can get are either C or D, depending on their preference types. If students only choose practical schools, we would observe few students choose school A or B as their first choice. If students always choose their most preferred schools even when these schools are not practical, we would observe few students choose school C or D as their first choice. In fact, the third ranked students choose the four schools almost equally likely as their first choice, with lightly higher frequency for school A. This shows that student behavior is a combination of being practical and being idealistic.

A similar behavior pattern happens in the low correlation environment. Despite students have zero chance of getting into school B, 39% of the first choices are school B because school B is the most preferred school for 50% of students. Students still choose school C or D even though these two schools are not their favorite, only because these schools are practical.

Table A.10: Distribution of the 3rd-ranked students' first choices

	High Correlation			
	A	B	C	D
Best can get	0	0	50%	50%
Most preferred	50.00%	50.00%	0	0
Actual	29.33%	24.67%	24.67%	21.00%
	Low Correlation			
	A	B	C	D
Best can get	16.67%	0	33.33%	50.00%
Most preferred	50.00%	50.00%	0	0
Actual	28.67%	39.00%	27.00%	25.33%

A.4 Experimental Instructions

The following instructions are translated from the original Chinese version. Instructions for the IM Dynamic mechanism high correlation environment are presented first. Instructions for the Boston mechanism and the DA mechanism are identical except for the subsection of allocation methods and for Review Questions #1, #9, #10, #11. Thus, only these subsections are presented. Instructions for the low correlation environments are identical except for preference type II, hence they are omitted. The Chinese version of instructions and the instructions for the low correlation environment are available from the authors upon request.

A.4.1 Instructions for the IM Dynamic Mechanism, High Correlation Environment

Instructions

Please turn off your cell phone. Thank you!

This is an experiment in the economics of decision making. In this experiment, we simulate a procedure to allocate students to schools. The procedure, payment rules, and student allocation method are described below. Please read them carefully. The amount of money you earn will depend upon the decisions you make and on the decisions other people make. Do not communicate with each other during the experiment. If you have questions at any point during the experiment, raise your hand and the experimenter will help you.
The experiment consists of 20 periods. Participants will be divided into several groups, and every group has four people.

Procedure

- In this experiment, four schools are available for each group. Each school has one slot. Each school slot is allocated to one participant.

- In each group, there are two type I participants, and two type II participants.

  Your payoff amount depends on the school you are assigned to at the end of each period. Payoff amounts of type I are outlined in the following table. These amounts reflect the desirability of the school.

Slot Received at School	A	B	C	D
Payoff to Type I	16	11	7	5

For type I participants, the table is explained as follows:
You will be paid 16 points if you hold a slot of School A at the end of a period.
You will be paid 11 points if you hold a slot of School B at the end of a period.
You will be paid 7 points if you hold a slot of School C at the end of a period.
You will be paid 5 points if you hold a slot of School D at the end of a period.

Payoff amounts of type II are outlined in the following table.

Slot Received at School	A	B	C	D
Payoff to Type I	11	16	5	7

For type II participants, the table is explained as follows:
You will be paid 11 points if you hold a slot of School A at the end of a period.
You will be paid 16 points if you hold a slot of School B at the end of a period.
You will be paid 5 points if you hold a slot of School C at the end of a period.
You will be paid 7 points if you hold a slot of School D at the end of a period.
In each period, you will be randomly assigned to be type I or type II.

- The experiment consists of 20 periods. You will be randomly matched into groups of four at the beginning of each period, which means the group composition is not fixed.

- Your priority of getting into schools is changed every five periods. The participant who ranks the first has the highest priority in any schools; the participant who ranks the second has the second highest priority in any schools; the participant who ranks the third has the third highest priority in any schools; and the participant who ranks the fourth has the fourth highest priority in any schools.

- Note that your allocation in each period is independent of your allocations in other periods.

- In addition, the upper left corner of the screen will show the current period number, and the remaining time in this period. When a new period starts, you will see a history of your choices, rankings and payoffs in previous periods.

- At the end of the experiment, 4 out of 20 periods will be randomly selected for payment. Your total payment will be the sum of payments from the allocation game and your payment from the review questions. You will get 1 Yuan for every 1 point you earned from the allocation game and 0.25 Yuan for every review question answered correctly. You will be paid in private and you are under no obligation to tell others how much you earn.

Allocation Method

- All participants enter an online system to apply for schools at the same time. Each participant can apply for one school at a time and see other group members' real-time applications. Before the system closes, you can change your application as many times as you want.
- The system will be open for half a minute (30 seconds) in each period.
- The system has different closing times for students with different rankings. At 15 seconds, the system will be closed to students who are ranked the first and the second. These students will log out automatically. After 15 seconds, students who are ranked the third and the fourth can still continue their applications. At 30 seconds, the system will be closed to students who are ranked the third and the fourth.
- If a school receives more applications than its capacity by the time the system closes, the applicant with the highest priority will be admitted. The remaining applicants will not be admitted to any schools.
- Below is the interface of the system. The upper left corner of the screen shows your ranking. The left side of the screen shows your payoffs from getting into different schools. The right side of the screen displays everyone's application at the moment. As shown, the student who ranks the first applies for School A, the student who ranks the second applies for School B, the student who ranks the third applies for School C, and the student who ranks the fourth applies for School D. If you want to change your application, just click the red button behind the school you want to apply to, and the system will update automatically. The upper right corner of the screen displays the remaining time in this period. The bottom of the screen shows your decisions and results of all previous periods.

Review Questions

The following example has the same number of students and schools as the actual decisions you will make. Please work out the allocation of this example for Review Question #1. We will go through this example after you submit your answer.

Feel free to refer to the experimental instructions before you answer any question. Each correct answer is worth 0.25 Yuan, and will be added to your total earnings. You can earn up to 4.25 Yuan for the Review Questions.

1. In this example, there are four students, 1-4, and four schools, A, B, C and D. The priority of students is: student 1 $\succ$ student 2 $\succ$ student 3 $\succ$ student 4. By the time the system closes, students' applications are as follows:

Student	Application at the deadline
Student 1	B
Student 2	C
Student 3	D
Student 4	C

Please fill student 1, 2, 3, 4 into the school that admits them in Review Question #1. If a school did not admit any students, fill in 0.

School	A	B	C	D
Student				

Review Question 2-14

2. How many participants are there in your group each period including you?
3. True or false: You will be matched with the same three participants each period.
4. How many different preference types are there?
5. True or false: Your priority is determined by your ranking.
6. True or false: Your ranking is fixed for the entire 20 periods.
7. True or false: The participant who ranks the first has the highest priority.
8. True or false: Other things being equal, a low lottery number is better than a high lottery number.
9. True or false: You are only allowed to revise your application up to a certain number of times.
10. True or false: You can't change your application once the system closes.
11. True or false: By the time the system closes, if the school you apply to has another applicant whose priority is higher than you, you will not be admitted by any schools.
12. How many participants are there in a group with the preference type A≻B≻C≻D?
13. How many participants are there in a group with the preference type B≻A≻D≻C?
14. True or false: Your preference of schools is fixed for all 20 periods.

You will have 5 minutes to go over the instructions at your own pace. Feel free to earn as much as you can. Are there any questions?

A.4.2 Instructions for the Boston Mechanism

……

Allocation Method

- Step 1.

  a An application to the first ranked school is sent for each participant.

  b Each school accepts the student with highest priority in that school. These students and their assignments are removed from the system. The remaining applications for each respective school are rejected.

- Step 2.

  a The rejected applications are sent to his/her second choice.

  b If a school is still vacant, then it accepts the student with the highest priority and rejects the remaining applications.

- Step 3.

  a The application of each participant who is rejected by his/her top two choices is sent to his/her third choice

  b If a school is still vacant, then it accepts the student with the highest priority and rejects the remaining applications.

- Step 4.

  a Each remaining participant is assigned a slot at his/her last choice.

Note that the allocation is final in each step.

……

Review Questions

1. In this example, there are four students, 1-4, and four schools, A, B, C and D. The priority of students is: student 1 $\succ$ student 2 $\succ$ student 3 $\succ$ student 4. The students submit the following school rankings:

   Please fill student 1, 2, 3, 4 into the school that admits them in Review Question #1. If a school did not admit any students, fill in 0.

9 True or False: If you are accepted by a school of your choice, the schools ranked below are irrelevant.

	1st Choice	2nd Choice	3rd Choice	4th Choice
Student 1	D	A	C	B
Student 2	D	A	B	C
Student 3	A	B	C	D
Student 4	A	D	B	C

School	A	B	C	D
Student				

10 True or False: If you are not rejected at a step, then you are accepted into that school.

11 True or False: The allocation is final at the end of each step.

. . . . . .

A.4.3 Instructions for the DA mechanism

Allocation Method

- An application to the first ranked school is sent for each participant.
- Throughout the allocation process, a school can hold no more applications than its capacity. If a school receives more applications than its capacity, then it temporarily retains the student with the highest priority and rejects the remaining students.
- Whenever an applicant is rejected at a school, his or her application is sent to the next highest ranked school.
- Whenever a school receives new applications, these applications are considered together with the retained application for that school. Among the retained and new applications, the one with the highest priority is temporarily on hold.
- The allocation is finalized when no more applications can be rejected. Each participant is assigned a slot at the school that holds his/her application at the end of the process.

Note that the allocation is temporary in each step until the last step.

A.5 Chinese College Admission Mechanism

In this table, we present the mechanisms used across different provinces in China. These mechanisms do not include tier 0, ethnic minority, military, art, PE students, or students who are qualified for the national anti-poverty program. All numbers listed below are for four-year college admissions; 2-year college admissions are omitted. Semicolons are used to separate different tiers. Commas are used to separate different groups within each tier. Some provinces have three tiers, while others have two tiers.

Table A.11: Chinese College Admission Mechanism by Provinces in 2016

Province	Mechanism Type	Sequence	No. of Application in 2016
Beijing	symmetric parallel	$(6;6;6;\cdots)$	61200
Tianjin	symmetric parallel	$(9,9;9,9;9;\cdots)$	60000
Jiangsu	symmetric parallel	$(5;5;5;\cdots)$	360400
Zhejiang	symmetric parallel	$(5;5;5;\cdots)$	307400
Shandong	symmetric parallel	$(6;6;\cdots)$	710000
Henan	symmetric parallel	$(6;6;6;\cdots)$	820000
Hubei	symmetric parallel	$(9;9;\cdots)$	361000
Hunan	symmetric parallel	$(5;5;5\cdots)$	401600
Guangxi	symmetric parallel	$(6;6;\cdots)$	330000
Hainan	symmetric parallel	$(6;6;6;\cdots)$	60400
Chongqing	symmetric parallel	$(6;6;\cdots)$	249000
Sichuan	symmetric parallel	$(6;6;\cdots)$	570000
Yunnan	symmetric parallel	$(5;5;5;\cdots)$	281100
Tibet	symmetric parallel	$(10;10;10;\cdots)$	23976
Shănxi	symmetric parallel	$(6;6;6;\cdots)$	328000
Gansu	symmetric parallel	$(6;6;6;\cdots)$	296000
Ningxia	symmetric parallel	$(4;4;4;\cdots)$	69100
Shanghai	asymmetric parallel	$(1$(peking U or Tsinghua U)$;10;\cdots)$	51000
Anhui	asymmetric parallel	$(6;6;4;\cdots)$	509900
Fujian	asymmetric parallel	$(6;10;\cdots)$	175000
Jiangxi	asymmetric parallel	$(6;8;\cdots)$	360600
Guangdong	asymmetric parallel	$(7,4;7,4;\cdots)$	733000
Guizhou	asymmetric parallel	$(6;8;8;\cdots)$	373800
Hebei	asymmetric parallel	$(5;10;\cdots)$	423100
Shanxi	asymmetric parallel	$(5,5,5;8,5,5;\cdots)$	330000
Liaoning	asymmetric parallel	$(7;9;\cdots)$	218200
Heilongjiang	asymmetric parallel	$(5,1;5,1;5,1;\cdots)$	197000
Xinjiang	asymmetric parallel	$(6;6;1;\cdots)$	166100
Jilin	parallel&sequential	$(5,2$(sequential)$;1,7,1,3;1,6;\cdots)$	148000
Qinghai	parallel&sequential	$(2$(sequential)$;2$(sequential)$;5;\cdots)$	44600
Inner Mongolia	dynamic adjustment		201100

APPENDIX B

Appendix for Chapter 3

B.1 Instructions

High_Small Treatment

Instructions

(Please mute or turn off your phone. Thanks!)

This is an experiment in the economics of decision making. In this experiment, you will be asked to make investment decisions. The amount of money you earn will depend on the decisions you make and on the decisions other people make. Please do not communicate with other people during the experiment. If you have questions at any point during the experiment, raise your hand and the experimenter will help you.

All transactions during the experiment will be calculated in terms of points: every 8 points equals 1 dollar in cash (8 points = $1). For example, if you earn 16 points during this session, you will be paid 2 dollars in cash at the end of the experiment. If your total earnings do not equal an integer, it will be rounded towards the nearest integer.

The experiment will consist of three parts: the review questions, the investment game, and the survey. The total amount of points you earn during the session will be the sum of points you earn from each of the three parts:

- At the end of the instructions, you will be asked to complete a set of review questions. You will get 1.6 points ($0.2) for every question you answer correctly.
- The investment game will consist of 20 rounds. At the end of the experiment, the computer will randomly select 1 round out of the 20 that you will get paid for. Each round is equally likely to be selected for payment.
- A lottery game will determine your payment from the final part of the experiment.

The Investment Game

The investment game consists of 20 rounds. At the beginning of each round, you will be randomly matched with 3 other participants to form a group of 4. The composition of the groups will change every round. You will not be informed of the identity of the other group members. At the beginning of the round, each participant will receive 10 points. You will decide how to divide 10 points between two investment opportunities. The return from the group account depends on the actions of your fellow group members. While the return from the private account depends only on your choices. You can invest any integer amount between 1 and 10 points, inclusive, in the group account. Meanwhile, other group members will independently make similar decisions. No one will be informed of the others' decisions until everyone submits his or her decision.
At the end of each round (after all decisions are submitted), you will see:

- Your investment decision
- The investment decisions of the other members in your group
- Your earnings

Then the next round will start. You will be randomly matched with another three participants and receive a new endowment of 10 points.
Earnings calculation Your total earnings will be the sum of the earnings from your private account and a fourth of the group account:

- *Earnings from the private account*: You will earn 1 point for every point you keep in the private account. For example, if you keep 5 points in the private account, your earnings from the private account will be 5 points.
- *Earnings from the group account*: The earnings from the group account will depend on the investments of all group members, and will be shared equally among the group members. Each group member will receive one quarter $(1/4)$ of the total return from the group account. The exact earnings earned from the group account can be accessed through the calculator on the bottom right corner of the screen.

The formula for calculating your earnings is as follows:

$$\text{Your total earnings} = (10 - x_i) + 15 \times \left(\sum_{i=1}^{4} x_i^{-2}\right)^{-1/2},$$

where x_i is each group member's investment level.
For your convenience, you can use the calculator instead of the formula to calculate your earnings.

Using the calculator to compute your earnings
The calculator on the bottom right of the screen can help you calculate your earnings. To activate the calculator, simply fill in a hypothetical value for your own investment and for other group members' investments in the group account. The calculator will then display your earnings from the group account, your earnings from the private account, and your total earnings.
Here are some facts about earnings from the group account:

1. The more you and others invest, the higher the return will be.

2. Consider two levels for your investment in the group account, low investment and high investment. Next, increase both the low and the high investment by 1 point. The total return will increase in both cases; however, the increase is smaller in the case of the higher investment level.

3. When you increase your investment in the group account, the total return will not increase at a constant rate. The rate of increase depends on all group members' investment levels in the group account.

4. For the same average investment in the group account, the total return will be higher if everyone invests similar amounts in the group account.

5. In general, it is a good idea to use the calculator to understand exactly what your earnings will be based on different investment scenarios.

Examples: we will go through three examples to illustrate how the investment game works. Please take five minutes to go through the examples at your own pace.

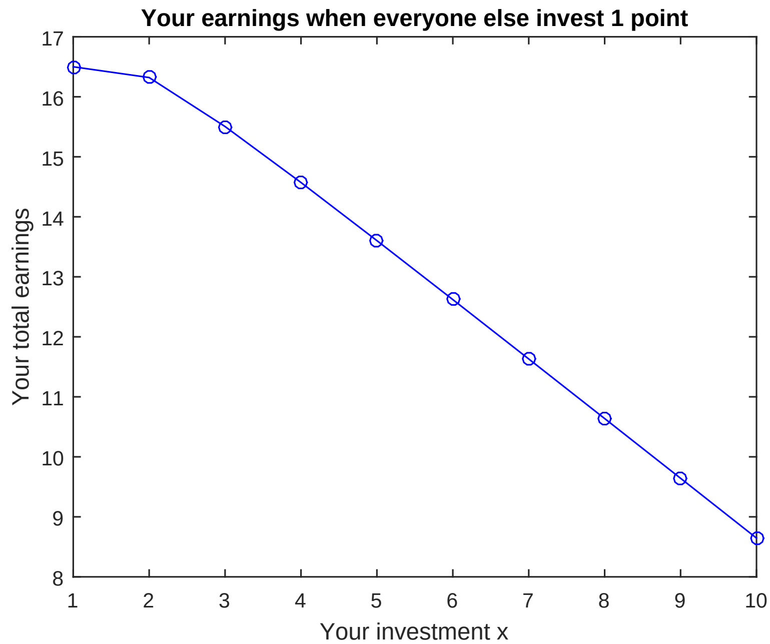
1. The first example demonstrates the special case when everyone else invests 1 point in the group account.

   (a) If everyone else in your group invests 1 point in the group account, and you invest 1 point, the total return from the group account would be 30.00 points. As a result, your earnings from the group account would be 7.50 points (a quarter of the total return). Your earnings from the private account would be 9 points, and your total earnings would be 16.50 points.

   (b) If everyone else in your group invests 1 point in the group account, and you invest 5 points, the total return from the group account would be 34.41 points. As a result, your earnings from the group account would be 8.60 points (a quarter of the total return). Your earnings from the private account would be 5 points, and your total earnings would be 13.60 points.

(c) If everyone else in your group invests 1 point in the group account, and you invest 10 points, the total return from the group account would be 34.58 points. As a result, your earnings from the group account would be 8.65 points (a quarter of the total return). Your earnings from the private account would be 0 points, and your total earnings would be 8.65 points.

(d) The following graph shows your total earnings if you invest x points when everyone else in your group invests 1 point in the group account.



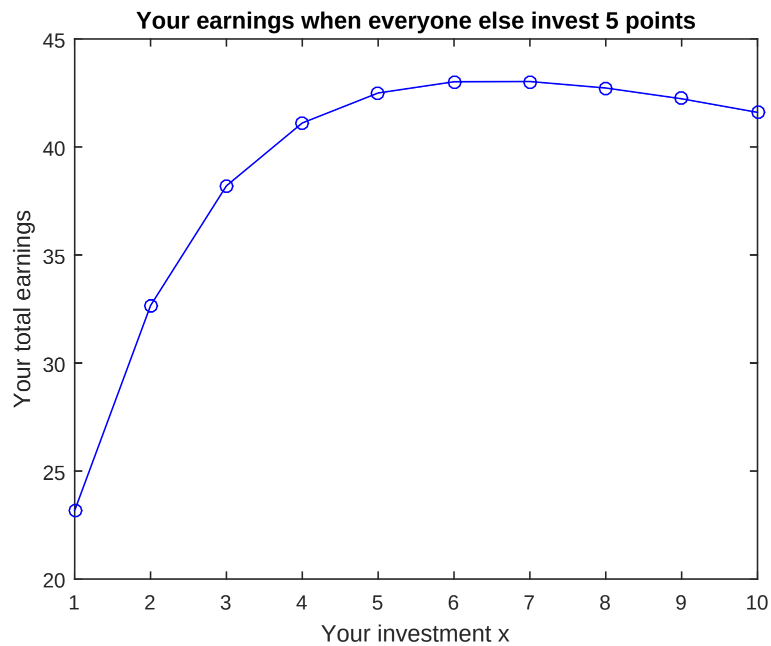
2. The second example demonstrates the special case when everyone else invests 5 points in the group account.

(a) If everyone else in your group invests 5 points in the group account, and you invest 1 points, the total return from the group account would be 56.69 points. As a result, your earnings from the group account would be 14.17 points (a quarter of the total return). Your earnings from the private account would be 9 points, and your total earnings would be 23.17 points.

(b) If everyone else in your group invests 5 points in the group account, and you invest 5 points, the total return from the group account would be 150 points. As a result, your earnings from the group account would be 37.50 points (a quarter of the total return). Your earnings from the private account would be 5 points, and your total earnings would be 42.50 points.

(c) If everyone else in your group invests 5 points in the group account, and you invest 10 points, the total return from the group account would be 166.41 points. As a result, your earnings from the group account would be 41.60 points (a quarter of the total return). Your earnings from the private account would be 0 points, and your total earnings would be 41.60 points.

(d) The following graph shows your total earnings if you invest x points when everyone else in your group invests 5 points in the group account.



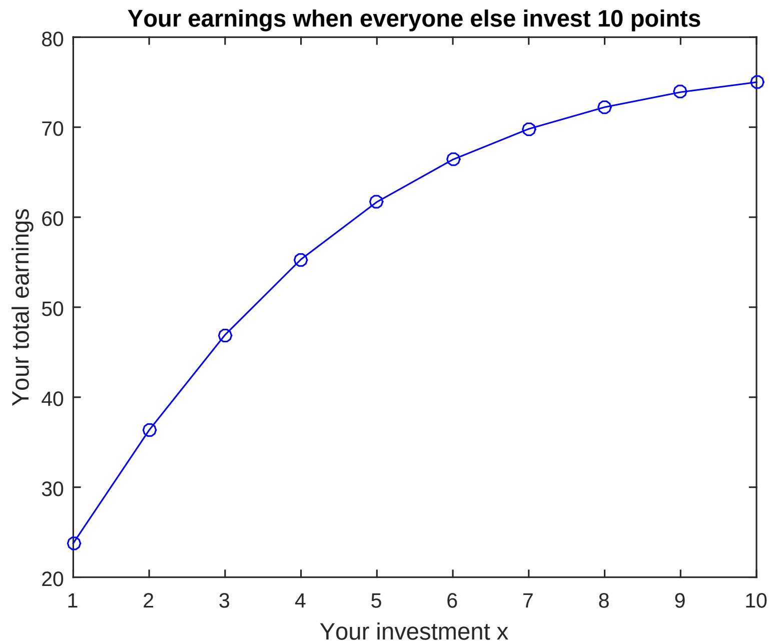
3. The third example demonstrates the special case when everyone else invests 10 points in the group account.

   (a) If everyone else in your group invests 10 points in the group account, and you invest 1 points, the total return from the group account would be 59.12 points. As a result, your earnings from the group account would be 14.78 points (a quarter of the total return). Your earnings from the private account would be 9 points, and your total earnings would be 23.78 points.

   (b) If everyone else in your group invests 10 points in the group account, and you invest 5 points, the total return from the group account would be 226.78 points. As a result, your earnings from the group account would be 56.69 points (a quarter of the total return). Your earnings from the private account would be 5 points, and your total earnings would be 61.69 points.

   (c) If everyone else in your group invests 10 points in the group account, and you invest 10 points, the total return from the group account would be 300 points. As a result, your earnings from the group account would be 28.98 points (a quarter of group earnings). Your earnings from the private account would be 75 points, and your total earnings would be 75 points.

(d) The following graph shows your total earnings if you invest x points when everyone else in your group invests 10 points in the group account.



Note that even though in all three examples we assume that everyone else in your group invests the same amount in the group account, in the actual game, they might invest different amounts in the group account.

Keep in mind that other people in your group face the same problem as you. Their earnings from the group account and private account follows the same rule as yours.

When you finish reading, click on the “next” button on the screen and proceed to the review questions. Please raise your hand if you have any questions.

High_Large Treatment

Instructions

(Please mute or turn off your phone. Thanks!)

This is an experiment in the economics of decision making. In this experiment, you will be asked to make investment decisions. The amount of money you earn will depend on the decisions you make and on the decisions other people make. Please do not communicate with other people during the experiment. If you have questions at any point during the experiment, raise your hand and the experimenter will help you.

All transactions during the experiment will be calculated in terms of points: every 2 points equals 1 dollar in cash (2 points = $1). For example, if you earn 4 points during this session, you will be paid 2 dollars in cash at the end of the experiment. If your total earnings do not equal an integer, it will be rounded towards the nearest integer.

The experiment will consist of three parts: the review questions, the investment game, and the survey. The total amount of points you earn during the session will be the sum of points you earn from each of the three parts:

- At the end of the instructions, you will be asked to complete a set of review questions. You will get 0.4 points ($0.2) for every question you answer correctly.
- The investment game will consist of 20 rounds. At the end of the experiment, the computer will randomly select 1 round out of the 20 that you will get paid for. Each round is equally likely to be selected for payment.
- A lottery game will determine your payment from the final part of the experiment.

The Investment Game

The investment game consists of 20 rounds. At the beginning of each round, you will be randomly matched with 9 other participants to form a group of 10. The composition of the groups will change every round. You will not be informed of the identity of the other group members. At the beginning of the round, each participant will receive 10 points. You will decide how to divide 10 points between two investment opportunities. The return from the group account depends on the actions of your fellow group members. While the return from the private account depends only on your choices. You can invest any integer amount between 1 and 10 points, inclusive, in the group account. Meanwhile, other group members will independently make similar decisions. No one will be informed of the others' decisions until everyone submits his or her decision.

At the end of each round (after all decisions are submitted), you will see:

- Your investment decision
- The investment decisions of the other members in your group
- Your earnings

Then the next round will start. You will be randomly matched with another nine participants and receive a new endowment of 10 points.

Earnings calculation Your total earnings will be the sum of the earnings from your private account and 10% of the group account:

- *Earnings from the private account*: You will earn 1 point for every point you keep in the private account. For example, if you keep 5 points in the private account, your earnings from the private account will be 5 points.
- *Earnings from the group account*: The earnings from the group account will depend on the investments of all group members, and will be shared equally among the group members. Each group member will receive 10 percent $(1/10)$ of the total return from the group account. The exact earnings earned from the group account can be accessed through the calculator on the bottom right corner of the screen.

The formula for calculating your earnings is as follows:

$$\text{Your total earnings} = (10 - x_i) + 15 \times (\sum_{i=1}^{10} x_i^{-2})^{-1/2},$$

where x_i is each group member's investment level.

For your convenience, you can use the calculator instead of the formula to calculate your earnings.

Using the calculator to compute your earnings

The calculator on the bottom right of the screen can help you calculate your earnings. To activate the calculator, simply fill in a hypothetical value for your own investment and for other group members' investments in the group account. The calculator will then display your earnings from the group account, your earnings from the private account, and your total earnings.

Here are some facts about earnings from the group account:

1. The more you and others invest, the higher the return will be.
2. Consider two levels for your investment in the group account, low investment and high investment. Next, increase both the low and the high investment by 1 point. The total return will increase in both cases; however, the increase is smaller in the case of the higher investment level.

3. When you increase your investment in the group account, the total return will not increase at a constant rate. The rate of increase depends on all group members' investment levels in the group account.

4. For the same average investment in the group account, the total return will be higher if everyone invests similar amounts in the group account.

5. In general, it is a good idea to use the calculator to understand exactly what your earnings will be based on different investment scenarios.

Examples: we will go through three examples to illustrate how the investment game works. Please take five minutes to go through the examples at your own pace.

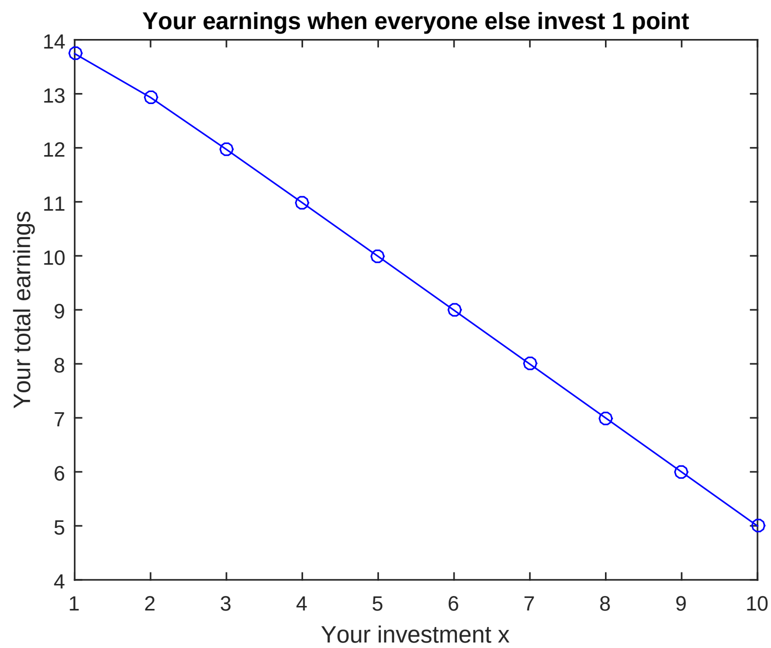
1. The first example demonstrates the special case when everyone else invests 1 point in the group account.

   (a) If everyone else in your group invests 1 point in the group account, and you invest 1 point, the total return from the group account would be 47.43 points. As a result, your earnings from the group account would be 4.74 points (10 percent of the total return). Your earnings from the private account would be 9 points, and your total earnings would be 13.74 points.

   (b) If everyone else in your group invests 1 point in the group account, and you invest 5 points, the total return from the group account would be 49.89 points. As a result, your earnings from the group account would be 4.99 points (10 percent of the total return). Your earnings from the private account would be 5 points, and your total earnings would be 9.99 points.

   (c) If everyone else in your group invests 1 point in the group account, and you invest 10 points, the total return from the group account would be 49.97 points. As a result, your earnings from the group account would be 5 points (10 percent of the total return). Your earnings from the private account would be 0 points, and your total earnings would be 5 points.

   (d) The following graph shows your total earnings if you invest x points when everyone else in your group invests 1 point in the group account.



2. The second example demonstrates the special case when everyone else invests 5 points in the group account.

   (a) If everyone else in your group invests 5 points in the group account, and you invest 1 points, the total return from the group account would be 128.62 points. As a result, your earnings from the group account would be 12.86 points (10 percent of the total return). Your earnings from the private account would be 9 points, and your total earnings would be 21.86 points.

   (b) If everyone else in your group invests 5 points in the group account, and you invest 5 points, the total return from the group account would be 237.17 points. As a result, your earnings from the group account would be 23.72 points (a quarter of the total return). Your earnings from the private account would be 5 points, and your total earnings would be 28.72 points.

   (c) If everyone else in your group invests 5 points in the group account, and you invest 10 points, the total return from the group account would be 246.60 points. As a result, your earnings from the group account would be 24.66 points (10 percent of the total return). Your earnings from the private account would be 0 points, and your total earnings would be 24.66 points.

   (d) The following graph shows your total earnings if you invest x points when everyone else in your group invests 5 points in the group account.

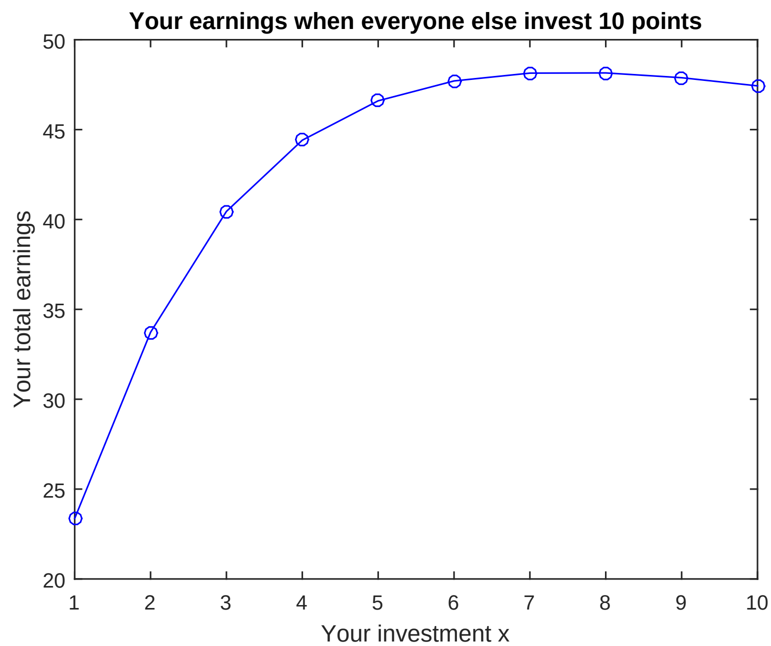
3. The third example demonstrates the special case when everyone else invests 10 points in the group account.

   (a) If everyone else in your group invests 10 points in the group account, and you invest 1 points, the total return from the group account would be 143.67 points. As a result, your earnings from the group account would be 14.37 points (10 percent of the total return). Your earnings from the private account would be 9 points, and your total earnings would be 23.37 points.

   (b) If everyone else in your group invests 10 points in the group account, and you invest 5 points, the total return from the group account would be 416.03 points. As a result, your earnings from the group account would be 41.60 points (10 percent of the total return). Your earnings from the private account would be 5 points, and your total earnings would be 46.60 points.

   (c) If everyone else in your group invests 10 points in the group account, and you invest 10 points, the total return from the group account would be 474.34 points. As a result, your earnings from the group account would be 47.43 points (10 percent of group earnings). Your earnings from the private account would be 75 points, and your total earnings would be 47.43 points.

   (d) The following graph shows your total earnings if you invest x points when everyone else in your group invests 10 points in the group account.



Note that even though in all three examples we assume that everyone else in your group invests the same amount in the group account, in the actual game, they might invest different amounts in the group account.
Keep in mind that other people in your group face the same problem as you. Their earnings from the group account and private account follows the same rule as yours.
When you finish reading, click on the "next" button on the screen and proceed to the review questions. Please raise your hand if you have any questions.

B.2 Additional Tables and Graphs

This section documents extra tables and graphs. Table B.1 reports the comparison between our theoretical prediction and the experimental results in Fenig et al. (2018).

Table B.1: Theoretical Predictions Validated by Fenig et al. (2018)

Study	Complementarity (γ)	Group Return (β)	Group Size	Lower Bound	Our Theoretical Prediction	Their Observed Trend
Fenig,	0.70	0.4	4	8.48	low	low
Gallipoli	0.65	0.4	4	5.48	low	medium
and Halevy	0.58	0.4	4	3.54	high	high
(2018)	0.54	0.4	4	2.93	high	high

Table B.2 reports the summary of statistics on gender, age, risk attitude (Holt and Laury), ethnicity, student status, major, and household income. We do not find statistical significant difference in any of these characteristics, except in risk attitude. We further use pairwise t-test with

Holm–Bonferroni correction to find which two treatments are different in risk attitude. It turns out that the High_Large treatment is 0.852 more risk averse than the Low_Four treatment using the Holt and Laury risk measure (p-value= 0.042, after multiple hypotheses testing correction). Results from any other pairwise comparisons are not significant.

Table B.2: Summary of Statistics and Balanceness Check

Feature		High Complementarity Small	High Complementarity Large	Low Complementarity Small	Low Complementarity Large	Balanceness Check P-Value
Male		0.351	0.431	0.424	0.345	0.745 (Z Test)
Age		21.123	20.672	20.915	20.690	0.772 (ANOVA)
Risk Attitude		5.649	6.276	5.424	5.845	0.043** (ANOVA)
Full Time Student		0.930	0.948	0.814	0.897	0.081 (Z Test)
Undergraduate		0.877	0.862	0.898	0.931	0.657 (Z Test)
Ethnicity	Caucasion	0.456	0.448	0.492	0.466	$0.721(\chi^2$ Test)
	Asian	0.404	0.431	0.339	0.448	
	Black	0.070	0.069	0.051	0.017	
	Hispanic	0.018	0.000	0.051	0.052	
	Multi-racial	0.053	0.052	0.068	0.017	
Major	Arts/Humanities/Education	0.123	0.086	0.017	0.034	$0.145(\chi^2$ Test)
	Business/Management (including MBA)	0.070	0.121	0.136	0.103	
	Math/Science/Engineering/Computer Science	0.246	0.310	0.492	0.259	
	Medical/Nursing (but not pre-med)	0.018	0.052	0.085	0.172	
	Economics/Politics/Psychology	0.281	0.241	0.169	0.207	
	Other Social Science	0.070	0.052	0.017	0.052	
	No Major or Pre-College	0.035	0.017	0.000	0.017	
	Other	0.158	0.121	0.085	0.155	
Household Income	Below $50,000	0.263	0.293	0.169	0.138	$0.085(\chi^2$ Test)
	$50,001 - $100,000	0.246	0.241	0.305	0.224	
	$100,001 - $200,000	0.316	0.207	0.390	0.293	
	above $200,000	0.175	0.259	0.136	0.345	

Table B.3 reports the average contribution and standard deviation at the session level.

Table B.3: Average Contribution by Sessions

Complementarity	Group Size	Session	Average Contribution	std
Low	Small	1	2.870	2.531
		2	5.055	2.931
		3	4.210	3.080
	Large	1	9.072	1.598
		2	9.570	1.347
		3	9.400	1.622
High	Small	1	7.690	2.354
		2	9.650	1.058
		3	8.908	1.505
	Large	1	4.562	2.351
		2	2.792	1.874
		3	4.612	2.325

We examine the effect of risk attitude on the level of contribution. After controlling for period and others' complementary sum, we find risk attitude does not have a significant effect on the contribution level in any treatments. Table B.4 shows the regression statistics from OLS regressions.

Table B.4: Effect of Risk Attitude on Contribution

Contribution	(1) Low_Small	(2) Low_Large	(3) High_Small	(4) High_Large	(5) Low_Large + High_Small	(6) Low_Small + High_Large
Risk	0.051	0.068	0.079	0.014	0.080	0.046
	(0.085)	(0.048)	(0.103)	(0.111)	(0.048)	(0.061)
OtherCompSum	0.153	0.168**	-1.952	-0.472***	0.014	0.017
	(0.108)	(0.019)	(1.112)	(0.003)	(0.012)	(0.084)
Period	-0.032	0.040*	0.119**	-0.045*	0.110***	-0.059***
	(0.019)	(0.011)	(0.025)	(0.014)	(0.014)	(0.012)
Constant	2.955*	1.316	7.183***	5.225**	7.123***	4.292***
	(0.900)	(0.619)	(0.219)	(1.106)	(0.366)	(0.734)
Observations	1,200	1,200	1,200	1,200	2,400	2,400
R-squared	0.023	0.210	0.175	0.103	0.175	0.017

Standard errors clustered at session level.
*** p<0.01, ** p<0.05, * p<0.1

An alternative measure to compare the predicted contribution with the real contribution is to use the Euclidean distance. We calculate the Euclidean distance between the predicted contribution distribution and the real contribution distribution. The Euclidean distance is defined as

$$d(\hat{\mathbf{p}}, \mathbf{p}) = \sqrt{\sum_{i=1}^{10} (\hat{p}_i - p_i)^2}, \tag{B.1}$$

where $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_1, \cdots; \hat{p}_{10})$ is the predicted contribution distribution, $\hat{p}_i$ is the predicted probability of contributing i. $\mathbf{p} = (p_1, p_1, \cdots; p_{10})$ is the real contribution distribution; p_i is the real probability of contributing i. The result is in Table B.5.

The Euclidean distance gives us the same results as the cosine similarity. Among the three models, the conditional cooperation model strictly dominate the best response model in all treatments. Between the conditional cooperation model and the fictitious play model, the conditional cooperation model has superior performance in the Low_Large treamtent, but the fictitious play model has superior performance in the Low_Small treatment. However, neither model is as good as the benchmark where players repeat what they did in period ten. In sum, none of our models predict better than simply using period ten contribution level.

Table B.5: Euclidean Distance Between Predicted and Real Contribution Distribution

Complementarity	Group Size	Conditional Cooperation	Best Response	Fictitious Play	Period 10	NE
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low	Small	0.642	1.121	0.379	0.311	0.568
	Large	0.067	0.205	0.766	0.067	0.205
High	Small	0.308	0.346	0.379	0.308	0.216
	Large	0.357	1.021	0.317	0.240	0.880

B.3 Classify players into different types

We observe that there are strong individual patterns in the data. A subject's contribution in the previous period can predict her contribution in the current period better than all the three models we have tried (Table B.6). We are curious how interactions between different types shape the contribution dynamics. Following Fischbacher et al. (2001); Fischbacher and Gächter (2010), we do a similar classification on our data.

Table B.6: Effect of the previous period contribution

	Low Complementarity		High Complementarity	
	Small	Large	Small	Large
Contribution	(1)	(2)	(3)	(4)
last period contribution	0.662***	0.488**	0.741***	0.663***
	(0.012)	(0.058)	(0.023)	(0.060)
Constant	1.363**	4.928**	2.432**	1.288**
	(0.236)	(0.605)	(0.324)	(0.138)
Observations	1,140	1,140	1,140	1,140
R-squared	0.430	0.321	0.645	0.447

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Because we didn't elicit subjects' belief on other players' contribution level in each round, we use other players' average contribution level in the previous period as subjects' belief. We can also use data in calculator as subjects' belief, but because the usage of calculator is not common, and highly concentrated in the first several rounds, we will lose the majority of our observation if we only depend on calculator data.

We first try to classify subjects' behavior by running a regression between their contribution and the average contributions of others in their group in the previous period. If the slope of regression is positive and the p-value of the slope is less than 5%, we classify this subject as a "conditional cooperator". If the slope is negative and the p-value is less than 5%, we classify this subject as an "inverse conditional cooperator". When the p-value is larger than 5%, if a subject always contributes more than 8 points out of 10, we call her a "unconditional cooperator"; in the contrary, if she always contributes less than 3 points out of 10, we call her an "free rider".

This classification rule gives us 45.83% of conditional cooperators, 6.25% unconditional cooperators, 5.83% of free riders, 4.17% of inverse conditional cooperators, while leaving 37.92% of subjects unclassified.

Because the large proportion of unclassified subjects, we are not satisfied with this classification method. For each unclassified subject, we did a scatter plot between their contribution and others' average contribution in the previous round. Going through each graph, we figured that the first 5 rounds are exploratory for most subjects, and they start to form a strategy after trying several random values in the first several rounds. Because of this, we re-run the classification task by excluding the first 5 rounds.

We find that 35.83% of subjects are contributing at the same amount in at least 14 rounds out of 15 rounds, regardless their group members' contribution in the previous round. Among them, 60% of them are unconditional cooperators, 11.63% of them are free riders, while the rest 27.91% are constantly contributing in the middle range. 31.67% subjects are conditional cooperators. 4.58% are unconditional cooperator but do not always contribute to the same amount. 2.92% are free riders who do not always contribute to the same amount. 2.08 are inverse conditional cooperators. This classification method leaves us 24.58% of unclassified subjects. Among unclassified subjects, jumping around 2 or 3 contribution levels is a common strategy, while the rest are more like random walk.

The third method that I use to classify subjects is a direct comparison between the first five rounds average and the last five rounds average using the permutation test. Because in the two treatments where the theoretical prediction is to contribute everything, we observe that the average contribution level increases in all sessions, we therefore classify subjects whose last five periods contribution higher than the first 5 periods'

Table B.7: Proportion of each type

Method	Cond Coop	Uncond Coop	Free-rider	Inverse Coop	Unclassified
OLS (all period)	45.83%	6.25%	5.83%	4.17%	37.92%
OLS (last 10 periods)	31.67%	26.25%	7.08%	2.08%	32.92%
Permutation Test	32.08%	26.25%	8.33%	1.25%	33.75%

Table B.8: Proportion of types in each treatment

Permutation Test (First 5 vs. Last 5)						
Complementarity	Group Size	Cond Coop	Uncond Coop	Free rider	Invese Coop	Unclassified
Low	Small	0.15	0.02	0.20	0.03	0.60
	Large	0.27	0.60	0.00	0.00	0.13
High	Small	0.55	0.35	0.00	0.00	0.10
	Large	0.32	0.02	0.13	0.02	0.52

OLS (Last 10 rounds)						
Complementarity	Group Size	Cond Coop	Uncond Coop	Free rider	Invese Coop	Unclassified
Low	Small	0.23	0.03	0.17	0.00	0.57
	large	0.15	0.68	0.00	0.03	0.13
High	Small	0.53	0.33	0.00	0.02	0.12
	Large	0.35	0.00	0.12	0.03	0.50

OLS (All rounds)						
Complementarity	Group Size	Cond Coop	Uncond Coop	Free rider	Invese Coop	Unclassified
Low	Small	0.20	0.00	0.17	0.02	0.62
	Large	0.55	0.15	0.00	0.08	0.22
High	Small	0.67	0.08	0.00	0.07	0.18
	Large	0.42	0.02	0.07	0.00	0.50

APPENDIX C

Appendix for Chapter 4

C.1 Destination Region vs. Destination Location

Two flexibility features exist on the platform. One is to specify a precise destination location. I call this feature the "destination location" feature. The other is to specify a coarse location preference, i.e., a destination region. I call this feature the "destination region" feature. Drivers can use the destination location features three times a day, and use the destination region feature unlimited. I expect both substitution and complementary effect between the two features. In this subsection, I are going to explore the usage patterns of the two different features.

The usage of destination location feature is prevalent. 72.931% of drivers in our trip-completed sample completed destination location trips, while 26.144% of them completed destination region trips during the experimental period. The usage patterns of the two features are different. Destination location trips are more likely to be end-of-session trips. 47.633% of direct destination trips are end-of-session trips, whereas only 14.565% of destination region trips are end-of-session trips.

The waiting time before a trip is much longer for destination location trips than destination region trips. The average waiting time for ordinary trips without any destination specification is 6.850 minutes. Setting a destination region slightly increases the average waiting time to 7.506 minutes, whereas setting a destination location increases the average waiting time to 11.879 minutes, and the p-values from t-tests are less than 0.001 for any pair-wise comparison among the three.

We next explore the transitional probability of the usage of direct destination feature and destination region feature. The transitional probability is defined as the proportion of drivers who use the same feature on their next working day among drivers who use a feature at the current working day. I use the transitional probability to measure how sticky the habit is of using each feature. It turns out that the usage of destination region feature is more sticky than the usage of direct destination feature. The average transitional probability of the destination region feature is 65.672% , whereas the average transitional probability of the direct destination feature is 55.678% during the experimental period excluding the first two weeks.

Finally, I find a small but significant substitution effect between the two features. The treatment driver in the trip-completed sample complete an average of 0.488 destination location trips daily, compared with an average of 0.529 direct destination location trips for the control group drivers. The difference is significant with a p-value less than 0.001 under T test.

BIBLIOGRAPHY

Abdulkadiroğlu, A., Che, Y.-K., and Yasuda, Y. (2011). Resolving conflicting preferences in school choice: The "boston mechanism" reconsidered. *The American Economic Review*, 101(1):399–410.

Abdulkadiroglu, A., Pathak, P., Roth, A. E., and Sonmez, T. (2006). Changing the Boston school choice mechanism. Technical report, National Bureau of Economic Research.

Abdulkadiroglu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *The American Economic Review*, 93(3):729–747.

Ai, W., Chen, R., Chen, Y., Mei, Q., and Phillips, W. (2016). Recommending teams promotes prosocial lending in online microfinance. *Proceedings of the National Academy of Sciences*, 113(52):14944–14948.

Ai, W., Chen, Y., Mei, Q., Ye, J., and Zhang, L. (2019a). The gig economy, teams and productivity: A field experiment at a ride-sharing platform. *Working Paper*.

Ai, W., Chen, Y., Mei, Q., Ye, J., and Zhang, L. (2019b). Putting organization into the gig economy: A field experiment at a ride-sharing platform. working paper.

Altonji, J. G. and Paxson, C. H. (1988). Labor supply preferences, hours constraints, and hours-wage trade-offs. *Journal of labor economics*, 6(2):254–276.

Altonji, J. G. and Usui, E. (2007). Work hours, wages, and vacation leave. *ILR Review*, 60(3):408–428.

Andreoni, J. (1993). An experimental test of the public-goods crowding-out hypothesis. *The American Economic Review*, 83(5):1317–1327.

Ariely, D., Ockenfels, A., and Roth, A. E. (2005). An experimental analysis of ending rules in internet auctions. *RAND Journal of Economics*, pages 890–907.

Avery, C., Lee, S., and Roth, A. E. (2019). College admissions as non-price competition: The case of south korea. *Working Paper*.

Baek, J. and Shore, J. (2020). Forum size and content contribution per person: A field experiment. *Management Science*, 66(12):5906–5924.

Balinski, M. and Sönmez, T. (1999). A tale of two mechanisms: student placement. *Journal of Economic Theory*, 84(1):73–94.

Biró, P. (2012). University admission practices-hungary. *www. matching-in-practice. eu accessed July*, 21:2014.

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387 – 424.

Bo, I. and Hakimov, R. (2018). The iterative deferred acceptance mechanism. *Working Paper*.

Böheim, R. and Taylor, M. P. (2004). Actual and preferred working hours. *British Journal of Industrial Relations*, 42(1):149–166.

Braun, S., Dwenger, N., and Kübler, D. (2010). Telling the truth may not pay off: An empirical study of centralized university admissions in germany. *The BE Journal of Economic Analysis & Policy*, 10(1).

Bryan, M. L. (2007). Free to choose? differences in the hours determination of constrained and unconstrained workers. *Oxford Economic Papers*, 59(2):226–252.

Budish, E., Cramton, P., and Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621.

Butler, B. (2001). Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. *Information Systems Research*, 12(4):346–362.

Bó, I. and Hakimov, R. (2019). Iterative versus standard deferred acceptance: Experimental evidence. *The Economic Journal*.

Calsamiglia, C. and Güell, M. (2018). Priorities in school choice: The case of the boston mechanism in barcelona. *Journal of Public Economics*, 163:20–36.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained school choice: An experimental study. *The American Economic Review*, 100(4):1860–1874.

Camerer, C., Babcock, L., Loewenstein, G., and Thaler, R. (1997). Labor supply of new york city cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112(2):407–441.

Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, 60(1):31–51.

Castillo, J. C., Knoepfle, D., and Weyl, G. (2017). Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 241–242. ACM.

Chamberlin, J. R. (1978). ACTION : SOME EXPERIMENTAL RESULTS ’. 29.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental economics*, 14(1):47–83.

Chen, M. K., Chevalier, J. A., Rossi, P. E., and Oehlsen, E. (2019). The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*, 127(6):2735–2794.

Chen, M. K. and Sheldon, M. (2016). Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. working paper.

Chen, R. and Chen, Y. (2011). The Potential of Social Identity for Equilibrium Selection. *The American Economic Review*, 101(6):2562–2589.

Chen, R., Chen, Y., Liu, Y., and Mei, Q. (2017). Does team competition increase pro-social lending? Evidence from online microfinance. *Games and Economic Behavior*, 101:311–333.

Chen, Y. and He, Y. (2017). Information acquisition and provision in school choice: an experimental study. Technical report, Working Paper.

Chen, Y., Jiang, M., and Kesten, O. (2015). Chinese College Admissions Reforms: Experimental and Empirical Evaluations. *Working Paper*.

Chen, Y., Jiang, M., Kesten, O., Robin, S., and Zhu, M. (2018a). Matching in the large: An experimental study. *Games and Economic Behavior*, 110:295–317.

Chen, Y. and Kesten, O. (2017). Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy*, 125(1):99–139.

Chen, Y. and Kesten, O. (2019). Chinese college admissions and school choice reforms: An experimental study. *Games and Economic Behavior*, 115:83–100.

Chen, Y. and Sönmez, T. (2002). Improving Efficiency of On-Campus Housing: An Experimental Study. *American Economic Review*, 92(5):1669–1686.

Chen, Y. and Sönmez, T. (2006). School choice: an experimental study. *Journal of Economic Theory*, 127(1):202–231.

Chen, Y., YeckehZaare, I., and Zhang, A. F. (2018b). Real or bogus: Predicting susceptibility to phishing with economic experiments. *PLoS ONE*, 13(6).

Cook, C., Diamond, R., Hall, J., List, J. A., and Oyer, P. (2018). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Technical report, National Bureau of Economic Research.

Crawford, V. P. and Meng, J. (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review*, 101(5):1912–32.

Dabbish, L., Kraut, R., and Patton, J. (2012). Communication and commitment in an online game team. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 879–888.

Diederich, J., Goeschl, T., and Waichman, I. (2016). Group size and the (in) efficiency of pure public good provision. *European Economic Review*, 85:272–287.

Drago, R., Wooden, M., and Black, D. (2009). Who wants and gets flexibility? changing work hours preferences and life events. *ILR Review*, 62(3):394–414.

Dubins, L. E. and Freedman, D. A. (1981). Machiavelli and the Gale-Shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494.

Dur, U., Hammond, R. G., and Morrill, T. (2018). Identifying the harm of manipulable school-choice mechanisms. *American Economic Journal: Economic Policy*, 10(1):187–213.

Echenique, F., Wilson, A. J., and Yariv, L. (2016). Clearinghouses for two-sided matching: An experimental study. *Quantitative Economics*, 7(2):449–482.

Ehlers, L. and Massó, J. (2007). Incomplete information and singleton cores in matching markets. *Journal of Economic Theory*, 136(1):587–600.

Ergin, H. and Sönmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics*, 90(1):215–237.

Ergin, H. I. (2002). Efficient resource allocation on the basis of priorities. *Econometrica*, 70(6):2489–2497.

Euwals, R. (2001). Female labour supply, flexibility of working hours, and job mobility. *The Economic Journal*, 111(471):120–134.

Farber, H. S. (2005). Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of political Economy*, 113(1):46–82.

Farber, H. S. (2015). Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *The Quarterly Journal of Economics*, 130(4):1975–2026.

Farrell, D. and Greig, F. (2016). Paychecks, paydays, and the online platform economy. In *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, volume 109, pages 1–40. JSTOR.

Fayol, M., Largy, P., and Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology*, 47(2):437–464.

Featherstone, C. R. and Niederle, M. (2016). Boston versus deferred acceptance in an interim setting: An experimental investigation. *Games and Economic Behavior*, 100:353–375.

Fenig, G., Gallipoli, G., Halevy, Y., and Others (2018). Piercing the Payo Function Veil: Tracing Beliefs and Motives. Technical report.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fischbacher, U. and Gächter, S. (2010). Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *The American Economic Review*, 100(1):541–556.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Ge, Y., Knittel, C. R., MacKenzie, D., and Zoepf, S. (2016). Racial and gender discrimination in transportation network companies. Technical report, National Bureau of Economic Research.

Gielen, A. C. (2008). Working hours flexibility and order workers' labor supply. *Oxford economic papers*, 61(2):240–274.

Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2002). Quantal Response Equilibrium and Overbidding in Private-Value Auctions. *Journal of Economic Theory*, 104(1):247–272.

Guttman, J. M. (1986). Matching Behavior and Collective Action: Some Experimental Evidence. *Journal of Economic Behavior & Organization*, 7(2):171–198.

Hakimov, R. and Kübler, D. (2021). Experiments on centralized school choice and college admissions: a survey. *Experimental Economics*, 24(2):434–488.

Hall, J. V., Horton, J. J., and Knoepfle, D. T. (2017). Labor market equilibration: Evidence from uber. *URL http://john-joseph-horton. com/papers/uber_price. pdf, working paper*.

Hall, J. V. and Krueger, A. B. (2018). An analysis of the labor market for uber's driver-partners in the united states. *ILR Review*, 71(3):705–732.

Halperin, B., Ho, B., List, J. A., and Muir, I. (2019). Toward an understanding of the economics of apologies: evidence from a large-scale natural field experiment. Technical report, National Bureau of Economic Research.

Hassidim, A., Marciano, D., Romm, A., and Shorrer, R. I. (2017). The mechanism is truthful, why aren't you? *American Economic Review*, 107(5):220–24.

Hassidim, A., Romm, A., and Shorrer, R. I. (2016). " strategic" behavior in a strategy-proof environment. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 763–764.

Hill, E. J., Erickson, J. J., Holmes, E. K., and Ferris, M. (2010). Workplace flexibility, work hours, and work-life conflict: finding an extra day or two. *Journal of Family Psychology*, 24(3):349.

Hill, E. J., Hawkins, A. J., Ferris, M., and Weitzman, M. (2001). Finding an extra day a week: The positive influence of perceived job flexibility on work and family life balance. *Family relations*, 50(1):49–58.

Holt, C. A. and Laury, S. K. (2002). Risk Aversion and Incentive Effects. *The American Economic Review*, 92(5):1644–1655.

Isaac, R. M. and Walker, J. M. (1988). Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. *Quarterly Journal of Economics*, 103(1):179–200.

Isaac, R. M., Walker, J. M., and Williams, A. W. (1994). Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups. *Journal of Public Economics*, 54(1):1–36.

Jiang, M. (2014). When do stable matching mechanisms fail? the role of standardized tests in college admissions. *Working Paper*.

Kamada, Y. and Kandori, M. (2011). Asynchronous revision games. *Working Paper*.

Karau, S. J. and Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4):681.

Katz, L. F. and Krueger, A. B. (2019). The rise and nature of alternative work arrangements in the united states, 1995-2015. *ILR Review*, 72(2):382–416.

King, A. G. (1978). Industrial structure, the flexibility of working hours, and women's labor force participation. *The Review of Economics and Statistics*, pages 399–407.

Klijn, F., Pais, J., and Vorsatz, M. (2013). Preference intensities and risk aversion in school choice: a laboratory experiment. *Experimental Economics*, 16(1):1–22.

Klijn, F., Pais, J., and Vorsatz, M. (2018). Static versus dynamic deferred acceptance in school choice: Theory and experiment. *Games and Economic Behavior*.

Kojima, F. and Ünver, M. U. (2014). The "boston" school-choice mechanism: an axiomatic approach. *Economic Theory*, 55(3):515–544.

Kübler, D. (2012). University admission practices—germany.

Laury, S. K. and Holt, C. A. (2008). Chapter 84 voluntary provision of public goods: Experimental results with interior nash equilibria. volume 1 of *Handbook of Experimental Economics Results*, pages 792–801. Elsevier.

Laury, S. K., Walker, J. M., and Williams, A. W. (1999). The voluntary provision of a pure public good with diminishing marginal returns. *Public Choice*, 99(1):139–160.

Ledyard, J. (1995). Public goods: A survey of experimental research. In Kagel, J. H. and Roth, A. E., editors, *The Handbook of Experimental Economics*, volume 1. Princeton University Press, Princeton, New Jersey.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–87.

Lien, J. W., Zheng, J., and Zhong, X. (2016). Preference submission timing in school choice matching: testing fairness and efficiency in the laboratory. *Experimental Economics*, 19(1):116–150.

Lien, J. W., Zheng, J., and Zhong, X. (2017). Ex-ante fairness in the boston and serial dictatorship mechanisms under pre-exam and post-exam preference submission. *Games and Economic Behavior*, 101:98–120.

Liu, M., Brynjolfsson, E., and Dowlatabadi, J. (2018). Do digital platforms reduce moral hazard? the case of uber and taxis. Technical report, National Bureau of Economic Research.

Liu, T., Wan, Z., and Yang, C. (2019). The efficiency of a dynamic decentralized two-sided matching market. working paper.

Marwell, G. and Ames, R. E. (1979). Experiments on the provision of public good I: Resources, interest, group size, and the free-rider problem. *American Journal of Sociology*, 84:1336–1360.

Monderer, D. and Shapley, L. S. (1996). Potential Games. *Games and Economic Behavior*, 14:124–143.

Nosenzo, D., Quercia, S., and Sefton, M. (2015). Cooperation in small groups: the effect of group size. *Experimental Economics*, 18(1):4–14.

Ockenfels, A. and Roth, A. E. (2006). Late and multiple bidding in second price internet auctions: Theory and evidence concerning different rules for ending an auction. *Games and Economic behavior*, 55(2):297–320.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.

Pais, J. and Pintér, Á. (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior*, 64(1):303–328.

Pan, S. (2019). The instability of matching with overconfident agents. *Games and Economic Behavior*, 113:396–415.

Pathak, P. A. and Sönmez, T. (2013). School admissions reform in chicago and england: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review*, 103(1):80–106.

Pereda, M., Capraro, V., and Sánchez, A. (2019). Group size effects and critical mass in public goods games. (March):1–10.

Presser, H. B. (1995). Job, family, and gender: Determinants of nonstandard work schedules among employed americans in 1991. *Demography*, 32(4):577–598.

Ravenelle, A. J. (2019). *Hustle and gig: struggling and surviving in the sharing economy*. Univ of California Press.

Rebitzer, J. B. and Taylor, L. J. (1995). Do labor markets provide enough short-hour jobs? an analysis of work hours and work incentives. *Economic Inquiry*, 33(2):257–273.

Rees-Jones, A. (2018). Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *Games and Economic Behavior*, 108:317–330.

Rees-Jones, A. and Skowronek, S. (2018). An experimental investigation of preference misrepresentation in the residency match. *Proceedings of the National Academy of Sciences*, 115(45):11471–11476.

Reynolds, J. (2004). When too much is not enough: Actual and preferred work hours in the united states and abroad. *Sociological Forum*, 19(1):89–120.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628.

Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of political Economy*, 92(6):991–1016.

Roth, A. E. and Ockenfels, A. (2002). Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon auctions on the internet. *American economic review*, 92(4):1093–1103.

Roth, A. E. and Peranson, E. (1997). The Effects of the Change in the NRMP Matching Algorithm. *JAMA: The Journal of the American Medical Association*, 278(9):729–732.

Roth, A. E. and Peranson, E. (1999). The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American economic review*, 89(4):748–780.

Roth, A. E., Sönmez, T., et al. (2005). A kidney exchange clearinghouse in new england. *American Economic Review*, 95(2):376–380.

Roth, A. E., Sönmez, T., and Ünver, M. U. (2004). Kidney exchange. *The Quarterly Journal of Economics*, 119(2):457–488.

Senesky, S. (2005). Testing the intertemporal labor supply model: are jobs important? *Labour Economics*, 12(6):749–772.

Shank, D. B., Kashima, Y., Saber, S., Gale, T., and Kirley, M. (2015). Dilemma of dilemmas: how collective and individual perspectives can clarify the size dilemma in voluntary linear public goods dilemmas. *PloS one*, 10(3):e0120379.

Shorrer, R. I. and Sóvágó, S. (2018). Obvious mistakes in a strategically simple college admissions environment: Causes and consequences. *Available at SSRN 2993538*.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, 1(1):161–176.

Smith, A. (2016). Gig work, online selling and home sharing. *Pew Research Center*, November.

Stephenson, D. G. (2016). Continuous feedback in school choice mechanisms. *Working Paper*.

Stewart, M. B. and Swaffield, J. K. (1997). Constraints on the desired hours of work of british men. *The Economic Journal*, 107(441):520–535.

Sweeney, J. W. (1974). Altruism, the free rider problem and group size. *Theory and decision*, 4(3-4):259–275.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.

Ui, T. (2000). A shapley value representation of potential games. *Games and Economic Behavior*, 31(1):121 – 135.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80:234–248.

Weimann, J., Brosig-Koch, J., Heinrich, T., Hennig-Schmidt, H., and Keser, C. (2019). Public good provision by large groups–the logic of collective action revisited. *European Economic Review*, 118:348–363.

Wu, B. and Zhong, X. (2014). Matching mechanisms and matching quality: Evidence from a top university in china. *Games and Economic Behavior*, 84:196–215.

Yang, W., Liu, W., Viña, A., Tuanmu, M.-N., He, G., Dietz, T., and Liu, J. (2013). Nonlinear effects of group size on collective action and resource outcomes. *Proceedings of the National Academy of Sciences*, 110(27):10916–10921.

Zhang, X. M. and Zhu, F. (2011). Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review*, 101(4):1601–1615.