Essays on Strategic Behavior and Equilibrium Selection in Two-Sided Matching Markets

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Ahrash Dianat
Master of Arts
George Mason University, 2011
Bachelor of Science
Rochester Institute of Technology, 2009

Director: Marco Castillo, Associate Professor
Department of Economics

Summer Semester 2015
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my loving parents, Sohail and Mitra.

ACKNOWLEDGEMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen. My advisor, Dr. Castillo, and the other members of my committee were of invaluable help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

ESSAYS ON STRATEGIC BEHAVIOR AND EQUILIBRIUM SELECTION IN TWO-SIDED MATCHING MARKETS

Ahrash Dianat, Ph.D.

George Mason University, 2015

Dissertation Director: Dr. Marco Castillo

This dissertation combines game theory with controlled laboratory experimentation to better understand the performance of two-sided matching markets. These markets are often organized as centralized clearinghouses, in which participants submit rank-order lists of their preferences to a central authority and then a particular algorithm determines the final outcome (i.e., who is paired with whom). We address two main questions. First, do market participants strategically misrepresent their preferences in these environments? Second, in markets with multiple equilibria, which equilibrium is more likely to be implemented?

CHAPTER 1

1 Introduction

We investigate how information about other agents' preferences affects strategic behavior in a dynamic version of the Gale-Shapley deferred acceptance algorithm.[1] The DA algorithm not only has important empirical applications, but it also offers an ideal environment to test how information affects behavior in markets.[2] Indeed, the incentives in the mechanism are such that one side of the market should ignore information about other agents' preferences while the other side of the market should use this information to their advantage.

In the algorithm, "proposers" have a dominant strategy of truth-telling and therefore this information is of no value. "Responders", on the other hand, often have room for strategic preference misrepresentation. In fact, information about other agents' preferences is a necessary condition for responders to calculate and coordinate on their optimal strategies. There is also an inherent tension between the interests of the two sides of the market: the best stable matching for the proposers is the worst stable matching for the responders, and vice versa.[3] The ability of responders to behave strategically by making use of the information available to them determines which particular stable matching arises. One should then predict that final outcomes will be more favorable to responders in the complete information environment.

[1]Henceforth, DA algorithm.

[2]For a seminal paper relating to school choice, see Abdulkadiroğlu and Sönmez (2003). For an application to entry-level labor markets, see Roth and Peranson (1999).

[3]A matching is said to be stable if no agent prefers remaining unmatched to her current allocation and no pair of agents mutually prefer each other to their current allocations.

To that end, we test how information affects the behavior of each side of the market and whether this is congruent with theoretical predictions. To do this, we implement a laboratory experiment in which information about other subjects' preferences is either fully available, partially available, or not available at all. In our baseline, subjects take part in an extensive form version of the DA mechanism under a variety of preference profiles and with complete information about others' preferences. Our treatments manipulate the information available to agents. In the extreme case, each subject only observes their own preference list; they are unaware of both the preferences of others as well as the underlying distribution from which those preferences are drawn. In the intermediate cases, in addition to their own preferences, subjects are either able to observe the preferences of all agents on their side of the market or of all agents on the opposing side of the market.

We find that the information structure has implications for the selection of final outcomes, but does not affect their stability. In particular, the average distance to the responders' preferred stable outcome is smallest in the complete information treatment. However, this result cannot be attributed to responders taking advantange of their profitable strategic opportunities. In fact, responders generally behave straightforwardly in all information treatments and accept the best offer that is available to them at any given time. The strategic preference misrepresentation that does occur involves subjects "truncating" their preferences and rejecting offers from very low-ranked match partners. Proposers, on the other hand, generally fail to play their dominant strategy in the complete information treatment and instead skip down their preference lists when making offers. This "skipping" behavior is sophisticated in the sense that proposers take into account how they are ranked by the other side of the market, forgoing offers to preferred match partners who are unlikely to accept. In line with this theory, the welfare loss from proposers' skipping behavior is modest.

Our results highlight the important role that behavioral biases can play in these environments. Understanding empirical regularities in participants' behavior is particularly crucial given that the interests of the two sides of the market are opposed on the question of equilibrium selection. Consider the National Resident Matching Program (NRMP), the entry-level job market for American physicians. The NRMP is organized as a centralized clearinghouse based on an algorithm that is very similar to the student-proposing DA algorithm. If medical students regularly engage in "skipping" behavior (e.g., by applying only to mid-ranked and low-ranked residency programs), then the aggregate effect is to push the final outcome further away from the student-optimal stable matching. This is an important consideration for policy-makers, who may have reasons to favor the welfare of one side of the market over another when designing matching mechanisms.

There is a large body of experimental work on two-sided matching (e.g. Chen and Sönmez, 2006; Ding and Schotter; Featherstone and Mayefsky, 2014; Featherstone and Niederle, 2014; Harrison and McCabe, 1989; Haruvy and Ünver, 2007; Pais and Pintér, 2008). However, this literature has almost exclusively studied the properties of static matching mechanisms, in which the strategy choice faced by laboratory subjects is which preference ordering to submit to the mechanism. We break from that tradition and employ a dynamic design where subjects are instead required to walk through the steps of the DA algorithm. Our work is inspired by and can be viewed as a natural extension of Echenique, Wilson, and Yariv (2014), who also study the performance of the DA mechanism using a dynamic implementation. In their experiment, they also report "skipping" behavior by proposers and straightforward play by responders.

However, our main contribution stems from our novel experimental design. By introducing treatments that alter the information available to subjects in a systematic way, our experimental design allows us to more carefully tease apart the driving forces behind subjects' behavior. In particular, we can analyze the merits of competing behavioral theories by observing their predictive power in environments where they should apply (i.e., when the relevant information is provided to subjects) and observing their lack of predictive power in environments where they should not apply (i.e., when the relevant information is not provided to subjects).

For motivation, consider the following hypothetical game discussed by Roth and Sotomayor (1992):

1. Actions in the market are organized in stages. Each stage is divided into two periods. Within each period, each man and woman must make decisions without knowing the decisions of other men and women in that period.

2. At the first period of the first stage, each man may make at most one proposal to any woman he chooses. (He is also free to make no proposals.) Proposals can only be made by men.

3. In the second period of the first stage, each woman who has received any proposals is free to reject any or all of them immediately. A woman may also keep at most one man "engaged" by not rejecting his proposal.

4. In the first period of any stage, any man who was rejected in the preceding stage may make at most one proposal to any woman he has not previously proposed to (and been rejected by). In the second period, each woman may reject any or all of these proposals, including that of any man who has proposed in an earlier stage and been kept engaged. A woman may keep at most one man engaged by not rejecting his proposal.

5. If, at the beginning of any stage, no man makes a proposal, then the market ends, and each man is matched to the woman he is engaged to. Men who are not

engaged to any woman, and women who are not engaged to any man, remain single.

Running the algorithm in real time (as described) has drawbacks but also confers several important advantages. Echenique et al. (2014) highlight many of these points but we also include them here for the sake of completeness.

The main drawback of our design is the lack of ecological validity.[4] In field settings such as school choice and the National Resident Matching Program (NRMP), the DA algorithm is implemented as a static mechanism. Requiring subjects to play the preference revelation game in the lab thus allows researchers to more credibly use experimental results to inform policy-making and institutional design in the real world. Another drawback concerns not having access to subjects' strategy choices. Modeling the DA mechanism as an extensive form game necessarily obscures our understanding of the subjects' strategies since behavior off the path of play is never observed.

However, our methodological choice also comes with several important advantages. First, in a static mechanism, it is unlikely that laboratory subjects can fully understand how the profile of stated preferences maps to the final matching computed by the algorithm. One of the advantages of our design involves transparency. Requiring that subjects walk through the steps of the procedure for themselves and make decisions along the way allows for a more concrete understanding of the strategic aspects of the game and the trade-offs associated with different strategy choices.[5] Another

[4]However, a dynamic version of the deferred acceptance procedure was used for some time in the entry-level job market for clinical psychologists. For details on the operation and evolution of that market, see Roth and Xing (1997).

[5]The ability of subjects to understand the connection between their actions and their payoffs can have profound implications for strategic behavior in the lab. The experimental literature on auctions confirms the importance of transparency in mechanism design settings by showing how subject behavior can systematically vary in theoretically equivalent institutions (Kagel, Harstad, and Levin, 1987).

argument for our design is based on the desire to reduce experimenter demand effects. Since the subjects' preferences are induced and they are literally handed preference lists by the experimenter before each round, merely asking the subjects to then provide preference lists as an input to a mechanism runs the risk of "giving the game away" or nudging them toward a certain behavior. In particular, subjects might infer that the experiment is testing whether or not they will report their preferences truthfully.

The argument for using laboratory experiments to test the performance of matching markets is compelling. In the field, individuals' true preferences are not observable and thus it is unclear to what extent agents are behaving strategically. As a consequence, while we can conjecture that the observed final outcomes are stable, it is unclear which particular stable matching is being implemented in markets with a multiplicity of stable outcomes. In addition, it is difficult in field settings to know precisely what information is common knowledge among market participants. The laboratory setting allows us to control for these features and to arrive at more robust conclusions.

2 Theoretical Background

Apart from being widely used in real-life allocation problems, the DA algorithm has many appealing theoretical properties. It was originally conceived as a constructive proof of the existence of a stable matching for any one-to-one matching market (Gale and Shapley, 1962). However, since then the non-cooperative game induced by the rules of the DA algorithm has received independent interest. The background provided here relates to the static implementation of the DA algorithm.[6] Throughout

[6]Echenique et al. (2014) show that the inclusion of two simple restrictions on strategies (stationarity and the congruence axiom) is sufficient to make the static and dynamic DA mechanisms

this analysis, we use the original convention from the classic "marriage" market of Gale and Shapley (1962).

We divide the market into two finite, disjoint sets M and W: $M = \{m_1, m_2, ..., m_n\}$ is the set of men and $W = \{w_1, w_2, ..., w_r\}$ is the set of women. Each agent has preferences over the agents on the other side of the market (as well as remaining single. The preferences of man m will be represented by an ordered list of preferences $P(m)$ on the set $W \cup \{m\}$. Similarly, the preferences of woman w will be represented by an ordered list of preferences $P(w)$ on the set $M \cup \{w\}$. For instance, the preferences of man m might be

$$P(m) = w_2, w_1, m, w_3, ..., w_r,$$

indicating that his first choice is to be matched with w_2, his second choice is to be matched with w_1, and his third choice is to remain single. This preference list can also be expressed more concisely as follows:

$$P(m) = w_2, w_1$$

where only the "acceptable" matches are listed (those individuals who are above the reservation option of remaining unmatched). Let $\mathbf{P}$ denote the set of all preferences, one for each man and one for each woman. A matching μ is a one-to-one correspondence from the set $M \cup W$ onto itself of order two (that is, $\mu^2(x) = x$) such that if $\mu(m) \neq m$ then $\mu(m) \in W$ and if $\mu(w) \neq w$ then $\mu(w) \in M$. The intuition behind the order two requirement ($\mu^2(x) = x$) is that if man m is matched to woman w, then woman w is also matched to man m. Note that the definition also forces agents who

strategically equivalent.

are not single (matched to themselves) to be matched with a member of the opposite set.

An individual m is said to block a matching μ if he prefers remaining single rather than being matched to $\mu(m)$. A pair of agents (m, w) is said to block a matching μ if they are not matched to one another at μ but they prefer each other to their assignments at μ. A matching μ is said to be stable if it is not blocked by any individual or any pair of agents. A stable matching μ is called an M-optimal stable matching (denoted μ_M) if every man likes it at least as well as any other stable matching. A W-optimal stable matching can be defined analogously (denoted μ_W). In the lattice of stable matchings, the M-optimal stable matching is thus the "best" stable matching for the men and the W-optimal stable matching is the "best" stable matching for the women. Gale and Shapley (1962) proved the following result:

- **Theorem 1**: A stable matching exists for every marriage market. Furthermore, when all men and women have strict preferences, there always exist an M-optimal stable matching and a W-optimal stable matching.

To examine the strategic issues involved, we analyze the revelation game in which each man m with preferences $P(m)$ is faced with the strategy choice of what preference ordering $Q(m)$ to state, and likewise for the women. Denote the set of stated preference lists, one for each man and one for each woman, by $\mathbf{Q}$. The mechanism then computes a matching $\mu = h(\mathbf{Q})$, where h is the function that maps any set $\mathbf{Q}$ of stated preferences into a matching. A mechanism h that for any stated preferences $\mathbf{Q}$ produces a matching $h(\mathbf{Q})$ that is stable with respect to the submitted preferences is called a stable mechanism. If $h(\mathbf{Q})$ produces the M-optimal stable matching with respect to $\mathbf{Q}$, then h is called the M-optimal stable mechanism. Roth (1982) proved an important negative result:

- **Theorem 2**: When there are at least two agents on each side of the market, no stable matching mechanism exists which always makes stating the true preferences a dominant strategy for every agent.

However, Roth (1982) also showed that it is often possible to arrange the market such that only one side faces difficult strategic questions.

- **Theorem 3**: The M-optimal stable mechanism makes it a dominant strategy for each man to state his true preferences.

Combining these results suggests that, under the M-optimal stable mechanism, it is the women who will sometimes have a profitable deviation by misrepresenting their true preferences. In particular, the women will generally have an incentive to follow a "truncation" strategy in which they submit a preference list that ranks the men in the same order as their true preference list but that leaves off all individuals below a certain threshold (even though being matched to those individuals would still be preferable to remaining unmatched). One focal Nash equilibrium of this matching game involves the men truthfully reporting their preferences and each woman w truncating her true preference list by leaving off all individuals below $\mu_W(w)$, her W-optimal stable match partner. In this way, the final outcome of the mechanism will actually be the W-optimal stable matching. More generally, Roth (1984) proved that any stable matching with respect to agents' true preferences can be achieved by a set of strategies that forms a Nash equilibrium in the revelation game induced by the M-optimal stable mechanism.

The behaviors discussed above have natural analogues in our dynamic design. In the static DA algorithm, both proposers and responders are regarded as "truthful" if they submit their actual preference lists to the centralized mechanism. In the dynamic implementation, the equivalent of this truth-telling strategy involves different heuristics for the two sides of the market.[7] A truthful proposer would make offers by straightforwardly moving down his preference list (from his first choice, to his second choice, to his third choice, and so on). Alternatively, since any potential match partner is above the reservation value of remaining unmatched in our experiment (\$0), a truthful responder would simply accept the best offer among all the available offers at any given point in time. Truncation would then be observed whenever an unmatched responder rejects an offer (thereby pretending that the offer is below their reservation value).

Our formulation is, of course, merely one approach among the full spectrum of incomplete information models that could be considered. You could imagine a matching environment where one side of the market is not even fully aware of their own preferences. For instance, a firm interviewing a potential employee may only have a signal of the applicant's quality based on the interview. The true quality would only be revealed ex-post once the employee begins work. This situation is analogous to the decision problem a bidder faces in a common value auction. Incomplete information about others' preferences could also be modeled more explicitly by allowing the distribution of preferences to be common knowledge and solving for the Bayesian-Nash equilibrium of the resulting non-cooperative matching game. This approach has been analyzed by Roth (1989). He shows that, although results concerning dominant and dominated strategies do generalize to the incomplete information model, the results concerning Nash equilibria do not. In a seminal paper, Roth and Rothblum (1999) discuss the issue of what practical advice can be given to market participants in a DA setting in the field. They show that the scope of profitable preference misrepresentation is greatly reduced in situations involving strategic uncertainty or incomplete

[7]Echenique et al. (2014) refer to truthful behavior in the dynamic DA mechanism as "straightforward" behavior.

information. In particular, any non-truncation strategy is stochastically dominated by a truncation strategy under quite general conditions on responders' beliefs. This insight has obvious consequences for predicting the behavior of responders in our experimental design.

A final point needs to be addressed: namely, what precisely constitutes stability in incomplete information matching environments?[8] The convention to be used in this paper is that of ex-post stability. After the dynamics of the process have played out, a final matching arises. This matching will be considered stable if, conditional on revealing all information about others' preferences, the matching meets the definition of complete information stability. That is, having preferences become common knowledge should not allow for the formation of a blocking pair. Given the information needed for subjects to compute Nash equilibrium behavior, the choice of modeling technique for incomplete information has substantive behavioral implications.

3 Experimental Design

Experimental subjects were drawn from the population of undergraduate students at George Mason University. A total of 176 students participated in the experiment. There were 16 subjects in each experimental session. Upon arriving at the lab, the subjects were randomly and equally divided into two disjoint groups: foods and colors.[9] There were thus 8 foods and 8 colors in each session. After the subjects signed the consent form, experimental instructions were handed out and subjects were given 10-15 minutes to privately read the instructions.[10] Once that time had

[8] For instance, Liu, Mailath, Postlewaite, and Samuelson (2014) have recently introduced a characterization of stability for incomplete information matching markets that is similar in spirit to the game-theoretic notion of rationalizability.

[9] These neutral labels were borrowed from the methodologically similar work of Echenique et al. (2014).

[10] The experimental instructions for all four treatments are included at the end of the paper.

elapsed, the experimenter re-entered the room to highlight the key elements of the instructions and answer any questions.

The subjects were then given a short quiz to test their understanding of the experiment. The quiz was incentivized: if a subject answered all questions correctly, \$3 was added to their final payment.[11] The experimenter and a lab assistant then went around the room and graded each subject's quiz in private. The subjects were informed of whether they received the \$3 bonus (by answering all the questions correctly) or not. The experimenter then publicly went over the solutions to the quiz questions in order to guarantee that everyone knew the correct answers prior to the start of the experiment. A short post-experiment survey was administered at the end of the session, with a \$2 bonus if subjects took the time to fill it out. All subjects completed the survey. The survey included basic demographic and educational questions in addition to questions intended to elicit feedback about the experiment itself.

Before the start of each round, the subjects were given a sheet of paper with the relevant payoff information. Appendix A contains an example of the payoff information that food APPLE would observe in all of the different treatments for the same market. When the payoffs are presented in a matrix, the first number in each cell corresponds to the monetary payoff of the "row" player (the food) and the second number in each cell corresponds to the monetary payoff of the "column" player (the color). In the incomplete information and incomplete-proposer treatments, subjects were told that just because they would earn \$x from a particular match does not imply that their match partner would earn the same amount: in fact, the partner could earn either less than \$x, exactly \$x, or more than \$x.

[11] Prior to handing out the instructions, subjects were informed of both the quiz as well as the incentive to answer questions correctly.

At the end of the experiment, one of the rounds was randomly chosen and subjects were paid according to their final match partners in that round. Each session lasted approximately two and one half hours and each subject automatically received $7 as a show-up payment. Including the show-up payment, subject earnings ranged from between $15 to $36. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

In the actual experiment, subjects went through the steps of the hypothetical game discussed earlier (Roth and Sotomayor, 1992). The foods were the "proposers" and the colors were the "responders" in the matching game. At the start of each market, all the foods and colors were unmatched. The market was organized in stages where foods and colors took turns in making decisions. In each stage, unmatched foods were required to make one offer to a color they had not previously made an offer to. Colors then viewed all the offers they had received in that stage and they were allowed to tenatively accept at most one of those offers (including any tentatively accepted offer they were still holding). The game ends when there are no foods left to make offers. This can happen because (1) all foods are matched or because (2) the only unmatched foods have already been rejected by all of the colors. The tentative matches that are in place when the game ends become the final matches for that market. Sample screens from the experimental interface are shown in Appendix B.

Each experimental session consisted of between 5-9 markets.[12] Each round was a separate matching market in which subjects played the extensive form version of the DA mechanism outlined earlier. Both the roles of the subjects (food vs. color) and their particular identities (say, food **APPLE** or color **BLACK**) were fixed throughout the experiment. However, subjects did not play the same matching market more

[12]Since the DA mechanism is implemented synchronously, there is substantial variation in the amount of time it takes for an experimental market to converge to a final outcome.

than once. Nine different preference profiles were used throughout the experiment. Two of these correspond to markets with a unique stable outcome characterized by positive assortative matching. In the seven other markets, there are either two or five economy-wide stable matchings. The economy-wide stable matchings are not disjoint: in all markets, each individual has either one, two, or three stable match partners.

Appendix C provides more details about the characteristics of the markets that were tested. Many of the markets obtain a multiplicity of stable outcomes through minor variations. In addition, proposers' stable match partners are often clustered near the top of their preference lists. These features are appealing in the sense that they effectively stack the deck against stability: seemingly inconsequential "skipping" behavior by proposers can have a pronounced effect and eliminate the possibility of observing stable outcomes. In several of the markets, responders also have a stable match partner near the bottom of their preference lists. In a similar spirit, seemingly harmless truncation by responders can in fact prevent certain stable outcomes from emerging.

To illustrate, consider a hypothetical matching game with four firms (f_1, f_2, f_3, f_4) and four workers (w_1, w_2, w_3, w_4) and the following preferences:

$$P(f_1) = w_1^*, w_2^*, w_3, w_4 \qquad P(w_1) = f_2^*, f_3, f_4, f_1^*$$
$$P(f_2) = w_2^*, w_1^*, w_3, w_4 \qquad P(w_2) = f_1^*, f_3, f_4, f_2^*$$
$$P(f_3) = w_3^*, w_4^*, w_1, w_2 \qquad P(w_3) = f_4^*, f_1, f_2, f_3^*$$
$$P(f_4) = w_4^*, w_3^*, w_1, w_2 \qquad P(w_4) = f_3^*, f_1, f_2, f_4^*$$

An agent's achievable match partners are denoted by asterisks. In this example, *any* skipping behavior by a firm is costly since it prevents the firm-optimal stable matching from being realized. In addition, minimal truncation behavior (i.e., rejecting the offer

Table 1: A total of 11 experimental sessions were conducted. For each session, the information structure, the number of subjects, the number of markets that subjects played, and the order in which the markets were presented to subjects is shown above.

Session	No. Subjects	No. Markets	Order of Markets
Complete Information 1	16	7	H,A,I,F,B,G,D
Complete Information 2	16	7	A,B,C,D,F,G,E
Complete Information 3	16	8	A,B,C,D,E,F,G,H
Incomplete Information 1	16	7	A,B,C,D,E,F,G
Incomplete Information 2	16	9	A,B,C,D,E,F,G,H,I
Inc. Information-Proposer 1	16	5	A,B,C,F,H
Inc. Information-Proposer 2	16	7	A,B,C,D,E,F,H
Inc. Information-Proposer 3	16	8	A,B,C,D,E,F,G,H
Inc. Information-Responder 1	16	8	A,B,C,D,E,F,G,H
Inc. Information-Responder 2	16	8	A,B,C,D,E,F,G,H
Inc. Information-Responder 3	16	8	A,B,C,D,E,F,G,H

from the fourth-ranked firm) will bring about the worker-optimal stable matching. Table 1 provides a summary of the details of each experimental session.

For the theoretical results pertaining to the complete information environment, only the ordinal preferences of the subjects play a role. However, the need to pay experimental subjects forces us to impose a particular cardinal structure on preferences. In our experiment, the payoff differential between a subject's nth and $(n+1)st$ choice match partners is fixed at \$3. We also use the convention that remaining unmatched results in a payoff of \$0. Since other experimental work has shown that altering the cardinal preference structure can substantially affect strategic behavior and final market outcomes, we opted to keep the cardinal representation fixed throughout all the markets that were tested (Echenique et al., 2014).

The baseline treatment is a replication of the environment found in Echenique et al. (2014): subjects participate in the dynamic DA mechanism with complete

Table 2: The experimental treatments are shown below.

Treatment	Description
Complete	The preferences of all agents are common knowledge (each agent can see 16 preference lists).
Incomplete	The preferences of all agents are private information (each agent can only see their own preference list).
Incomplete-Proposer	Proposers can see the preferences of all agents on their side of the market (8 preference lists). Responders have complete information (16 preference lists).
Incomplete-Responder	Proposers can see their own preferences as well as the preferences of all agents on the opposite side of the market (9 preference lists). Responders have complete information (16 preference lists).

Table 3: The experimental hypotheses are shown for all information treatments.

Session	Proposer Behavior	Responder Behavior	Stability	Selection
Complete Information	truth-telling	truncation	Y	responder-optimal
Incomplete Information	truth-telling	truth-telling	Y	proposer-optimal
Incomplete-Proposer	truth-telling	truncation	Y	responder-optimal
Incomplete-Responder	truth-telling	truncation	Y	responder-optimal

information. Our manipulation involves the information structure. We employ a between-subjects design in which each subject participates in only one of four information environments. More details on the information structures that subjects face is given in Table 2. The experimental instructions are also provided at the end of the paper.

4 Hypotheses

We list our experimental hypotheses below. They are also summarized in Table 3. In all cases, the null hypothesis corresponds to strictly rational behavior (i.e., the null hypothesis predicts that subjects will play their optimal equilibrium strategies in the treatments where they have the necessary information to do so).

- **Hypothesis 1**: The level of information will not affect proposer behavior. Proposers will play their dominant strategy of truth-telling in all four information structures.

- **Hypothesis 2**: Responders will truncate their preference lists when they have complete information about the preferences of other agents. In the incomplete information treatment, responders will behave truthfully and not strategically reject offers.

- **Corollary 1**: In terms of final outcomes, stability will be unaffected by the level of information that is available to subjects.

- **Corollary 2**: In terms of final outcomes, the selection of stable outcomes will be affected by the level of information that is available to subjects. The proposer-optimal stable matching will emerge in the incomplete information treatment. In all other treatments, final outcomes will be "closer" to the responder-optimal stable matching.

5 Results

Aggregate Outcomes

We begin by analyzing aggregate outcomes. First, stability is not the norm in our experimental markets: 46% (38/82) of the markets culminated in an economy-wide stable matching. Fifty-five percent (12/22) of final outcomes are stable in the complete information treatment, 50% (8/16) in the incomplete information treatment, 25% (5/20) in the incomplete-proposer treatment, and 54% (13/24) in the incomplete-responder treatment. The distribution of stable outcomes is not significantly different by treatment ($\chi^2(3) = 4.9359$, p-value = 0.177).

In order to assess the empirical relevance of stability in our experiments, we need to deal with two issues. First, we would like to know the severity of the observed deviations from stability in our data. To that end, we define a metric for the space of all matchings. This allows us to measure the distance between an observed outcome in our experimental data and a particular stable outcome. This measure is also used to assess the existence of treatment effects at the aggregate level. Let $\mathcal{M}$ denote the set of all matchings and let W denote the set of all workers. Consider an arbitrary matching $\mu \in \mathcal{M}$ and an arbitrary worker $w \in W$. Define $F(\mu(w))$ as the position of $\mu(w)$ in the ordinal preference list of worker w. If w is matched to her most preferred firm at μ, then $F(\mu(w)) = 1$. If w is matched to her least preferred firm at μ, then $F(\mu(w)) = 4$. For simplicity, if w is unmatched we let $F(\mu(w)) = F(w) = 5$. Thus, $|F(\mu(w)) - F(\mu'(w))|$ is the absolute distance in ranking between $\mu(w)$ and $\mu'(w)$ according to the preferences of worker w. We can then define the distance from μ to μ' as the sum of this measure for all the workers in the market. More formally, the distance $d : \mathcal{M} \times \mathcal{M} \longrightarrow \Re_+$ between two matchings μ and μ' is defined as $d(\mu, \mu') = \sum_{w \in W} |F(\mu(w)) - F(\mu'(w))|$. Intuitively, we are defining the distance between two outcomes as the sum of the absolute distance between each worker's match partners at those outcomes (according to the worker's ordinal preferences).

Second, we need to establish the power of our experimental design to detect stability. For this purpose, we simulate the distance to stability for all experimental markets under the assumption of independent and uniformly random behavior by agents. We repeat this exercise 10,000 times. Appendix D shows the distribution of our metric corresponding to this procedure, alongside the distribution from our experimental data.

Table 4 reports the average distances to the nearest stable outcome and to the W-optimal stable outcome across the four treatments. From this, it is clear that (1)

Table 4: Average distance measures across treatments.

Treatment	Avg. Dist. to Stability	Avg. Dist. to W-optimal	# of Markets
Complete Info.	5.3	6.9	22
Incomplete Info.	5.9	13.6	16
Incomplete-Proposer	12.1	14.1	20
Incomplete-Responder	5.2	9.0	24

outcomes are farthest from stability in the incomplete-proposer treatment and that (2) outcomes are closest to the W-optimal stable matching in the complete information and incomplete-responder treatments.

Figure 1 shows the distribution of the distance to the nearest stable outcome by treatment. We conduct a Kruskal Wallis test to assess the presence of treatment effects in distance to stability. We find that the distance to the nearest stable outcome is statistically different between the four treatments at conventional levels ($\chi^2(3) = 7.560$, p-value = 0.0560). To further test for the robustness of this result, we remove the incomplete information treatment from our analysis. We do so because the remaining three treatments have an important similarity: responders have access to the same level of information about other agents' preferences (complete information). This should then give the best chance for the success of the null hypothesis. Restricting attention to these three treatments, we still find that there are statistically significant treatment differences in our measure of distance to stability ($\chi^2(2) = 7.119$, p-value = 0.0285). Hence, we conclude that the level of information about other agents' preferences affects the magnitude of the deviations from stability.

We now turn to equilibrium selection across treatments. Figure 2 shows the distribution of the distance to the W-optimal stable outcome by treatment. This measure is also statistically different between the four treatments at conventional levels (Kruskal-Wallis test, p-value = 0.0957). Since responders can engage in meaningful

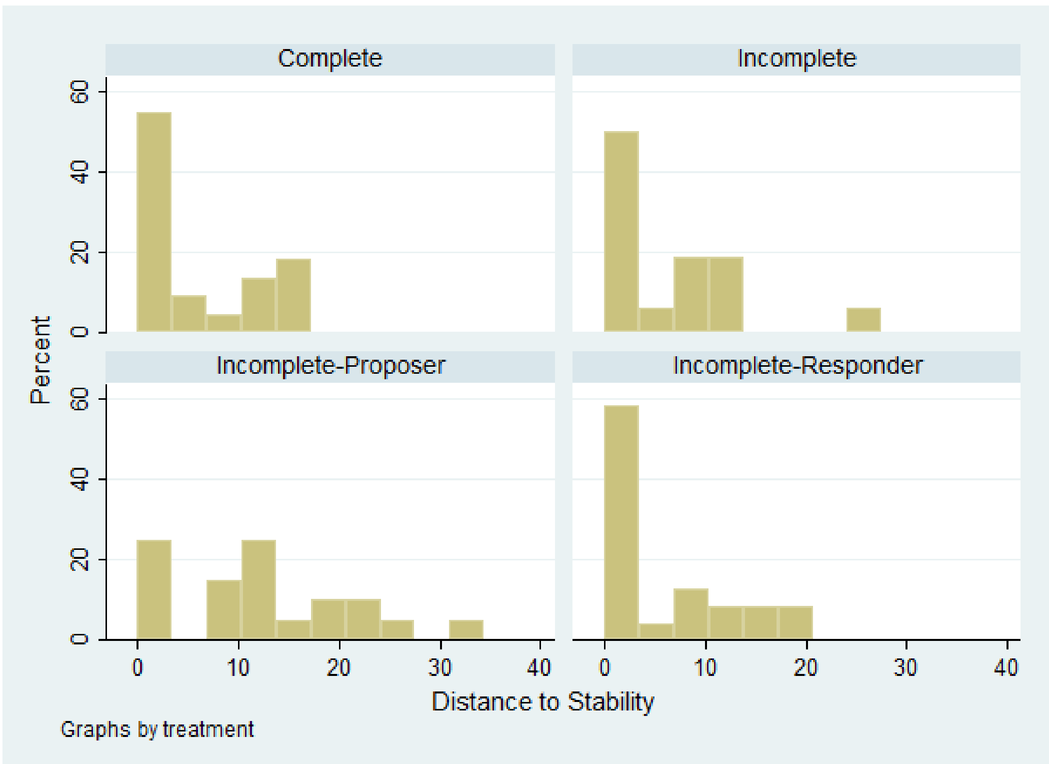


Figure 1: Distance to the nearest stable outcome across treatments.

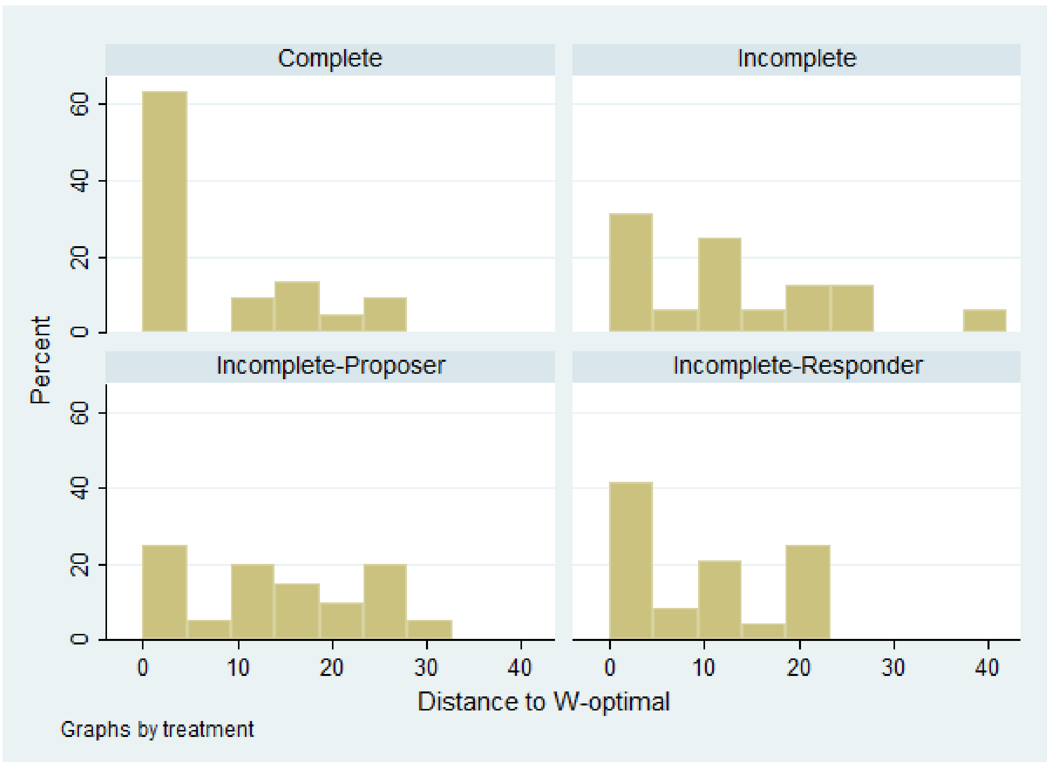


Figure 2: Distance to the W-optimal stable outcome across treatments.

strategic behavior only in treatments in which they possess complete information, we also use a Wilcoxon-Mann-Whitney test to assess whether the distance to the W-optimal stable matching is statistically smaller in the three treatments where responders have complete information. We cannot reject the null hypothesis that the distance to the W-optimal stable matching is statistically indistinguishable between the incomplete information treament and the remaining treatments (z-score = -1.170, p-value = 0.2419).

In the next section, we investigate behavioral explanations for these findings when discussing individual behavior.

Individual Behavior

Responders

We start by reporting the behavior of responders. In the static DA mechanism, responders have an incentive to misrepresent their preferences in markets with more than one achievable match partner. This misrepresentation can take two forms: truncation (submitting a shortened preference list that otherwise maintains the order of the true preferences) and manipulation (submitting a preference list that switches the order of preference between at least two match partners). In our dynamic implementation, a truncation of preferences is observed anytime an unmatched responder rejects an offer. Intuitively, an agent who truncates her preferences is pretending that she would rather be unmatched than be matched to less preferred partners. A manipulation of preferences, on the other hand, is observed anytime (1) a currently matched responder accepts a less-preferred offer or (2) an unmatched responder accepts a less-preferred offer when facing two or more offers.

An agent who truncates her preferences faces a balance of risks. By truncating,

Table 5: Responder behavior across treatments.

Treatment	Manipulation	Percent	Truncation	Percent	Number of Decisions
Complete Info.	5	1.6%	18	5.8%	310
Incomplete Info.	1	0.4%	26	11.0%	236
Incomplete-Proposer	5	1.3%	52	13.7%	380
Incomplete-Responder	7	2.1%	25	7.6%	328
Total	18	1.4%	121	9.6%	1,254

there is a greater likelihood of remaining unmatched, but there is also a greater likelihood of obtaining a more preferred match partner conditional on matching. When all other agents behave straightforwardly, there is an "optimal" truncation strategy: a responder can be matched with her most preferred achievable partner by rejecting all offers from lower-ranked partners. However, this strategy requires for a responder to calculate or otherwise know the identity of her most preferred achievable partner. Absent this information, responders face the risk of "over-truncating" and remaining unmatched.[13] Thus, meaningful truncation behavior is only possible in the three treatments in which responders have complete information about all other subjects' preferences.[14] In addition, merely identifying the existence of a profitable truncation opportunity requires for an agent to know that she has more than one achievable partner. Similarly, this information is only available with complete information about other subjects' preferences.

Table 5 presents the behavior of responders by treatment. We observe that neither manipulation nor truncation of preferences is common. Of a total of 1,256 decision problems[15] faced by responders, 18 (1.4%) involve a manipulation of preferences and

[13] Over-truncation refers to the situation where a responder rejects an offer from her most preferred achievable partner.

[14] Since strategic uncertainty remains about other agents' behavior in our experimental design, even playing the optimal truncation strategy does not guarantee for a responder to be matched with her most preferred achievable partner.

[15] By "decision problem", we refer to any instance in which a responder is called upon to act in the dynamic DA mechanism. In other words, either (1) an unmatched responder who must evaluate

121 (9.6%) involve a truncation of preferences. The results indicate that the distribution of truncations is affected by the information structure ($\chi^2(3) = 14.4001$, p-value $= 0.002$), but that the distribution of manipulations is not ($\chi^2(3) = 2.9466$, p-value $= 0.400$).

Regarding responders' willingness to engage in strategic behavior, we find no statistically significant difference between the incomplete information treatment and the other three treatments combined ($\chi^2(3) = 0.6239$, p-value $= 0.430$).[16] This indicates that strategic preference misrepresentation is not systematically related to the available information. Since purposeful strategic behavior requires information about the preferences of other agents, this suggests that responders' non-straightforward behavior is driven by other heuristics.

In particular, it is natural to investigate whether truncation behavior is driven by a simple threshold strategy in which responders are more likely to reject offers that yield low payoffs. Figure 3 shows truncation rates for different payoff levels. Twenty-percent of offers that would have yielded a payoff of \$3 are rejected by responders. However, only 1% of offers that would have yielded a payoff of \$21 are rejected. We use an extension of the Wilcoxon rank-sum procedure to non-parametrically test for a trend in truncation rates across payoff levels. We find that truncation rates are decreasing by the profitability of the offer (p-value < 0.001).

Table 6 presents an OLS regression of a truncation dummy variable on the round of the experiment, a dummy variable for the three treatments where responders have complete information, the rank of an agent's most preferred achievable partner,[17] the

one or more offers or (2) a tentatively matched responder who must evaluate one or more alternative offers.

[16]For this test, we pool all manipulations and truncations together.

[17]This measure captures the riskiness of truncation. It can be thought of as the likelihood of mistakenly over-truncating.

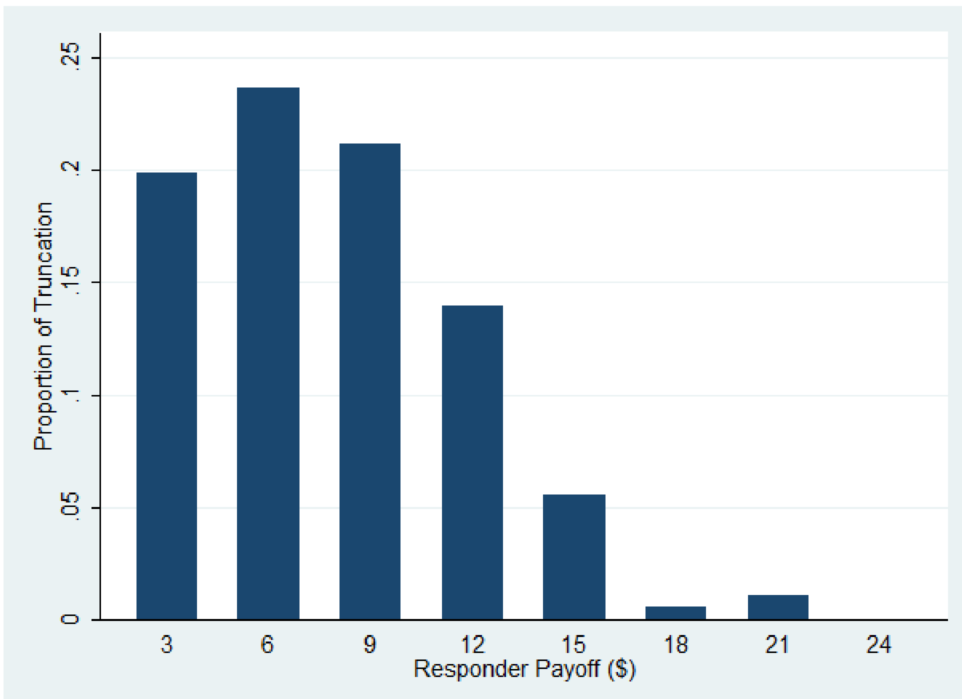


Figure 3: Truncation rates according to the profitability of offers.

span of the core,[18] and the rank of an agent's most preferred offer in a given decision problem. The regression indicates that responders are sensitive to the profitability of truncation (i.e., the span of the core). However, responders are not sensitive to the riskiness of truncation (i.e., the rank of their most preferred achievable partner in their preference list).

The regression results also confirm our earlier intuition: the rank of an offer is a statistically significant predictor of truncation behavior. In particular, receiving an offer from a proposer who is ranked one spot lower in preference (i.e., lowering the payoff by \$3) makes a responder 4% more likely to reject the offer. The fact that the coefficient estimate on the complete information dummy variable is neither economically nor statistically significant further supports the conclusion that responders are not making use of the relevant information for truncation decisions but rather following a simpler heuristic.

Proposers

In the static DA mechanism, proposers have a dominant strategy of truthful preference revelation. In the dynamic DA mechanism, the analogue of truthful behavior is *straightforward* behavior (i.e., when an agent makes offers in the order of her true preferences). We will focus our analysis on initial offers (i.e., the first offers made by proposers in each experimental market).[19]

Of the 656 initial offers made by proposers across all the experimental markets, 365 (56%) were made to proposers' first choice match partners. This represents a non-trivial departure from the weakly dominant strategy. The breakdown by treatment

[18] The span of the core is defined as the ordinal distance between an individual's most preferred and least preferred achievable partners.

[19] This is because we do not have access to subjects' complete strategies. For all markets, however, we do have data on the first offer made by each proposer.

Table 6: OLS regression of truncation on market features. Standard errors are clustered at the individual level.

	(1)
VARIABLES	Truncation
Round of the experiment	-0.00444
	(0.00383)
Complete information	-0.00977
	(0.0387)
Rank of best achievable partner	-0.00245
	(0.00590)
Span of the core	0.0346***
	(0.0105)
Rank of best offer	0.0399***
	(0.00528)
Constant	-0.0934
	(0.0404)
Observations	1,254

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

Table 7: Proposers' initial offers across treatments.

Treatment	Offers Made to First Choice	Percent	Number of Skips Mean	Number of Skips Std. Dev.	Total Number of Offers
Complete Info.	81	46%	1.2	1.5	176
Incomplete Info.	98	77%	0.39	0.81	128
Incomplete-Proposer	74	46%	1.46	1.77	160
Incomplete-Responder	112	58%	0.89	1.34	192

is shown in Table 7. Interestingly, the incomplete information treatment has the highest proportion of dominant-strategy play in terms of initial offers. We find that the failure to play the dominant strategy varies across treatments ($\chi^2(3) = 35.5771$, p-value < 0.001). This is not only due to the fact that proposers play their dominant strategy more frequently in the incomplete information treatment. The distribution of dominant-strategy play is significantly different across the remaining three treatments as well ($\chi^2(2) = 7.2797$, p-value $= 0.026$).

A natural measure of the extent of a deviation from the dominant strategy is the ordinal distance between a worker's initial offer and her most preferred match partner. We refer to this measure as a "skip". A skip of 0 is equivalent to dominant-strategy play and a skip of 7 indicates that a proposer's initial offer was made to her least preferred match partner (i.e., the proposer skipped over 7 more preferred matches). Table 6 also shows the mean and standard deviation of this measure for the four treatments. According to this measure, there are also differences in dominant-strategy play across treatments ($\chi^2(21) = 60.9252$, p-value < 0.001).

Figure 4 shows the distribution of the number of skips across treatments. The empirical distributions of first offers suggest that dominant-strategy play is most common in the incomplete information treatment. Indeed, we can reject the hypothesis that the average number of skips in the incomplete information treatment is equal to the average number of skips in the remaining three treatments (t-statistic = 5.5313, p-value < 0.001). We also reject this hypothesis in pairwise comparisons against the incomplete information treatment (IC v C, t-statistic = 5.6573, p-value < 0.001; IC v IP, t-statistic = 6.3459, p-value < 0.001; IC v IR, t-statistic = 3.7895, p-value = 0.0002). Finally, the incomplete information treatment has less skipping than the other three treatments according to non-parametric trend tests (test for IC,IR,IP,C: z-score = 5.79, p-value < 0.001; test for IC, IP, IR, C: z-score = 3.81, p-value < 0.001).

The fact that truth-telling by proposers is only a *weakly* dominant strategy implies the existence of a class of skipping behavior that is not harmful (i.e., does no worse than truth-telling). In fact, we know that an individual agent can do no better than to be matched with her most preferred achievable partner via a Nash equilibrium strategy profile. Thus, in equilibrium, there is no welfare loss from a proposer skipping past her most preferred match partner and instead making her first offer to her most preferred *achievable* match partner. We speak of "consequential" skipping to refer to the situation where a proposer makes her first offer to an individual ranked below her most preferred achievable partner.[20]

We find that 53% (155/291) of skips are consequential. Figure 5 shows the distribution of consequential skipping across treatments.[21] The proportion of inconsequential skips is highest in the complete information treatment: 60% (57/95). Importantly, proposers can determine whether skipping is consequential or not only by calculating the set of stable matchings. In other words, for proposers to know whether skipping entails welfare losses, they must have full information about other agents' preferences. Thus, purposeful skipping is only possible in the complete information

[20]This construction is a bit loose since proposers who skip will sometimes have the option of correcting their mistake by making their next offer to a *more* preferred individual. That is, proposers are not forced to monotonically move down their preference lists when making offers.

[21]Figure 4 presents data on consequential skipping *conditional on skipping*. Dominant-strategy play (a "skip" of 0) is not included in the graphs.

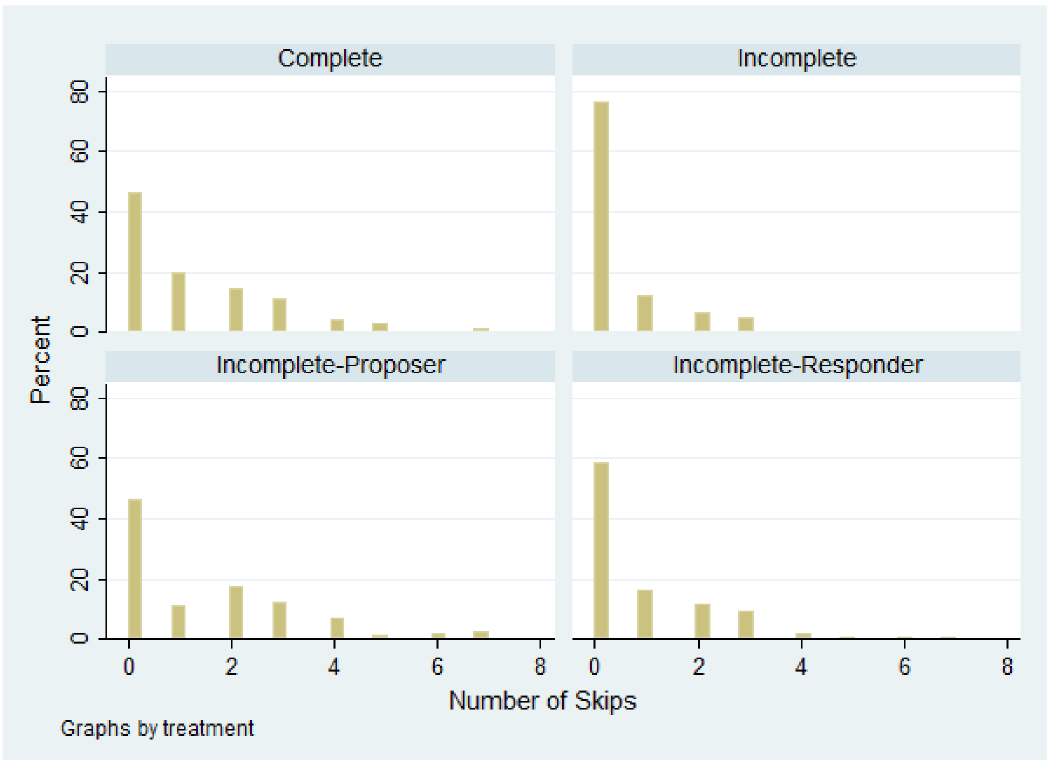


Figure 4: The number of skips across treatments.

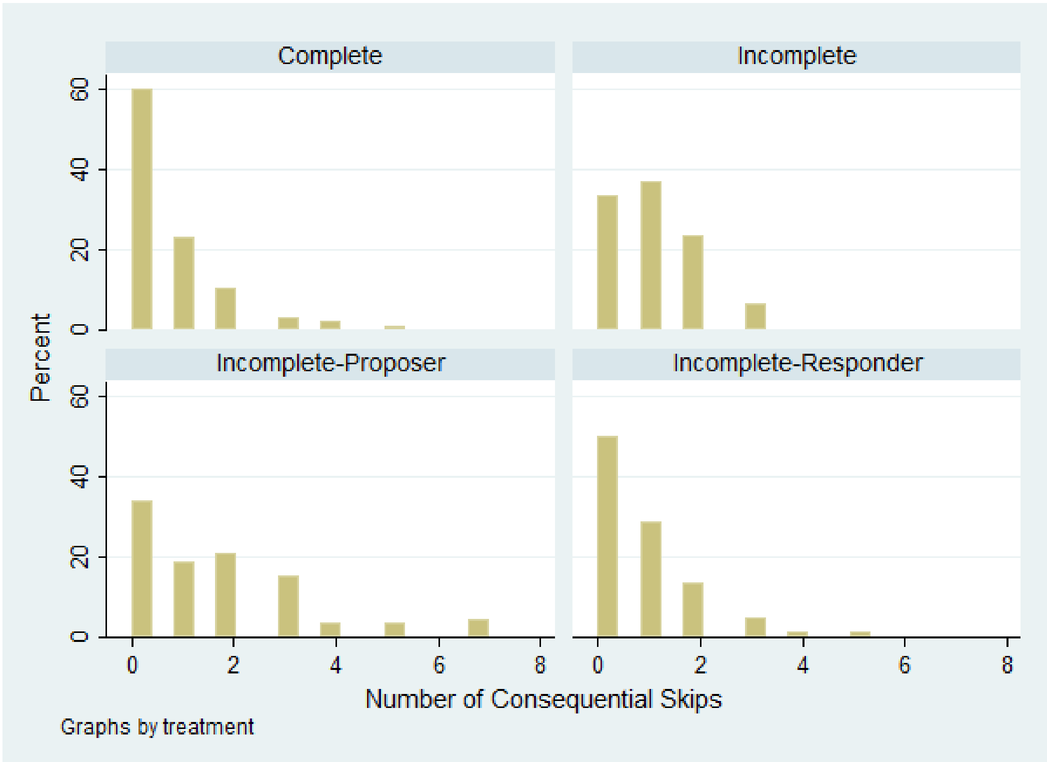


Figure 5: The number of consequential skips across treatments.

treatment. We find that the distribution of consequential skipping depends on the information structure that subjects face (Kruskal-Wallis test, p-value = 0.0001). In fact, this difference persists if we compare the distribution of consequential skipping in the complete information treatment against the pooled data from the remaining three treatments (Kruskal-Wallis test, p-value = 0.0010).

The stark difference in proposers' behavior between the complete and incomplete information treatments suggests that information affects the play of dominant strategies. By altering the information that is available to proposers, our experimental design allows us to decompose these information effects. Echenique et al. (2014), for instance, suggest that proposers might internalize the probability of rejection when

Table 8: OLS regressions of the number of skips for each treatment. The regressions include fixed effects at the individual level.

Dependent Variable: Number of Skips	(1) Complete	(2) Incomplete	(3) I-Proposer	(4) I-Responder
Round of the experiment	-0.0620	-0.0315*	-0.0691	0.0125
	(0.0413)	(0.0169)	(0.0506)	(0.0411)
Rank in responder's preference	0.254***	0.0264	-0.0440	0.142***
	(0.0469)	(0.0222)	(0.0406)	(0.0444)
Number of competitors	0.0550	0.00644	-0.0268	-0.0312
	(0.0360)	(0.0147)	(0.0486)	(0.0466)
Number of inconsequential skips	0.153*	-0.0617	-0.0509	0.0868
	(0.0877)	(0.0368)	(0.0662)	(0.0865)
Constant	-0.0103	0.501***	2.06***	0.136
	(0.340)	(0.130)	(0.238)	(0.268)
Observations	176	128	160	192

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

deciding to make an offer. In other words, a proposer might conjecture (correctly or not) that her offer is more likely to be rejected the lower she is ranked in the responder's preference list. Another possible explanation is that proposers might avoid making offers to their favorite match partner if there are other proposers competing for the same individual (i.e, other proposers agree on who the "best" partner is). Both of these theories would suggest skipping if agents are averse to rejection or otherwise sensitive to the probability of rejection. The former theory requires that proposers have information about the preferences of responders. The latter theory, however, requires that proposers have information about the preferences of other proposers.

We now use our experimental design to tease apart what drives proposers' behavior. Table 8 presents a regression analysis of skipping behavior as a function of four variables: the round of the experiment, a proposer's rank in her most preferred partner's preferences, the number of proposers who share the same most preferred partner, and the number of available inconsequential skips.[22] Importantly, and depending on

[22]Column 1 presents the result for the complete information treatment, column 2 for the incomplete information treatment, column 3 for the incomplete-proposer treament, and column 4 for the

the particular treatment, not all of the explanatory variables can be observed by the experimental subjects. However, the same explanatory variables are included in all the regressions as a way to test for placebo effects.[23]

We do not find evidence for the theory that proposers take into account how many competitors they have for their most preferred partner. The coefficient estimate on the number of competitors is statistically insignificant for all experimental treatments. However, we do find support for the hypothesis advanced by Echenique et al. (2014): when deciding on initial offers, proposers take into account how they are ranked by the other side of the market. A proposer's rank in the preference list of her most preferred partner is a statistically significant predictor of skipping behavior in both treatments where this information is available (complete and incomplete-responder). Moreover, this variable is not a statistically significant predictor of skipping in the treatments where this information is not available (incomplete and incomplete-proposer).

In addition, we find that the availability of opportunities for inconsequential skipping is significantly and positively correlated with skipping in the complete information treatment. This lends further support to our earlier finding that harmless skipping is most pronounced in the complete information treatment. Notably, this is also the only experimental treatment in which it is theoretically possible for a subject to calculate the identity of her most preferred achievable partner.

---

incomplete-responder treatment.

[23]For instance, the coefficient estimate on the number of competitiors should not be significant in the incomplete information or incomplete-responder treatments. Similarly, the coefficient estimate on the rank in the responder's preference list should not be significant in the incomplete information or incomplete-proposer treatments.

6 Conclusion and Discussion

We investigate the impact of information on strategic behavior in a decentralized matching market that mimics the Gale-Shapley deferred acceptance procedure. Rather than have subjects submit preference lists to a centralized clearinghouse, our experimental design requires subjects to make sequential decisions in an extensive-form game. This dynamic implementation has the advantage of reducing experimenter demand effects and making the strategic tensions of the game more transparent to subjects.

We find that information does not affect the stability of final outcomes, but it does affect selection among stable outcomes. This result is driven largely by a change in the behavior of one side of the market. When information is available about other agents' preferences, proposers often fail to play their dominant strategy of truth-telling and instead skip past more preferred partners when making offers. This skipping behavior is sophisticated in the sense that proposers take into account how they are ranked by the other side of the market. Responders, on the other hand, play similarly in all information treatments and accept the best offer that is available to them at any given time. This straightforward behavior is in contrast to the theoretical prediction that responders should be engaging in strategic preference misrepresentation in markets with more than one stable matching.

Our results highlight the important role that behavioral biases play in these environments. Much like market participants can fail to recognize the existence of profitable strategic opportunities (e.g., responders' straightforward behavior), they can also fail to recognize the *lack* of profitable strategic opportunities (e.g., proposers' skipping behavior). The ability to successfully absorb "behavioral" agents is a critical component of sound market design. In the context of the DA mechanism, there are already protective features in place that bound the losses of proposers who behave sub-optimally. Indeed, the fact that their best course of action is weakly dominant necessarily allows for inconsequential skipping in a large class of markets.[24]

There is another compelling reason to incorporate behavioral insights into mechanism design. One of the stated advantages of strategy-proof mechanisms in the field is that, under theoretical conditions, they permit the observation of true preferences. This in turn allows researchers to make welfare statements about market participants. By documenting the prevalence of strategic behavior in a strategy-proof environment, our results suggest a note of caution in this regard.[25] However, more work needs to be done to investigate the extent to which findings from laboratory studies of matching markets generalize to the field.

[24] Rees-Jones (2014) suggests that this type of tolerance for behavioral faults is one reason for the success and persistence of the DA mechanism in the field.

[25] Sub-optimal strategic behavior in the DA mechanism has also been documented empirically by Echenique et al. (2014) and Rees-Jones (2014).

Appendix A

Complete Information

	BLACK	BLUE	GREEN	PINK	PURPLE	RED	WHITE	YELLOW
APPLE	\$9,\$9	\$15,\$24	\$18,\$18	\$21,\$12	\$24,\$9	\$3,\$15	\$6,\$21	\$12,\$24
BANANA	\$21,\$18	\$6,\$21	\$9,\$3	\$24,\$3	\$3,\$24	\$18,\$3	\$15,\$15	\$12,\$21
CHERRY	\$12,\$12	\$24,\$12	\$6,\$24	\$21,\$18	\$3,\$18	\$18,\$21	\$9,\$24	\$15,\$9
GRAPE	\$12,\$3	\$21,\$18	\$6,\$6	\$3,\$21	\$18,\$21	\$24,\$12	\$9,\$3	\$15,\$3
KIWI	\$18,\$21	\$3,\$15	\$15,\$21	\$24,\$6	\$9,\$3	\$6,\$6	\$12,\$9	\$21,\$15
MANGO	\$12,\$6	\$6,\$3	\$9,\$9	\$3,\$15	\$24,\$12	\$21,\$18	\$18,\$18	\$15,\$6
PEACH	\$15,\$15	\$9,\$9	\$21,\$15	\$12,\$24	\$3,\$6	\$6,\$9	\$24,\$6	\$18,\$18
PEAR	\$9,\$24	\$18,\$6	\$12,\$12	\$6,\$9	\$15,\$15	\$3,\$24	\$21,\$12	\$24,\$12

Incomplete Information

	PURPLE	PINK	GREEN	BLUE	YELLOW	BLACK	WHITE	RED
APPLE	\$24	\$21	\$18	\$15	\$12	\$9	\$6	\$3

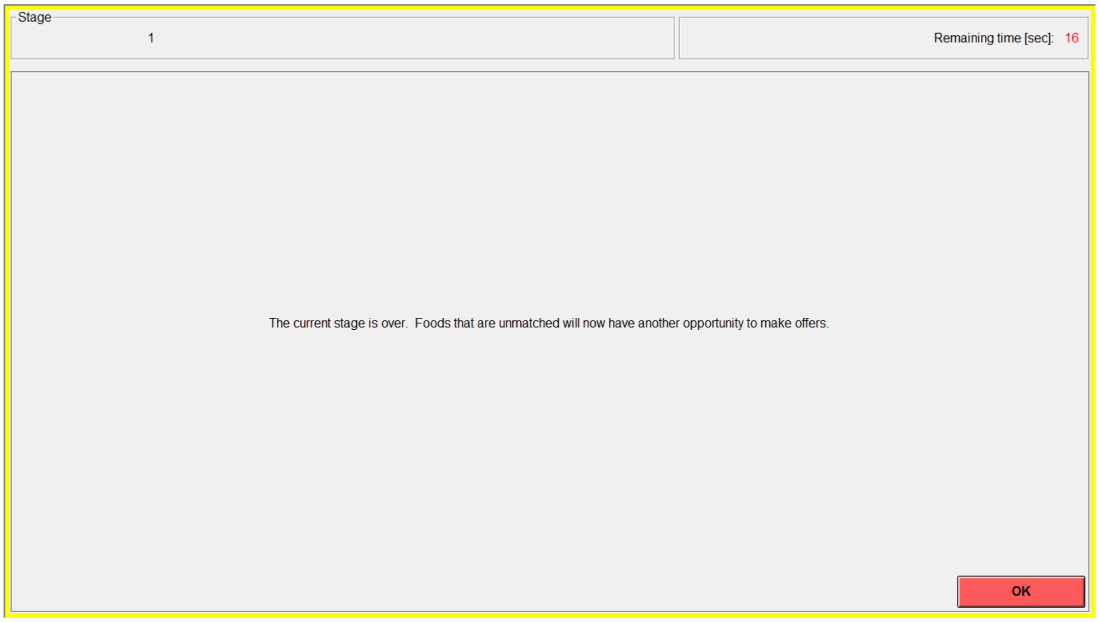
Incomplete-Proposer

	BLACK	BLUE	GREEN	PINK	PURPLE	RED	WHITE	YELLOW
APPLE	\$9	\$15	\$18	\$21	\$24	\$3	\$6	\$12
BANANA	\$21	\$6	\$9	\$24	\$3	\$18	\$15	\$12
CHERRY	\$12	\$24	\$6	\$21	\$3	\$18	\$9	\$15
GRAPE	\$12	\$21	\$6	\$3	\$18	\$24	\$9	\$15
KIWI	\$18	\$3	\$15	\$24	\$9	\$6	\$12	\$21
MANGO	\$12	\$6	\$9	\$3	\$24	\$21	\$18	\$15
PEACH	\$15	\$9	\$21	\$12	\$3	\$6	\$24	\$18
PEAR	\$9	\$18	\$12	\$6	\$15	\$3	\$21	\$24

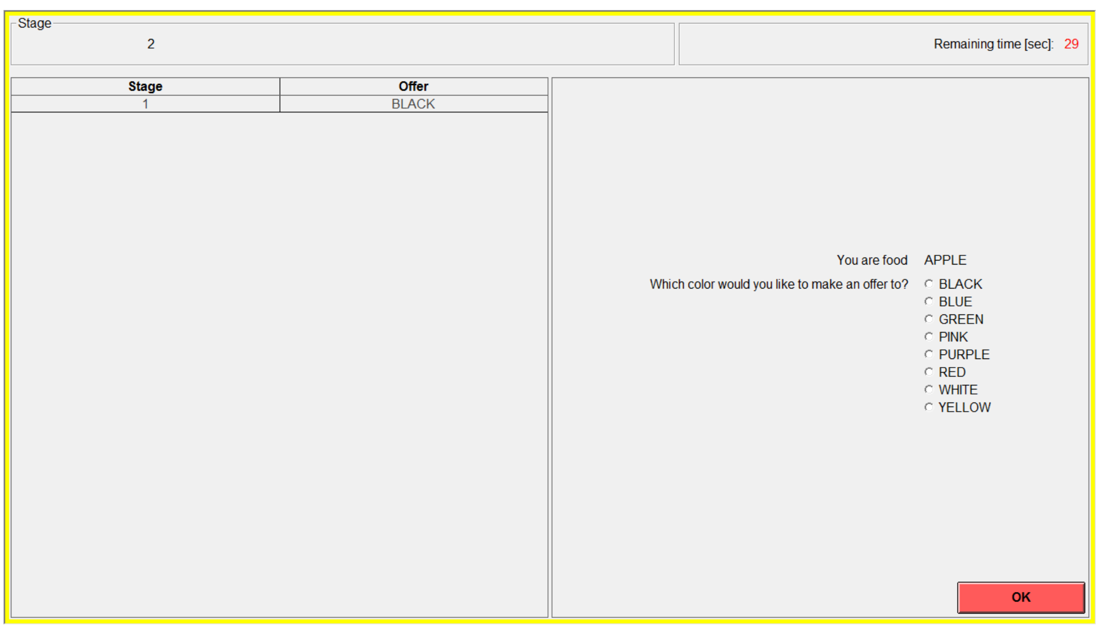
Incomplete-Responder

	BLACK	BLUE	GREEN	PINK	PURPLE	RED	WHITE	YELLOW
APPLE	\$9,\$9	\$15,\$24	\$18,\$18	\$21,\$12	\$24,\$9	\$3,\$15	\$6,\$21	\$12,\$24
BANANA	\$18	\$21	\$3	\$3	\$24	\$3	\$15	\$21
CHERRY	\$12	\$12	\$24	\$18	\$18	\$21	\$24	\$9
GRAPE	\$3	\$18	\$6	\$21	\$21	\$12	\$3	\$3
KIWI	\$21	\$15	\$21	\$6	\$3	\$6	\$9	\$15
MANGO	\$6	\$3	\$9	\$15	\$12	\$18	\$18	\$6
PEACH	\$15	\$9	\$15	\$24	\$6	\$9	\$6	\$18
PEAR	\$24	\$6	\$12	\$9	\$15	\$24	\$12	\$12

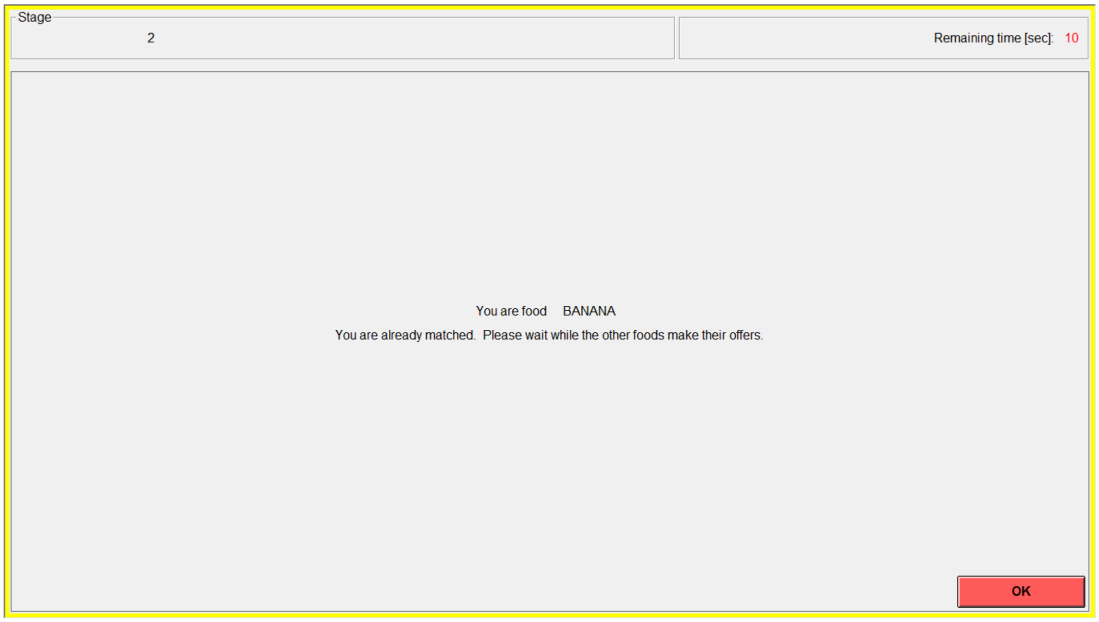
Appendix B



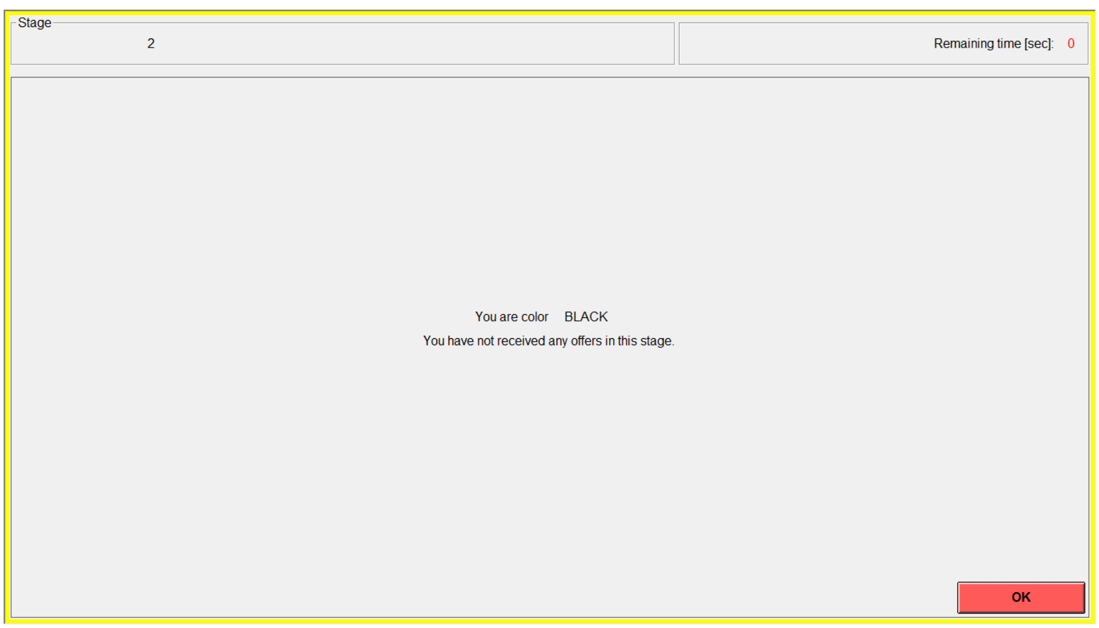
The end of the first stage of offers in the DA algorithm.



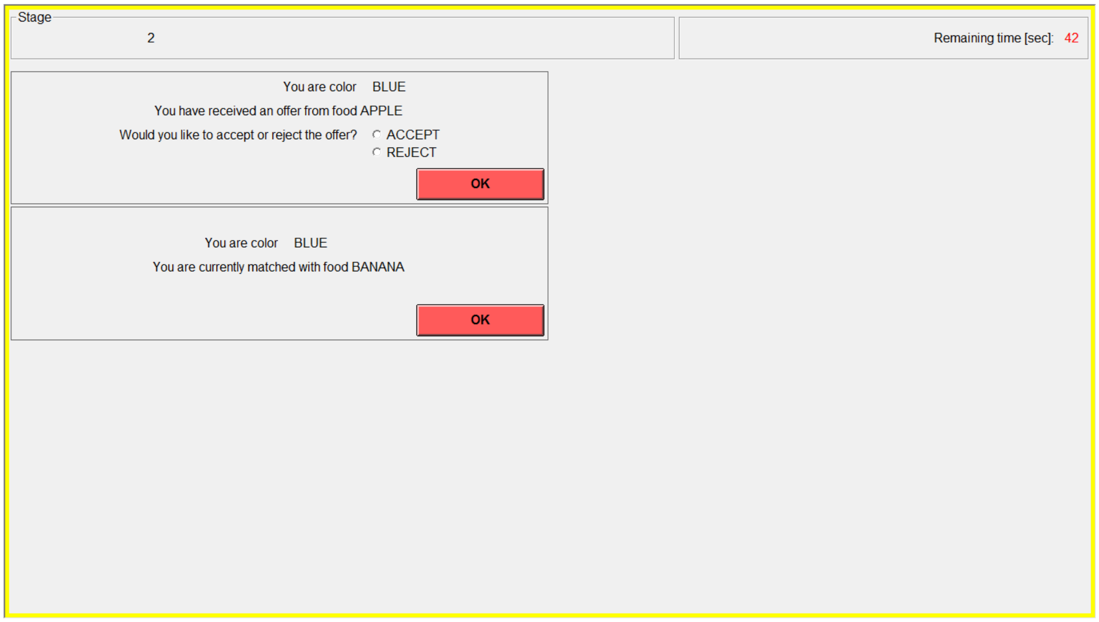
Food APPLE chooses an offer in the second stage of the DA algorithm.



Food BANANA is tentatively matched and cannot make an offer in the second stage of the DA algorithm.



Color BLACK has not received any offers in the second stage of the DA algorithm.



Color BLUE is tentatively matched and has received a new offer in the second stage of the DA algorithm.

Appendix C

Market Characteristics

Market	Stable Matches	Turns to Converge with Truth-Telling
A	2	10
B	5	5
C	2	10
D	5	3
E	2	5
F	1	8
G	1	8
H	5	5
I	2	3

Ordinal Preference Profiles and Stable Matchings

When there are multiple stable outcomes listed below, the first matching corresponds to the proposer-optimal stable matching while the last matching corresponds to the responder-optimal stable matching. Asterisks in an individual's preference list are used to denote the identities of that individual's achievable match partners.

Market A

$P(f_1) = c_2, c_4, c_6^*, c_8, c_1, c_7, c_3, c_5$

$P(f_2) = c_6, c_2, c_5^*, c_8^*, c_1, c_7, c_3, c_4$

$P(f_3) = c_4, c_8, c_1^*, c_3, c_7, c_5, c_6, c_2$

$P(f_4) = c_5, c_6, c_7, c_8^*, c_1, c_3^*, c_2, c_4$

$P(f_5) = c_7^*, c_3, c_8, c_1, c_4, c_2, c_6, c_5$

$P(f_6) = c_8, c_7, c_2^*, c_5^*, c_3, c_1, c_4, c_6$

$P(f_7) = c_5, c_4, c_3^*, c_2^*, c_8, c_1, c_7, c_6$

$P(f_8) = c_4^*, c_1, c_6, c_7, c_8, c_3, c_2, c_5$

$P(c_1) = f_3^*, f_5, f_1, f_7, f_8, f_6, f_4, f_2$

$P(c_2) = f_7^*, f_4, f_3, f_6^*, f_1, f_8, f_5, f_2$

$P(c_3) = f_2, f_4^*, f_3, f_8, f_6, f_1, f_7^*, f_5$

$P(c_4) = f_8^*, f_3, f_6, f_1, f_4, f_7, f_5, f_2$

$P(c_5) = f_3, f_1, f_6^*, f_2^*, f_8, f_5, f_7, f_4$

$P(c_6) = f_1^*, f_2, f_7, f_5, f_8, f_3, f_6, f_4$

$P(c_7) = f_8, f_5^*, f_2, f_7, f_3, f_1, f_6, f_4$

$P(c_8) = f_1, f_2^*, f_4^*, f_5, f_3, f_7, f_8, f_6$

$$\mu_1 = (f_1, c_6), (f_2, c_5), (f_3, c_1), (f_4, c_8), (f_5, c_7), (f_6, c_2), (f_7, c_3), (f_8, c_4)$$
$$\mu_2 = (f_1, c_6), (f_2, c_8), (f_3, c_1), (f_4, c_3), (f_5, c_7), (f_6, c_5), (f_7, c_2), (f_8, c_4)$$

Market B

$P(f_1) = c_6^*, c_2, c_5^*, c_8, c_1, c_7, c_3, c_4$

$P(f_2) = c_4, c_8, c_1, c_3^*, c_7^*, c_5, c_6, c_2$

$P(f_3) = c_5^*, c_6^*, c_7, c_8^*, c_1, c_3, c_2, c_4$

$P(f_4) = c_7^*, c_3^*, c_8, c_1, c_4, c_2, c_6, c_5$

$P(f_5) = c_8^*, c_7, c_2, c_5, c_3, c_1, c_4, c_6^*$

$P(f_6) = c_5, c_4^*, c_3, c_2, c_8, c_1, c_7, c_6$

$P(f_7) = c_4, c_1^*, c_6, c_7, c_8, c_3, c_2, c_5$

$P(f_8) = c_2^*, c_4, c_6, c_8, c_1, c_7, c_3, c_5$

$P(c_1) = f_7^*, f_4, f_3, f_6, f_1, f_8, f_5, f_2$

$P(c_2) = f_2, f_4, f_3, f_8^*, f_6, f_1, f_7, f_5$

$P(c_3) = f_8, f_3, f_6, f_1, f_4^*, f_7, f_5, f_2^*$

$P(c_4) = f_3, f_1, f_6^*, f_2, f_8, f_5, f_7, f_4$

$P(c_5) = f_1^*, f_2, f_7, f_5, f_8, f_3^*, f_6, f_4$

$P(c_6) = f_8, f_5^*, f_2, f_7, f_3^*, f_1^*, f_6, f_4$

$P(c_7) = f_1, f_2^*, f_4^*, f_5, f_3, f_7, f_8, f_6$

$P(c_8) = f_3^*, f_5^*, f_1, f_7, f_8, f_6, f_4, f_2$

$$\mu_1 = (f_1, c_6), (f_2, c_3), (f_3, c_5), (f_4, c_7), (f_5, c_8), (f_6, c_4), (f_7, c_1), (f_8, c_2)$$

$$\mu_2 = (f_1, c_6), (f_2, c_7), (f_3, c_5), (f_4, c_3), (f_5, c_8), (f_6, c_4), (f_7, c_1), (f_8, c_2)$$

$$\mu_3 = (f_1, c_5), (f_2, c_3), (f_3, c_6), (f_4, c_7), (f_5, c_8), (f_6, c_4), (f_7, c_1), (f_8, c_2)$$

$$\mu_4 = (f_1, c_5), (f_2, c_7), (f_3, c_6), (f_4, c_3), (f_5, c_8), (f_6, c_4), (f_7, c_1), (f_8, c_2)$$

$$\mu_5 = (f_1, c_5), (f_2, c_7), (f_3, c_8), (f_4, c_3), (f_5, c_6), (f_6, c_4), (f_7, c_1), (f_8, c_2)$$

Market C

$P(f_1) = c_4^1, c_8, c_1, c_3, c_7, c_5, c_6, c_2$

$P(f_2) = c_5, c_6^*, c_7, c_8, c_1, c_3, c_2, c_4$

$P(f_3) = c_7^*, c_3, c_8, c_1, c_4, c_2, c_6, c_5$

$P(f_4) = c_8, c_7, c_2, c_5, c_3^*, c_1^*, c_4, c_6$

$P(f_5) = c_5^*, c_4, c_3, c_2, c_8, c_1, c_7, c_6$

$P(f_6) = c_4, c_1^*, c_6, c_7, c_8, c_3^*, c_2, c_5$

$P(f_7) = c_2, c_4, c_6, c_8^*, c_1, c_7, c_3, c_5$

$P(f_8) = c_6, c_2^*, c_5, c_8, c_1, c_7, c_3, c_4$

$P(c_1) = f_2, f_4^*, f_3, f_8, f_6^*, f_1, f_7, f_5$

$P(c_2) = f_8^*, f_3, f_6, f_1, f_4, f_7, f_5, f_2$

$P(c_3) = f_3, f_1, f_6^*, f_2, f_8, f_5, f_7, f_4^*$

$P(c_4) = f_1^*, f_2, f_7, f_5, f_8, f_3, f_6, f_4$

$P(c_5) = f_8, f_5^*, f_2, f_7, f_3, f_1, f_6, f_4$

$P(c_6) = f_1, f_2^*, f_4, f_5, f_3, f_7, f_8, f_6$

$P(c_7) = f_3^*, f_5, f_1, f_7, f_8, f_6, f_4, f_2$

$P(c_8) = f_7^*, f_4, f_3, f_6, f_1, f_8, f_5, f_2$

$$\mu_1 = (f_1, c_4), (f_2, c_6), (f_3, c_7), (f_4, c_3), (f_5, c_5), (f_6, c_1), (f_7, c_8), (f_8, c_2)$$

$$\mu_2 = (f_1, c_4), (f_2, c_6), (f_3, c_7), (f_4, c_1), (f_5, c_5), (f_6, c_3), (f_7, c_8), (f_8, c_2)$$

Market D

$P(f_1) = c_5^*, c_6, c_7, c_8, c_1, c_3, c_2, c_4$

$P(f_2) = c_7^*, c_3^*, c_8, c_1, c_4, c_2, c_6, c_5$

$P(f_3) = c_8^*, c_7^*, c_2^*, c_5, c_3, c_1, c_4, c_6$

$P(f_4) = c_5, c_4, c_3^*, c_2, c_8^*, c_1, c_7, c_6$

$P(f_5) = c_4, c_1^*, c_6^*, c_7, c_8, c_3, c_2, c_5$

$P(f_6) = c_2^*, c_4, c_6, c_8, c_1^*, c_7, c_3, c_5$

$P(f_7) = c_6^*, c_2, c_5, c_8, c_1^*, c_7^*, c_3, c_4$

$P(f_8) = c_4^*, c_8, c_1, c_3, c_7, c_5, c_6, c_2$

$P(c_1) = f_8, f_3, f_6^*, f_1, f_4, f_7^*, f_5^*, f_2$

$P(c_2) = f_3^*, f_1, f_6^*, f_2, f_8, f_5, f_7, f_4$

$P(c_3) = f_1, f_2^*, f_7, f_5, f_8, f_3, f_6, f_4^*$

$P(c_4) = f_8^*, f_5, f_2, f_7, f_3, f_1, f_6, f_4$

$P(c_5) = f_1^*, f_2, f_4, f_5, f_3, f_7, f_8, f_6$

$P(c_6) = f_3, f_5^*, f_1, f_7^*, f_8, f_6, f_4, f_2$

$P(c_7) = f_7^*, f_4, f_3^*, f_6, f_1, f_8, f_5, f_2^*$

$P(c_8) = f_2, f_4^*, f_3^*, f_8, f_6, f_1, f_7, f_5$

$$\mu_1 = (f_1, c_5), (f_2, c_7), (f_3, c_8), (f_4, c_3), (f_5, c_1), (f_6, c_2), (f_7, c_6), (f_8, c_4)$$

$$\mu_2 = (f_1, c_5), (f_2, c_7), (f_3, c_8), (f_4, c_3), (f_5, c_6), (f_6, c_2), (f_7, c_1), (f_8, c_4)$$

$$\mu_3 = (f_1, c_5), (f_2, c_3), (f_3, c_7), (f_4, c_8), (f_5, c_6), (f_6, c_2), (f_7, c_1), (f_8, c_4)$$

$$\mu_4 = (f_1, c_5), (f_2, c_3), (f_3, c_7), (f_4, c_8), (f_5, c_1), (f_6, c_2), (f_7, c_6), (f_8, c_4)$$

$$\mu_5 = (f_1, c_5), (f_2, c_3), (f_3, c_2), (f_4, c_8), (f_5, c_6), (f_6, c_1), (f_7, c_7), (f_8, c_4)$$

Market E

$P(f_1) = c_7, c_3^*, c_8, c_1^*, c_4, c_2, c_6, c_5$

$P(f_2) = c_8, c_7^*, c_2, c_5, c_3, c_1, c_4, c_6$

$P(f_3) = c_5^*, c_4, c_3, c_2, c_8, c_1, c_7, c_6$

$P(f_4) = c_4^*, c_1, c_6, c_7, c_8, c_3, c_2, c_5$

$P(f_5) = c_2^*, c_4, c_6, c_8, c_1, c_7, c_3, c_5$

$P(f_6) = c_6^*, c_2, c_5, c_8, c_1, c_7, c_3, c_4$

$P(f_7) = c_4, c_8, c_1^*, c_3^*, c_7, c_5, c_6, c_2$

$P(f_8) = c_5, c_6, c_7, c_8^*, c_1, c_3, c_2, c_4$

$P(c_1) = f_3, f_1^*, f_6, f_2, f_8, f_5, f_7^*, f_4$

$P(c_2) = f_1, f_2, f_7, f_5^*, f_8, f_3, f_6, f_4$

$P(c_3) = f_8, f_5, f_2, f_7^*, f_3, f_1^*, f_6, f_4$

$P(c_4) = f_1, f_2, f_4^*, f_5, f_3, f_7, f_8, f_6$

$P(c_5) = f_3^*, f_5, f_1, f_7, f_8, f_6, f_4, f_2$

$P(c_6) = f_7, f_4, f_3, f_6^*, f_1, f_8, f_5, f_2$

$P(c_7) = f_2^*, f_4, f_3, f_8, f_6, f_1, f_7, f_5$

$P(c_8) = f_8^*, f_3, f_6, f_1, f_4, f_7, f_5, f_2$

$$\mu_1 = (f_1, c_3), (f_2, c_7), (f_3, c_5), (f_4, c_4), (f_5, c_2), (f_6, c_6), (f_7, c_1), (f_8, c_8)$$
$$\mu_2 = (f_1, c_1), (f_2, c_7), (f_3, c_5), (f_4, c_4), (f_5, c_2), (f_6, c_6), (f_7, c_3), (f_8, c_8)$$

Market F

$P(f_1) = c_1^*, c_2, c_3, c_4, c_5, c_6, c_7, c_8$

$P(f_2) = c_1, c_2^*, c_3, c_4, c_5, c_6, c_7, c_8$

$P(f_3) = c_1, c_2, c_3^*, c_4, c_5, c_6, c_7, c_8$

$P(f_4) = c_1, c_2, c_3, c_4^*, c_5, c_6, c_7, c_8$

$P(f_5) = c_1, c_2, c_3, c_4, c_5^*, c_6, c_7, c_8$

$P(f_6) = c_1, c_2, c_3, c_4, c_5, c_6^*, c_7, c_8$

$P(f_7) = c_1, c_2, c_3, c_4, c_5, c_6, c_7^*, c_8$

$P(f_8) = c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8^*$

$P(c_1) = f_1^*, f_2, f_3, f_4, f_5, f_6, f_7, f_8$

$P(c_2) = f_1, f_2^*, f_3, f_4, f_5, f_6, f_7, f_8$

$P(c_3) = f_1, f_2, f_3^*, f_4, f_5, f_6, f_7, f_8$

$P(c_4) = f_1, f_2, f_3, f_4^*, f_5, f_6, f_7, f_8$

$P(c_5) = f_1, f_2, f_3, f_4, f_5^*, f_6, f_7, f_8$

$P(c_6) = f_1, f_2, f_3, f_4, f_5, f_6^*, f_7, f_8$

$P(c_7) = f_1, f_2, f_3, f_4, f_5, f_6, f_7^*, f_8$

$P(c_8) = f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8^*$

$$\mu_1 = (f_1, c_1), (f_2, c_2), (f_3, c_3), (f_4, c_4), (f_5, c_5), (f_6, c_6), (f_7, c_7), (f_8, c_8)$$

Market G

$P(f_1) = c_8, c_7, c_6, c_5, c_4, c_3, c_2, c_1^*$

$P(f_2) = c_8, c_7, c_6, c_5, c_4, c_3, c_2^*, c_1$

$P(f_3) = c_8, c_7, c_6, c_5, c_4, c_3^*, c_2, c_1$

$P(f_4) = c_8, c_7, c_6, c_5, c_4^*, c_3, c_2, c_1$

$P(f_5) = c_8, c_7, c_6, c_5^*, c_4, c_3, c_2, c_1$

$P(f_6) = c_8, c_7, c_6^*, c_5, c_4, c_3, c_2, c_1$

$P(f_7) = c_8, c_7^*, c_6, c_5, c_4, c_3, c_2, c_1$

$P(f_8) = c_8^*, c_7, c_6, c_5, c_4, c_3, c_2, c_1$

$P(c_1) = f_8, f_7, f_6, f_5, f_4, f_3, f_2, f_1^*$

$P(c_2) = f_8, f_7, f_6, f_5, f_4, f_3, f_2^*, f_1$

$P(c_3) = f_8, f_7, f_6, f_5, f_4, f_3^*, f_2, f_1$

$P(c_4) = f_8, f_7, f_6, f_5, f_4^*, f_3, f_2, f_1$

$P(c_5) = f_8, f_7, f_6, f_5^*, f_4, f_3, f_2, f_1$

$P(c_6) = f_8, f_7, f_6^*, f_5, f_4, f_3, f_2, f_1$

$P(c_7) = f_8, f_7^*, f_6, f_5, f_4, f_3, f_2, f_1$

$P(c_8) = f_8^*, f_7, f_6, f_5, f_4, f_3, f_2, f_1$

$$\mu_1 = (f_1, c_1), (f_2, c_2), (f_3, c_3), (f_4, c_4), (f_5, c_5), (f_6, c_6), (f_7, c_7), (f_8, c_8)$$

Market H

$P(f_1) = c_5, c_4^*, c_3^*, c_2^*, c_8, c_1, c_7, c_6$

$P(f_2) = c_4, c_1^*, c_6, c_7^*, c_8^*, c_3, c_2, c_5$

$P(f_3) = c_2^*, c_4^*, c_6^*, c_8, c_1, c_7, c_3, c_5$

$P(f_4) = c_6^*, c_2^*, c_5^*, c_8, c_1, c_7, c_3, c_4$

$P(f_5) = c_4, c_8^*, c_1^*, c_3^*, c_7, c_5, c_6, c_2$

$P(f_6) = c_5^*, c_6^*, c_7^*, c_8, c_1, c_3, c_2, c_4$

$P(f_7) = c_7, c_3^*, c_8^*, c_1, c_4^*, c_2, c_6, c_5$

$P(f_8) = c_8, c_7^*, c_2, c_5^*, c_3, c_1^*, c_4, c_6$

$P(c_1) = f_8^*, f_5^*, f_2^*, f_7, f_3, f_1, f_6, f_4$

$P(c_2) = f_1^*, f_2, f_4^*, f_5, f_3^*, f_7, f_8, f_6$

$P(c_3) = f_3, f_5^*, f_1^*, f_7^*, f_8, f_6, f_4, f_2$

$P(c_4) = f_7^*, f_4, f_3^*, f_6, f_1^*, f_8, f_5, f_2$

$P(c_5) = f_2, f_4^*, f_3, f_8^*, f_6^*, f_1, f_7, f_5$

$P(c_6) = f_8, f_3^*, f_6^*, f_1, f_4^*, f_7, f_5, f_2$

$P(c_7) = f_3, f_1, f_6^*, f_2^*, f_8^*, f_5, f_7, f_4$

$P(c_8) = f_1, f_2^*, f_7^*, f_5^*, f_8, f_3, f_6, f_4$

$$\mu_1 = (f_1, c_4), (f_2, c_1), (f_3, c_2), (f_4, c_6), (f_5, c_8), (f_6, c_5), (f_7, c_3), (f_8, c_7)$$

$$\mu_2 = (f_1, c_3), (f_2, c_7), (f_3, c_4), (f_4, c_2), (f_5, c_1), (f_6, c_6), (f_7, c_8), (f_8, c_5)$$

$$\mu_3 = (f_1, c_3), (f_2, c_8), (f_3, c_6), (f_4, c_2), (f_5, c_1), (f_6, c_7), (f_7, c_4), (f_8, c_5)$$

$$\mu_4 = (f_1, c_2), (f_2, c_7), (f_3, c_4), (f_4, c_5), (f_5, c_3), (f_6, c_6), (f_7, c_8), (f_8, c_1)$$

$$\mu_5 = (f_1, c_2), (f_2, c_8), (f_3, c_6), (f_4, c_5), (f_5, c_3), (f_6, c_7), (f_7, c_4), (f_8, c_1)$$

Market I

$P(f_1) = c_8^*, c_7, c_2, c_5, c_3, c_1, c_4, c_6$

$P(f_2) = c_5, c_4, c_3^*, c_2, c_8, c_1, c_7, c_6$

$P(f_3) = c_4^*, c_1, c_6, c_7, c_8, c_3, c_2, c_5$

$P(f_4) = c_2^*, c_4, c_6^*, c_8, c_1, c_7, c_3, c_5$

$P(f_5) = c_6^*, c_2^*, c_5, c_8, c_1, c_7, c_3, c_4$

$P(f_6) = c_4, c_8, c_1^*, c_3, c_7, c_5, c_6, c_2$

$P(f_7) = c_5^*, c_6, c_7, c_8, c_1, c_3, c_2, c_4$

$P(f_8) = c_7^*, c_3, c_8, c_1, c_4, c_2, c_6, c_5$

$P(c_1) = f_1, f_2, f_7, f_5, f_8, f_3, f_6^*, f_4$

$P(c_2) = f_8, f_5^*, f_2, f_7, f_3, f_1, f_6, f_4^*$

$P(c_3) = f_1, f_2^*, f_4, f_5, f_3, f_7, f_8, f_6$

$P(c_4) = f_3^*, f_5, f_1, f_7, f_8, f_6, f_4, f_2$

$P(c_5) = f_7^*, f_4, f_3, f_6, f_1, f_8, f_5, f_2$

$P(c_6) = f_2, f_4^*, f_3, f_8, f_6, f_1, f_7, f_5^*$

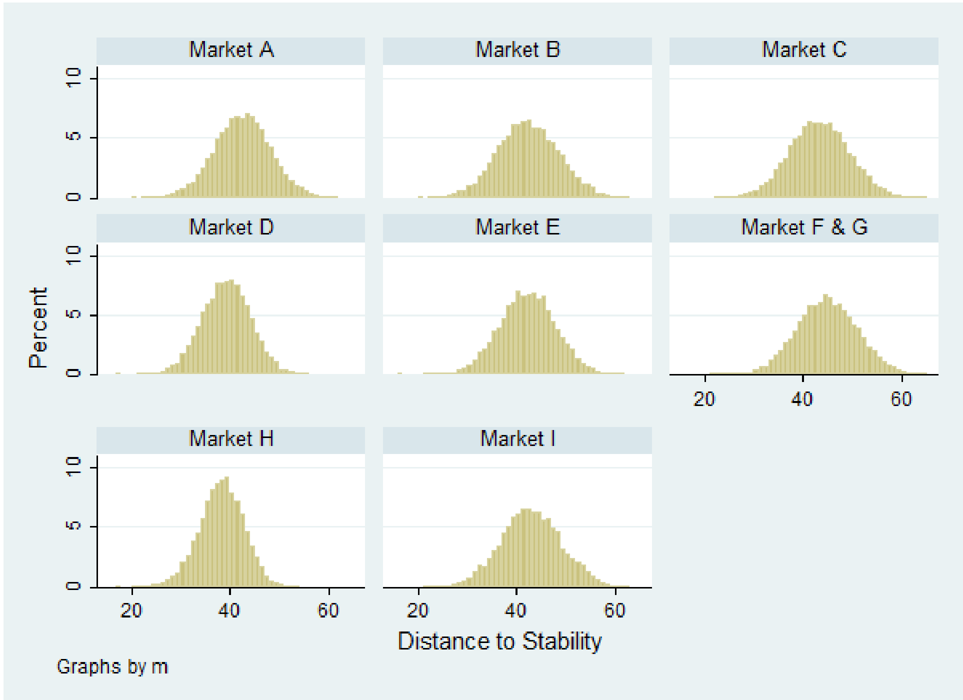
$P(c_7) = f_8^*, f_3, f_6, f_1, f_4, f_7, f_5, f_2$

$P(c_8) = f_3, f_1^*, f_6, f_2, f_8, f_5, f_7, f_4$

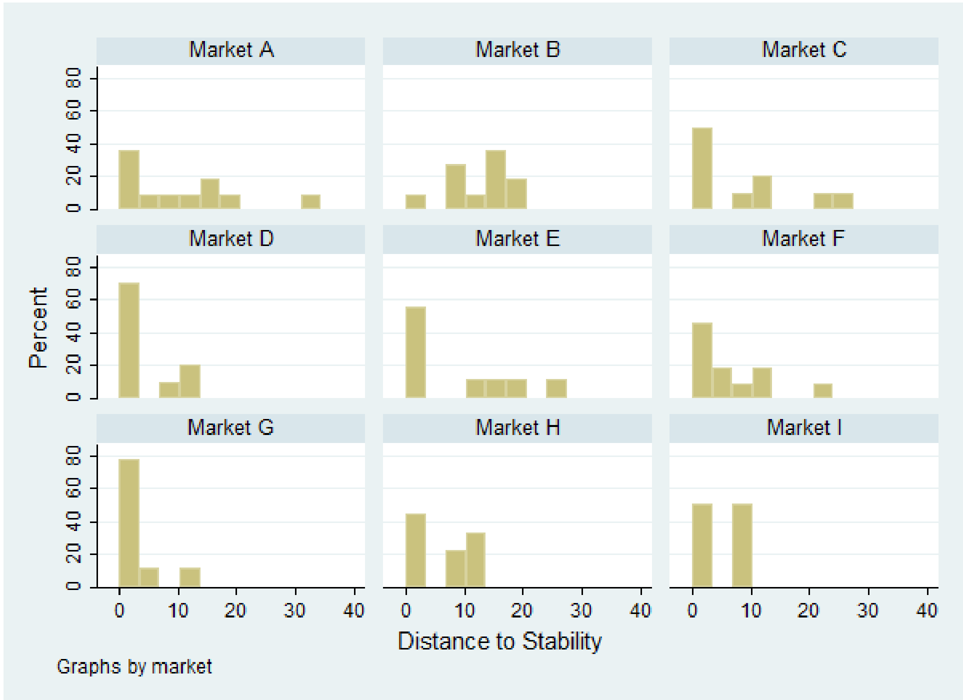
$$\mu_1 = (f_1, c_8), (f_2, c_3), (f_3, c_4), (f_4, c_2), (f_5, c_6), (f_6, c_1), (f_7, c_5), (f_8, c_7)$$

$$\mu_2 = (f_1, c_8), (f_2, c_3), (f_3, c_4), (f_4, c_6), (f_5, c_2), (f_6, c_1), (f_7, c_5), (f_8, c_7)$$

Appendix D



The distance to stability across markets from 10,000 simulations. The simulations were conducted assuming independent and uniformly random play by agents.



The distance to stability across markets from the experimental data.

REFERENCES

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

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

Tingting Ding and Andrew Schotter. Matching and chatting: An experimental study of the impact of network communication on school-matching mechanisms. *Working Paper*.

Federico Echenique, Alistair J Wilson, and Leeat Yariv. Clearinghouses for two-sided matching: An experimental study. *Working Paper*, 2014.

Clayton R Featherstone and Eric Mayefsky. Why do some clearinghouses yield stable outcomes? experimental evidence on out-of-equilibrium truth-telling. *Working Paper*, 2014.

Clayton R Featherstone and Muriel Niederle. Improving on strategy-proof school choice mechanisms: An experimental investigation. *Working Paper*, 2014.

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

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, pages 9–15, 1962.

Glenn W Harrison and Kevin A McCabe. Stability and preference distortion in resource matching: An experimental study of the marriage market. *Research in Experimental Economics*, 8, 1989.

Ernan Haruvy and M Utku Ünver. Equilibrium selection and the role of information in repeated matching markets. *Economics Letters*, 94(2):284–289, 2007.

John H Kagel, Ronald M Harstad, and Dan Levin. Information impact and allocation rules in auctions with affiliated private values: A laboratory study. *Econometrica*, pages 1275–1304, 1987.

Qingmin Liu, George J Mailath, Andrew Postlewaite, and Larry Samuelson. Stable matching with incomplete information. *Econometrica*, 82(2):541–587, 2014.

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

Alex Rees-Jones. When "strategy" hurts: Suboptimal behavior in strategy-proof mechanisms. *Working Paper*, 2014.

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

Alvin E Roth. Misrepresentation and stability in the marriage problem. *Journal of Economic Theory*, 34(2):383–387, 1984.

Alvin E Roth. Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior*, 1(2):191–209, 1989.

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

Alvin E Roth and Uriel G Rothblum. Truncation strategies in matching markets–in search of advice for participants. *Econometrica*, 67(1):21–43, 1999.

Alvin E Roth and Marilda A Oliveira Sotomayor. *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press, 1992.

Alvin E Roth and Xiaolin Xing. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of political Economy*, 105(2):284–329, 1997.

CHAPTER 2

1 Introduction

Two-sided matching theory has informed the design of institutions in areas as diverse as kidney exchange (Roth, Sönmez, and Ünver, 2004), entry-level labor markets (Roth and Peranson, 1999), and school choice (Abdulkadiroğlu and Sönmez, 2003). These institutions often operate as centralized clearinghouses, in which participants submit rank-order lists of their preferences and then a particular algorithm selects the final outcome (i.e., who is paired with whom). In this context, a widely-used matching algorithm is the Gale-Shapley deferred acceptance algorithm.[1,2] In the DA algorithm, the market is divided into "proposers" and "receivers." This algorithm has an important property: if all agents submit their true preferences, then the resulting outcome is stable and is also the most preferred stable outcome for the proposing side of the market.[3] In the DA algorithm, it is well-known that the proposers have a dominant strategy of truth-telling (Dubins and Freedman, 1981). The receivers, on the other hand, might have incentives to misrepresent their preferences to produce a more favorable outcome for themselves (Gale and Sotomayor, 1985).

We investigate whether - and under what conditions - receivers behave strategically in the preference-revelation game induced by the DA algorithm. We focus attention on a particular class of strategic behavior: truncation strategies (i.e., submitting a shortened preference list that otherwise maintains the order of the true

[1]The algorithm was first introduced by Gale and Shapley (1962).

[2]Henceforth, DA algorithm.

[3]A matching is said to be stable if no agent prefers remaining unmatched to her current allocation and no pair of agents mutually prefer each other to their current allocations.

preferences). This emphasis arises for two reasons. First, truncation strategies are intuitively appealing and simple for agents to implement. Second, when evaluating an agent's profitable misrepresentation opportunities in the DA algorithm, it suffices to restrict attention to truncation strategies. In other words, misrepresenting one's true preferences in a manner other than truncation can do no better than what can be achieved via truncation (Roth and Peranson, 1999).

To make progress on this question, we first characterize the conditions under which a receiver *acting on her own* can secure a match that is no worse than her most preferred achievable partner.[4] Our proposition is a straightforward extension of a classic result: in markets with more than one stable matching, there will be an incentive for some receiver to truncate her preferences whenever all other agents report their preferences truthfully (Gale and Sotomayor, 1985). Although an agent's optimal truncation is a function of the profile of other agents' reported preferences, we show that no direct knowledge of the strategies of other agents is required. Rather, it merely suffices for other receivers to be constrained to truncation strategies to be able to calculate the best response. In general, the optimal truncation strategy will change if other receivers are allowed to play more general misrepresentation strategies since these strategies can substantially alter the set of stable matchings. In those situations, it is possible to be optimally truncating with respect to the true preferences but sub-optimally truncating with respect to the reported preferences. While not breaking new ground, this result is methodologically important for our experimental design. It allows us to construct environments that maintain the key interactive features of matching markets while essentially reducing optimal truncation to a decision-theoretic problem.

[4]Two individuals are said to be achievable for each other if they are paired at some stable matching.

Even after removing this aspect of strategic uncertainty, there are two practical difficulties that present themselves with respect to optimal truncation. First, an agent might be unable to identify the existence of a profitable opportunity to misrepresent her preferences. Second, an agent might over-truncate her preferences and remain unmatched (her worst possible outcome).[5] In a laboratory experiment, we investigate whether truncation depends on the magnitude of the potential monetary gains from truncation as well as the rank of the most preferred achievable partner in an agent's preferences. The first measure is important since it is only when a profitable opportunity exists that an agent has an incentive to truncate her preferences. The second measure is important since the rank of the most preferred achievable partner determines the likelihood of remaining unmatched by mistakenly over-truncating.

To mirror the theoretical conditions, the experiment is conducted in an environment with complete information about other agents' preferences. The proposing side of the market is automated to play its dominant strategy of truthful preference revelation. The experimental subjects play in the role of the receivers and they are restricted to either truth-telling or truncation strategies. Importantly, since we have removed the element of strategic uncertainty over other players' actions, the *only* risk associated with truncation in our environment comes from over-truncating. Our experiment tests whether agents truncate their preferences in situations that are the most conducive to truncation behavior.[6] Ideally, the simplicity of our environment would provide insight into the reasons why market participants choose to either behave straightforwardly or strategically.

[5]Over-truncation refers to the situation where an agent truncates "too much" and leaves her most preferred achievable partner off her submitted rank-order list.

[6]Another factor that makes truncation more attractive in our experiment is the linearity in subject payoffs. In most applications, it is reasonable to expect a discontinuity in utility between matching with one's least preferred partner and remaining unmatched. If the risk of remaining unmatched is important in our environment, then it is likely to be even more important in field settings where remaining unmatched is very costly.

We find truth-telling to be the most common strategy in our experimental markets: 56% (511/920) of submitted rank-order lists are identical to subjects' true preferences. We also find that truncation is not sensitive to considerations of profitability, but is sensitive to the rank of the most preferred achievable partner. This result is robust to alternative specifications. We consider this to be remarkable given the difficulty in identifying achievable match partners even in small markets.

Regarding aggregate outcomes, 88% (203/230) of our experimental markets culminate in stable outcomes. This is not only due to the fact that truth-telling is common, but also because over-truncation is rare.[7] We also find that final outcomes are closer to the receiver-optimal stable matching than to the proposer-optimal stable matching. However, this result is not entirely surprising. Since strategic behavior has positive spillover effects in our environment, the receiver-optimal stable outcome can be attained when only a subset of agents truncates its preferences optimally.

A useful benchmark to measure the success of centralized matching clearinghouses is their ability to produce stable outcomes.[8] The hallmark of a stable matching mechanism is that, for any profile of reported preferences, it produces an outcome that is stable with respect to the reported preferences. However, understanding *which* stable outcome arises in markets with multiple stable outcomes is no less important than the question of *whether* a stable outcome arises. The issue of equilibrium selection has important welfare consequences since the interests of the two sides of the market are diametrically opposed on the question of which stable matching to implement.[9] This is a relevant consideration for policymakers, who may have reasons to favor the

[7]Due to the constrained nature of our strategy space, over-truncation is the only way to observe instability in final outcomes.

[8]Mechanisms that produce unstable outcomes necessarily give some participants an incentive to seek out alternative match partners after the market closes. In fact, centralized clearinghouses based on unstable matching mechanisms often perform no better than the decentralized markets that they replace (Roth, 1991).

[9]This result is a consequence of the fact that the set of stable matchings is a lattice.

welfare of one side of the market over another when designing matching markets. In May 1997, for instance, the National Resident Matching Program (NRMP) switched from the hospital-proposing version of the DA algorithm to the student-proposing version over concerns that the original design unduly favored hospitals at the expense of students.

Empirically, it is also important to determine whether the DA mechanism approaches strategy-proofness in practice. By providing a level playing field for all participants, regardless of their institutional knowledge or strategic reasoning abilities, strategy-proof mechanisms can help assuage the concerns of market participants and promote market "thickness". If receivers generally play truth-telling strategies even in situations where there are gains from preference misrepresentation, then this fact could partially explain the success and persistence of the DA mechanism in the field.

Finally, our work highlights the complementarity between controlled laboratory experimentation and market design. In the field, data on submitted rank-order lists is often available but participants' underlying preferences are not observed. This makes the extent of strategic behavior difficult to estimate. By allowing us to directly control for subjects' preferences and other market features, the laboratory setting is ideally suited for answering questions related to both strategic behavior and equilibrium selection. Our results show that subjects respond to market features, but not necessarily in the ways suggested by theory. In particular, the finding that subjects respond to the riskiness of strategic behavior suggests that behavioral insights can play an important role in the field of market design.

There is a growing body of experimental work studying the performance of centralized matching mechanisms in the lab. However, much of this experimental literature focuses on the DA algorithm as it relates to the school choice problem (e.g. Chen and Sönmez, 2006; Ding and Schotter; Featherstone and Niederle, 2014; Pais and Pintér, 2008). In these studies, strategic agents exist only on the proposing side of the market. There has been relatively little experimental work done on the strategies pursued by the *receiving* side of the market.[10] This is an important gap to fill: only the receivers in the DA algorithm face substantive strategic questions. In addition, their ability to behave strategically - either in isolation or as a group - can have large effects on market outcomes and participants' welfare.[11]

Our work is most closely related to Featherstone and Mayefsky (2014), which to the best of our knowledge is the only laboratory experiment studying the DA algorithm to automate the proposing side of the market in order to focus exclusively on the strategies pursued by the receiving side. They interpret "out-of-equilibrium truth-telling" as a reason for the success and persistence of the DA mechanism despite being manipulable in theory. However, our paper departs from their design in that we introduce a novel experimental framework with which to study truncation strategies. The advantage of our approach lies in the fact that, by studying a restricted version of the same problem, we have created an environment in which some form of truncation is always a best response. Although our work addresses the optimal truncation problem in a complete information environment, it can also be viewed in the same spirit as Roth and Rothblum (1999), which addresses the question of what practical advice can be given to market participants in the context of a centralized matching clearinghouse based on the DA algorithm. They show that any non-truncation strategy is stochastically dominated by a truncation strategy in symmetric, incomplete

[10]Existing studies, however, report high rates of truth-telling by receivers (Echenique, Wilson, and Yariv, 2014; Featherstone and Mayefsky, 2014; Harrison and McCabe, 1989).

[11]However, it should be noted that "core convergence" results for large matching markets imply that there is limited scope for strategic behavior in this context. We will return to this issue when discussing the implications of our main findings.

information environments.[12]

A critical question that remains is whether truncation behavior has theoretical or empirical relevance. While the literature on "core convergence" suggests that there is little scope for strategic misrepresentation in large markets (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Lee, 2014), there are important qualifications to these results. Coles and Shorrer (2014), for instance, show that while the utility gain from optimal truncation may be small, the optimal degree of truncation can still remain quite large. In fact, when an agent has uniform beliefs regarding the reported preferences of others, the optimal truncation approaches 100% of her list as the size of the market grows.[13]

The paper is organized as follows. Section 2 provides theoretical background, Section 3 describes our experimental design, Section 4 presents results, and Section 5 discusses broader implications and concludes.

2 Theory

In this section, we introduce the theoretical framework that informs our experimental design. We first review some basic results from two-sided matching theory that are necessary for this purpose. For a more detailed survey, see Roth and Sotomayor (1992).[14] Consider two finite, disjoint sets M and W, where M is the set of men and W is the set of women. Each agent has complete and transitive preferences over the agents on the other side of the market (as well as remaining single). The preferences of man m will be represented by an ordered list of preferences $P(m)$ on

[12] Ehlers (2008) generalizes this result from deferred acceptance mechanisms to a much larger class of mechanisms.

[13] "Uniform beliefs" refers to the case where an agent believes that the reported preferences of others are chosen uniformly and randomly from the set of all possible full-length reported preference lists.

[14] Proofs of most of the cited results can be found there.

the set $W \cup \{m\}$. Similarly, the preferences of woman w will be represented by an ordered list of preferences $P(w)$ on the set $M \cup \{w\}$. We write $w \succ_m w'$ to denote that m prefers w to w', and $w \succeq_m w'$ to denote that m likes w at least as much as w'. Similarly, we can write $m \succ_w m'$ and $m \succeq_w m'$. Woman w is said to be ***acceptable*** to man m if he likes her at least as much as remaining single (i.e., $w \succeq_m m$). Similarly, m is acceptable to w if $m \succeq_w w$.

Let $\mathbf{P}$ denote the set of all preferences, one for each man and one for each woman. A marriage market is denoted by the triplet $(M, W, \mathbf{P})$. A matching is a function $\mu : M \cup W \longrightarrow M \cup W$ such that

1. for any $m \in M$, $\mu(m) \in W \cup \{m\}$
2. for any $w \in W$, $\mu(w) \in M \cup \{w\}$
3. for any $m \in M$, $w \in W$, $\mu(m) = w$ if and only if $\mu(w) = m$

Throughout the analysis, we also distinguish between market-wide matchings (represented by μ) and a given individual's match partner. For woman w at the matching μ, her match partner is represented by $\mu(w)$. For each individual, their preference over two alternative matchings corresponds exactly to their preference over their match partners at the two matchings.

A matching μ is ***individually rational*** if every individual is matched to an acceptable partner. A pair of agents (m, w) is said to ***block*** a matching μ if they are not matched to one another at μ but they prefer each other to their assignments at μ (i.e., $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$). A matching μ is ***stable*** if it is individually rational and not blocked by any pair of agents. A stable matching is called an ***M-optimal stable matching*** (denoted μ_M) if every man likes it at least as well as any other stable matching. A ***W-optimal stable matching*** can be defined analogously (denoted μ_W). The M-optimal stable matching is thus the "best" stable matching for the men and the W-optimal stable matching is the "best" stable matching for the women. A man m and a woman w are said to be ***achievable*** for each other in a marriage market $(M, W, \mathbf{P})$ if they are matched to each other at some stable matching. For woman w, $\mu_W(w)$ is her most preferred achievable partner.

Gale and Shapley (1962) proved the following result:

Theorem 1: *A stable matching exists for every marriage market.*

In their constructive proof of the existence of stable matchings, Gale and Shapley (1962) developed a "deferred acceptance" procedure that produces one of the two extremal stable matchings for any preference profile. In their algorithm, the market is divided into two groups: "men" (proposers) and "women" (receivers). Initially, all the men and women are unmatched. The algorithm then goes through several stages where men and women take turns in making decisions. In a generic stage, each unmatched man makes an offer to his most preferred woman among the set of women that he has not previously made an offer to. Each woman then views all the offers she has received in that stage and tentatively accepts her most preferred offer among the new offers and any tentatively accepted offer that she is still holding from a previous stage. The algorithm ends when there are no men left to make offers. This can happen because (1) all men are matched or because (2) the only unmatched men have already been rejected by all of the women. The tentative matches that are in place when the algorithm ends become the final matches. This leads directly to the following result:

Theorem 2: *When all men and women have strict preferences, there always ex-*

ist an M-optimal stable matching and a W-optimal stable matching. The matching produced by the deferred acceptance algorithm with men proposing is the M-optimal stable matching. The W-optimal stable matching is the matching produced by the algorithm when the women propose.

A related result, often referred to as the "lone wolf" theorem, will prove useful later in our analysis:

Theorem 3: *In a market* $(M, W, \boldsymbol{P})$ *with strict preferences, the set of people who are single is the same for all stable matchings.*

To examine the strategic issues involved in two-sided matching markets, we analyze the preference-revelation game in which each man m with preferences $P(m)$ is faced with the strategy choice of what preference ordering $Q(m)$ to state, and likewise for the women. Denote the set of stated preference lists, one for each man and one for each woman, by $\mathbf{Q}$. The mechanism then computes a matching $\mu = h(\mathbf{Q})$, where h is the function that maps any set $\mathbf{Q}$ of stated preferences into a matching. A mechanism h that for any stated preferences $\mathbf{Q}$ produces a matching $h(\mathbf{Q})$ that is stable with respect to the stated preferences is called a stable mechanism. If $h(\mathbf{Q})$ produces the M-optimal stable matching with respect to $\mathbf{Q}$, then h is called the M-optimal stable mechanism. The next theorem highlights an important negative result:

Theorem 4: *No stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent.*

However, it is possible to arrange the market in such a way that only one side faces strategic questions. This is summarized by the following theorem:

Theorem 5: *The M-optimal stable mechanism makes it a dominant strategy for each man to state his true preferences.*

Combining these results suggests that, under the M-optimal stable mechanism, it is the women who will sometimes have a profitable deviation by misrepresenting their true preferences. This is formalized below:

Corollary 1: *When preferences are strict and the M-optimal stable mechanism is employed, there will be an incentive for some woman to misrepresent her preferences whenever more than one stable matching exists.*

Consider a marriage market characterized by $(M, W, \mathbf{P})$ in which preferences are strict and there is more than one stable matching. Let μ_M denote the M-optimal stable matching and μ_W denote the W-optimal stable matching under the true preferences $\mathbf{P}$. With slight abuse of notation, we denote the last man on the preference list $P(w)$ of woman w by $\underline{P}(w)$. Furthermore, we confine attention to markets in which each agent prefers being married to remaining single (all men are acceptable to all women and vice versa) and $|M| = |W|$. Without loss of generality, the theoretical results below are framed in terms of the incentives facing the women in the revelation game induced by the man-proposing deferred acceptance algorithm. A symmetric argument holds for men when the woman-proposing deferred acceptance algorithm is used.

We will find it useful to define two classes of strategies for the women in this market:

Definition 1: *A* ***truncation*** *of a preference list $P(w)$ containing k acceptable men is a list $P'(w)$ containing $k' \leq k$ acceptable men such that the k' elements of $P'(w)$ are the first k' elements of $P(w)$, in the same order.*

Definition 2: *A* ***manipulation*** *of a preference list $P(w)$ is any list that is not a truncation of $P(w)$.*

A truncation strategy involves misrepresenting your preferences by shortening the list of acceptable matches without changing their order. For convenience, we allow for truth-telling to trivially satisfy the definition of a truncation strategy. A manipulation strategy involves misrepresenting preferences by changing the order of preference between at least two men (regardless of the length of the list). We now define three particular types of truncation strategies that are central to our analysis:

Definition 3: *An* ***over-truncation*** *of a preference list $P(w)$ is a truncation of $P(w)$ that does not contain $\mu_W(w)$, the most preferred achievable partner of woman w.*

Definition 4: ***Optimal truncation*** *of a preference list $P(w)$ is a truncation of $P(w)$ that contains $\mu_W(w)$ but does not contain any men who are ranked below $\mu_W(w)$.*[15]

Definition 5: *An* ***under-truncation*** *of a preference list $P(w)$ is a truncation of*

[15] To be clear, it is still possible for an agent to achieve the optimal equilibrium result (being matched to her most preferred achievable partner) without optimal truncation. However, it is convenient to define optimal truncation in this manner.

$P(w)$ that contains $\mu_W(w)$ but also contains at least one man who is ranked below $\mu_W(w)$.

By submitting a truncated preference list, an agent is effectively telling the mechanism to play a threshold strategy on her behalf (i.e., to reject all offers below a certain cutoff). With truncation, agents face a balance of risks: the likelihood of remaining unmatched increases, while conditional on matching the likelihood of being matched to a more favorable partner increases. The risks associated with truncation can arise from two sources: over-truncation and uncertainty regarding other agents' actions. This is a subtle point that deserves clarification. Optimal truncation requires an agent to possess a great deal of information on the preferences of other agents and the ability to calculate or otherwise identify her most preferred achievable partner. A mistake in this calculation could result in over-truncation. If an agent over-truncates, then this opens up the possibility of remaining unmatched.[16]

However, even if an agent is able to correctly identify her most preferred achievable partner and then truncate optimally, it is still possible for her to remain unmatched depending on the actions of other agents. If other agents distort their preferences in a manner that changes the set of stable outcomes, then it is possible for a particular woman to be optimally truncating with respect to the *true* preferences but over-truncating with respect to the *stated* preferences. This naturally leads to the question of what restrictions need to be placed on other agents' strategies to prevent this from happening.

In the context of the man-proposing deferred acceptance algorithm, we can now state and prove the following results: Consider a marriage market in which prefer-

[16] An agent who over-truncates is hurting herself but is also helping the other agents on her side of the market. Thus, even with over-truncation, it is possible for an agent to be matched if at least one other agent over-truncates.

ences are strict and there is more than one stable matching. Suppose that $|M| = |W|$ and all men are acceptable to all women (and vice versa). Let $\mathbf{Q}$ be a profile of stated preferences in which each man states his true preferences, and each woman w states a list $Q(w)$ that constitutes a truncation of $P(w)$ but not an over-truncation. Then the following statements are true:

1. No woman w will remain single.

   Proof. See Appendix A. $\square$

2. The set of stable matchings under $\mathbf{Q}$ is a subset of the set of stable matchings under $\mathbf{P}$.

   Proof. See Appendix A. $\square$

3. Each woman w can truncate in such a way as to be matched to $\mu_W(w)$, her most preferred achievable partner under the true preferences $\mathbf{P}$.

   Proof. See Appendix A. $\square$

If we do not restrict attention to Nash equilibrium strategy profiles (and hence permit outcomes that are unstable with respect to the agents' true preferences), we can make an even stronger statement regarding the conditions under which it is advisable for an individual agent to play a truncation strategy. This is formalized in the following proposition: Consider a marriage market in which preferences are strict and there is more than one stable matching. Suppose that $|M| = |W|$ and all men are acceptable to all women (and vice versa). Let $\mathbf{Q}$ be a set of preferences in which each man states his true preferences, and each woman in $W \setminus \{w\}$ states a list that is *not* a manipulation of her true preferences. Then woman w can truncate her preference list in such a way as to be matched to a man she likes at least as much as $\mu_W(w)$, her most preferred achievable partner under the true preferences $\mathbf{P}$.[17]

Proof. See Appendix A. $\square$

An alternative characterization of Proposition 2 is as a dominant-strategy result for a modified matching game with a pruned strategy space. Suppose a woman found herself playing the preference-revelation game induced by the man-proposing DA algorithm. Suppose further that the woman knew the identity of her most preferred achievable partner. Would it be advisable for this woman to truncate her preferences by leaving off all men ranked below her most preferred achievable partner? In general, the answer to this question would depend on the woman's risk attitudes and beliefs about other agents' actions. However, Proposition 2 provides the conditions on other players' strategies such that the answer to this question is unambiguously yes.

Proposition 2 is at the heart of our experimental design. Our environment contains automated, truthful proposers and also constrains the set of strategies that are available to receivers. This approximation of a decision-theoretic setting allows us to conveniently test for truncation behavior without worrying about the need to coordinate behavior with other agents and the heterogeneity of beliefs over other agents' actions. As detailed in the next section, our experimental design systematically manipulates the profitability of truncation (i.e., the magnitude of the monetary gain from truncation) and the riskiness of truncation (i.e., the likelihood of over-truncation).

There are two points worth emphasizing. First, we should only expect *behavioral* agents to be responsive to the magnitude of the monetary gain from truncation and the likelihood of over-truncation. Sophisticated agents who have the ability to calculate the set of stable outcomes should only be responsive to the *existence* of a

[17]This proposition is a straightforward extension of a result from Gale and Sotomayor (1985).

profitable strategic opportunity: they should (optimally) truncate their preferences only if they have more than one achievable partner. Second, these features that we identify are only relevant under the conditions imposed in our experiment. If an agent faces strategic uncertainty about the actions of other agents, then these notions lose much of their value. For instance, it could be the case that an agent is optimally truncating with respect to the true preferences but over-truncating with respect to the stated preferences.

In the context of our experiment, we can now cast the following hypotheses:

Hypothesis 1: *Truncation behavior will be increasing in the profitability of truncation.*

Hypothesis 2: *Truncation behavior will be decreasing in the riskiness of truncation.*

3 Experimental Design

In the experiment, the two sides of the market are labeled "firms" (proposers) and "workers" (receivers). Each experimental market consists of four subjects. Each experimental session contains either one or two parallel experimental markets (thus each session consists of either four or eight subjects). The roles of the firms are automated: they are programmed to play their dominant strategy of truth-telling. Fixing the behavior of firms in this fashion is necessary in order to test our main proposition. Each subject is randomly assigned to the role of one of the four workers. Their assigned role remains the same throughout the experiment.

Upon arriving at the lab, subjects read and sign an informed consent document.

The experimenter then reads aloud the experimental instructions.[18] Before the experiment begins, each subject is required to work through a demonstration of the DA algorithm and answer relevant questions. We use a hypothetical set of reported preferences that includes examples of both truth-telling and truncation. To secure comprehension, we do not proceed with the actual experiment until all the subjects complete the demonstration and answer the questions correctly. The relevant screen shots from the demonstration are included in Appendix B.

Subjects play 10 rounds of the preference-revelation game induced by the firm-proposing DA algorithm. In each round, subjects observe the payments that they (and the other subjects) will receive from matching with the different firms. They also observe the order in which the firms will be making offers to match with the workers in the DA algorithm. The action that subjects take in each round is to choose which message (i.e., ranking of the firms) to submit to the computer to be used in the matching process.[19] Subjects are required to spend a minimum of three minutes on this task in each round: if all subjects are done sooner than that, they still have to wait until the full three minutes have elapsed. A representative screen that subjects face during the experiment is shown in Figure 1.

At the end of each round, subjects are informed of the identity of their match partner and their payoff for that round. While each subject's role remains the same throughout the 10 rounds, each round corresponds to a different matching market (i.e., the agents are endowed with a different set of preferences). The particular preference profiles used in the experiment are included in Appendix C. At the end of the experiment, 1 of the 10 rounds is randomly selected and subjects are paid based

[18]The full set of instructions is included after the appendices.

[19]Importantly, the terminology of "preferences" is never used in the experimental instructions or the experimental interface. Subjects' true preferences are referred to as payments and subjects' reported preferences are referred to as submitted messages or rankings. This caution was taken to reduce experimenter demand effects.

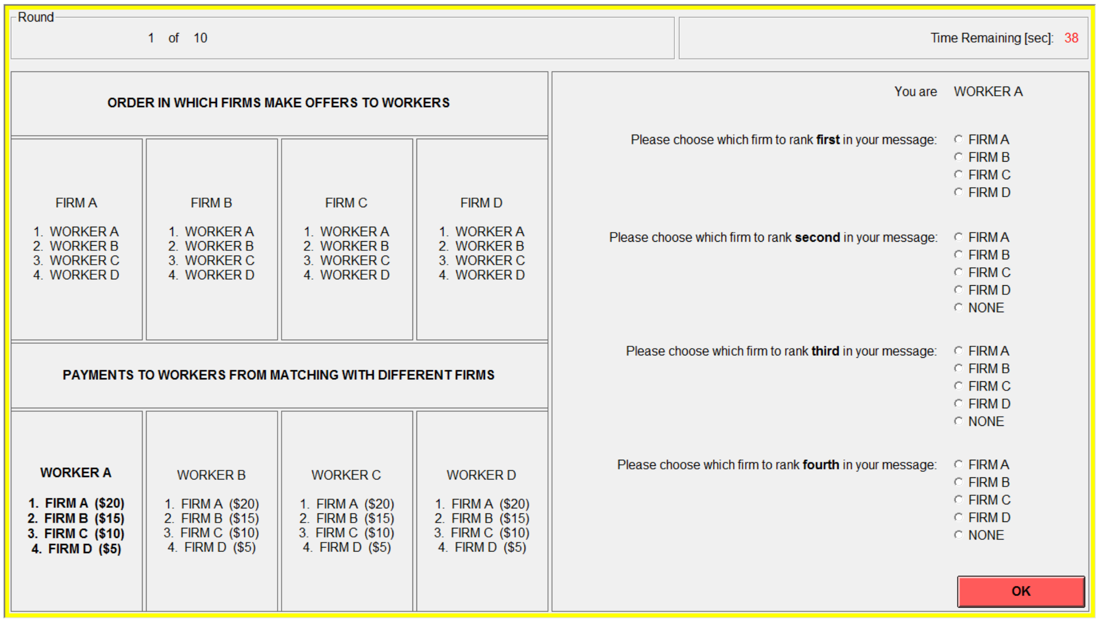


Figure 1: An example of our experimental interface. This is the screen that WORKER A observes in Round 1 of the experiment.

on their match partners in that round (in addition to a fixed $5 show-up payment).[20] Matched subjects earn anywhere from $5 to $20 in increments of $5 (depending on whether they matched with their most preferred, second most preferred, third most preferred, or least preferred firm). Unmatched subjects earn $0 for that round.

Proposition 2 establishes the optimality of truncation when other workers refrain from manipulating their preferences (i.e., they do not switch their order of preference between firms). In that sense, strategic behavior in the context of the DA algorithm can be viewed as a coordination problem: a worker can best-respond by truncating her preferences only if other workers are also truncating (or truth-telling). In

[20]The choice of payment procedure is still an open question in the field of experimental economics. Advantages and disadvantages of competing approaches, and the theoretical conditions under which they can be justified, are discussed in Azrieli, Chambers, and Healy (2014).

Table 1: Our within-subject experimental design varies the strategic incentives that subjects face in terms of the profitability and riskiness of truncation.

	risky	not risky
not profitable	$P(w) = f_2, f_4, f_1, {f_3}^*$	$P(w) = {f_2}^*, f_4, f_1, f_3$
profitable	$P(w) = f_2, {f_4}^*, {f_1}^*, f_3$	$P(w) = {f_2}^*, {f_4}^*, {f_1}^*, {f_3}^*$

our experiment, we solve this coordination problem by restricting subjects to either truth-telling or truncation. Thus, in our 4x4 experimental markets, each subject has four pure strategies. Their decision problem consists of choosing the length of their submitted rank-order list.[21] Subjects who attempted to submit a manipulation of their preferences or who left an empty position in the middle of their rank-order list received appropriate error messages on their screens.

However, even controlling for the behavior of other agents, optimal truncation is still a practical challenge. For an agent to optimally truncate, it requires (1) the ability to identify the existence of a profitable strategic opportunity and (2) the ability to identify her most preferred achievable partner. Having controlled for other agents' behavior, the *only* risk associated with truncation in our experiment is the possibility of over-truncation (which could result in remaining unmatched). We use a within-subject experimental design to investigate whether truncation behavior is correlated with the profitability and riskiness of truncation.

By **profitability**, we refer to the ordinal distance between a worker's most preferred and least preferred achievable firms in her preference list (i.e., the span of the core).[22] If a worker has a unique achievable firm, then truncation can do no better

[21]The same number of mouse clicks was required to submit a rank-order list, regardless of length. This was done to ensure that subjects do not perceive truncation to be (marginally) more convenient or easier than truth-telling.

[22]This measure is distinct from the number of achievable partners that an agent has. Rather, it can be thought of as a measure of the potential monetary gain from optimal truncation compared

- and in fact can do worse - than truth-telling. If a worker has multiple achievable firms, then her optimal strategy is to submit a truncation of her true preferences by leaving off all firms that are ranked below her most preferred achievable firm.[23]

By **riskiness**, we refer to the ranking of a worker's most preferred achievable firm in her preference list. If a worker's most preferred achievable firm coincides with her most preferred firm overall, then there is no possibility of mistakenly over-truncating. On the other hand, if a worker's most preferred achievable firm coincides with her least preferred firm overall, then *any* truncation is an over-truncation and carries with it the possibility of remaining unmatched.

Table 1 illustrates how our within-subject experimental design systematically varies the profitability and riskiness of truncation. Note that the worker's achievable firms are denoted by asterisks (*) in her preference list. In the top left box of Table 1, it is both unprofitable and risky for worker w to submit a truncation of her true preferences. There is a unique achievable firm (so there is no benefit to misrepresenting preferences) and furthermore *any* truncation will be an over-truncation. In the bottom right box of Table 1, it is both profitable and risk-less for worker w to truncate her preferences. By truth-telling, worker w will be matched to firm f_3, but by optimally truncating her preferences worker w will be matched to firm f_2.[24] Since her most preferred achievable firm is also her most preferred firm overall, there is no chance of mistakenly over-truncating and remaining unmatched.

For convenience, we define indices for profitability and riskiness that we refer to throughout the remaining analysis. We measure the profitability of truncation

to truth-telling.

[23] In fact, any truncation that does not include a worker's second most preferred achievable firm will yield the same result. Thus, there is no unique optimal truncation strategy in certain markets.

[24] These statements assume that all other workers submit their true preferences. If other workers are truncating, this can sometimes result in worker w being matched to a more favorable partner than f_3 even if she behaves truthfully. In other words, truncation behavior in this environment has positive externalities.

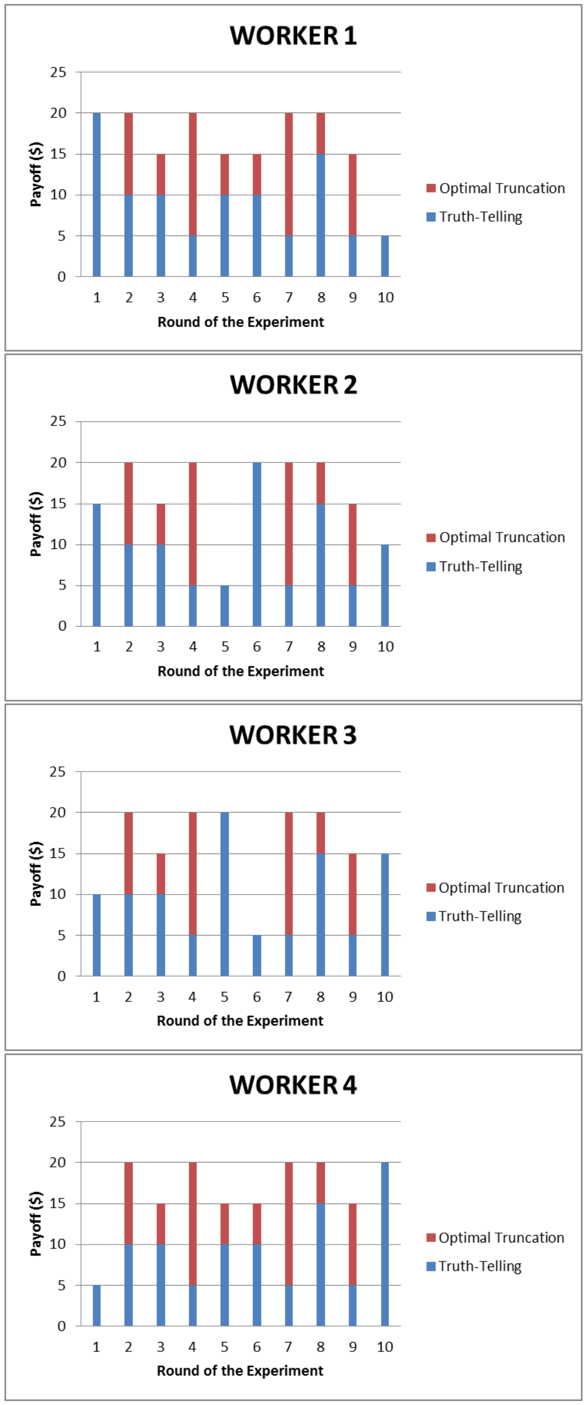


Figure 2: The payoff difference between truth-telling and optimal truncation across rounds of the experiment.

on an integer scale from 0-3, representing the ordinal distance between an agent's most preferred and least preferred achievable partners in her preference list. For a worker with a unique achievable firm, the profitability of truncation is thus coded as a "0". Similarly, we measure the riskiness of truncation on an integer scale from 1-4, representing the ranking of an agent's most preferred achievable partner in her preference list. For a worker whose most preferred achievable firm coincides with her most preferred firm overall, there is no possibility of over-truncation and the riskiness of truncation is coded as a "1".

Another characterization of profitability is given in Figure 2, which presents the payoff difference between truth-telling and optimal truncation across all experimental rounds. This difference is calculated under the assumption of truthful reporting by all other subjects. However, calculating the optimal truncation strategy is a difficult problem. It is natural to ask how profitability is perceived by a naive agent who chooses a truncation level (corresponding to a "cut point" in her preferences) and plays it consistently throughout all experimental rounds. Appendix D shows the expected payoffs in the experiment for this hypothetical subject in different roles.[25] Even for a subject who does not optimally best-respond, the strategic tension is apparent: on average, truncation will increase a subject's payoff up until the most extreme truncation strategy. Thus, our experimental design allows for significant amounts of truncation to be profitable and for the gains from truncation to be realized by naive agents.

Clearly, the measures that we use for profitability and riskiness are correlated. In fact, approximately 31% of the variability in the profitability index is shared with the riskiness index in our experimental markets. When discussing the results, we will

[25]As before, the expected payoff is calculated under the assumption of truthful reporting by all other subjects.

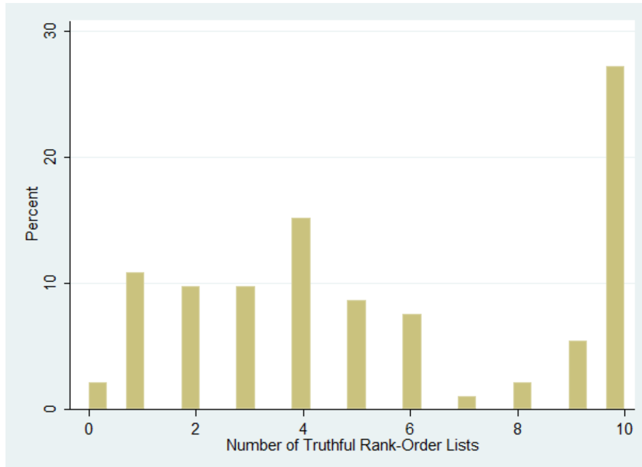


Figure 3: Distribution of the number of truthful reports.

use regression analysis to tease out the effects of the variables in isolation. However, a strength of our experimental design lies in the fact that profitable and risk-less opportunities for truncation are not clustered near the beginning nor end of the experiment, but rather are spread uniformly throughout. Appendix D lists key features of the markets in the order that they are presented to subjects.

Experiment Implementation

The experimental sessions were conducted from June-October 2014 at the ICES Experimental Economics Laboratory of George Mason University. A total of 92 subjects participated in the experiment. Experimental subjects were recruited via email from a pool of George Mason University undergraduates who had all previously registered to receive invitations for experiments. Each experimental session lasted approximately 90 minutes. Subject payments ranged from \$5 to \$25 (including a \$5 show-up payment). The experiment was programmed and conducted with the software z-Tree

(Fischbacher, 2007).

4 Experimental Results

The rest of this section proceeds as follows. We begin by analyzing individual behavior, and then move on to market-level outcomes.

Individual Behavior

We first analyze the basic decision of whether to report preferences truthfully or to behave strategically (i.e., truncate preferences). We find that truth-telling is common in our experimental markets: 56% (511/920) of submitted rank-order lists coincide with agents' true preferences. Since subjects have four pure strategies, uniformly random behavior would imply a truth-telling rate of 25%. We find that truth-telling occurs significantly more often than random chance would predict ($\chi^2(1) = 457.740$, p-value < 0.001). Each individual subject submits a total of 10 rank-order lists throughout the experiment (one for each experimental round). Figure 3 shows the distribution of the number of truthful rank-order lists in our experimental data. Twenty-seven percent (25/92) of subjects consistently reported their true preferences; the remaining subjects truncated their preferences in at least one round.

A natural question that emerges from this analysis is whether subjects become more strategic with market experience. Figure 4 shows truncation rates across the rounds of the experiment. We use an extension of the Wilcoxon rank-sum test to non-parametrically test for any trend in truncation rates across experimental rounds. We find that truncation rates systematically increase across the rounds of the experiment ($z = 5.090$, p-value < 0.001).[26]

[26]It should be noted that the incentive to truncate does not exist in all experimental rounds.

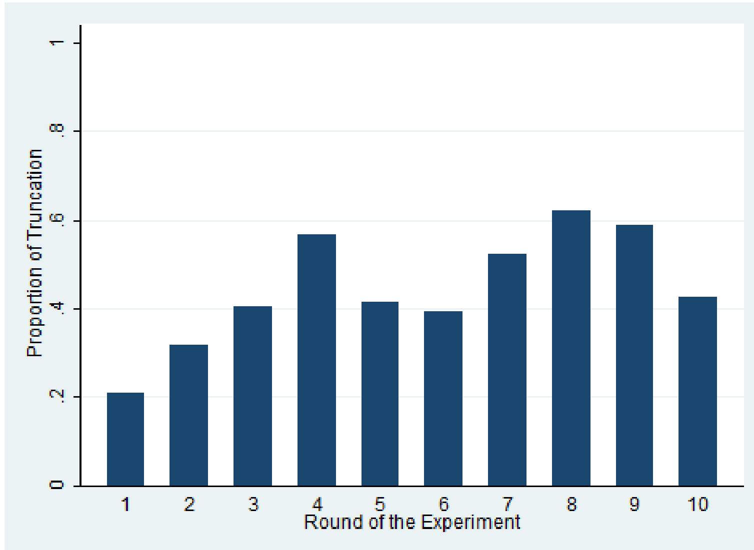


Figure 4: Truncation rates across the rounds of the experiment.

Tables 2 and 3 show a more detailed breakdown of the strategies found in our experimental data. Table 2 shows the breakdown according to the length of the submitted rank-order list (i.e., how many firms were included in the ranking), while Table 3 shows the breakdown according to the degree of truncation (i.e., over-truncation, optimal truncation, and under-truncation). For this latter purpose, it is convenient to classify *all* strategies as either over-truncation, optimal truncation, or under-truncation.[27] In both tables, we include the distribution of strategies derived from random behavior (i.e., if subjects were to randomize uniformly among their four pure strategies) alongside the distribution from our experimental data. When comparing the empirical distribution with the random distribution, we find a significant dif-

However, this result is unchanged when we confine attention to cases where the span of the core is non-zero (z = 4.270, p-value < 0.001).

[27] If an agent's unique achievable partner is ranked last in her preferences, then truth-telling qualifies as optimal truncation. In all other cases, truth-telling constitutes under-truncation.

Table 2: Distribution of subjects' strategies (based on the length of the submitted rank-order list). The distributions from both the experimental data and derived from uniformly random behavior are included.

	Frequency	Percent	Random Percent
1 Firm	99	10.76	25.00
2 Firms	165	17.93	25.00
3 Firms	145	15.76	25.00
4 Firms	511	55.54	25.00
Total	920	100	100

Table 3: Distribution of subjects' strategies (based on the degree of truncation). The distributions from both the experimental data and derived from uniformly random behavior are included.

	Frequency	Percent	Random Percent
Over-Truncation	30	3.26	18.75
Optimal Truncation	247	26.85	25.00
Under-Truncation	643	69.89	56.25
Total	920	100	100

ference for both cases (Table 2: $\chi^2(3) = 467.7$, p-value < 0.001; Table 3: $\chi^2(2) = 149.410$, p-value < 0.001). The entire difference comes from experimental subjects behaving more conservatively: the rate of the most extreme truncation strategy (only one ranked firm) is 57% less than random behavior would dictate and the rate of over-truncation is 83% less than random behavior would dictate.

We now examine whether strategic behavior is sensitive to considerations of profitability and riskiness. For the remainder of our analysis, we differentiate truncation from truth-telling. In other words, a truncation is observed whenever an agent's submitted rank-order list contains strictly less than four firms. Figure 5 shows the

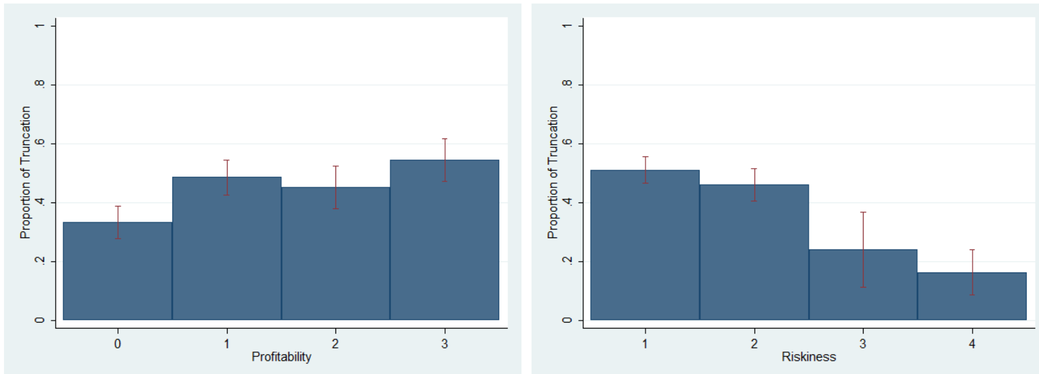


Figure 5: Proportion of truncation according to the profitability and riskiness of truncation.

proportions of truncation in varying environments of profitability and riskiness.[28] We find that there is a statistically significant relationship between truncation and both measures (profitability: $\chi^2(3) = 23.025$, p-value < 0.001; riskiness: $\chi^2(3) = 45.876$, p-value < 0.001). Furthermore, non-parametric trend tests show that truncation is increasing in profitability and decreasing in riskiness (profitability: p-value < 0.001; safety: p-value < 0.001).

We estimate an OLS regression model of a dummy variable for truncation on relevant market features: the index for profitability (0-3), the index for riskiness (1-4), and the round of the experiment (1-10).[29] We find that the riskiness of truncation and market experience are the only significant predictors of truncation. In particular, moving the most preferred achievable firm down one rank in preference decreases

[28]Since profitability and riskiness are correlated, a more illuminating 3-D graph of the sensitivity of truncation to strategic incentives is shown in Appendix D. The "holes" in the graph to the right of the main diagonal correspond to profitability-riskiness ordered pairs that are impossible to construct in 4x4 matching markets.

[29]The OLS regression results are shown in Table 4. By using individual-specific fixed effects, our estimation rules out the effects of "naive" truncators who are playing identical strategies in each round.

the probability of truncation by 0.11 and an additional round of market experience increases the probability of truncation by 0.3. We also estimate probit and conditional logit regression models of the dummy variable for truncation on the same set of regressors.[30] The main results remain unchanged. The coefficient estimates for both the riskiness of truncation and the round of the experiment are still significant in the directions predicted by theory.

We conduct several checks for the robustness of our results. First, we replace our index for profitability with a dummy variable for situations where truncation is profitable (i.e., whenever an agent has more than one achievable partner). Even if agents are not responsive to the magnitude of the monetary gains from truncation, it is possible that they are responsive to the *existence* of a profitable strategic opportunity. We find that profitability measured in this manner is also not significant at conventional levels.

Second, we add an "average rank" variable to the regressions. For a particular worker in a given market, average rank is defined as the average of the ordinal position of that worker in the firms' preference lists. Thus, average rank would be "1" for a worker who is ranked first by all of the firms and "4" for a worker who is ranked last by all of the firms. Since determining the riskiness of truncation according to our measure requires that a worker have knowledge of the identity of her most preferred achievable firm, average rank has appeal as a plausible heuristic that agents might instead use in this setting. We find that the coefficient estimate on average rank is only significant for the OLS regression specification. However, the significance of riskiness as a predictor of truncation behavior still remains.

We also explore the possibility of an alternative heuristic. If a worker is ranked first by a particular firm, then the worker can secure that match as a lower bound

[30]The probit and conditional logit regression results are included in Appendix D.

Table 4: The table reports results from OLS regressions with individual-specific fixed effects.

VARIABLES	(1) Truncation	(2) Truncation	(3) Truncation	(4) Truncation
Profitability	0.00728		0.0251	0.00370
	(0.0125)		(0.0158)	(0.0165)
More than one achievable partner		0.0362		
		(0.0334)		
Riskiness	-0.110***	-0.105***	-0.0722***	-0.115***
	(0.0155)	(0.0157)	(0.0225)	(0.0222)
Average rank in firms' preferences			-0.0485*	
			(0.0274)	
Ranked first by top three				-0.0121
				(0.0369)
Round of the experiment	0.0304***	0.0303***	0.0299***	0.0299***
	(0.00576)	(0.00578)	(0.00571)	(0.00546)
Constant	0.460***	0.437***	0.495***	0.481***
	(0.0435)	(0.0464)	(0.0484)	(0.0686)
Observations	920	920	920	920
Number of individuals	92	92	92	92

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

by including the firm in her reported preference list. This implies that if a worker is ranked first by one of her top three firms, then it is safe to exclude her least preferred firm from her reported preference list. To test whether this line of reasoning is predictive of truncation in our experimental data, we add a dummy variable for whether a worker is ranked first by one of her top three firms. For all regression specifications, the coefficient estimate on this dummy variable is not significant while the significance of our riskiness measure remains.

Finally, we investigate whether the degree of truncation is responsive to the strategic incentives that we identify. For this purpose, we estimate an ordered logit regression model of the number of firms included in an agent's submitted rank-order list on the same set of regressors.[31] We find that the round of the experiment and the

[31] Results from the ordered logit regression are shown in Table 5. We exclude the dummy variable for whether a worker is ranked first by one of her top three choices since that line of reasoning would not capture any truncation beyond the fourth choice.

Table 5: The table reports results from ordered logit regressions.

Dependent Variable: Number of Firms in Submitted List VARIABLES	(1)	(2)	(3)
Profitability	0.130**		0.0870
	(0.0543)		(0.0682)
More than one achievable partner		0.117	
		(0.164)	
Riskiness	0.736***	0.671***	0.647***
	(0.0965)	(0.0918)	(0.124)
Average rank in firms' preferences			0.126
			(0.139)
Round of the experiment	-0.160***	-0.159***	-0.159***
	(0.0252)	(0.0250)	(0.0251)
Observations	920	920	920

Standard errors are shown in parentheses and are clustered at the individual level.
*** p<0.01, ** p<0.05, * p<0.1

rank of the most preferred achievable firm matter once again. Surprisingly, we now also have that increasing the monetary gains from truncation makes subjects slightly more likely to lengthen their submitted rank-order lists. This finding is the opposite of what theory predicts. However, the significance of profitability disappears both when average rank is included in the regression and when our index for profitability is replaced with a dummy variable. The other findings remain unchanged in these alternative specifications.

Aggregate Outcomes

Stability is the norm in our experimental markets: 88% (203/230) of final outcomes are stable. Assuming uniformly random behavior by subjects, only 51% of final outcomes are expected to be stable. The high incidence of stability in our data is due to the fact that over-truncation is the only way to observe instability. As shown earlier, over-truncation is rare in our experiment. Three percent (30/920) of submitted preference lists constitute over-truncation and this occurs in 27 of the 230 markets. Table 6 shows a more detailed distribution of the stability of final outcomes. The worker-optimal stable matching arose in 24% (56/230) of markets, suggesting a limited ability on the part of our experimental subjects to play the "right" kind of truncation strategy.

To make welfare statements about the two sides of the market, we need a meaningful way to measure the "distance" from an observed outcome in our experimental data to a particular stable outcome. To that end, we first define a metric for the space of all matchings. Let $\mathcal{M}$ denote the set of all matchings and let W denote the set of all workers. Consider an arbitrary matching $\mu \in \mathcal{M}$ and an arbitrary worker $w \in W$. Define $F(\mu(w))$ as the position of $\mu(w)$ in the ordinal preference list of worker w. If w is matched to her most preferred firm at μ, then $F(\mu(w)) = 1$. If w is matched to her least preferred firm at μ, then $F(\mu(w)) = 4$. For simplicity, if w is unmatched we let $F(\mu(w)) = F(w) = 5$. Thus, $|F(\mu(w)) - F(\mu'(w))|$ is the absolute distance in ranking between $\mu(w)$ and $\mu'(w)$ according to the preferences of worker w. We can then define the distance from μ to μ' as the sum of this measure for all the workers in the market. More formally, the distance $d : \mathcal{M} \times \mathcal{M} \longrightarrow \Re_+$ between two matchings μ and μ' is defined as $d(\mu, \mu') = \sum_{w \in W} |F(\mu(w)) - F(\mu'(w))|$. Intuitively, we are defining the distance between two outcomes as the sum of the absolute distance between each worker's match partners at those outcomes (according to the worker's ordinal preferences).

Figure 6 shows the distances to the worker-optimal and firm-optimal stable matchings across all experimental markets. According to our metric, the average distance to the worker-optimal stable matching is 2.16 and the average distance to the firm-optimal stable matching is 3.42. We find that final outcomes are significantly closer to the worker-optimal stable matching than to the firm-optimal stable matching (one-sided t-test, p-value < 0.001). However, this result should be interpreted with caution:

Table 6: Distribution of final outcomes.

	Frequency	Percent
Firm-Optimal Stable	48	20.87
Worker-Optimal Stable	56	24.35
Intermediate Stable	60	26.09
Unique Stable	39	16.96
Unstable	27	11.74
Total	230	100

it is *not* due to the fact that a majority of agents optimally truncates its preferences. In this context, truncation behavior has positive externalities: non-strategic agents can benefit from other agents' truncation behavior. Thus, the worker-optimal stable matching can be observed if even a small subset of agents optimally truncates its preferences.

5 Discussion and Conclusion

The paper investigates the ability of agents to strategically misrepresent their preferences in two-sided matching markets. We use a controlled laboratory experiment that allows us to construct environments that are conducive to truncation behavior. We find that subjects do not truncate their preferences more often when it is profitable to do so. They do, however, truncate less often when it is dangerous (i.e., when there is a risk of "over-truncating" and remaining unmatched).

Our results suggest that agents in matching markets may respond to strategic incentives, but not necessarily in the ways that are predicted by theory. In particular, this implies that eliminating profitable opportunities for strategic behavior might not be sufficient to induce participants to reveal their true preferences. In other words,

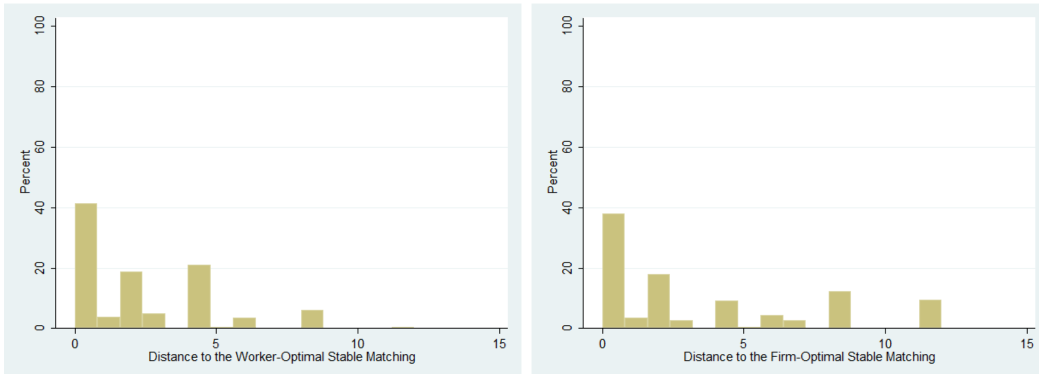


Figure 6: Distributions of the distances to the worker-optimal and firm-optimal stable matchings.

just as agents can fail to recognize profitable strategic opportunities, they can also fail to recognize the *lack* of profitable strategic opportunities.[32] Our work also highlights the importance of understanding behavioral biases and heuristics when designing matching markets. In fact, for the proposing side of the market, it has been shown that the DA algorithm already possesses protective features that bound the losses of agents who behave sub-optimally (Rees-Jones, 2014). However, more work needs to be done to understand the extent to which this tolerance of behavioral faults applies to the receivers in the DA algorithm.

An open question remains as to the efficacy of strategic behavior in this context. We argue that the ability of agents to engage in strategic behavior is important because it affects equilibrium selection - and hence welfare - in these environments. However, there is both computational and theoretical evidence suggesting otherwise. In May 1997, the NRMP transitioned from the hospital-proposing version of the DA algorithm to the student-proposing version. When analyzing the data from the

[32]Relatedly, there is evidence that proposers also misunderstand the incentives in the DA algorithm and engage in suboptimal behavior (Echenique et al., 2014; Rees-Jones, 2014).

NRMP transition, it has been shown that very few participants would have received different matches from the two algorithms. This has been cited as evidence of the fact that the set of stable matchings is small (Roth and Peranson, 1999).[33] As a consequence, it is argued that there is little room for strategic misrepresentation of preferences in this environment.

However, the comparison of match outcomes in the NRMP transition is based on agents' submitted rank-order lists and not on their underlying preferences. If the students were optimally truncating their preferences in the original NRMP, then reversing the roles in the DA algorithm (with the same set of reported preferences) would still produce the student-optimal stable outcome. Moreover, there is evidence suggesting that the submitted rank-order lists differ substantially from the underlying preferences. Echenique et al. (2014) note the high incidence of matches between residents and their top-ranked hospitals. This suggests that either preferences have a strong negative correlation in this market, or more likely that the stated preferences are different from the true preferences. Thus, the span of the core might be small for the stated preferences but not for the true preferences.

Even if the theoretical incentives to behave strategically vanish in larger markets (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Lee, 2014), survey data from the field suggests that strategic behavior still persists. In March 2014, the NRMP surveyed the directors of all programs participating in the residency match.[34] Across all specialties, the average number of applicants interviewed was 96 and the average number of applicants ranked was 77. It should be noted that this is merely

[33] In a field setting such as the NRMP, the set of stable matchings can plausibly be small for several reasons. First, there might be a high degree of positive correlation in agents' preferences. In the extreme case of perfect correlation, there is a unique stable matching. Second, there are practical limits on the number of interviews that can be conducted between hospitals and medical students. This restriction is at the heart of many of the "core convergence" results.

[34] The results of the 2014 NRMP Program Director Survey can be found here: http://www.nrmp.org/wp-content/uploads/2014/09/PD-Survey-Report-2014.pdf

suggestive of truncation and not definitive evidence. It is quite plausible that in many of these instances, the residency programs would genuinely prefer to leave a positon vacant rather than hire a low-quality applicant. The 2013 NRMP Applicant Survey is more conclusive. When asked about different strategies used in creating their rank-order lists, 29% of US senior applicants and 53% of independent applicants answered no to the claim "I ranked all programs that I was willing to attend."[35] This is particularly surprising since the students have a dominant strategy of truth-telling in this environment.

The evidence from the field has natural analogues to our experimental data. The fact that a non-trivial proportion of medical students admit to strategic considerations speaks again to the idea that agents might be misrepresenting their preferences even in environments in which it is unprofitable to do so. Similarly, our finding that subjects take into account the riskiness of truncation suggests that submitted NRMP lists should be shorter on average for top-ranked residency programs.

Our results also suggest natural directions for future work. We have shown that truncation behavior has the flavor of a coordination game: optimal truncation can essentially be reduced to a decision-theoretic problem only if other agents are also truncating. In our experiment, we overcome the need for coordination by exogenously imposing a constraint on the strategy space and making the constraint common knowledge. It would be worthwhile to investigate whether agents can endogenously coordinate on truncation strategies in an unconstrained environment and also whether their truncation behavior depends on the size of the market. In addition to the benefit of increased ecological validity, this environment also allows for a more direct test of the empirical content of truncation strategies.

[35] The results of the 2013 NRMP Applicant Survey can be found here: http://www.nrmp.org/wp-content/uploads/2013/08/applicantresultsbyspecialty2013.pdf

Appendix A: Proofs

Proof of Proposition 1(a): Consider μ_W, the W-optimal stable matching with respect to the true preferences $\mathbf{P}$. At μ_W, woman w is matched to $\mu_W(w) \in M$.[36] Clearly the matching μ_W is still individually rational under $\mathbf{Q}$.[37] Also, the matching μ_W still admits no blocking pairs under $\mathbf{Q}$ since there are now fewer possible blocking pairs. Thus, μ_W is stable with respect to the stated preferences $\mathbf{Q}$. We can conclude from Theorem 3 that w must be matched at all stable matchings with respect to the stated preferences $\mathbf{Q}$. Since the man-proposing deferred acceptance algorithm produces the M-optimal stable matching with respect to $\mathbf{Q}$, w is matched by the algorithm.

Proof of Proposition 1(b): Suppose that μ is a stable matching with respect to the stated preferences $\mathbf{Q}$. Then μ is individually rational and not blocked by any pair of agents under $\mathbf{Q}$. Clearly, the construction of the preference profile $\mathbf{P}$ does not create any new blocking opportunities. To see this, note that for each man m added to $Q(w)$ in order to construct $P(w)$, we have that $\mu(w) \succ_w m$.[38] Thus, μ is also a stable matching with respect to the true preferences $\mathbf{P}$.

Proof of Proposition 1(c): Let $\underline{Q}(w) = \mu_W(w)$. Denote by μ'_M the M-optimal stable matching with respect to the submitted preferences $\mathbf{Q}$. By Proposition 1(a), we know that woman w will not be single at μ'_M. Thus, w has to be matched to a man she likes at least as much as $\mu_W(w)$. Suppose that w is matched to a man she

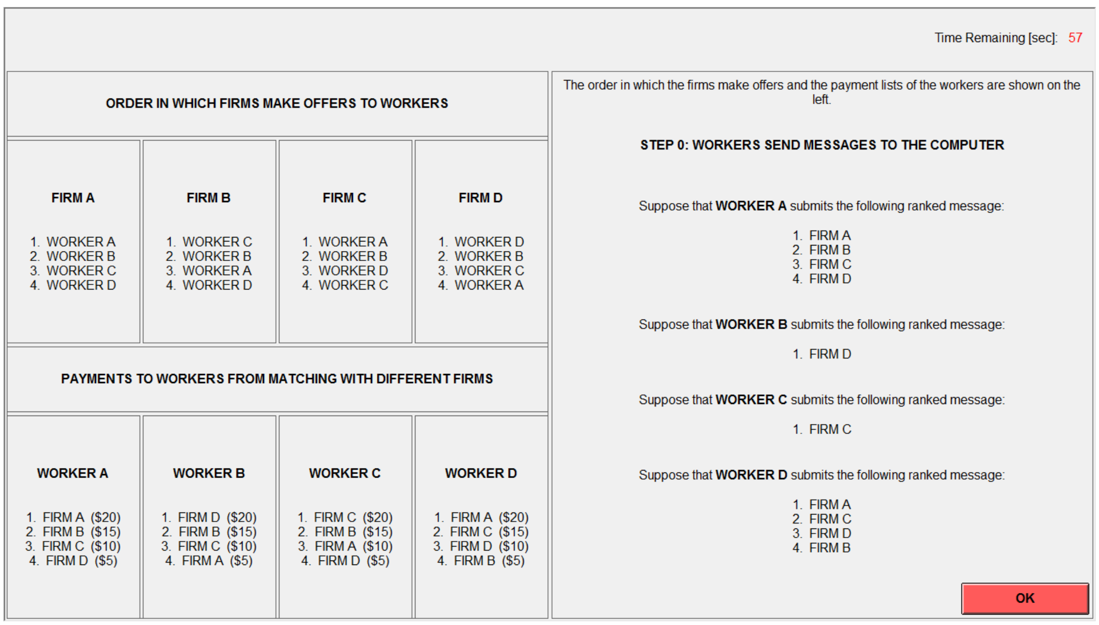
[36]Woman w is not single at μ_W since $|M| = |W|$ and all men are acceptable to all women (and vice versa).
[37]Since no woman is over-truncating, note that $\mu_W(w)$ is included on the list $Q(w)$ for each w.
[38]Recall that w is not single at μ by Proposition 1(a).

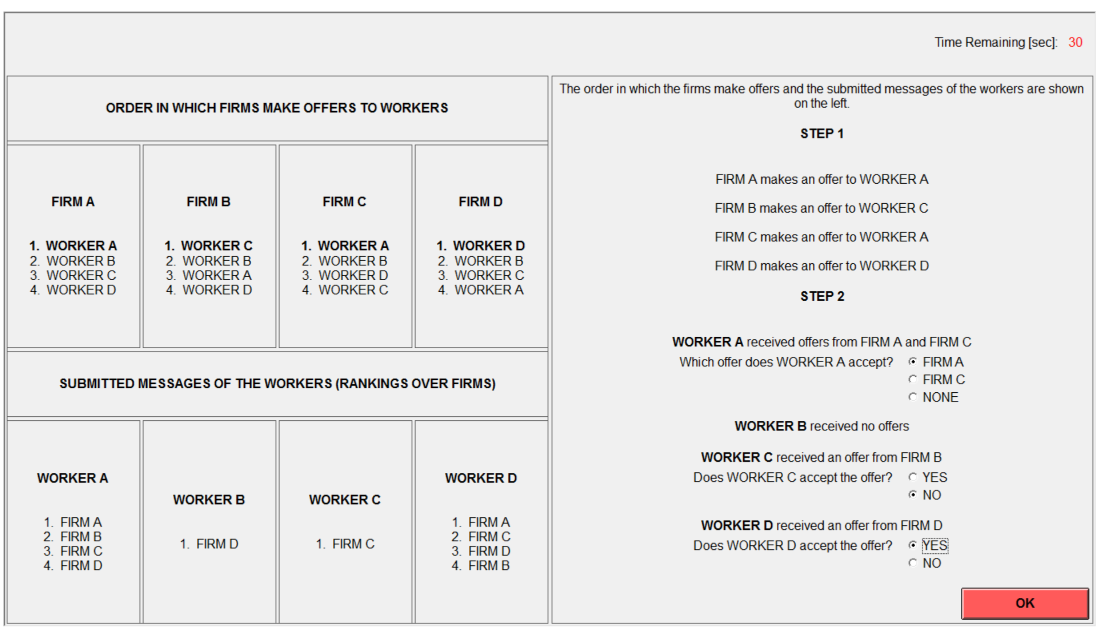
strictly prefers to $\mu_W(w)$. Denote this man by m. By Proposition 1(b), we know that μ'_M is also stable with respect to the true preferences $\mathbf{P}$. Thus, m is achievable for w and $m \succ_w \mu_W(w)$. We have now arrived at a contradiction since $\mu_W(w)$ is the most preferred achievable mate of w. Therefore, w must be matched to $\mu_W(w)$.

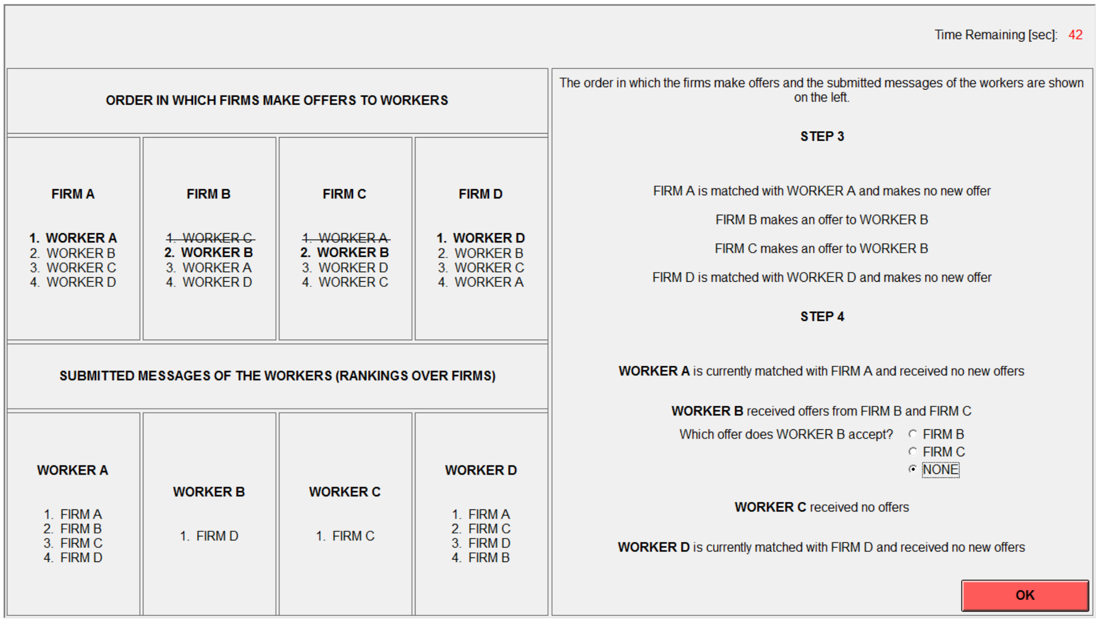
Proof of Proposition 2: Suppose that woman w submits a preference list $Q(w)$ that is a truncation of $P(w)$ such that $\underline{Q}(w) = \mu_W(w)$. If none of the other women over-truncates, then woman w will be matched to $\mu_W(w)$ by Proposition 1(c). So the remaining case to be considered involves at least one of the other women over-truncating. Without loss of generality, suppose that some non-empty subset $T \subseteq W \setminus \{w\}$ over-truncates. Let $M' = \{\mu_W(w') : w' \in T\}$. Each man $m' \in M'$ is now available to make offers to other women in the deferred acceptance algorithm. Clearly woman w can only benefit from the availability of these men. If man m' now makes an offer to match with woman w, the algorithm would only accept the offer on the woman's behalf if it were the case that $m' \succ_w \mu_W(w)$.

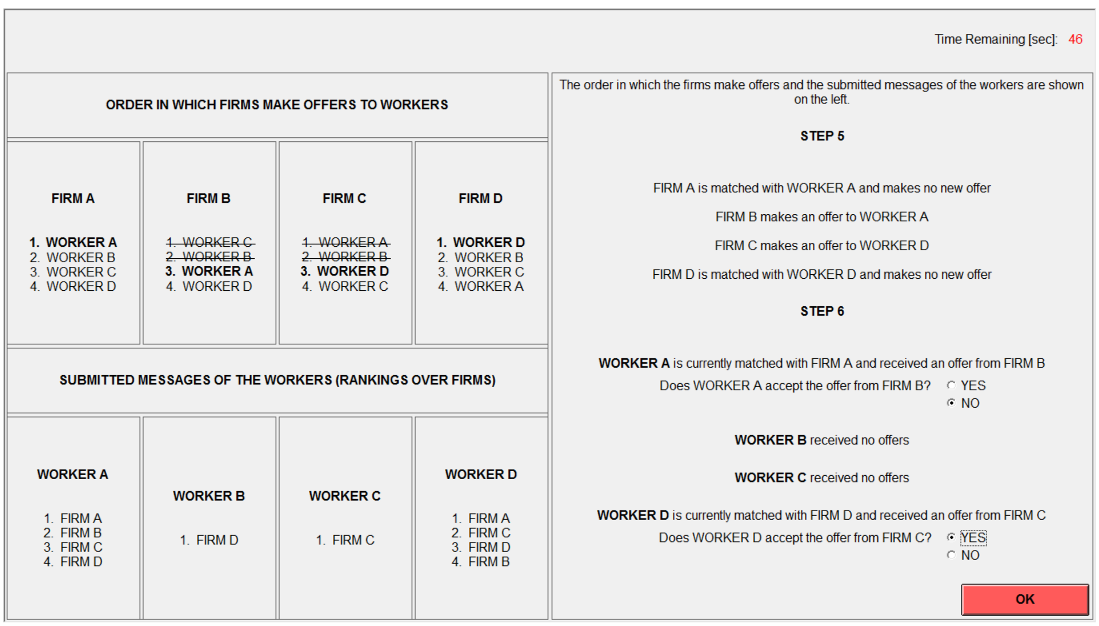
Appendix B: Demonstration of the DA Algorithm

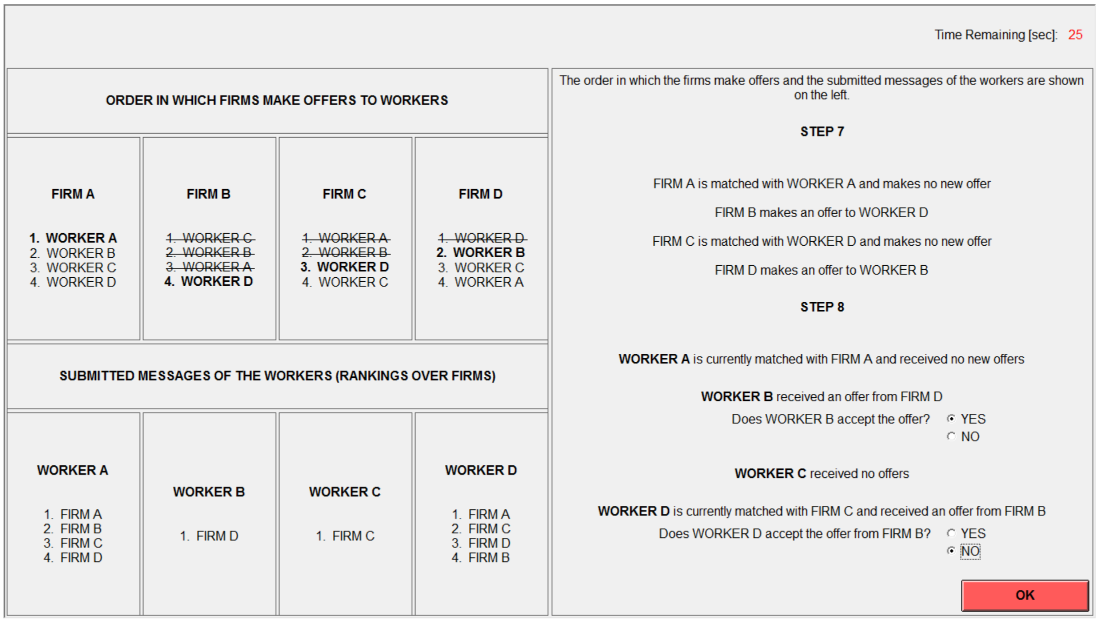
Before the experiment begins, all subjects are required to work through a demonstration of the DA algorithm and correctly answer a series of questions. The relevant screen shots are shown below (with the correct answers already selected).

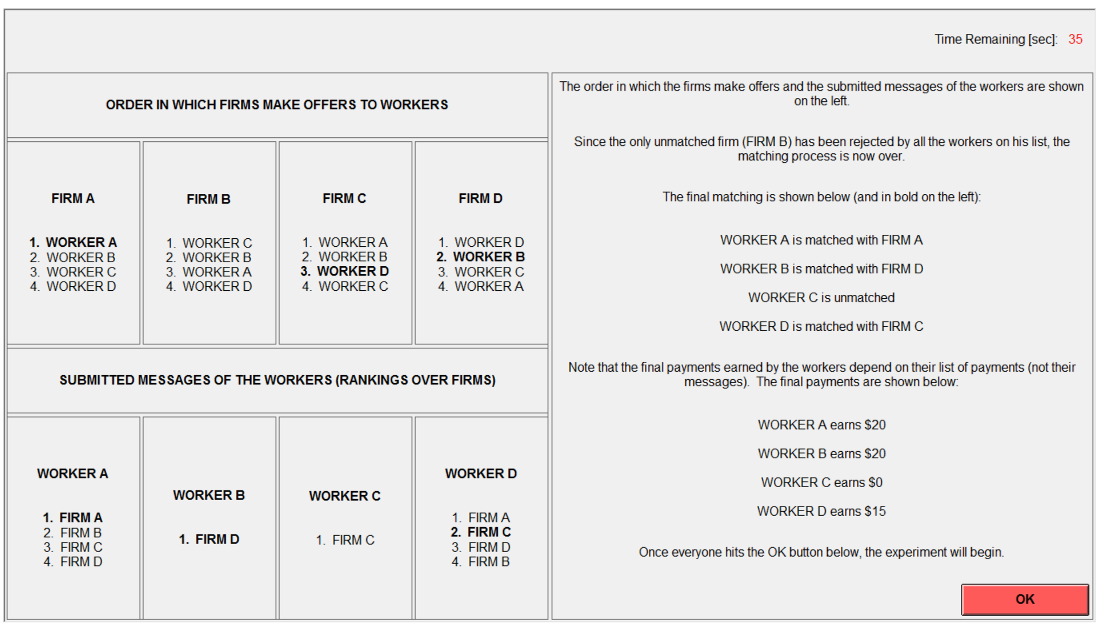





Time Remaining [sec]: 42
ORDER IN WHICH FIRMS MAKE OFFERS TO WORKERS
FIRM A
1. WORKER A
2. WORKER B
3. WORKER C
4. WORKER D
FIRM B
~~1. WORKER C~~
2. WORKER B
3. WORKER A
4. WORKER D
FIRM C
~~1. WORKER A~~
2. WORKER B
3. WORKER D
4. WORKER C
FIRM D
1. WORKER D
2. WORKER B
3. WORKER C
4. WORKER A
SUBMITTED MESSAGES OF THE WORKERS (RANKINGS OVER FIRMS)
WORKER A
1. FIRM A
2. FIRM B
3. FIRM C
4. FIRM D
WORKER B
1. FIRM D
WORKER C
1. FIRM C
WORKER D
1. FIRM A
2. FIRM C
3. FIRM D
4. FIRM B
The order in which the firms make offers and the submitted messages of the workers are shown on the left.
STEP 3
FIRM A is matched with WORKER A and makes no new offer
FIRM B makes an offer to WORKER B
FIRM C makes an offer to WORKER B
FIRM D is matched with WORKER D and makes no new offer
STEP 4
WORKER A is currently matched with FIRM A and received no new offers
WORKER B received offers from FIRM B and FIRM C
Which offer does WORKER B accept? FIRM B
FIRM C
NONE
WORKER C received no offers
WORKER D is currently matched with FIRM D and received no new offers
OK


Time Remaining [sec]: 46
ORDER IN WHICH FIRMS MAKE OFFERS TO WORKERS
FIRM A
1. WORKER A
2. WORKER B
3. WORKER C
4. WORKER D
FIRM B
~~1. WORKER C~~
~~2. WORKER B~~
3. WORKER A
4. WORKER D
FIRM C
~~1. WORKER A~~
~~2. WORKER B~~
3. WORKER D
4. WORKER C
FIRM D
1. WORKER D
2. WORKER B
3. WORKER C
4. WORKER A
SUBMITTED MESSAGES OF THE WORKERS (RANKINGS OVER FIRMS)
WORKER A
1. FIRM A
2. FIRM B
3. FIRM C
4. FIRM D
WORKER B
1. FIRM D
WORKER C
1. FIRM C
WORKER D
1. FIRM A
2. FIRM C
3. FIRM D
4. FIRM B
The order in which the firms make offers and the submitted messages of the workers are shown on the left.
STEP 5
FIRM A is matched with WORKER A and makes no new offer
FIRM B makes an offer to WORKER A
FIRM C makes an offer to WORKER D
FIRM D is matched with WORKER D and makes no new offer
STEP 6
WORKER A is currently matched with FIRM A and received an offer from FIRM B
Does WORKER A accept the offer from FIRM B? YES
NO
WORKER B received no offers
WORKER C received no offers
WORKER D is currently matched with FIRM D and received an offer from FIRM C
Does WORKER D accept the offer from FIRM C? YES
NO
OK


Time Remaining [sec]: 25
ORDER IN WHICH FIRMS MAKE OFFERS TO WORKERS
FIRM A
1. WORKER A
2. WORKER B
3. WORKER C
4. WORKER D
FIRM B
~~1. WORKER C~~
~~2. WORKER B~~
~~3. WORKER A~~
4. WORKER D
FIRM C
~~1. WORKER A~~
~~2. WORKER B~~
3. WORKER D
4. WORKER C
FIRM D
~~1. WORKER D~~
2. WORKER B
3. WORKER C
4. WORKER A
SUBMITTED MESSAGES OF THE WORKERS (RANKINGS OVER FIRMS)
WORKER A
1. FIRM A
2. FIRM B
3. FIRM C
4. FIRM D
WORKER B
1. FIRM D
WORKER C
1. FIRM C
WORKER D
1. FIRM A
2. FIRM C
3. FIRM D
4. FIRM B
The order in which the firms make offers and the submitted messages of the workers are shown on the left.
STEP 7
FIRM A is matched with WORKER A and makes no new offer
FIRM B makes an offer to WORKER D
FIRM C is matched with WORKER D and makes no new offer
FIRM D makes an offer to WORKER B
STEP 8
WORKER A is currently matched with FIRM A and received no new offers
WORKER B received an offer from FIRM D
Does WORKER B accept the offer? YES NO
WORKER C received no offers
WORKER D is currently matched with FIRM C and received an offer from FIRM B
Does WORKER D accept the offer from FIRM B? YES NO
OK


Time Remaining [sec]: 35
ORDER IN WHICH FIRMS MAKE OFFERS TO WORKERS
FIRM A
1. WORKER A
2. WORKER B
3. WORKER C
4. WORKER D
FIRM B
1. WORKER C
2. WORKER B
3. WORKER A
4. WORKER D
FIRM C
1. WORKER A
2. WORKER B
3. WORKER D
4. WORKER C
FIRM D
1. WORKER D
2. WORKER B
3. WORKER C
4. WORKER A
SUBMITTED MESSAGES OF THE WORKERS (RANKINGS OVER FIRMS)
WORKER A
1. FIRM A
2. FIRM B
3. FIRM C
4. FIRM D
WORKER B
1. FIRM D
WORKER C
1. FIRM C
WORKER D
1. FIRM A
2. FIRM C
3. FIRM D
4. FIRM B
The order in which the firms make offers and the submitted messages of the workers are shown on the left.
Since the only unmatched firm (FIRM B) has been rejected by all the workers on his list, the matching process is now over.
The final matching is shown below (and in bold on the left):
WORKER A is matched with FIRM A
WORKER B is matched with FIRM D
WORKER C is unmatched
WORKER D is matched with FIRM C
Note that the final payments earned by the workers depend on their list of payments (not their messages). The final payments are shown below:
WORKER A earns $20
WORKER B earns $20
WORKER C earns $0
WORKER D earns $15
Once everyone hits the OK button below, the experiment will begin.
OK

Appendix C: Ordinal Preference Profiles

An agent's achievable match partners are denoted by asterisks ($*$) in her preference list.

Round 1

$P(f_1) = w_1^*, w_2, w_3, w_4$ $\quad$ $P(w_1) = f_1^*, f_2, f_3, f_4$

$P(f_2) = w_1, w_2^*, w_3, w_4$ $\quad$ $P(w_2) = f_1, f_2^*, f_3, f_4$

$P(f_3) = w_1, w_2, w_3^*, w_4$ $\quad$ $P(w_3) = f_1, f_2, f_3^*, f_4$

$P(f_4) = w_1, w_2, w_3, w_4^*$ $\quad$ $P(w_4) = f_1, f_2, f_3, f_4^*$

Round 2

$P(f_1) = w_1^*, w_2^*, w_3, w_4$ $\quad$ $P(w_1) = f_2^*, f_3, f_1^*, f_4$

$P(f_2) = w_2^*, w_1^*, w_3, w_4$ $\quad$ $P(w_2) = f_1^*, f_3, f_2^*, f_4$

$P(f_3) = w_3^*, w_4^*, w_1, w_2$ $\quad$ $P(w_3) = f_4^*, f_1, f_3^*, f_2$

$P(f_4) = w_4^*, w_3^*, w_1, w_2$ $\quad$ $P(w_4) = f_3^*, f_1, f_4^*, f_2$

Round 3

$P(f_1) = w_4^*, w_2^*, w_3, w_1$ $\quad$ $P(w_1) = f_3, f_2^*, f_4^*, f_1$

$P(f_2) = w_3^*, w_1^*, w_4, w_2$ $\quad$ $P(w_2) = f_4, f_1^*, f_3^*, f_2$

$P(f_3) = w_2^*, w_4^*, w_1, w_3$ $\quad$ $P(w_3) = f_3, f_4^*, f_2^*, f_1$

$P(f_4) = w_1^*, w_3^*, w_2, w_4$ $\quad$ $P(w_4) = f_4, f_3^*, f_1^*, f_2$

Round 4

$P(f_1) = w_1^*, w_2^*, w_3, w_4$

$P(f_2) = w_2^*, w_1^*, w_3, w_4$

$P(f_3) = w_3^*, w_4^*, w_1, w_2$

$P(f_4) = w_4^*, w_3^*, w_1, w_2$

$P(w_1) = f_2^*, f_3, f_4, f_1^*$

$P(w_2) = f_1^*, f_3, f_4, f_2^*$

$P(w_3) = f_4^*, f_1, f_2, f_3^*$

$P(w_4) = f_3^*, f_1, f_2, f_4^*$

Round 5

$P(f_1) = w_3^*, w_1, w_4, w_2$

$P(f_2) = w_3, w_4^*, w_1^*, w_2$

$P(f_3) = w_3, w_1^*, w_4^*, w_2$

$P(f_4) = w_3, w_4, w_1, w_2^*$

$P(w_1) = f_1, f_2^*, f_3^*, f_4$

$P(w_2) = f_1, f_3, f_2, f_4^*$

$P(w_3) = f_1^*, f_2, f_3, f_4$

$P(w_4) = f_1, f_3^*, f_2^*, f_4$

Round 6

$P(f_1) = w_2^*, w_4, w_1, w_3$

$P(f_2) = w_2, w_1^*, w_4^*, w_3$

$P(f_3) = w_2, w_4^*, w_1^*, w_3$

$P(f_4) = w_2, w_1, w_4, w_3^*$

$P(w_1) = f_1, f_3^*, f_2^*, f_4$

$P(w_2) = f_1^*, f_2, f_3, f_4$

$P(w_3) = f_1, f_3, f_2, f_4^*$

$P(w_4) = f_1, f_2^*, f_3^*, f_4$

Round 7

$P(f_1) = w_1^*, w_2^*, w_3^*, w_4^*$ $\quad$ $P(w_1) = f_2^*, f_3^*, f_4^*, f_1^*$

$P(f_2) = w_2^*, w_3^*, w_4^*, w_1^*$ $\quad$ $P(w_2) = f_3^*, f_4^*, f_1^*, f_2^*$

$P(f_3) = w_3^*, w_4^*, w_1^*, w_2^*$ $\quad$ $P(w_3) = f_4^*, f_1^*, f_2^*, f_3^*$

$P(f_4) = w_4^*, w_1^*, w_2^*, w_3^*$ $\quad$ $P(w_4) = f_1^*, f_2^*, f_3^*, f_4^*$

Round 8

$P(f_1) = w_1^*, w_2^*, w_3, w_4$ $\quad$ $P(w_1) = f_2^*, f_1^*, f_3, f_4$

$P(f_2) = w_2^*, w_1^*, w_3, w_4$ $\quad$ $P(w_2) = f_1^*, f_2^*, f_3, f_4$

$P(f_3) = w_3^*, w_4^*, w_1, w_2$ $\quad$ $P(w_3) = f_4^*, f_3^*, f_1, f_2$

$P(f_4) = w_4^*, w_3^*, w_1, w_2$ $\quad$ $P(w_4) = f_3^*, f_4^*, f_1, f_2$

Round 9

$P(f_1) = w_4^*, w_2^*, w_3, w_1$ $\quad$ $P(w_1) = f_3, f_2^*, f_1, f_4^*$

$P(f_2) = w_3^*, w_1^*, w_4, w_2$ $\quad$ $P(w_2) = f_4, f_1^*, f_2, f_3^*$

$P(f_3) = w_2^*, w_4^*, w_1, w_3$ $\quad$ $P(w_3) = f_3, f_4^*, f_1, f_2^*$

$P(f_4) = w_1^*, w_3^*, w_2, w_4$ $\quad$ $P(w_4) = f_4, f_3^*, f_2, f_1^*$

Round 10

$P(f_1) = w_4^*, w_3, w_2, w_1$

$P(f_2) = w_4, w_3^*, w_2, w_1$

$P(f_3) = w_4, w_3, w_2^*, w_1$

$P(f_4) = w_4, w_3, w_2, w_1^*$

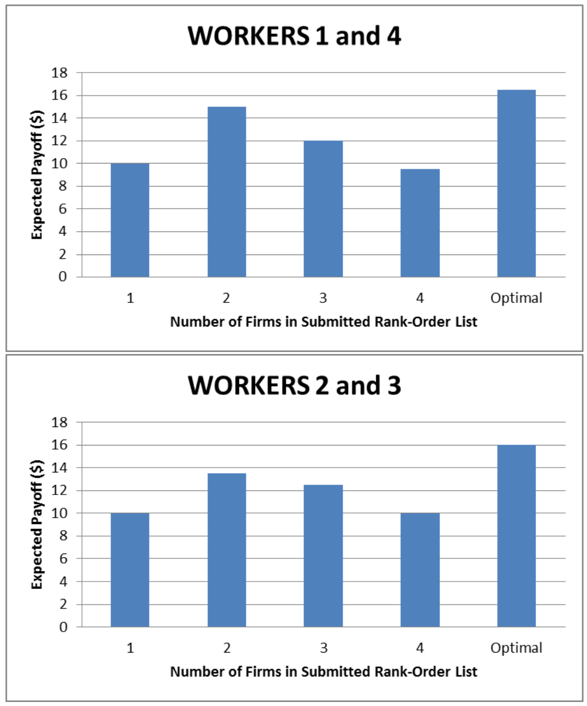
$P(w_1) = f_1, f_2, f_3, f_4^*$

$P(w_2) = f_1, f_2, f_3^*, f_4$

$P(w_3) = f_1, f_2^*, f_3, f_4$

$P(w_4) = f_1^*, f_2, f_3, f_4$

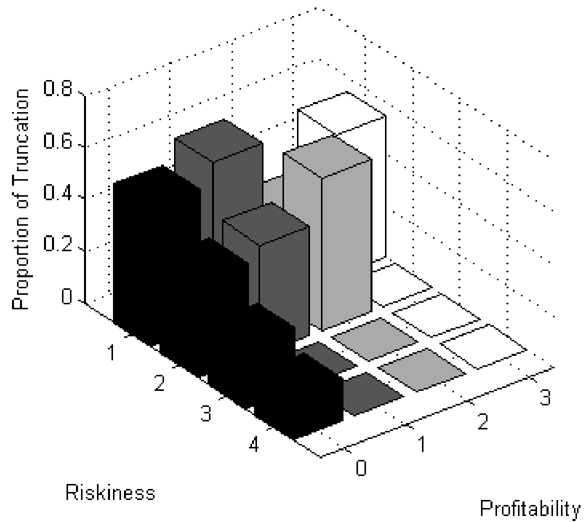
Appendix D: Additional Figures and Tables



The expected payoff of different truncation levels.

Round	Worker	Number of Achievable Partners	Profitability	Riskiness
1	A	1	0	1
1	B	1	0	2
1	C	1	0	3
1	D	1	0	4
2	A,B,C,D	2	2	1
3	A,B,C,D	2	1	2
4	A,B,C,D	2	3	1
5	A	2	1	2
5	B	1	0	4
5	C	1	0	1
5	D	2	1	2
6	A	2	1	2
6	B	1	0	1
6	C	1	0	4
6	D	2	1	2
7	A,B,C,D	4	3	1
8	A,B,C,D	2	1	1
9	A,B,C,D	2	2	2
10	A	1	0	4
10	B	1	0	3
10	C	1	0	2
10	D	1	0	1

The strategic incentives faced by workers throughout the experiment.



Proportion of truncation across varying strategic incentives.

Probit Regression of Truncation by Market Features

	(1)	(2)	(3)	(4)
VARIABLES	Truncation	Truncation	Truncation	Truncation
Profitability	0.00908		0.0210	0.00713
	(0.0129)		(0.0180)	(0.0177)
More than one achievable partner		0.0562		
		(0.0394)		
Riskiness	-0.123***	-0.115***	-0.0980***	-0.125***
	(0.0193)	(0.0194)	(0.0289)	(0.0258)
Average rank in firms' preferences			-0.0328	
			(0.0346)	
Ranked first by top three				-0.00645
				(0.0391)
Round of the experiment	0.0325***	0.0325***	0.0322***	0.0322***
	(0.00654)	(0.00660)	(0.00650)	(0.00630)
Observations	920	920	920	920

The table reports marginal effects from probit regressions.

Standard errors are shown in parentheses and are clustered at the individual level.

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

Conditional Logit Regression of Truncation by Market Features

VARIABLES	(1) Truncation	(2) Truncation	(3) Truncation	(4) Truncation
Profitability	0.0355		0.182	0.0208
	(0.110)		(0.155)	(0.134)
More than one achievable partner		0.254		
		(0.254)		
Riskiness	-0.851***	-0.814***	-0.551**	-0.872***
	(0.135)	(0.128)	(0.260)	(0.173)
Average rank in firms' preferences			-0.384	
			(0.289)	
Ranked first by top three				-0.0531
				(0.279)
Round of the experiment	0.238***	0.238***	0.235***	0.236***
	(0.0357)	(0.0356)	(0.0358)	(0.0377)
Observations	650	650	650	650
Number of individuals	65	65	65	65

The table reports results from conditional logit regressions with individual-specific fixed effects.

Standard errors are shown in parentheses.

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

REFERENCES

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

Yaron Azrieli, Christopher P Chambers, and Paul J Healy. Incentives in experiments: A theoretical analysis. *Working Paper*, 2014.

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

Peter A Coles and Ran I Shorrer. Optimal truncation in matching markets. *Games and Economic Behavior*, 87:591–615, 2014.

Tingting Ding and Andrew Schotter. Matching and chatting: An experimental study of the impact of network communication on school-matching mechanisms. *Working Paper*.

Lester E Dubins and David A Freedman. Machiavelli and the gale-shapley algorithm. *American Mathematical Monthly*, 88(7):485–494, 1981.

Federico Echenique, Alistair J Wilson, and Leeat Yariv. Clearinghouses for two-sided matching: An experimental study. *Working Paper*, 2014.

Lars Ehlers. Truncation strategies in matching markets. *Mathematics of Operations Research*, 33(2):327–335, 2008.

Clayton R Featherstone and Eric Mayefsky. Why do some clearinghouses yield stable outcomes? experimental evidence on out-of-equilibrium truth-telling. *Working Paper*, 2014.

Clayton R Featherstone and Muriel Niederle. Improving on strategy-proof school choice mechanisms: An experimental investigation. *Working Paper*, 2014.

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

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, pages 9–15, 1962.

David Gale and Marilda Sotomayor. Ms. machiavelli and the stable matching problem. *American Mathematical Monthly*, 92(4):261–268, 1985.

Glenn W Harrison and Kevin A McCabe. Stability and preference distortion in resource matching: An experimental study of the marriage market. *Research in Experimental Economics*, 8, 1989.

Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–62. Society for Industrial and Applied Mathematics, 2005.

Fuhito Kojima and Parag A Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–627, 2009.

SangMok Lee. Incentive compatibility of large centralized matching markets. *Working Paper*, 2014.

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

Alex Rees-Jones. When "strategy" hurts: Suboptimal behavior in strategy-proof mechanisms. *Working Paper*, 2014.

Alvin E Roth. A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the united kingdom. *American Economic Review*, 81(3):415–440, 1991.

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

Alvin E Roth and Uriel G Rothblum. Truncation strategies in matching markets–in search of advice for participants. *Econometrica*, 67(1):21–43, 1999.

Alvin E Roth and Marilda A Oliveira Sotomayor. *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press, 1992.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119(2):457–488, 2004.

CHAPTER 3

1 Introduction

In recent decades, two-sided matching clearinghouses have been successfully used in a variety of allocation problems (Abdulkadiroğlu and Sönmez, 2003; Roth and Peranson, 1999; Sönmez and Switzer, 2013; Abdulkadiroğlu and Sönmez, 1998). In a matching clearinghouse, participants submit rank-order lists of their preferences to a central authority, which uses the submitted lists to calculate a particular outcome (i.e., who is matched with whom). These environments induce a non-cooperative game in which an agent's strategy choice is which preference ordering to report to the central authority.

Both theory and practice suggest the use of *stable* matching mechanisms where no single individual nor coalition of individuals have an incentive to deviate from the final allocation that is produced by the mechanism. When attention is confined to stable outcomes, the interests of the two sides of the market are opposed: the best stable matching for one side of the market is the worst stable matching for the other side of the market, and vice versa.[1]

However, in markets with multiple stable outcomes, the question of which stable outcome arises becomes equally important. This is a relevant consideration for policymakers, who may have reasons to favor the welfare of one side of the market over another when designing matching mechanisms. An example of the sensitivity of policymakers to participants' welfare is provided by the National Resident Matching

[1]This divergence of interests is a consequence of the fact that the set of stable matchings is a lattice.

Program (NRMP), the entry-level labor market for American physicians. In May 1997, the NRMP altered the algorithm that was being used over concerns that the original design unduly favored hospitals at the expense of students.

We apply insights from the literature on equilibrium selection in coordination games to better characterize the outcomes that emerge in two-sided matching markets. First, we focus on symmetric equilibria and highlight the tension between the familiar concepts of payoff-dominance and risk-dominance in our environment. Second, we present a simple behavioral model of strategic preference misrepresentation that provides useful comparative statics predictions and argues for the selection of asymmetric equilibria.

At the heart of our theoretical results is the inherent trade-off between improving one's match partner and the risk of remaining unmatched. As either the potential gains from strategic behavior decrease or the utility loss from remaining unmatched increases, our equilibrium selection criteria eliminate a certain type of preference misrepresentation. This suggests that our results will be sensitive to the details of the institutional environment that is being studied, particularly the assumptions on agents' cardinal preferences.

2 Theoretical Background

Equilibrium Selection

Let $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ denote a standard game, where N is a non-empty, finite set of players, S_i is the non-empty, finite set of pure strategies available to player i, and $u_i : \times_{j \in N} S_j \longrightarrow \Re$ is the von Neumann-Morgenstern utility function of player i. We denote the set of all possible pure strategy profiles by $S = \times_{j \in N} S_j$.

A Nash equilibrium in pure strategies is then a strategy profile $s^* \in S$ such that $u_i(s^*) \geq u_i(s^*_{-i}, s_i)$ for all $i \in N$ and for all $s_i \in S_i$.

We begin my summarizing basic definitions regarding the desirable properties that equilibria may possess in normal-form games.[2]

Definition: Let r and t be two Nash equilibrium strategy profiles. We say that r ***payoff-dominates*** t if $u_i(r) > u_i(t)$ for all $i \in N$.

Definition: Let r be a Nash equilibrium strategy profile. We say that r is ***payoff-dominant*** if there is no other Nash equilibrium strategy profile that payoff-dominates r.

Clearly, payoff-dominance is an attractive feature for a Nash equilibrium to possess. In games with multiple equilibria, the presence of a payoff-dominant equilibrium can serve as a focal point for coordination. Intuitively, allowing pre-play communication between the players would likely further improve the chances of coordination by allowing the players to enter into a self-enforcing agreement.

However, there are arguments against the use of payoff-dominance as an equilibrium selection device (Aumann, 1990). In particular, payoff-dominance ignores the possibility of "mistakes" or perturbations in other players' strategy choices. When faced with this strategic uncertainty, a payoff-dominated equilibrium may actually appear more compelling if it guarantees a higher minimum payoff in the counter-factual situation where the other player deviates from the proposed equilibrium.

For motivation, consider the game shown in Figure 1.[3] The two pure-strategy

[2] These definitions can be found in Harsanyi and Selten (1988).
[3] This game is taken from Harsanyi and Selten (1988).

	A	B
A	**9,9**	$0,8$
B	$8,0$	**8,8**

Figure 1: A game where payoff-dominance and risk-dominance conflict.

Nash equilibria are shown in bold. Although the equilibrium (A, A) payoff-dominates the equilibrium (B, B), it is not unreasonable to expect a conservative player to choose B. In fact, playing B becomes a best-response for either player if their assessment of the probability with which the other player chooses B is greater than $\frac{1}{9}$.

This has inspired the notion of risk-dominance in 2×2 games. For the generic game shown in Figure 2, we have the following definition:

Definition: U ***risk-dominates*** V if

$$(a_{11} - a_{21})(b_{11} - b_{12}) > (a_{22} - a_{12})(b_{22} - b_{21}) \tag{1}$$

In other words, equilibrium U risk-dominates equilibrium V if the product of the players' utility losses of deviating from U is greater than that of deviating from V.

Although the concept of risk-dominance was originally constructed for 2×2 games, it is straightforward to extend the logic to any finite, normal-form game. Suppose the game $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ has two pure-strategy Nash equilibria: $r = (r_1, r_2, ...r_n)$ and $t = (t_1, t_2, ..., t_n)$.

	U_2	V_2
U_1	a_{11}, b_{11}	a_{12}, b_{12}
V_1	a_{21}, b_{21}	a_{22}, b_{22}

Figure 2: The two pure-strategy Nash equilibria are $U = (U_1, U_2)$ and $V = (V_1, V_2)$.

Definition: r ***risk-dominates*** t if

$$\prod_{i \in N} [u_i(r) - u_i(t_i, r_{-i})] > \prod_{i \in N} [u_i(t) - u_i(r_i, t_{-i})] \tag{2}$$

In this more general definition, equilibrium r risk-dominates equilibrium t if the product of each player i's utility loss of deviating from r_i to t_i is greater than that of deviating from t_i to r_i. However, the ability of risk-dominance to select a unique equilibrium in finite normal-form games is limited by the fact that the risk-dominance relationship can be cyclical (Morris, Rob, and Shin, 1995). Thus, it is possible to have pairwise comparisons among three equilibria A, B, and C such that A risk-dominates B, B risk-dominates C, and C risk-dominates A.

Two-Sided Matching Markets

Consider two finite, disjoint sets M and W, where M is the set of men and W is the set of women. Each agent has complete and transitive preferences over the agents on the other side of the market (as well as remaining single). The preferences of man m will be represented by an ordered list of preferences P_m on the set $W \cup \{m\}$. Similarly, the preferences of woman w will be represented by an ordered list of preferences P_w on the set $M \cup \{w\}$. We write $w \succ_m w'$ to denote that m prefers w to w', and $w \succeq_m w'$

to denote that m likes w at least as much as w'. Similarly, we can write $m \succ_w m'$ and $m \succeq_w m'$. Woman w is said to be ***acceptable*** to man m if he likes her at least as much as remaining single (i.e., $w \succeq_m m$). Similarly, m is acceptable to w if $m \succeq_w w$.

Let $\mathbf{P}$ denote the set of all preferences, one for each man and one for each woman. A ***marriage market*** is denoted by the triplet $(M, W, \mathbf{P})$. A matching is a function $\mu : M \cup W \longrightarrow M \cup W$ such that

1. for any $m \in M$, $\mu(m) \in W \cup \{m\}$

2. for any $w \in W$, $\mu(w) \in M \cup \{w\}$

3. for any $m \in M$, $w \in W$, $\mu(m) = w$ if and only if $\mu(w) = m$

Throughout the analysis, we also distinguish between market-wide matchings (represented by μ) and a given individual's match partner. For woman w at the matching μ, her match partner is represented by $\mu(w)$. For each individual, their preference over two alternative matchings corresponds exactly to their preference over their match partners at the two matchings.

A matching μ is ***individually rational*** if every individual is matched to an acceptable partner. A pair of agents (m, w) is said to ***block*** a matching μ if they are not matched to one another at μ but they prefer each other to their assignments at μ (i.e., $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$). A matching μ is ***stable*** if it is individually rational and not blocked by any pair of agents. A stable matching is called an ***M-optimal stable matching*** (denoted μ_M) if every man likes it at least as well as any other stable matching. A ***W-optimal stable matching*** can be defined analogously (denoted μ_W). The M-optimal stable matching is thus the "best" stable matching for the men and the W-optimal stable matching is the "best" stable matching for the women. A man m and a woman w are said to be ***achievable*** for each other in a marriage market $(M, W, \mathbf{P})$ if they are matched to each other at some stable matching. For woman w, $\mu_W(w)$ is her most preferred achievable partner.

To examine the strategic issues involved in two-sided matching markets, we analyze the preference-revelation game in which each man m with preferences P_m is faced with the strategy choice of what preference ordering Q_m to state, and likewise for the women. Denote the set of stated preference lists, one for each man and one for each woman, by $\mathbf{Q}$. The mechanism then computes a matching $\mu = h(\mathbf{Q})$, where h is the function that maps any set $\mathbf{Q}$ of stated preferences into a matching. A mechanism h that for any stated preferences $\mathbf{Q}$ produces a matching $h(\mathbf{Q})$ that is stable with respect to the stated preferences is called a stable mechanism. If $h(\mathbf{Q})$ produces the M-optimal stable matching with respect to $\mathbf{Q}$, then h is called the M-optimal stable mechanism.

In the strategic game induced by the M-optimal stable mechanism, it is well-known that the men have a dominant strategy of truth-telling (Dubins and Freedman, 1981). The women, on the other hand, might have incentives to misrepresent their preferences to produce a more favorable outcome for themselves (Gale and Sotomayor, 1985). We will find it useful to define two classes of strategies for the women in this market:

Definition: *A **truncation** of a preference list P_w containing k acceptable men is a list P'_w containing $k' \leq k$ acceptable men such that the k' elements of P'_w are the first k' elements of P_w, in the same order.*

Definition: *A **manipulation** of a preference list P_w is any list that is not a truncation of P_w.*

A truncation strategy involves misrepresenting your preferences by shortening the list of acceptable matches without changing their order. For convenience, we allow for truth-telling to trivially satisfy the definition of a truncation strategy. A manipulation strategy involves misrepresenting preferences by changing the order of preference between at least two men (regardless of the length of the list). We now define two particular types of truncation strategies that are central to our analysis:

Definition: *An **over-truncation** of a preference list P_w is a truncation of P_w that does not contain $\mu_W(w)$, the most preferred achievable partner of woman w.*

Definition: ***Optimal truncation** of a preference list P_w is a truncation of P_w that contains $\mu_W(w)$ but does not contain any men who are ranked below $\mu_W(w)$.*

3 Results

Our immediate goal is to more precisely characterize the nature of the different equilibria that can arise in the strategic game induced by the M-optimal stable mechanism. Although our focus is on symmetric equilibria, we will discuss the implications of asymmetric equilibria as well. Ideally, we would like to pave the way toward a theory or solution concept that argues for the selection of one equilibrium over another.

For these purposes, the concepts of payoff-dominance and risk-dominance will prove useful. To apply these concepts, we first need to model the preferences of market participants from a cardinal perspective. Let $u_m : W \cup \{m\} \longrightarrow \Re$ denote the von Neumann-Morgenstern utility function for man $m \in M$, and $u_w : M \cup \{w\} \longrightarrow \Re$ denote the von Neumann-Morgenstern utility function for woman $w \in W$.[4] We now

[4]To avoid the case of indifference, we will impose the restriction that these functions are all

consider a simple example that illustrates the tension between payoff-dominance and risk-dominance in this strategic environment.

Example: Consider a marriage market with two men and two women characterized by the following ordinal preferences:

$P(m_1) = w_1, w_2$	$P(w_1) = m_2, m_1$
$P(m_2) = w_2, w_1$	$P(w_2) = m_1, m_2$

Suppose that the von Neumann-Morgenstern utilities for this economy are as follows:

$u_{m_1}(w_1) = 3$	$u_{w_1}(m_2) = 3$
$u_{m_1}(w_2) = 2$	$u_{w_1}(m_1) = 2$
$u_{m_1}(m_1) = 0$	$u_{w_1}(w_1) = 0$
$u_{m_2}(w_2) = 3$	$u_{w_2}(m_1) = 3$
$u_{m_2}(w_1) = 2$	$u_{w_2}(m_2) = 2$
$u_{m_2}(m_2) = 0$	$u_{w_2}(w_2) = 0$

If the men are constrained to truth-telling, then the preference-revelation game can be represented by the following normal-form, where w_1 is the row player and w_2 is the column player. The seven pure-strategy Nash equilibria are shown in bold.

one-to-one.

	m_1, m_2	m_2, m_1	m_1	m_2
m_1, m_2	$2, 2$	**2,2**	$2, 0$	**2,2**
m_2, m_1	$2, 2$	$2, 2$	**3,3**	$2, 2$
m_1	$2, 2$	**2,2**	$2, 0$	**2,2**
m_2	**3,3**	$0, 2$	**3,3**	$0, 2$

This game can be simplified to the following reduced normal-form, where the strategies are classified as either truth-telling, truncation, or manipulation. As before, the pure-strategy Nash equilibria are shown in bold.

	$Truth$	$Truncate$	$Manipulate$
$Truth$	$2, 2$	**3,3**	$2, 2$
$Truncate$	**3,3**	**3,3**	$0, 2$
$Manipulate$	$2, 2$	$2, 0$	**2,2**

Note that manipulating one's preference list is weakly dominated by reporting one's true preference list in this example. In fact, in the game induced by the M-optimal stable mechanism, any strategy in which a woman does not list her true first choice at the head of her list is weakly dominated (Roth and Sotomayor, 1992). Thus, while all manipulation strategies in 2×2 marriage markets are weakly dominated, there exist un-dominated manipulation strategies in more general marriage markets. There are several points worth emphasizing regarding the equilibria in this game:

1. There is no Nash equilibrium in which both women report their true preferences.

2. There is a Nash equilibrium in which both women manipulate their preferences.

3. The payoff-dominant Nash equilibria all involve one woman truncating her preferences and the other woman either truncating or reporting her true preferences.

4. The Nash equilibrium in which both women manipulate their preferences risk-dominates the Nash equilibrium in which both women truncate their preferences.

The question that we now address is whether any of these findings generalize from our simple example to generic marriage markets. We answer in the affirmative in all cases. All the propositions below are framed in the context of the preference-revelation game induced by the M-optimal stable mechanism. We confine attention to markets in which each agent prefers being married to remaining single (all men are acceptable to all women and vice versa) and $|M| = |W| = n$. Throughout, we assume that all men are playing their dominant strategy of truth-telling.

Proposition 1. *Consider a marriage market in which preferences are strict and there is more than one stable matching. There is no Nash equilibrium in which all women report their true preferences.*

Proof. Consider a candidate Nash equilibrium strategy profile in which all women report their true preferences: $p = (p_1, p_2, ..., p_n)$. By Theorem 4.6 of Roth and Sotomayor (1992), there exists one woman who can profitably misrepresent her preferences when all other agents report their true preferences. This means that there exists a woman $w \in W$ such that $u_w(q_w, p_{-w}) \geq u_w(p)$ for some $q_w \in \mathcal{P}_w$. Therefore, p is not a Nash equilibrium. $\square$

Proposition 2. *Consider a marriage market in which preferences are strict and all women have more than one achievable partner. There is a Nash equilibrium in which all women manipulate their preferences.*

Proof. Consider μ_M, the M-optimal stable matching with respect to the true preferences. Suppose that each woman w submits the following preference list: $Q(w) = \mu_M(w)$. This is clearly a preference manipulation since $\mu_M(w)$ is not at the head of any woman's true preference list.[5] Furthermore, this is also a Nash equilibrium by Theorem 4.15 of Roth and Sotomayor (1992). $\square$

Proposition 3. *Consider a marriage market in which preferences are strict and there is more than one stable matching. There is a payoff-dominant Nash equilibrium in which all women truncate their preferences.*

Proof. Suppose that each woman w truncates her preference list by leaving off all men ranked below $\mu_W(w)$, her most preferred achievable partner with respect to the true preferences. Denote this strategy profile by t. By Theorem 4.17 of Roth and Sotomayor (1992), t is a Nash equilibrium and it produces the matching μ_W. Suppose that another Nash equilibrium c payoff-dominates t. Denote the matching that c produces by μ'. Since c payoff-dominates t, we know that $u_w(c) > u_w(t)$ for all $w \in W$. In other words, we have that $\mu'(w) \succ_w \mu_W(w)$ for all $w \in W$. Since c is a Nash equilibrium, we know by Theorem 4.16 of Roth and Sotomayor (1992) that the matching μ' is also stable with respect to the true preferences. This is a contradiction, since μ_W is the most preferred stable matching by the women. Thus, there is no other Nash equilibrium that payoff-dominates t. We conclude that t is payoff-dominant. $\square$

[5] If $\mu_M(w)$ were at the head of any woman's true preference list, then this contradicts the assumption that all women have more than one achievable partner.

Proposition 4. *Consider a marriage market in which preferences are strict and there is more than one stable matching. There is a payoff-dominant Nash equilibrium in which a subset of women truncates its preferences.*

Proof. To prove this claim, we construct a strategy profile where a subset of women truncates its preferences that still produces the matching μ_W. By the same argument given in the proof of Proposition 3, this strategy profile will constitute a payoff-dominant Nash equilibrium.

Since there is more than one stable matching, there exists a woman $w \in W$ such that $\mu_M(w) \neq \mu_W(w)$. Denote by Q the profile of stated preferences in which woman w reports her true preferences and each woman $v \in W \setminus w$ truncates her preference list by leaving off all men ranked below $\mu_W(v)$. Let μ' denote the M-optimal stable matching with respect to the stated preferences Q. By Proposition 1 of Castillo and Dianat (2014), μ' is also stable with respect to the true preferences P. Thus, for each woman $v \in W \setminus w$, $\mu'(v) = \mu_W(v)$.[6] Since the set of individuals who are unmatched is the same at all stable matchings, we must have that $\mu'(w) \in M$. We conclude that $\mu'(w) = \mu_W(w)$. □

Proposition 5. *Consider a marriage market in which preferences are strict and all women have more than one achievable partner. If $u_w(\mu_M(w)) - u_w(w) > u_w(\mu_W(w)) - u_w(\mu_M(w))$ for all $w \in W$, then there exists a Nash equilibrium in weakly dominated manipulation strategies that risk-dominates the Nash equilibrium in which all women optimally truncate their preferences.*

Proof. Denote by r the Nash equilibrium in which each woman w submits the preference list $Q(w)$ that ranks $\mu_M(w)$ in the first position but otherwise leaves her

[6] Note that we cannot have $\mu'(v) \succ_v \mu_W(v)$, since that contradicts the fact that μ_W is the W-optimal stable matching. We also cannot have $\mu_W(v) \succ_v \mu'(v)$, since that implies $\mu'(v) = v$ and contradicts the fact that the set of individuals who are unmatched is the same at all stable matchings.

true preferences $P(w)$ unchanged. Denote by t the Nash equilibrium in which each woman w plays the optimal truncation strategy where all men ranked below $\mu_W(w)$ are deemed unacceptable.

Consider the equilibrium r and let μ be the matching that is produced. Clearly, we have that $\mu(w) = \mu_M(w)$ for all $w \in W$. Suppose that woman v deviates to strategy t_v and let μ' be the new matching that is produced. We argue that $\mu'(v) = v$ (i.e., woman v remains unmatched).[7] Thus, $u_w(r) = u_w(\mu_M(w))$ and $u_w(t_w, r_{-w}) = u_w(w)$ for all $w \in W$.

Similarly, consider the equilibrium t and let μ be the matching that is produced. Clearly, we have that $\mu(w) = \mu_W(w)$ for all $w \in W$. Suppose that woman v deviates to strategy r_v and let μ' be the new matching that is produced. We argue that $\mu'(v) = \mu_M(v)$.[8] Thus, $u_w(t) = u_w(\mu_W(w))$ and $u_w(r_w, t_{-w}) = u_w(\mu_M(w))$ for all $w \in W$.

We now know that $u_w(\mu_M(w)) - u_w(w) > u_w(\mu_W(w)) - u_w(\mu_M(w))$ is equivalent to $u_w(r) - u_w(t_w, r_{-w}) > u_w(t) - u_w(r_w, t_{-w})$. Since this is true for all $w \in W$, we have that

$$\prod_{w \in W} [u_w(r) - u_w(t_w, r_{-w})] > \prod_{w \in W} [u_w(t) - u_w(r_w, t_{-w})] \tag{3}$$

We conclude that equilibrium r risk-dominates equilibrium t. $\square$

Proposition 5 demonstrates the existence of a symmetric equilibrium in manipulation strategies that risk-dominates the symmetric equilibrium in truncation strategies. This result holds for markets in which the "risk-dominance condition" is satisfied (i.e.,

[7]To see this, note that woman v now rejects the offer from man $\mu_M(v)$ at some stage of the DA algorithm. All future offers that man $\mu_M(v)$ makes in later stages of the DA algorithm will also be rejected, since each woman $w \in W \setminus \{v\}$ prefers $\mu_M(w)$ to $\mu_M(v)$ according to their submitted preferences.

[8]To see this, note that man $\mu_M(v)$ prefers woman v to his match partner at μ (since woman v is his M-optimal stable match partner).

$u_w(\mu_M(w)) - u_w(w) > u_w(\mu_W(w)) - u_w(\mu_M(w)))$. This condition has a natural and intuitive interpretation: the manipulation equilibrium risk-dominates the truncation equilibrium when the utility loss from "over-truncation" (i.e., $u_w(\mu_M(w)) - u_w(w)$) exceeds the utility gain from "optimal truncation" (i.e., $u_w(\mu_W(w)) - u_w(\mu_M(w))$).

Although this condition may seem restrictive, we argue that it is likely to be satisfied in real-world settings for two reasons. First, having a large utility loss between matching with less-preferred partners and being unmatched reflects the fact that remaining unassigned might be particularly distasteful in field settings.[9] Second, the theoretical literature on "core convergence" shows that the span of the core is decreasing in market size. This suggests that there are limited gains from strategic preference misrepresentation in large markets (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Lee, 2014).

The existence of a payoff-dominant equilibrium in truncation strategies raises the question of whether it is reasonable to expect this equilibrium to arise. We have shown that, under plausible conditions, risk-dominance argues in favor of a different equilibrium. However, this obscures a more fundamental reason to doubt the emergence of the truncation equilibrium: to implement the optimal truncation strategy, it is necessary for each woman w to know the identity of $\mu_W(w)$. This requires the ability to calculate the set of stable matchings. As suggested by Roth and Sotomayor (1992), this is an unreasonable assumption for the real-world environments where this theory applies.

Absent this degree of strategic sophistication, truncation behavior carries with it the risk of "over-truncating" and remaining unmatched. We illustrate the salience of this risk through a simple behavioral model that allows for the possibility of mistakes

[9] In the NRMP, for example, remaining unassigned could mean no longer having the option to pursue a career in medicine.

in the calculation of the optimal truncation strategy. Suppose that a woman playing the optimal truncation strategy "trembles" and over-truncates her preference list with probability ϵ.[10] We can now state and prove the following result:

Proposition 6. *If* $\epsilon > \frac{u_w(\mu_W(w)) - u_w(\mu_M(w))}{u_w(\mu_W(w)) - u_w(w)}$ *and all women* $v \in W \setminus w$ *report their true preferences, then truth-telling yields a higher expected payoff than truncation for woman* w.

Proof. Let p_w denote the strategy in which woman w reports her true preferences and t_w denote the strategy in which woman w optimally truncates her preferences with a tremble probability of ϵ. For all $w \in W$, we have that

$$u_w(p) = u_w(\mu_M(w)) > (1-\epsilon)u_w(\mu_W(w)) + \epsilon u_w(w) = u_w(t_w, p_{-w})$$

$\square$

Define ϵ^* as the threshold tremble probability that makes woman w indifferent between truth-telling and optimal truncation (when all other women report their true preferences). In other words,

$$u_w(\mu_M(w)) = (1-\epsilon^*)u_w(\mu_W(w)) + \epsilon^* u_w(w)$$

Solving for ϵ^*, we have that

$$\epsilon^* = \frac{u_w(\mu_W(w)) - u_w(\mu_M(w))}{u_w(\mu_W(w)) - u_w(w)}$$

We can now demonstrate the following comparative statics properties:

[10] Thus, if there are n strategies that constitute over-truncation, then each over-truncation strategy is played with probability $\frac{\epsilon}{n}$.

Proposition 7. *The threshold tremble ϵ^* is increasing in the utility of matching with the most preferred achievable partner.*

Proof.

$$\frac{\partial \epsilon^*}{\partial u_w(\mu_W(w))} = \frac{u_w(\mu_M(w)) - u_w(w)}{[u_w(\mu_W(w)) - u_w(w)]^2} > 0$$

□

Proposition 8. *The threshold tremble ϵ^* is decreasing in the utility of matching with the least preferred achievable partner.*

Proof.

$$\frac{\partial \epsilon^*}{\partial u_w(\mu_M(w))} = \frac{-1}{u_w(\mu_W(w)) - u_w(w)} < 0$$

□

Proposition 9. *The threshold tremble ϵ^* is increasing in the utility of remaining unmatched.*

Proof.

$$\frac{\partial \epsilon^*}{\partial u_w(w)} = \frac{u_w(\mu_W(w)) - u_w(\mu_M(w))}{[u_w(\mu_W(w)) - u_w(w)]^2} > 0$$

□

There are several important advantages to directly incorporating a probability of error into the strategic model. First, for a sufficiently high ϵ, truth-telling can be explained through payoff-maximizing behavior. This is in line with experimental studies that demonstrate high truth-telling rates by women in the M-proposing DA mechanism (Castillo and Dianat, 2014; Echenique, Wilson, and Yariv, 2014; Featherstone and Mayefsky, 2014). Second, allowing for trembles clarifies the welfare implications of asymmetric equilibria. While the women prefer truncation equilibria to non-truncation equilibria, each individual woman prefers not to bear the truncation risk herself.[11]

This incentive to free-ride on other agents' truncation is implied in our model. Let p denote the strategy profile in which all women report their true preferences, and t denote the strategy profile in which all women optimally truncate their preferences. Clearly, for all $w \in W$ we have that

$$u_w(p_w, t_{-w}) = u_w(\mu_W(w)) > (1-\epsilon)u_w(\mu_W(w)) + \epsilon u_w(w) = u_w(t)$$

4 Discussion and Conclusion

In this paper, we investigated equilibrium selection in the context of two-sided matching clearinghouses. For a large class of markets, the familiar concepts of payoff-dominance and risk-dominance argue for the selection of different equilibria. While there is no equilibrium that supports truth-telling for all agents, payoff-dominance suggests a symmetric equilibrium in which participants *truncate* their preference lists while risk-dominance suggests a symmetric equilibrium in which participants *manipulate* their preference lists. By introducing a simple behavioral model that incorporates "trembles" in agents' strategies, we then derive several useful comparative statics predictions.

Since our analysis exploits the inherent trade-off between improving one's match partner and the risk of remaining unmatched, our results are necessarily sensitive to the assumptions placed on agents' cardinal utilities. The importance of cardinal payoff structures has also been demonstrated experimentally by Battalio, Samuelson,

[11] In asymmetric truncation equilibria of the Bayesian game where agents submit preferences to the M-optimal stable mechanism, the women who truncate less receive higher payoffs (Coles and Shorrer, 2014).

and Van Huyck (2001), who provide evidence that the pecuniary incentive to play a best-response strategy (i.e., optimization premium) significantly affects the behavior of laboratory subjects in coordination games. In our environment, when the optimization premium is small or the utility loss from remaining unmatched is large, the symmetric truncation equilibrium becomes less attractive since agents prefer to free-ride on others' truncation behavior.

Our findings suggest directions for future work. Rather than allowing for trembles in players' strategy choices, we can model the strategic environment with a noisy payoff structure. This is the approach to equilibrium selection taken by the "global games" literature (e.g. Carlsson and Van Damme, 1993). In addition, evolutionary approaches can help us to further refine the set of equilibria to those that are resistant to small mutations in the population.[12]

[12] For an introduction to the evolutionary approach to games, see Gintis (2000) or Samuelson (1998).

REFERENCES

Atila Abdulkadiroğlu and Tayfun Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–702, 1998.

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

Robert Aumann. Communication need not lead to nash equilibrium. *Hebrew University of Jerusalem Working Paper*, 1990.

Raymond Battalio, Larry Samuelson, and John Van Huyck. Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, 69(3):749–764, 2001.

Hans Carlsson and Eric Van Damme. Global games and equilibrium selection. *Econometrica*, pages 989–1018, 1993.

Marco Castillo and Ahrash Dianat. Truncation strategies in two-sided matching markets: Theory and experiment. *Working Paper*, 2014.

Peter A Coles and Ran I Shorrer. Optimal truncation in matching markets. *Games and Economic Behavior*, 87:591–615, 2014.

Lester E Dubins and David A Freedman. Machiavelli and the gale-shapley algorithm. *American Mathematical Monthly*, 88(7):485–494, 1981.

Federico Echenique, Alistair J Wilson, and Leeat Yariv. Clearinghouses for two-sided matching: An experimental study. *Working Paper*, 2014.

Clayton R Featherstone and Eric Mayefsky. Why do some clearinghouses yield stable outcomes? experimental evidence on out-of-equilibrium truth-telling. *Working Paper*, 2014.

David Gale and Marilda Sotomayor. Ms. machiavelli and the stable matching problem. *American Mathematical Monthly*, 92(4):261–268, 1985.

Herbert Gintis. *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton University Press, 2000.

John C Harsanyi and Reinhard Selten. *A general theory of equilibrium selection in games*. MIT Press, 1988.

Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–62. Society for Industrial and Applied Mathematics, 2005.

Fuhito Kojima and Parag A Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–627, 2009.

SangMok Lee. Incentive compatibility of large centralized matching markets. *Working Paper*, 2014.

Stephen Morris, Rafael Rob, and Hyun Song Shin. p-dominance and belief potential. *Econometrica*, pages 145–157, 1995.

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

Alvin E Roth and Marilda A Oliveira Sotomayor. *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press, 1992.

Larry Samuelson. *Evolutionary games and equilibrium selection*. MIT Press, 1998.

Tayfun Sönmez and Tobias B Switzer. Matching with (branch-of-choice) contracts at the united states military academy. *Econometrica*, 81(2):451–488, 2013.

BIOGRAPHY

Ahrash Dianat was born in Rochester, New York in 1987. He graduated from Pittsford Mendon High School, Pittsford, New York, in 2005. He received his Bachelor of Science degree from the Rochester Institute of Technology in 2009. He then received his Master of Arts in Economics from George Mason University in 2011.