On The Possibility of Manipulation in Two-Sided Matching*

Sherif G. Salem

Queen Mary, University of London, School of Economics and Finance, Mile End Road, London E1 4NS, UK.

American University in Cairo, School of Business, Economics department, AUC Avenue, P.O. Box 74 New Cairo 11835, Egypt.

And The British University in Egypt, Economics department, Misr Ismailia Desert Road El Sherouk City, P.O. Box 43, Cairo, Egypt.

E-mail address: gamalsalem.sherif@gmail.com

Abstract

We study a two-sided matching model where individuals have limited computational abilities. We show that in this case, a stable matching mechanism is not immune to manipulation via preferences. However, the task of manipulation for individuals with limited computational abilities is harder than for fully rational players.

Keywords: Bounded rationality, Stability, Two-sided matching.

JEL Classification Numbers: C78, D61, D03.

1 Introduction

Economics has been always concerned with analysing markets, but designing them is a relatively new field. Economic design includes designing incentives systems, contracts, auctions and trading platforms to generate socially optimal outcomes, to name a few. We focus here on a specific branch of economic design: matching theory. The theory came mainly in response to some market failures

---

*I would like to thank Andriy Zapechelnyuk and Christopher Tyson for helpful comments and suggestions.

1 Cell: +012-521-3344.

2 See Milgrom (2011) for a short survey on market design.
which led to the introduction of some centralized market clearing mechanisms (e.g. Roth and Xing 1994). Generally speaking, matching theory can be divided into two branches: (a) matching objects with agents like colleges with students, Kidneys with transplant patients, and so on. (b) Matching two different types of agents like women with men or employers with employees, and so on. Also, there is another classification: (a) one-to-one matching where each agent on one side of the market is matched with only one agent from the other side (e.g. men and women) and (b) many-to-one matching where a member of one side of the market can be matched with many members from the other side (e.g. firm and workers). It is widely regarded that matching theory originated by the seminal Gale and Shapley (1962) paper which described an algorithm (satisfying a nice criterion called stability) to allocate students to colleges. Nevertheless, it was not until 1984 when economists started getting really interested in Gale and Shapley (1962) when Alvin Roth published a paper (Roth 1984) showing that the clearinghouse mechanism used to match doctors with the residency program in the United States (Now called the National Resident Matching program) is in fact equivalent to Gale and Shapley (1962) algorithm. In recent years the theory has been extended to cover many market applications. For example Abdulkadiroglu and Sönmez (2003), Abdulkadiroglu, Pathak, and Roth (2005), and Abdulkadiroglu, Pathak, Roth, and Sönmez (2005) study the design of matching students with schools and colleges mainly in U.S. Similar studies for reforming the school matching system in different countries can be found in Biró (2008) for Hungary, Feng (2005) for China, Ting (2007) for Taiwan; and Selim and Salem (2010) for Egypt. Also research on designing organ transplantation networks (like kidney exchange) can be found in Roth, Sönmez, and Ünver (2004), and Roth, Sönmez, and Ünver (2005).

[3] There is also few literature on many-to-many matching, for example Echenique and Oviedo (2006).
The central solution concept used in the matching literature is the notion of stability. Simply, a matching is stable if no agent from any side of the market is matched to unacceptable partner and if there exists no pair of agents who would prefer to be matched to each other than to their current mates. One reason to focus on stable outcomes is because otherwise there will exist always a pair of agents (or just one) who would like to break the final match once announced. Another reason is an empirical one, Roth (2002) makes a comparison between different markets that historically adopted centralized clearinghouses (e.g. American medical markets, British Regional Medical Markets, Canadian Lawyers Markets). He finds that markets which adopted stable matching mechanisms have mostly succeeded while those who adopted not stable matching mechanisms have mostly failed. A stronger evidence about the success of stable matchings can be found in the markets of new physicians in Britain (Roth 1991). Another property that economists would like to have when designing a matching mechanism is non-manipulability i.e. the mechanism is immune to any kind of manipulation. Unfortunately, Roth (1982) shows, for the one-to-one matching case, that we cannot have both properties in one mechanism. More specifically, he shows that for any stable mechanism one can find an example where there exists at least one agent who have an incentive to manipulate his or her preferences to end up matched with a better mate. In many-to-one matching markets there are similar strategic concerns, Sönmez (1997, 1999) shows two more different kinds of manipulation of stable matchings when matching students with colleges: a) manipulation via capacities, when colleges benefit by underreporting their quotas; and b) manipulation via pre-arranged matches when it is in the interest of colleges and students to agree to match before receiving their final matches from the clearinghouse. Now, the inevitable question is why are stable mechanisms popular and why do they perform well in real markets despite that they are not immune to manipulation?
One approach to address this problem is by imposing exogenous restrictions on participants’ preferences that produce the desirable matching (e.g., Alcalde and Barbera (1994) for preference manipulation, and Kojima (2007), Kesten (2006), and Konishi and Unver (2006) for manipulation via pre-arranged matches and capacity manipulation). The other approach is motivated by noting that Roth (1982) impossibility theorem works well in small markets where all men and women (possibly) know each other. However, we can expect some problems to emerge when moving to real-life large markets. For example, in NYC there are 500 school programs and 90,000 students to be matched every year (Abdulkadiroglu, Pathak, and Roth 2009). Another example is the National Resident Matching Program (NRMP) which is the matching system applied in the U.S. to match medical residents (over 20,000 students) and hospital residency programs (3,000 to 4,000 hospitals) every year.

The motivation of studying large markets is due to Roth and Peranson (1999), they note after conducting a series of simulations using NRMP data and also using randomly generated data that only few hospitals and students can benefit from manipulating their preference lists (or via capacity manipulation for hospitals). For instance, in 1996, only 21 students from 24,749 students can benefit from submitting false preference lists assuming the data are the true preferences. More specifically, Roth and Peranson (1999) show that very few players have more than one stable matching (and hence can benefit from manipulation). They conjecture that in a probabilistic environment, when a market gets larger the probability of each player getting matched to the same partner in all stable matchings is one. This is known as the core convergence property. Immorlica and Mahdian (2005) formally prove it for the one-to-one matching case where each college has a quota of one. Kojima and Pathak (2008) generalize it to the many-to-one matching case where colleges have quotas of more than one.
In this paper we present a different approach to address the possibility of manipulation in large markets by assuming that players are *boundedly rational* in a computational sense. We show that assuming bounded rationality is not enough to rule out profitable manipulation, but makes it *more* difficult for players to manipulate. Our framework is different from Roth and Peranson (1999); and Immorlica and Mahdian (2005). The main difference is that the lack of manipulability in the previous papers is due to the core convergence property derived in a probabilistic setting, while our results are derived from the bounded rationality assumption without assuming any probabilistic environment.

The paper proceeds as follows. Section 2 presents the marriage model. Section 3 defines bounded rationality and presents the main result. Section 4 shows why it can be difficult for boundedly rational participants to manipulate a stable DA mechanism. Section 5 concludes.

### 2 The Marriage Model

The two sides of the marriage market are men and women. Men and women are gathered in two disjoint finite sets, $M := \{m_1, m_2, ..., m_n\}$ for men and $W := \{w_1, w_2, ..., w_l\}$ for women. Each man has a preference list $P(m)$ over $W \cup \{m\}$ and similarly each woman has a preference list $P(w)$ over $M \cup \{w\}$. If $P(m) := w_2, w_1, m$, then player $m$ prefers to be matched to $w_2$, then to $w_1$, then to remain single, and the same for $P(w) := m_2, m_1, m$. Let the set of all preference lists be denoted by $\mathbf{P} := \{P(m_1), P(m_2), ..., P(m_n), P(w_1), P(w_2), ..., P(w_l)\}$ for all $n \in N$ men and $l \in L$ women. Preferences are assumed to be rational (complete and transitive) and strict (no indifferences). In case if a man (woman) is indifferent between two choices, a *tie breaking* rule can be introduced that indicates how a man (woman) can rank his (her) choices based on a specific criterion like the age of the mate\(^4\) (e.g. younger is better). A marriage market results in a set of matchings between men and women plus singles (if any).

\(^4\) The Turkish placement office matching students with colleges in Turkey applies student age as a tie breaking rule when there are some students with the same test score (Balinski and Sönmez 1999).
Definition 1. A matching $\mu$ is a function mapping from $M \cup W$ to itself that associates with every individual $p$ a mate $\mu(p)$ in $M \cup W$ and satisfies: (i) $\mu(m) = w$ if and only if $\mu(w) = m$. (ii) If $\mu(m) \neq m$ (individual $m$ not matched to himself), then $\mu(m) \in W$. Similarly, if $\mu(w) \neq w$, then $\mu(w) \in M$.

Hence if woman $w_1$ is matched to man $m_1$, then man $m_1$ is matched to woman $w_1$. Also, a player who is not matched to a mate from the opposite sex remains single (or self-matched). We write the matching between $M := \{m_1, m_2, m_3\}$ and $W := \{w_1, w_2, w_3, w_4\}$ as the set of matching pairs $\mu := \{m_1 \leftrightarrow w_2; m_2 \leftrightarrow w_3; m_3 \leftrightarrow w_1; w_4 \rightarrow w_4\}$. It means simply that $m_1$ is matched to $w_2$, $m_2$ is matched to $w_3$, $m_3$ is matched to $w_1$, and $w_4$ remains single. However, we need to be picky when choosing which matching outcomes are reasonable from agents’ viewpoint.

The matching literature requires that matching outcome satisfies a criterion called stability. The idea is quite simple, a matching is stable if no agent is matched to a mate he or she does not accept and if there exists no pair of agents who prefer to be matched to each other than their current mates. The first requirement is called individual rationality and the second is the no blocking pair requirement.

Definition 2. The matching $\mu$ is individually rational if no agent is matched to unacceptable mate.

This is straightforward because we allow agents to remain single, hence if an agent is faced with an unacceptable matching he or she can simply reject the proposal and be self matched. It can be seen as minimal requirement for a solution concept. Clearly, whatever agents’ preferences are there exists always at least one individual rational matching since any matching that makes all agents single is individually rational. Nevertheless, individual rationality is still not satisfactory alone. Consider a male $m$ who remains self-matched (which is his least preference) under the matching $\mu$ and assume there exists a female $w$ who prefers $m$ to its current matching. Then they can break their current matching and be matched together. This motivates the following definition:
Definition 3. The matching $\mu$ is blocked by a pair of agents $(m, w)$ if they prefer each other to their current mates i.e. $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$.

Where $\succ_p$ is the preference relation of agent $p$. The pair $(m, w)$ is called a **blocking pair**. Now we have two criteria for a stable matching.

Definition 4. The matching $\mu$ is stable if it is individually rational and there are no blocking pairs.

As mentioned before, Stability is a very demanding concept from both a theoretical and empirical point of view. Two important questions arise in the context, does a stable matching always exist? If the answer is yes, how to find it? Gale and Shapley (1962) answered both questions. The first question has a positive answer (theorem 1 below) which is fortunate since as, Kojima and Troyan (2010) notes, that there are many economic problems with no solutions. For example, non-cooperative games may not have pure strategy Nash equilibrium and Cooperative games may have empty core.

**Theorem 1.** (Gale and Shapley 1962) A stable matching exists for every marriage market.

One of the (possible) multiple stable matchings can be found using the following algorithm.

**Men Proposing Deferred Acceptance Algorithm** (Gale and Shapley 1962)

Imagine a situation where there exists a centralized matchmaker who collects information about the preferences of all players and then carry out the following steps:

**Step 0:** If some agents have weak preferences, introduce a tie breaking rule.

**Step 1:** (a) Each man $m$ proposes his 1’st choice from the list of his acceptable choices (if any).

---

\[5\] See the introduction for more details.
(b) Each woman $w$ rejects any unacceptable proposals, and if she receives more than one acceptable proposal she “holds$^6$” her top choice and rejects other proposals.

$\vdots$

Step $k$: (a) Each man who was rejected at step $(k-1)$ will make a new proposal to his most preferred mate that has not rejected him in any period before. If no acceptable mates remain, he makes no new proposals.

(b) Each woman holds her most preferred acceptable offer to date and rejects the rest.

Stop: The algorithm terminates when no new proposals are made by matching each woman to the man (if any) whose proposal she is holding.

If for any matching $\mu$, $m_i$ and $w_j$ are matched at a step $k$ of the DA we write $\mu: = (k)\ m_i \longleftrightarrow w_j$. The matching resulting from the men proposing deferred acceptance is denoted $\mu_M$ while if we reverse the players’ role such that women start proposing to men the resulting matching is denoted by $\mu_W$[7]. A stable matching $\mu$ is said to be M-optimal stable matching (W-optimal stable matching) if every man (woman) considers it at least as good as any other stable matching $\mu'$. Although both matchings are stable, men prefer $\mu_M$ to $\mu_W$ while women prefer the opposite.

Theorem 2. (Gale and Shapley 1962) When all men and women have strict preferences, there always exists an M-optimal stable matching, and a W-optimal stable matching. Furthermore, the matching $\mu_M$ ($\mu_W$) produced by men (women) proposing DA algorithm is the M (W)-optimal stable matching.

The DA algorithm is appealing because it has been successful in labor market applications like the National Resident Matching Program (NRMP) and the Boston and NYC school systems[8].

---

[6] Roth and Sotomayor (1991) refers to this as being “engaged”.

[7] Here we just focus on the Men Proposing DA, so $\mu_m$ is simply denoted $\mu$.

[8]
2.1 Strategic Behavior

Until now we have nice results about the existence of a stable matching (Theorem 1) and the way to compute it (the DA algorithm), however this is not enough to make the DA a good matching mechanism because of the incentives consideration. If the DA does not give incentive for players to truthfully report their true preferences, then the DA may not be stable with respect to the true preferences. We focus here on the case of a central matchmaker who collects participants’ preferences which are used as input to an algorithm that produces the final matching. Roth and Sotomayor (1990) show that some of the results, by the revelation principle, also apply to decentralized markets when there is no matchmaker.

The final matching will depend on the list of preferences that players will state (or reveal) to the central clearinghouse. The vector of stated preferences is denoted by $Q := \{Q(m_1), Q(m_2), \ldots, Q(m_n), Q(w_1), Q(w_2), \ldots, Q(w_l)\}$ where $Q(m_i)$ is the revealed preference list for man $i$ and $Q(w_j)$ is the revealed preference list for woman $j$. A marriage market now is a tuple $(M, W, P, Q)$. Note that the matchmaker produces a matching as a function of the revealed and not the true preferences. Therefore the outcome of the matching $\mu$ is a function $(h)$, called the matching mechanism, that describes the output of the matchmaker’s algorithm given the vector of stated preferences $(Q)$, i.e. $\mu = h(Q)$. A matching mechanism is stable if $h(Q)$ is always stable with respect to $Q$.

**Definition 5.** A strategy $Q(p)$ for an agent $p$ is said to be a dominant strategy if it is a best response to all possible sets of strategy choices $Q_{-p}$ by the other agents.

An important desideratum in market design is strategy proofness.

**Definition 6.** A matching mechanism is strategy proof if it makes it a dominant strategy for each player to state his or her true preferences in the strategic game it induces.

---

Fortunately the men proposing DA satisfies strategy proofness for men. This is quite intuitive since the proposing side (men in this case) have no incentive to misreport their preferences since the algorithm will assign a man $m_i$ to his most preferred choice $w_j$ if $m_i$ is acceptable to $w_j$ (see theorem 2). However, this is not the same for women who have an incentive to manipulate their preference lists. This is shown in the following impossibility theorem.

**Theorem 3.** (Roth 1982) *There is no stable and strategy proof matching mechanism.*

Therefore there exists at least one agent who prefers *not* to reveal his or her true preferences to the matchmaker. To show this, it is sufficient to find a marriage market where no stable matching makes truthful revelation a dominant strategy, which is shown by Roth (1982).

## 3 The Case Of Bounded Rationality

Until now we have maintained two crucial assumptions about the kind of players we have and the environment of the matching market. The first assumption that players are *fully rational* and the second assumption is that of *complete information* of the marriage market. Roth and Rothblum (1999) conclude from a labor market model with incomplete information that potential strategic manipulation can be reduced but not eliminated, and hence Roth (1982) result still holds. Our goal here is to relax the assumption of full rationality while keeping that of complete information. A good motivation for this is Jones (1999) statement:

> Findings from behavioral organization theory, behavioral decision theory, survey research, and experimental economics leave no doubt about the failure of rational choice as a descriptive model of human behavior. But this does not mean that people and their politics are irrational. Bounded rationality asserts that decision makers are intendedly rational; that is, they are goal oriented and adaptive, but because of human cognitive and emotional architecture, they sometimes fail, occasionally in important decisions.
Herbert Simon was the first to criticize the traditional use of rationality assumption in economics, and he was actually the one who coined the term *bounded rationality* (e.g. Simon 1955, 1972). Since then, the theory of bounded rationality has been wide in scope and there are many models describing different states of human behavior (especially the economic one). Some models investigate decision making (e.g. Tyson 2008), others examine knowledge (e.g. Geanakoplos 1989 and Hintikka 1962) and limited memory (e.g. Piccione and Rubinstein 1997), to name a few.9

Our focus will be on a specific type of bounded rationality which is: *limited computational ability* of players (e.g. Tsang 2008; Rubinstein 1998). The main idea is based on observing that full rationality does not require only complete information, but it requires costly computation. Moreover, as Rubinstein (1998) notes that even if computation is not costly still there are some computational *impossibilities* which prevent players from achieving full rationality. Consider the following Traveling sales man problem:

Suppose a merchant has to visit his customers, who are located far apart. Traveling from one customer to another involves a cost, which may vary depending on the time and distance to travel. Some customers may not be available at all times. Suppose the merchant wants to plan an itinerary that visits all his 100 customers, with the objective to minimize traveling costs and satisfying all the customers’ availability constraints. A “rational” merchant would attempt to find the optimal itinerary in the above problem. The amount of computation required to find the optimal itinerary in this problem is nontrivial ... Given a fixed amount of planning time, one may not be able to find the optimal itinerary (i.e. an itinerary with the minimal traveling cost) (Tsang 2008).

Then to emphasis the importance of the computational aspect, Tsang adds:

It is also worth noting that computation itself involves a cost. Knowledge acquisition (e.g. to find out the traveling costs between two cities) could also involve costs. A rational agent should not only minimize traveling cost. It should attempt to minimize the traveling cost plus the cost of computation and knowledge acquisition.

---

Since we assume complete information in the marriage market, players do not have knowledge acquisition costs. Nevertheless, computational costs is still a hindering factor from achieving full rationality.

3.1 Preferences Manipulation And Computational Complexity

In this section we introduce a form of computational complexity, in the context of the marriage market, that might prevents players from optimally calculating their payoffs and check whether this affects their ability of manipulating their preference lists successfully. A common problem in studying bounded rationality is how to quantify it. This made some researchers like Selten (2001) to pinpoint that the best way to understand bounded rationality is to resort to qualitative reasoning. Nevertheless, Tsang (2008) argues in his Computational Intelligence Determines Effective Rationality theory (CIDER) that the degree of rationality of an agent depends on his or her computational power, especially if we consider the case when decisions can be encoded in heuristics and algorithms. In two sided matching the final outcome of any matching mechanism depends, inter alia, on how the algorithm used by the clearinghouse works. Therefore a rational agent should perform two basic assignments before choosing which strategy to play:

1. Calculate his or her payoff under the final matching in the case of submitting his or her true preference list.

2. Calculate his or her payoff under the final matching in the case of submitting a more profitable false preference list.

Recall that according to Roth (1982) theorem, if we have a stable mechanism then there exists at least one agent who can profitably misrepresent his or her preferences assuming that the others report the truth. This is an existence result, but is it feasible in real life market applications? In other words, does there exist an agent who can perform all the computations required no matter how complex they are to make a rational choice between submitting the true or a false preference list. To see the point consider the following example.

Example 1. Consider a 2×2 marriage market with the following preference lists.

Note that this requires not only complete information about other players’ preferences but also complete understanding of how the algorithm works, which we assume.
The resulting M-optimal stable matching is \( \mu : (1) \ m_1 \leftrightarrow w_1; \ m_2 \leftrightarrow w_2 \). Now consider the following manipulation by \( w_1 \) such that \( Q(w_1) := m_2 \), then the resulting stable matching \( v : (2) \ m_1 \leftrightarrow w_2; (3) \ m_2 \leftrightarrow w_1 \), which results in higher payoff for \( w_1 \) (with respect to the true preferences). In a 2×2 market the task of performing a successful manipulation is easy, since there is no much computation required to perform. Nonetheless in a 20×20 market we can expect that the computational task will get more difficult for players.

### 3.1.1 Measuring Complexity

The (CIDER) theory suggests to model bounded rationality by limiting individuals’ computational abilities. Although statements like “an agent is 60% rational” is not plausible, but a natural criterion for measuring complexity is: the \( k \) number of DA steps that players have to follow to calculate their payoffs. For instance in example 1, the DA of \( \mu \) has one step while for \( v \) has 3 steps. So, it is natural to consider \( v \) more complex relative to \( \mu \) in example 1.

**Definition 7.** A matching \( \mu \) is more complex than \( v \) for player \( p \) if \( |DA_\mu| > |DA_v| \).

Where \( |DA_\mu| \) denotes the number of DA steps associated with a matching \( \mu \). Note that \( \mu \) and \( v \) are any two different matchings whether for two different markets or for the same market. A player \( p \) is boundedly rational if he or she has \( k \)-bounded computational capacity i.e. can only trace up to \( k \) DA steps for any matching. But if the number of DA steps exceed \( k \) than \( p \) submits his or her true preference list.

**Definition 8.** (Bounded Rationality) A player \( p \) is boundedly rational if when \( |DA_\mu| > k \) for any matching \( \mu \), then \( P(p) = Q(p) \).

\( P(p) = Q(p) \) means that player \( p \)’s true preference list is the same as the submitted one. The idea is that if a player has bounded computational capacity of 3 steps for instance, but then in a big market where he or she needs to compute more than 3 steps, then he or she simply submits the true preference list since there is no way to compute profitable misrepresentations. Notice that as \( k \to \infty \), player \( p \) becomes fully rational in the traditional sense.

---

[11] For simplicity, we drop the self-match choice notation from all men and women’s preference lists.
3.1.2 Preferences Manipulation In Large Markets

In this part we attempt to answer the following question: is it possible for boundedly rational individuals to successfully manipulate their preference lists in large markets? The problem we face in large markets is that the number of possible manipulations is a function of the number of players. Consider a $4 \times 4$ marriage market, then to know which manipulations are successful, each woman (in case of a men proposing DA) need to compute and compare all possible manipulations i.e. $4!+4 \times 3 \times 3+3+4=64$. Fortunately, we know that any successful manipulation can be replicated via a truncating strategy which makes the list of computation much shorter.

**Definition 9.** (Preference Truncation) A truncation of a woman $w_j$ preference list $P(w_j)$ containing $n$ acceptable men is a list $Q(w_j)$ containing $n' < n$ acceptable men, such that the $n'$ elements of $Q(w_j)$ are the first elements of $P(w_j)$ in the same order.

For example, if a woman has the following preferences, $P(w_j):= m_1, m_2, m_3, m_4$ and $w_j$ was originally matched to $m_3$. Then, submitting a truncated preference list at the matching point is $Q(w_j):= m_1, m_2, m_3$ and submitting a truncated preference list above the matching point is $Q(w_j):= m_1, m_2$. It turns out that preferences manipulation via truncation is at least as good as any other profitable manipulation (if any).

**Lemma 1.** (Roth and Vande Vate 1991) Consider an arbitrary stable mechanism, now fix the preferences of all players other than $p$. Suppose the mechanism produces $\mu$ under some report for $p$. Then there exists a truncation strategy $Q(p)$ that produces a matching $\mu'$ such that $\mu'$ is weakly preferred to $\mu$ according to $p$’s true preferences.

Intuitively, if truncation is above the matching point then lemma 1 implies that $\mu'$ is strongly preferred to $\mu$. Lemma 1 is very useful because now we can focus on a subclass of all possible manipulations. In the $4 \times 4$ marriage market we have now only $2^{16} = 2^{4} = 16$ strategies instead of 64. Returning to the question of existence of stable and strategy proof matching in large markets. It turns out that Roth (1982) existence result still holds even for extremely boundedly rational individuals.
Proposition 1. In an \( n \times l \) marriage market, such that \( n, l \geq 2 \), there exists no stable matching mechanism such that stating the true preferences is a dominant strategy for all players, including boundedly rational ones.

Proof. Here it is sufficient to prove one counterexample for a specific marriage market and show that no stable mechanism induce truth revelation as dominant strategy for all players, including boundedly rational ones. Consider the following \( n \times l \) marriage market with men proposing DA, assuming that women are extremely boundedly rational with \( k = 3 \)\(^{12}\). Let men have the following preferences:

\[
\text{for all } i \in N \setminus \{n-1, n\} \text{ and } j \in L \setminus \{l-1, l\}, P(m_i) := w_j, \ldots, w_1, \text{ such that } i = j, P(m_{n-1}) := w_l, w_{l-1}, \ldots, w_j. \text{ Let women have the following preferences: } P(w_j) := m_i, \ldots, m_n, \text{ such that } j = i, P(w_{l-1}) := m_{n-1}, m_n, \ldots, m_i \text{ and } P(w_l) := m_n, m_{n-1}, \ldots, m_i. \]

The stable matching resulting from applying the DA is \( \mu: (1) \) \( m_i \leftrightarrow w_j; m_{n-1} \leftrightarrow w_l; m_n \leftrightarrow w_{l-1} \) with one DA step generated. Note however that if \( Q(w_l) := m_n \), then the stable matching resulting is \( v: (1) \) \( m_i \leftrightarrow w_j; (2) m_{n-1} \leftrightarrow w_{l-1}; (3) m_n \leftrightarrow w_l \) with 3 DA steps generated. This misrepresentation does not only generates higher payoff for \( w_l \), but also is Pareto optimal. Since the model is symmetric we get same result for women proposing DA for the second stable matching, completing the proof. \( \square \)

Note that requiring \( k = 3 \) is a sufficiently low requirement since it is the same condition required for a successful manipulation in any \( 2 \times 2 \) market.

Remark 1. In any \( 2 \times 2 \) marriage market, if \( w_j \) can compute up to three DA steps, then she can gain by misreporting her preferences. To see this, check \( 2 \times 2 \) marriage markets with all different preferences ordering and all possible preferences manipulations. Notice that all manipulations lead to the same or lower payoff with the exception of the following case: \( P(m_1) := w_1, w_2, P(m_2) := w_2, w_1 \) and \( P(w_1) := m_2, m_1, P(w_2) := m_1, m_2 \). The resulting stable matching is \( \mu: (1) m_1 \leftrightarrow w_1; m_2 \leftrightarrow w_2. \) Now consider the following manipulations by \( w_1, Q(w_1) := m_2; \) and by \( w_2, \) \( Q(w_2) := m_1 \) then the resulting stable matching \( v: (2) m_1 \leftrightarrow w_2; (3) m_2 \leftrightarrow w_1 \), which results in higher payoff for \( w_1 \) and \( w_2 \) (with respect to the true preferences).

\(^{12}\) See remark 1 for an explanation.
Corollary 1. Increasing the number of players in a marriage market is not a sufficient condition for making the matching resulting from a successful manipulation via preferences more complex.

This can be seen from the examples mentioned in the proof of proposition 1 and remark 1. Note that letting the market size grow from the $2 \times 2$ market of remark 1 to the $n \times l$ market used in the proof of proposition 1 does not increase the computational task required for successful manipulation.

4 Two Sources Of Complexity

In spite of the existence result in proposition 1, it does not mean that it is easy for bounded rational players to carry on successful manipulations whenever possible. We have two main sources for this difficulty: firstly, because profitable preference truncation lead to bigger number of DA steps to compute (under certain conditions). Secondly, because the way players’ preferences are ordered can make some markets more manipulable than others (under certain conditions).

4.1 Profitable Preferences Truncation

From Lemma 1 we know that if there exists a profitable manipulation via preference list, it can be replicated via a truncation strategy. The aim here is to show that for bounded rational players this is also a problem since truncating preference lists, under some conditions, also increase the computational task required from players.

Lemma 2. (Roth 1982) Consider a market $(M,W,P,Q)$, if a player $p$ is unmatched at some stable matching then he or she is unmatched at every stable matching.

From lemma 1 and 2, we conjecture the following:

Conjecture 1. Let $\mu$ be a stable matching generated by $Q(p)$ and $\mu'$ be a stable matching generated by $Q'(p)$ such that $\mu'$ is preferred to $\mu$ by player $p$. Then $|DA_{\mu'}| > |DA_\mu|$.
**Proof.** Since the model is symmetric, then we can choose \( p = w \). Let \( Q(w_j) = m_1, ..., m_k, ..., m_n \) such that the resulting stable matching is \( \mu: w_j \rightarrow m_k \). From lemma 1, there exists a truncation \textit{above} the matching point\(^{13}\) \( Q'(w_j) := m_1, ..., m_k' \), resulting in a stable matching \( \mu' \) such that \( \mu' \succ_{w_j} \mu \), where \( m_k' \succ_{w_j} m_k \). Let \( w_j \) be unmatched under \( \mu' \), then by lemma 2 \( w_j \) will be unmatched also under \( \mu \) but then \( \mu \sim \mu' \), a contradiction. Hence \( w_j \) be matched under \( \mu' \). Since other players’ preferences are the same then \( m_k \) will propose to \( w_j \) as in \( \mu \), but now \( w_j \rightarrow m_i \) where \( m_i \in [m_1, m_k] \). Hence \( m_k \) will be rejected and propose to his next choice (if any) generating at least one more step such that \( |DA_{\mu'}| > |DA_{\mu}| \).

**Important remark about conjecture 1.** We have \textit{two conditions} for the conjecture to hold: (1) \( w_j \) is not the last choice available in \( m_k \)'s preference list. (2) The extra steps that make \( |DA_{\mu'}| \) greater than \( |DA_{\mu}| \) should be generated at the last step of the original \( DA_{\mu} \). In other words, if the extra generated proposal (due to \( w_j \) preferences truncation) of \( m_k \) in the matching \( \mu' \) does not occur at the last step of \( DA_{\mu} \), then it is not clear whether \( DA_{\mu'} \) will have more steps than \( DA_{\mu} \) or not.

### 4.2 Market Irregularity

What makes a market \( M_1 \) different from any other market \( M_2 \) (keeping the number of players constant) is how men and women rank each others in one market relative to the other market. Observing that these differences in preferences ordering among participants play a role in making number of \( DA \) steps computation more or less complex for boundedly rational participants motivates the following two definitions.

**Definition 10.** (Multiple Proposals) Multiple proposals for a player \( p \) take two forms: (a) if \( p \) is single then \( p \) receives at step \( k \), many proposals from at least two players \( q \) and \( q' \). (b) If \( p \) is engaged at step \( k \), then \( p \) receives another proposal from at least one player \( q' \) at any subsequent step.

\(^{13}\) The matching point is \( m_k \).
So, in (a) $p$ is single and receives simultaneous proposals from many players at the same time. But in (b) $p$ is engaged at some step $k$ and receives another proposal from one or more players at any subsequent step.\footnote{There is also the case when $p$ is single a receive multiple offers at two different steps, but we skip it to shorten the analysis.}

**Definition 11.** (Market Irregularity) A market $M$ is irregular of order $n$ (i.e. $M$ is $I(n)$), if there are $n$ DA steps such that at each step of the $n$ steps there exists at least one player who receives multiple proposals.

Therefore if a market is irregular of order 5 (i.e. $M$ is $I(5)$), then there are 5 DA steps such that there exists at least one player who receives multiple proposals at each step of the 5 steps. To get the intuition of how multiple proposals can complicate the computation, we give two examples for case (a) and (b) of definition 11. Take case (a) first:

**Example 2.** Consider a regular $3 \times 3$ market with no multiple proposals. $P(m_1) := w_2, w_1, w_3; P(m_2) := w_1, w_2, w_3; P(m_3) := w_3, w_1, w_2; P(w_1) := m_1, m_2, m_3; P(w_2) := m_2, m_1, m_3; P(w_3) := m_1, m_2, m_3$. The resulting $M$-optimal stable matching is $\mu := (1) m_1 \leftrightarrow w_2; m_2 \leftrightarrow w_1; m_3 \leftrightarrow w_3$. Here the market is irregular of order zero (i.e. $I(0)$), since no single woman receives multiple proposals at any step. Now consider a different $3 \times 3$ market. $P(m_1) := w_1, w_3, w_2; P(m_2) := w_1, w_2, w_3; P(m_3) := w_1, w_3, w_2; P(w_1) := m_2, m_1, m_3; P(w_2) := m_1, m_2, m_3; P(w_3) := m_3, m_2, m_1$. The resulting $M$-optimal stable matching is $\mu := (1) m_2 \leftrightarrow w_1; (2) m_3 \leftrightarrow w_3; (3) m_1 \leftrightarrow w_2$. Here we have more DA steps because $w_1$ and $w_3$ receives multiple proposals at step 1 and 2 respectively and hence the market is $I(2)$. Now take an example for case (b),

**Example 3.** Take the following market $P(m_1) := w_2, w_1, w_3; P(m_2) := w_1, w_2, w_3; P(m_3) := w_3, w_1, w_2; P(w_1) := m_3, m_1, m_4; P(w_2) := m_2, m_1, m_3; P(w_3) := m_1, m_2, m_3$. The resulting $M$-optimal stable matching is $\mu := (1) m_3 \leftrightarrow w_3; (2) m_2 \leftrightarrow w_2; (3) m_1 \leftrightarrow w_1$. Here the number of steps increased compared to the $I(0)$ market in example 2 because $w_2$ receives a second better proposal from $m_2$ at step 2. Hence the market is $I(1)$. It might be intuitive to see that under certain conditions as market irregularity increases the number of DA steps increase.
Conjecture 2. Let $M_1$ and $M_2$ be two different markets with same $n \times l$ number of players. Let $M_1$ be irregular of order $t$ with $\mu'$ stable matching and $M_2$ be irregular of order $r$ with $\mu$ stable matching such that $t > r$. Then $|DA_{\mu'}| > |DA_{\mu}|$.

**Proof.** Let $M_1$ be $I_1(t)$, and let $M_2$ be $I_2(t + s)$ such that $t + s = r$. Let $s = 1$, then from definition 11 we have two cases for player $p = w_j$\(^{[15]}\). Case (a): $w_j$ receives multiple proposals from $m_i$ and $m'_i$, then either $w_j \leftarrow m_i$ (hence $m'_i$ proposes to his next choice, if any) or $w_j \leftarrow m'_i$ (hence $m_i$ proposes to his next choice, if any) or $w_j \leftarrow w_j$ (hence $m_i$ and $m'_i$ propose to their next choices, if any). In the three cases we have at least one more DA step. In case (b) $w_j \leftarrow m_i$ at step $k$ and receives a proposal from $m'_i$ at any subsequent step such that $m'_i \succ_j m_i$, then we have one disengagement\(^{[16]}\) and one extra DA step. If not, then $m'_i$ proposes to his next choice (if any) and also we have one extra DA step. \(\square\)

Notice that conjecture 2 has the same *two limitations* mentioned in “remark about conjecture 1”. From conjectures 1 and 2, a corollary follows about manipulation in irregular markets.

**Corollary 2.** Let $M_1$ and $M_2$ be two different markets with same $n \times l$ number of players. Let $M_1$ be irregular of order $t$ with $\mu$ stable matching generated by $Q(p)$, and with $\mu'$ stable matching generated by $Q'(p)$ such that $\mu'$ is preferred to $\mu$ by player $p$. Also, let $M_2$ be irregular of order $r$ with $\nu$ stable matching generated by $Q(p)$, and with $\nu'$ stable matching generated by $Q'(p)$ such that $\nu'$ is preferred to $\nu$ by player $p$ and, where $r > t$. Then $|DA_{\nu'}| > |DA_{\mu'}|$.

In words, corollary 2 states that preferences truncation in more irregular markets generate more DA steps than in less irregular markets.

---

\(^{[15]}\) This is because the model is symmetric.

\(^{[16]}\) For all players $i, i' \in N$ and $j, j' \in L$, $m_i$ and $w_j$ are disengaged if at any step $k$ of the DA, $m_i$ and $w_j$ are engaged ($m_i \leftarrow w_j$) and then at any subsequent step, we have either $m_i$ engaged to another mate $w_j'$ or self-matched, and either $w_j$ engaged to another mate $m_i'$ or self-matched. In words, two players are disengaged if at any step they were matched tentatively (engaged) and then at a latter step they are not.
5 Discussion And Open Questions

The question of why the adaptation of different versions of the DA algorithm have been successful in professional labor markets, medical markets, education markets, and others is central to the theory of matching. The suggested answer by theory is that because the matchings produced under the DA are stable, however this is puzzling because there are a series of impossibility results in the literature about the non vulnerability of the DA to manipulation in different environments (e.g. Roth (1982), Roth and Rothblum (1999)). One approach to solve the puzzle is proposed by by Immorlica and Mahdian (2005) and Kojima and Pathak (2008) who show that the solution is to move to large markets. Large markets, under some conditions, guarantee that the set of participants who can profitably manipulate the DA mechanism is extremely small, but not completely eliminated. A more traditional approach is to assume some restrictions on participants' preferences that ensure some desirable properties to be produced by a mechanism (e.g. Alcalde and Barbera 1994).

In this paper we suggest a bounded rationality approach to address the problem. More specifically, by assuming that participants have $k$-limited computational ability (i.e. they can only trace limited $k$ DA steps). The final conclusion is that although we can not completely rule out profitable manipulation for the DA mechanism (proposition 1), the task of manipulation might be more difficult (under certain conditions) for boundedly rational players compared to fully rational players (conjectures 1 and 2). For the model to be consistent with Roth and Peranson (1999) empirical and experimental findings, we need to answer the following open question:

**Question 1:** If we assume that $k=3$ is the minimum condition required for boundedly rational individuals to manipulate a stable DA mechanism and if men and women have randomly generated preference lists according to an arbitrary distribution. What is the probability of finding a manipulable marriage market from all possible markets? If the answer is that the probability tends to zero as market size tends to infinity, that will explain Roth and Peranson (1999) results. Another open question is:
Question 2: Can we prove conjectures 1 and 2 after dropping the two required conditions mentioned before, such that $|DA_{\mu'}| \geq |DA_{\mu}|$? We have good reasons to believe that $|DA_{\mu'}| > |DA_{\mu}|$ stated in conjectures 1 and 2 will not hold globally, but do $|DA_{\mu'}| \geq |DA_{\mu}|$ holds?

References


