Red Nacional de Investigadores en Economía

Obvious Manipulations in Matching with and without Contracts

R. Pablo Arribillaga (UNSL/CONICET)

Eliana Pepa Risma (UNSL/CONICET)

DOCUMENTO DE TRABAJO N° 257

Julio de 2023

Citar como:

Arribillaga, R. Pablo y Eliana Pepa Risma (2023). Obvious Manipulations in Matching with and without Contracts. *Documento de trabajo RedNIE N°257.*

Obvious Manipulations in Matching with and without Contracts*

R. Pablo Arribillaga† Eliana Pepa Risma†

June 29, 2023

Abstract

In a many-to-one matching model, with or without contracts, where doctors' preferences are private information and hospitals' preferences are substitutable and public information, any stable matching rule could be manipulated for doctors. Since manipulations can not be completely avoided, we consider the concept of *obvious manipulations* and look for stable matching rules that prevent at least such manipulations (for doctors). For the model with contracts, we prove that: *(i)* the doctor-optimal matching rule is non-obviously manipulable and *(ii)* the hospital-optimal matching rule is obviously manipulable, even in the one-to-one model. In contrast to *(ii)*, for a many-to-one model without contracts, we prove that the hospital-optimal matching rule is not obviously manipulable.Furthermore, if we focus on quantile stable rules, then we prove that the doctor-optimal matching rule is the only non-obviously manipulable quantile stable rule.

JEL classification: D71, D72.

Keywords: obvious manipulations, matching, contracts

*We acknowledge financial support from UNSL through grants 032016, 030120, from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) through grant PIP 112-200801-00655, and from Agencia Nacional de Promoción Científica y Tecnológica through grant PICT 2017-2355.

†Instituto de Matemática Aplicada San Luis(CONICET-UNSL) and Departamento de Matemática, Universidad Nacional de San Luis, San Luis, Argentina, and RedNIE. Emails: rarribi@unsl.edu.ar and ebpepa@unsl.edu.ar@unsl.edu.ar

1 Introduction

In the two-sided many-to-one matching model with contracts, there is a bilateral market whose disjoint sides are typically referred to as doctors and hospitals. Each contract refers to a doctor-hospital pair, although there may exist two or more contracts involving the same agents under different contractual conditions. The classical matching models (without contracts) can be addressed as special cases where there exists at most one contract involving each pair of agents. The problem consists of assigning agents from one side of the market to agents on the opposite side, through some contracts. In the many-to-one model, each doctor is allowed to sign one contract at most, whereas hospitals can sign multiple contracts. A set of contracts that contains at most one contract involving a doctor is called *allocation* and is a possible outcome of the matching problem. All the agents have preferences defined over the set of contracts involving them in some allocation. A *matching rule* is a function that, for each preference profile declared by the agents, selects an allocation. Since two agents wishing to sign an existing contract are free to do it, and also the agents can unilaterally terminate previous contracts if they find it convenient, there are outcomes that are unstable. We will consider *stable allocations*, i.e., outcomes that are sustainable over time, supposing the market remains unchanged. Hatfield and Milgrom (2005) proved that if all hospitals have *substitutable* preferences,[1] then the set of stable allocations is a nonempty lattice whose maximum and minimum elements can be obtained through a generalization of the deferred acceptance algorithm introduced by Gale and Shapley (1962). As in models without contracts, these extreme points correspond to the unanimously most preferred stable allocation for the doctors (doctor-optimal stable allocation), and the unanimously most preferred stable allocation for the hospitals (hospital-optimal stable allocation)

In addition to stability, the non-manipulability of a matching rule also plays a central role in two-sided matching literature. An agent manipulates a matching rule if there exists a situation in which it obtains a better result for itself by declaring an alternative preference to his true one. In this paper, we assume that hospitals' preferences are public and only focus on the manipulations that can be performed by doctors, whose preferences are private information. It is well known that the doctor-optimal matching rule, *i.e.*, the

[1] Substitutability is a basic condition, widely used in matching literature, and means that the hospitals do not consider the contracts as complementary among themselves.

matching rule that for each preference profile for doctors, returns the doctor-optimal stable allocation, is the only candidate to be stable and non-manipulable. Roth (1982) shows that, in the classical one-to-one matching model without contracts, the only strategy-proof and stable matching rule is the doctor-optimal one. In many-to-one matching models, such kind of result depends on the characteristics of the (fixed and public) hospitals' preferences. In the context without contracts, if hospitals' preferences are *q-responsive* (this model is known as *college admission problem*) the doctor-optimal matching rule is also the only strategy-proof and stable matching rule (see Roth, 1985). However, if hospitals' preferences are substitutable, the doctor-optimal matching rule could fail to be strategy-proof and it might not exist a strategy-proof stable rule. This is shown by Hatfield and Milgrom (2005) and Martínez et al. (2004) for the matching models with and without contracts, respectively. Those papers add different conditions to substitutability to recover the strategy-proofness of the doctor-optimal matching rule in each context.[2]

Therefore, in the many-to-one matching model (with and without contracts) with substitutable preferences, any stable matching could be susceptible to manipulations. Given that manipulations can not be completely avoided in this context, we look for stable matching rules that at least prevent *obvious manipulations* (for doctors), as defined by Troyan and Morrill (2020). A manipulation is "obvious" if it is much easier for agents to recognize and execute successfully than others in a specific and formal sense. To formalize the word "obvious", it is necessary to specify how much information each agent has about other agents' preferences. Troyan and Morrill (2020) assume that each agent has complete ignorance in this respect and, therefore, each agent focuses on the set of all outcomes that can be chosen by the rule given its own report. Now, a manipulation is obvious if the best possible outcome under the manipulation is strictly better than the best possible outcome under truth-telling or the worst possible outcome under the manipulation is strictly better than the worst possible outcome under truth-telling. Complementary, the term "obvious" reflects that an agent could deduce that a mechanism is manipulable even if it does not fully know how the mechanism is defined (see Theorem 1 in Troyan and Morrill, 2020, for this interpretation). In the context of college admission, Troyan and Morrill (2020) prove that any stable matching rule is no-obviously manipulable. Our results show that such a statement can be partially extended to the context of matching with contracts and substitutable preferences, arising some significant differences between the models with and

[2] Martínez et al. (2004) add *q-separability* and Hatfield and Milgrom (2005) add the *law of aggregate demand*.

without contracts.

First, we prove that the doctor-optimal matching rule is non-obviously manipulable in the general context of a many-to-one matching model with contracts and substitutable preferences for hospitals. Hence, although there are no matching rules that are non-manipulable, at least there is a matching rule that is non-obviously manipulable in such context. This result can be seen as an extension to the model with contracts of the one obtained by Troyan and Morrill (2020) but its proof requires completely different arguments and techniques. Surprisingly, we show that the opposite result holds for the hospital-optimal matching rule which turns out to be obviously manipulable (for doctors) even in the particular context of a one-to-one matching model with contracts. This result is remarkable because it reveals a substantial difference between the models with and without contracts from the point of view of the strategic behavior of agents. In the context of the many-to-one classical matching model without contracts and with substitutable hospitals' preferences we prove that the hospital-optimal matching rule is non-obviously manipulable.

Finally, we focus on the class of quantile-stable matching rules (or mechanisms) defined in Chen et al. (2016) and studied in Chen et al. (2021) and Fernandez (2020), among others. It is a new and relevant class of matching rules that generate stable matchings that can be seen as a compromise between the two sides of the market. Two of the (extremal) quantile-stable rules are the optimal matching rules that assign the best stable outcome for one side and the worst for the other. Assume that given a profile, we have a market with k stable matching, Chen et al. (2016) show that if the hospitals' preferences are substitutes and satisfy the law of aggregate demand[3], then the set of contracts that assigns each doctor the i-th ($1 \leq i \leq k$) best stable matching outcome is a stable matching. Now given $q \in [0, 1]$, the $q-$quantile-stable matching rule is a rule that for each profile selects the $\lceil kq \rceil$-th best stable allocation for each doctor in such profile.[4] When $q = 0$, the $q-$quantile-stable matching rule is the doctor-optimal matching rule. Fernandez (2020) proposes the notion of regret-free truth-telling as a weakness of strategy-proofness. This is a notion which assumes that the agents obtain limited information about the preferences of the other agents by observing the outcome of the rule. That paper shows that, when both

[3]Law of aggregate demand estates that the number of contracts chosen by the agent either rises or stays the same if the set of available contracts increases.

[4]$\lceil x \rceil$ denotes the lowest positive integer equal to or larger than x.

doctors and hospitals are strategic, the doctor-optimal and hospital-optimal are the only regret-free truth-telling matching rules in the class of the quantile-stable. In line with that result, our last theorem states that the doctor-optimal is the only quantile-stable matching rule which is non-obvious manipulable when doctors are strategic.

Troyan et al. (2020) apply the notion of obvious manipulation in the context of one-sided matching markets. They present the notion of essential stable matching and show that no essentially stable mechanism is obviously manipulable. Other recent papers that study the notion of obvious manipulation, in contexts other than two-side matching markets are Aziz and Lam (2021) and Arribillaga and Bonifacio (2022) in the context of voting; Ortega and Segal-Halevi (2022) in cake-cutting and Psomas and Verma (2022) in allocation problems.

The rest of the paper is organized as follows. The model and the concepts of stability and obvious manipulations are introduced in Section 2. Section 3 presents the main results of our paper in three subsections. Subsection 3.1 contains the formal definition of the doctor-optimal matching rule and the proof that such matching rule is non-obvious manipulable in the general context with contracts and substitutable preferences. Subsection 3.2 contains the formal definition of the hospital-optimal matching rule and the proof that such matching rule is: (i) obvious manipulable in the context with contracts, even in the one-to-one model and (ii) non-obvious manipulable in the context without contracts and substitutable preferences. In Subsection 3.3 the quantile stable rules are introduced and it is shown that the doctor-optimal matching rule is the only non-obvious manipulable quantile stable rule. To conclude, some final remarks are gathered in Section 4.

2 Preliminaries

2.1 Matching model with contracts and strategic doctors

We consider a many-to-one matching model with contracts where particular markets have two disjoint sides: a set of doctors D and a set of hospitals H, both finite. The problem consists of assigning agents from one side of the market to agents on the opposite side but, unlike in the classical matching model without contracts, the contractual conditions (salary, schedules, work tasks, etc.) characterizing the relationship between two agents are not fixed beforehand. In every particular market, there is a finite universal set of contracts $\mathbf{X}$. Each contract $x \in \mathbf{X}$ is bilateral, involving exactly one doctor $x_D \in D$ and one hospital $x_H \in H$. The set $\mathbf{X}$ could contain two or more contracts relating the same pair of agents $(d, h) \in \mathbf{D} \times \mathbf{H}$, under different conditions. The classical matching model without contracts can be considered as a special case of this setting, where $\mathbf{X}$ contains one and only one contract involving each pair $(d, h) \in \mathbf{D} \times \mathbf{H}$.

In the many-to-one matching model that we study here, each hospital can sign many contracts and each doctor can sign one contract at most. An *allocation* is a subset of contracts meeting such requirements.

Definition 1 *A set of contracts $Z \subseteq \mathbf{X}$ is an **allocation** if $x \neq y$ implies $x_D \neq y_D$ for all $x, y \in Z$.*

Given a set of contracts $Y \subseteq \mathbf{X}$ and $i \in D \cup H$, we will denote by $A(Y)$ the set of all allocations which are subsets of Y; and by Y_i the set of all contracts in Y involving i. Note that the empty set (referring to a situation where no contract is signed) is an allocation and $\varnothing \in A(Y)$ for all $Y \subseteq \mathbf{X}$.

Given a set of contracts $\mathbf{X}$, a particular market is determined by a *preference relation*, over the set of allocations $A(\mathbf{X}_i)$,[5] for each agent $i \in D \cup H$. Such preferences are antisymmetric, transitive, and complete. Observe that $|Z| \leq 1$ for all $Z \in A(\mathbf{X}_d)$ and $d \in D$. Therefore, we might visualize a doctor's preference relation as an order on the contracts involving itself and the empty set. In our analysis, as is usual in the literature, we will assume that only one side of the market is strategic: the doctors; while hospitals' preferences are fixed and common knowledge.

An arbitrary preference for doctor d is denoted by P_d. The weak preference relation over $A(\mathbf{X}_i)$ associated to P_i will be denoted by R_i.[6] By $\mathcal{P}_d$ we will denote the set of all feasible preference relations that a doctor d has in a given market. A *preference profile* $P = (P_d)_{d \in D}$ will identify a preference relation for each doctor. By $\mathcal{P} = \prod_{i \in D} \mathcal{P}_i$ we will denote the set of all profiles of preferences that could take place in the given market. Finally, for each profile P and doctor $d \in D$, we will denote by P_{-d} the subprofile in $\mathcal{P}_{-d} = \prod_{i \in D \setminus \{d\}} \mathcal{P}_i$ obtained by deleting P_d from P.

Definition 2 *A **(matching) rule** is a function $\phi : \mathcal{P} \rightarrow A(\mathbf{X})$ that returns an allocation in $A(\mathbf{X})$ for each profile of preferences $P \in \mathcal{P}$.*

[5] Observe that every allocation in $A(\mathbf{X}_i)$ contains only contracts involving i.

[6] This is, for all $x, y \in A(\mathbf{X}_i)$, $x R_d y$ if and only if either $x = y$ or $x P_d y$.

Given $d \in D$, we will denote by $\phi_d(P)$ the (only) contract in $\phi(P)$ involving d.

2.2 Stability

An essential property to be considered in matching models is the stability of allocations. In order to introduce the notion of stability, it is necessary to consider the preferences of the hospitals over the subsets of contracts. As we mentioned before, each hospital $h \in H$ has an antisymmetric, transitive and complete *preference* relation over $A(\mathbf{X}_h)$ that we will denote by $\succ_h$. A preference profile for hospitals is denoted by $\succ := (\succ_h)_{h \in H}$. Along the paper $\succ$ is assumed to be arbitrary but fixed and known by doctors.

The **choice set of** $h \in H$ given a preference $\succ_h$ and $Y \subseteq \mathbf{X}$, is the subset of Y_h that h likes best according to $\succ_h$

$$C_h(\succ_h, Y) = \max_{\succ_h} A(Y_h)$$

Similarly, the **choice set of** $d \in D$ given a preference $P_d \in \mathcal{P}_d$ and $Y \subseteq \mathbf{X}$, is the subset of Y_d that d likes best according to P_d [7]

$$C_d(P_d, Y) = \max_{P_d} A(Y_d)$$

Remark 1 *To keep the notation simpler, we will omit describing the preferences used to obtain the choice sets unless this information were relevant and not obvious from the context. We will write* $C_h(Y)$ *and* $C_d(Y)$ *instead of* $C_h(Y, \succ_h)$ *and* $C_d(Y, P_d)$*, respectively.*

In addition, given $Y \subseteq \mathbf{X}$ we define $C_H(Y) = \cup_{h \in H} C_h(Y)$ and $C_D(Y) = \cup_{d \in D} C_d(Y)$.

Now, we are ready to introduce the concept of individually rational and stable allocation. Given $P \in \mathcal{P}$, an allocation $Y \in A(\mathbf{X})$ is **individually rational** at P if it does not contain unwanted contracts, *i.e.,*

$$C_D(Y) = C_H(Y) = Y.$$

Whenever two agents wish to sign a contract, they are free to do it, and they are also free to terminate previous contracts. Consequently, an allocation can be blocked by a

[7] Observe that for each $d \in D$, the choice set $C_d(Y, P_d)$ contains only the best element in Y_d according to P_d. However, we introduce this definition here to have a symmetric notation for hospitals and doctors which makes the exposition simpler.

contract that it does not include. Given $P \in \mathcal{P}$ and $Y \in A(\mathbf{X})$, the contract $x \in \mathbf{X} \setminus Y$ is a **blocking contract** at P for Y if

$$x \in C_{x_D}(Y \cup \{x\}) \cap C_{x_H}(Y \cup \{x\}).$$

Definition 3 *Let $P \in \mathcal{P}$. A set $Y \in A(\mathbf{X})$ is a **stable allocation** at P if*

(i) Y is individually rational at P;

(ii) There are no blocking contracts for Y at P.

Definition 4 *A rule $\phi : \mathcal{P} \to A(\mathbf{X})$ is **stable** if for all $P \in \mathcal{P}$, the allocation $\phi(P)$ is stable at P.*

Hatfield and Milgrom (2005) show that if all hospitals have *substitutable* preferences, then the set of stable allocations is nonempty. Substitutability is a basic condition, widely used in the matching literature, and means that the hospitals do not consider the contracts as complementary among themselves. In other words, no contract ceases being chosen by a hospital because another contract becomes unavailable for it.

Definition 5 *A preference $\succ_h$ of an agent $h \in H$ satisfy **substitutability** if $x \in C_h(W, \succ_h)$ implies $x \in C_h(Y, \succ_h)$ whenever $x \in Y_h \subseteq W_h \subseteq \mathbf{X}$.*

From now on $\succ_h$ will always satisfy substitutability for each $h \in H$.

Hatfield and Milgrom (2005) also show that if all hospitals have substitutable preferences, the set of stable allocations has a lattice structure whose extreme points are characterized as the *doctor-optimal stable allocation* and the *hospital-optimal stable allocation*. They are the allocation that assigns every doctor the best contract it could obtain through any stable allocation and the allocation which assigns every hospital the best set of contracts it could obtain through any stable allocation, respectively. Formal algorithms to compute the mentioned optimal stable allocations are included in Section 3. In the rest of the paper, we will focus particularly on the doctor-optimal rule and hospital-optimal rule which return the doctor-optimal and the hospital-optimal stable allocations at each $P \in \mathcal{P}$, respectively.

2.3 Manipulations and obvious manipulations

The concept of non-manipulability (or strategy-proofness) has played a central role in studying the strategic behavior of agents. A doctor manipulates a matching rule if there exists a situation in which this agent obtains a better result for him by declaring an alternative preference to his true one.

Definition 6 *Given a rule $\phi : \mathcal{P} \to A(\mathbf{X})$, an agent $d \in D$ and a preference $P_d \in \mathcal{P}_d$, the preference $P'_d \in \mathcal{P}_d$ is a **manipulation** of ϕ at P_d if there is a (sub)profile $P_{-d} \in \mathcal{P}_{-d}$such that*

$$\phi_d(P'_d, P_{-d}) \ P_d \ \phi_d(P_d, P_{-d}). \tag{1}$$

*A rule is **strategy-proof** if it has no manipulation.*

In the context of substitutable preferences in many-to-one matching models (with and without contracts), any stable matching is susceptible to manipulation (see Martínez et al., 2004; Hatfield and Milgrom, 2005). So, a weaker notion than strategy-proofness could be considered to study the strategic behavior of agents. Given that manipulations can not be completely avoided, this paper considers the concept of *obvious manipulations* introduced by Troyan and Morrill (2020) and looks for stable rules that at least avoid obvious manipulations which are, in some sense, more easily identifiable by agents. To describe such manipulations, it is important to specify how much information each agent has about other agents' preferences. Troyan and Morrill (2020) assume that each agent has complete ignorance in this respect and, therefore, agents focus on the set of outcomes that can be chosen by the rule given their own reports. Now, a manipulation is obvious if the best possible outcome under the manipulation is strictly better than the best possible outcome under truth-telling or the worst possible outcome under the manipulation is strictly better than the worst possible outcome under truth-telling. In the context of college admission, Troyan and Morrill (2020) prove that any stable rule is not obviously manipulable. We will prove that such a result could be partially extended to our context of matching with contracts and substitutable preferences.

Given a rule $\phi : \mathcal{P} \to A(\mathbf{X})$, a doctor $d \in D$ and a preference $P_d \in \mathcal{P}_d$, we define the **option set** left open by P_d at ϕ as

$$O^{\phi}(P_d) = \{\phi_d(P_d, P_{-d}) : P_{-d} \in \mathcal{P}_{-d}\}.$$

Given $Y \subseteq \mathbf{X}$, denote by $W_d(P_d, Y)$ to the worst alternative in Y_d according to preference P_d.

Definition 7 *Let $\phi : \mathcal{P} \to A(\mathbf{X})$ a rule, $i \in D$ let $P_d \in \mathcal{P}_d$, and let $P'_d \in \mathcal{P}_d$ be a manipulation of ϕ at P_d. Manipulation P'_d is* ***obvious*** *if*

$$W_d(P_d, O^{\phi}(P'_d)) \; P_d \; W_d(P_d, O^{\phi}(P_d)). \tag{2}$$

or

$$C_d(P_d, O^{\phi}(P'_d)) \; P_d \; C_d(P_d, O^{\phi}(P_d)). \tag{3}$$

Rule f is ***not obviously manipulable (NOM)*** *if it does not admit any obvious manipulation. In other case, f is* ***obviously manipulable (OM)***

3 Main results

In this section, we present the results of the paper. In all of them, we assume that hospitals' preferences are substitutable. For a clear presentation, we omit to mention this general hypothesis in each statement.

3.1 Doctor-optimal rule

Given a profile $P \in \mathcal{P}$, the doctor-optimal stable allocation can be defined by the *Doctor-proposing Deferred Acceptance algorithm (D-DA)*. At this algorithm doctors make offers: a doctor proposes its best contract (if any) among the set of contracts that have not been rejected during the previous steps, while a hospital accepts its choice set given the set of received offers. The algorithm stops when all offers are accepted. The output of the algorithm is the set of contracts (allocation) accepted by the hospitals in the final iteration. Formally:

The doctor-proposing deferred acceptance algorithm (D-DA)

Input:

A market $(\mathbf{X}, P)$.

Begin:

Set $X^1 = \mathbf{X}$ and $t := 1$.

Repeat:

Step 1: Determine the set of contracts that doctors offer in the iteration t, this is $O^t := C_D(X^t)$.

Step 2: Determine the set of contracts that are (provisionally) accepted by hospitals in the

iteration $t : C_H(O^t)$.
If $C_H(O^t) = O^t$, the algorithm stops with output $C_H(O^t)$.
If $C_H(O^t) \subsetneq O^t$, define $X^{t+1} = X^t - (O^t - C_H(O^t))$, set $t := t+1$ and repeat steps 1 and 2
End

Let $\overline{\phi} : \mathcal{P} \to A(\mathbf{X})$ denote the **doctor-optimal rule**, *i.e.*, the rule that returns the doctor-optimal stable allocation for each preference profile for doctors, $P \in \mathcal{P}$. Trivially, $\overline{\phi}$ is a stable rule.

Now, we can state and prove the first result of this section.

Theorem 1 *The doctor-optimal rule $\overline{\phi} : \mathcal{P} \to A(\mathbf{X})$ is NOM.*

The following two lemmas are necessary in order to prove Theorem 1. The first provides elementary properties of the choice sets when the hospitals' preferences are substitutable. The second states that, at the end of each iteration of D-DA, each hospital is assigned to its choice set from the set of all offers received up to that time.

Lemma 1 *For all $X, Y \subseteq \mathbf{X}$ and $i \in D \cup H$:*

(i) $C_i(Y) \subseteq X \subseteq Y$ *implies* $C_i(X) = C_i(Y)$.

(ii) $C_i(C_i(Y)) = C_i(Y)$.

(iii) If i's preferences satisfy substitutability, then $C_i(X \cup Y) = C_i(C_i(X) \cup Y)$.

Proof. (i) and (ii) are direct consequences of the definition of choice set.
(iii) Let $i \in D \cup H$. By substitutability, $X \cap C_i(X \cup Y) \subseteq C_i(X)$. Therefore, $C_i(X \cup Y) \subseteq C_i(X) \cup Y \subseteq X \cup Y$. This implies $C_i(X \cup Y) = C_i(C_i(X) \cup Y)$ according to (i). □

Lemma 2 *Given $P \in \mathcal{P}$, let O^t be as in D-DA definition and let define $O^t_A := \cup_{k=1}^{t} O^k$. Then,*

$$C_h(O^t) = C_h(O^t_A) \text{ for all } t = 1, ...T \tag{4}$$

and

$$\overline{\phi}_h(P) = C_h(O^T_A) \tag{5}$$

for all $h \in H$, where T is the number of iterations of the D-DA at P.

Proof. First, we prove (4). Given $P \in \mathcal{P}$, let t be a stage of the D-DA at P, such that $1 \leq t \leq T$. The proof is by induction on t. If $t = 1$ the proof is trivial. Now we assume that (4) holds for all $k < t$ and prove that it also holds for t. Let $x \in C_h(O^{t-1})$, then by definition of D-DA, $x \in O^{t-1} = C_D(X^{t-1})$ and $x \in X^t$. Therefore, as $X^t \subseteq X^{t-1}$, $x \in C_D(X^t) = O^t$. Hence, $C_h(O^{t-1}) \subseteq O^t$. This implies $C_h(O^t) = C_h(C_h(O^{t-1}) \cup O^t)$. Then, by induction hypothesis, $C_h(O^t) = C_h(C_h(O_A^{t-1}) \cup O^t)$ and, consequently, we have $C_h(O^t) = C_h(O_A^{t-1} \cup O^t)$ according to (iii) in Lemma 1. Therefore, by definition of O_A^t, $C_h(O^t) = C_h(O_A^t)$

The proof of (5) follows from (4) when $t = T$ and the definition of $\overline{\phi}_h$. □

Proof of Theorem 1 First, we prove that $\overline{\phi}$ does not admit any obvious manipulation in the sense of (2).

Given $d \in D$ and $P_d \in \mathcal{P}_d$, let $y = W_d(P_d, O^{\overline{\phi}}(P_d))$. As $\overline{\phi}$ is individually rational, $\{y\}R_d\varnothing$. As $\mathbf{X}$ is finite, there exists $\widehat{P}_{-d} \in \mathcal{P}_{-d}$ such that $\{y\} = \overline{\phi}_d(P_d, \widehat{P}_{-d})$. We denote $\widehat{X} = \overline{\phi}(P_d, \widehat{P}_{-d})$, so $\widehat{X}_d = \{y\}$.

Let O^t be the set of offers made in the Stage t of the D-DA at $(P_d, \widehat{P}_{-d})$ and let $O_A^t = \cup_{k=1}^t O^k$ for all $t = 1, \ldots, T$, where T the number of iterations of the D-DA at $(P_d, \widehat{P}_{-d})$.

<u>Claim 1:</u> Let $x = \min_{P_d}\{w \in \mathbf{X}_d : \{w\}P_d\{y\}\}$, then $\{x\}\ P_d\ W_d(P_d, O^{\overline{\phi}}(P'_d))$ for all $P'_d \in \mathcal{P}_d \setminus \{P_d\}$.[8]

As $\{x\}P_d\{y\} = \overline{\phi}_d(P_d, \widehat{P}_{-d})$, there exists $k' \in \{1, ..., T\}$ in which $x \in O^{k'}$ and $x \notin C_{x_H}(O^{k'})$. As d makes at most one offer in each stage of the algorithm,

$$[C_H(O^{k'})]_d = \varnothing. \tag{6}$$

Let $\widetilde{P}_{-d} \in \mathcal{P}_{-d}$ be such that for each $i \in D\backslash\{d\}$, $\widetilde{P}_i = [C_H(O^k)]_i, \varnothing$. We proved that $\{x\}P_d\overline{\phi}_d(P'_d, \widetilde{P}_{-d})$ for all $P'_d \in \mathcal{P}_d \setminus \{P_d\}$.

Assume that there exists $P'_d \in \mathcal{P}_d \setminus \{P_d\}$ such that $\overline{\phi}_d(P'_d, \widetilde{P}_{-d}) = \{z\}$ and $\{z\}R_d\{x\}$. Let T' be the number of iterations of the D-DA at $(P'_d, \widetilde{P}_{-d})$. W.l.o.g. we can assume that $T = T'$, in another case we add some artificial steps where each doctor offers the contract assigned by the D-DA. Let O'^t be the set of offers in the Stage t of the D-DA at $(P'_d, \widetilde{P}_{-d})$ and let $O'^t_A := \cup_{k=1}^t O'^k$. As $\{z\}R_d\{x\}P_d\{y\}R_d\varnothing$, $z \neq \varnothing$ and $z_H \in H$ is well defined. Therefore,

[8] If $\{w \in \mathbf{X}_d : \{w\}P_d\{y\}\} = \varnothing$, then $\overline{\phi}$ would not admit any obvious manipulation trivially.

by Lemma 2, $\overline{\phi}_{z_H}(P'_d, \widetilde{P}_{-d}) = C_{z_H}(O'^T_A)$ and then,

$$z \in C_{z_H}(O'^T_A) \tag{7}$$

Hence, as $z_D = d$, by substitutability, $z \in C_{z_H}([O'^T_A]_d)$. Furthermore, as $C_{z_H}([O'^T_A]_d)$ is in $A(([O^T_A]_d)_{z_H})$, $\left|C_{z_H}((O'^T_A)_d)\right| \leq 1$. Therefore,

$$C_{z_H}([O'^T_A]_d) = \{z\}. \tag{8}$$

By definition of $\widetilde{P}_{-d}$, $[O'^T_A]_i = [O'^1]_i = [C_H(O^{k'})]_i$ for each $i \in D \setminus \{d\}$. Then,

$$C_{z_H}(\cup_{i \in D\setminus\{d\}}[O'^T_A]_i) = C_{z_H}(\cup_{i \in D\setminus\{d\}}[C_H(O^k)]_i). \tag{9}$$

Hence by (6) and (9) ,

$$C_{z_H}(\cup_{i \in D\setminus\{d\}}[O'^T_A]_i) = C_{z_H}(C_H(O^{k'}))$$

Therefore, by (ii) in Lemma 1,

$$C_{z_H}(\cup_{i \in D\setminus\{d\}}[O'^T_A]_i) = C_{z_H}(O^{k'}) \tag{10}$$

Then, as $O'^T_A = [O'^T_A]_d \cup (\cup_{i \in D\setminus\{d\}}[O'^T_A]_i)$, by (iii) in Lemma 1,

$$C_{z_H}(O'^T_A) = C_{z_H}(C_{z_H}([O'^T_A]_d) \cup C_{z_H}(\cup_{i \in D\setminus\{d\}}[O'^T_A]_i)).$$

Hence by (8) and (10), $C_{z_H}(O'^T_A) = C_{z_H}(\{z\} \cup C_{z_H}(O^{k'}))$. Then, by (7),

$$z \in C_{z_H}(\{z\} \cup C_{z_H}(O^{k'})). \tag{11}$$

Furthermore, by (6) and Lemma 2,

$$z \notin C_{z_H}(O^{k'}_A). \tag{12}$$

As $\{z\}R_d\{x\}$ and $x \in O^{k'}$, $z \in O^{k'}_A$. Then, $O^{k'}_A = \{z\} \cup O^{k'}_A$. By (iii) in Lemma 1,

$$C_{z_H}(O^{k'}_A) = C_{z_H}(\{z\} \cup O^{k'}_A) = C_{z_H}(\{z\} \cup C_{z_H}(O^{k'}_A)) = C_{z_H}(\{z\} \cup C_{z_H}(O^{k'}))$$

Therefore, by (12), $z \notin C_{z_H}(\{z\} \cup C_{z_H}(O^{k'}))$ which contradicts (11).

Then, $\{x\}\ P_d\, \overline{\phi}_d(P'_d, \widetilde{P}_{-d})$ for all $P'_d \in \mathcal{P}_d \setminus \{P_d\}$ and the proof of Claim 1 is completed.

Now, by Claim 1 and definition of x, $\{y\}\ R_d\ W_d(P_d, O^{\overline{\phi}}(P'_d))$ for all $P'_d \in \mathcal{P}_d \setminus \{P_d\}$. Hence, $\overline{\phi}$ does not admit obvious manipulations in the sense of (2).

Second, we prove that $\overline{\phi}$ does not admit any obvious manipulation in the sense of (3).

Given $d \in D$ and $P_d \in \mathcal{P}_d$ let $\overline{x} \in X_d$ such that $\overline{x} = \max_{P_d}\{x \in \mathbf{X}_d : \{x\} \succ_{x_H} \varnothing\}$. Now let $\widetilde{P}_{-d} \in \mathcal{P}_{-d}$ be such that $\widetilde{P}_i = \varnothing$ for all $i \in D \setminus \{d\}$. Then, from the definition of D-DA, it follows that $\{\overline{x}\} = \overline{\phi}_d(P_d, \widetilde{P}_{-d})$. This implies $C_d(P_d, O^{\overline{\phi}}(P_d))\ R_d\ \{\overline{x}\}$.

Furthermore, given $P'_d \in \mathcal{P}_d \setminus \{P_d\}$, if $y \in X_d$ and $\varnothing \succ_{y_H} \{y\}$, as $\overline{\phi}$ is individually rational and $\succ_{y_H}$ is substitutable, we have that $y \notin O^{\overline{\phi}}(P'_d)$.[9] Therefore, for all $P'_d \in \mathcal{P}_d \setminus \{P_d\}$ we have $O^{\overline{\phi}}(P'_d) \subseteq \{x \in \mathbf{X}_d : \{x\}P_{x_H}\varnothing\}$ and, consequently, $\{\overline{x}\}\ R_d\ C_d(P_d, O^{\overline{\phi}}(P'_d))$. So, for every $P'_d \in \mathcal{P}_d \setminus \{P_d\}$, we obtain

$$C_d(P_d, O^{\overline{\phi}}(P_d))\ R_d\ \{\overline{x}\}\ R_d\ C_d(P_d, O^{\overline{\phi}}(P'_d))$$

This completes the proof of Theorem 1. □

3.2 The hospital-optimal rule

Similarly to the previous subsection, given a profile $P \in \mathcal{P}$, the hospital-optimal stable allocation can be defined by the *Hospital-proposing Deferred Acceptance Algorithm (H-DA)* where hospitals make offers. It is defined as the D-DA by interchanging the roles of D and H.

Let $\underline{\phi} : \mathcal{P} \to A(\mathbf{X})$ denote the **hospital-optimal rule**, *i.e.*, the rule that returns the hospital-optimal stable allocation for each preference profile for doctors, $P \in \mathcal{P}$. Trivially, $\underline{\phi}$ is a stable rule. In the context of college admission Troyan and Morrill (2020) show that $\underline{\phi}$ is NOM.[10] Surprisingly, such a result can not be extended to the context of matching with contracts. In fact, the opposite result holds: $\underline{\phi}$ is obviously manipulable even in the particular case of the one-to-one matching model with contracts. This result reveals a substantial difference between the models with and without contracts from the point of view of the strategic incentives of agents.

Theorem 2 *The hospital-optimal rule $\underline{\phi} : \mathcal{P} \to A(\mathbf{X})$ is OM, even in the one-to-one matching model with contracts.*[11]

[9] By sustitubility, $x \in C_{y_H}(Y)$ implies that $x \in C_{y_H}(\{x\})$. Hence $\{x\}P_{y_H}\varnothing$.

[10] They show that any stable is NOM.

[11] The one-to-one model can be addressed as a particular case of the many-to-one model where $\succ_h$ only has singleton sets as acceptable, for each $h \in H$.

Proof. It will be sufficient to show a particular market where a doctor $d \in D$ has an obvious manipulation at $\underline{\phi}$. Let $D = \{d_1, d_2\}$, $H = \{h_1, h_2\}$ and $\mathbf{X} = \{x, z, w\}$ be the sets of doctors, hospitals, and contracts, respectively. Suppose that $x_D = z_D = d_1$, $x_H = z_H = h_1$ and $w_D = d_2$, $w_H = h_2$. Assume that $\succ_{h_1} = x, y$ and $\succ_{h_2} = w$. Let P_2 be an arbitrary preference in $\mathcal{P}_{d_2}$. Let $P_1 \in \mathcal{P}_{d_1}$ such that $P_1 = y, x, \emptyset$. Following H-DA, h_1 offers its best contract x to doctor d_1 and this contract is accepted by d_1. Then, d_1 does not receive more offers, and h_1 does not make more offers. Then, the contract x is the only one contract of d_1 chosen by $\underline{\phi}$ at profile (P_1, P_2). Hence, $\underline{\phi}_{d_1}(P_1, P_2) = x$. Now, let $P_1' \in \mathcal{P}_{d_1}$ such that $P_1' := y, \emptyset$. Following H-DA, h_1 offers its best contract x to doctor d_1 and this contract is rejected by d_1. Then, h_1 offers its second-best contract y to doctor d_1 and this contract is accepted by d_1. Then, d_1 does not receive more offers, and h_1 does not make more offers. Then, the contract y is the only contract of d_1 chosen by $\underline{\phi}$ at profile (P_1', P_2). Hence, $\underline{\phi}_{d_1}(P_1', P_2) = y$. Therefore, $\underline{\phi}_{d_1}(P_1', P_2) = y P_1 x = \underline{\phi}_{d_1}(P_1, P_2)$, for any $P_2 \in \mathcal{P}_{d_2}$. Hence, P_1' is an obvious manipulation of $\underline{\phi}$ at P_d. □

Finally, in the context without contracts, an equivalent result to Theorem 1 can be obtained for $\underline{\phi}$. This extends the result in Troyan and Morrill (2020) to substitutable preferences.

Theorem 3 *Assume that for each pair $(d, h) \in D \times H$ there is at most one contract $x \in \mathbf{X}$ such that $x_D = d$ and $x_H = d$. Then, the hospital-optimal rule $\underline{\phi} : \mathcal{P} \to A(\mathbf{X})$ is NOM.*

Proof. First, we prove that $\underline{\phi}$ does not admit any obvious manipulation in the sense of (2). Given $d \in D$ and $P_d \in \mathcal{P}_d$, let $y = W(P_d, O^{\underline{\phi}}(P_d))$. As $\underline{\phi}$ is individually rational, $\{y\} R_d \emptyset$. As $\mathbf{X}$ is finite, there exists $\widehat{P}_{-d} \in \mathcal{P}_{-d}$ such that $\{y\} = \underline{\phi}_d(P_d, \widehat{P}_{-d})$. We denote $\widehat{X} = \underline{\phi}(P_d, \widehat{P}_{-d})$, so $\widehat{X}_d = \{y\}$.

Let $\widetilde{P}_{-d} \in \mathcal{P}_{-d}$ be such that for each $i \in D \backslash \{d\}$, $\widetilde{P}_i = \widehat{X}_i, \emptyset$.

<u>Claim:</u> $\widehat{X}_d R_d \underline{\phi}(P_d', \widetilde{P}_{-d})$ for all $P_d' \in \mathcal{P}_d \setminus \{P_d\}$.

Assume on contradiction that there exists $P_d' \in \mathcal{P}_d \setminus \{P_d\}$ such that $\underline{\phi}_d(P_d', \widetilde{P}_{-d}) = \{z\}$ and

$$\{z\} P_d \{y\}. \tag{13}$$

Let T' be the number of iterations of the H-DA in $(P_d', \widetilde{P}_{-d})$ and let $X^{T'}$ has in H-DA's definition. As $\{z\} P_d \{y\} R_d \emptyset$, $z \neq \emptyset$ and $z_H \in H$ is well defined. Furthermore, as $z_D = d$, by hypothesis, $z_H \neq y_H$. As $\widehat{X}$ is an allocation, $y \in \widehat{X}$ and $y_D = d$, it follows that $x_D \neq d$ for each $x \in \widehat{X}_{z_H}$. Hence, by definition of $\widetilde{P}_{-d}$, $\widetilde{P}_{x_D} := x, \emptyset$ for each $x \in \widehat{X}_{z_H}$. Then, contracts in $\widehat{X}_{z_H}$ are never rejected in the H-DA at $(P'_d, \widetilde{P}_{-d})$. Then, by definition of H-DA, $\widehat{X}_{z_H} \subseteq X^{T'}$. As $\underline{\phi}_d(P'_d, \widetilde{P}_{-d}) = \{z\}$, $z \in C_H(X^{T'})$. Then, by substitutability and the definition of choice set,

$$z \in C_{z_H}(z \cup \widehat{X}_{z_H}) = C_{z_H}(z \cup \widehat{X}) \tag{14}$$

But (13) and (14) contradict the stability of $\underline{\phi}(P_d, \widehat{P}_{-d})$ and the fact that $\underline{\phi}_d(P'_d, \widehat{P}_{-d}t) \neq z$. This finishes the proof of the Claim.

By Claim, $W(P_d, O^{\underline{\phi}}(P_d))P_d W(P_d, O^{\underline{\phi}}(P'_d))$ for all $P'_d \in \mathcal{P}_d \setminus \{P_d\}$. Therefore, $\underline{\phi}$ does not admit any obvious manipulation in the sense of (2). The proof that $\underline{\phi}$ does not admit any obvious manipulation in the sense of (3) is similar to the one given to $\overline{\phi}$ in Theorem 1 and therefore it is omitted. □

3.3 Quantile stable matchings

In this section, we introduce the class of quantile stable rules (or mechanisms) defined in Chen et al. (2016) and studied in Chen et al. (2021) and Fernandez (2020), among others. It is a new and relevant class of rules that generate stable matchings which can be seen as a compromise between both sides of the market. The extremal quantile-stable rules are the doctor-optimal and the hospital-optimal ones. They assign the best stable outcome for one side and the worst for the other.

To guarantee the existence of quantile-stable rules it is necessary to assume that hospitals' preferences satisfy the *law of aggregate demand*, in addition to substability.[12]. The law of aggregate demand establishes that the number of contracts chosen by the agent either rises or stays the same if the set of available contracts increases.

Definition 8 *A preference $\succ_h$ of an agent $h \in H$ satisfies the **law of aggregate demand** (LAD) if*

$$|C_h(Y, \succ_h)| \leq |C_h(Z, \succ_h)|$$

for all $Y \subseteq Z \subseteq \mathbf{X}$.

LAD is less restrictive than the property of q-responsiveness assumed in the college admission problem (see Pepa Risma, 2015) and trivially holds in the one-to-one matching model.

[12]see Chen et al. (2016).

Along this subsection, we assume that $\succ_h$ satisfies substitutability and the law of aggregated demand, for each $h \in H$.

Definition 9 *For each $q \in [0,1]$, the q-quantile stable rule $\phi^q : \mathcal{P} \to A(\mathbf{X})$ is defined as follows: given $P \in \mathcal{P}$, let k be the number of stable matchings under P (and $\succ$), then $\phi^q(P)$ is the allocation obtained by joining, for each doctor $d \in D$, d's $\lceil kq \rceil$-th best stable allocation in P according to order P_d.*[13]

Observe that $\phi^0 = \overline{\phi}$ and $\phi^1 = \underline{\phi}$ are the extreme quantile stable rules. Chen et al. (2016) show, that ϕ^q is a stable rule when $\succ_h$ satisfies substatibility and the law of aggregate demand for each $h \in H$.

Theorem 4 *A q-quantile stable rule $\phi^q : \mathcal{P} \to A(\mathbf{X})$ is NOM if and only if $q = 0$ i.e.$\phi^q = \overline{\phi}$ is the doctor-optimal rule, even in the one-to-one matching model with contracts.*

Proof. By Theorem 1, $\phi^0 = \overline{\phi}$ is NOM. Now assume that $q \in (0,1]$, we will show that a particular market exists where a doctor $d \in D$ has an obvious manipulation at ϕ^q. Let $D = \{d_1, d_2\}$, $H = \{h_1, h_2\}$ be the sets of doctors and hospitals, respectively. Let k be a positive integer such that $\lceil kq \rceil = 2$ and let $\mathbf{X} = \{x^1, x^2, \dots, x^k, w\}$ be the set of contracts, where $x^i_D = d_1$, $x^i_H = h_1$ for each $i = 1, \dots, k$ and $w_D = d_2$, $w_H = h_2$. Assume that $\succ_{h_1} = x^k, x^{k-1}, \dots, x^1$ and $\succ_{h_2} = w$. Let P_2 be an arbitrary preference in $\mathcal{P}_{d_2}$. Let $P_1 \in \mathcal{P}_{d_1}$ such that $P_1 = x^1, x^2, \dots, x^k$. Observe that the set of all stable allocations under P is $\cup_{j=1\cdots k}\{\{x^j, w\}\}$ if $wP_2\varnothing$ and $\cup_{j=1\cdots k}\{\{x^j\}\}$ if $\varnothing P_2 w$. As $\lceil kq \rceil = 2$, $\phi^q_{d_1}(P) = x^2$. Now, let $P'_1 \in \mathcal{P}_{d_1}$ such that $P'_1 = x^1$. Then, the set of all stable allocations under (P'_1, P_2) is $\{\{x^1, w\}\}$ if $wP_2\varnothing$, and $\{\{x^1\}\}$ if $\varnothing P_2 w$. Hence $\phi^q_{d_1}(P'_1, P_2) = x^1$. Therefore, $\phi^q_{d_1}(P'_1, P_2) = x^1 P_1 x^2 = \phi^q_{d_1}(P_1, P_2)$ for any $P_2 \in \mathcal{P}_{d_2}$. So, P'_1 is an obvious manipulation of ϕ^q at P_d. □

4 Final Remarks

Table 1 summarizes our main findings. Remember that we assume that hospitals' preferences are substitutable. When preferences are *responsive*[14] the results are the same as in the one-to-one model.

[13] Here, $\lceil x \rceil$ denotes the lowest positive integer equal to or larger than x.

[14] Preference $\succ_h$ is **responsive** with quota q, where q is a non negative integer number, if:

	NOM (for doctors)				
	Without Contracts		With Contracts		
	one-to-one †	many-to-one	one-to-one	many-to-one	
Doctor-optimal	*yes*	*yes*	*yes*	*yes*	Th.1
Hospital-optimal	*yes*	*yes*	*no*	*no*	Th.3 and Th.2
ϕ^q (q-quantile) $q \in (0,1)$	*yes*	?	*no*	*no*	Th.4

† These results are due to Troyan and Morrill (2020)

Table 1: *Summary of results.*

Two remarks arise from the table. First, there are some open questions left by the present paper: (i) Are all (quantile) stable rules NOM in the many-to-one matching model without contracts and substitutable preferences for hospitals? (ii) Are there other NOM stable rules besides the doctor-optimal rule in matching models with contracts and, if so, how might these rules be characterized? Second, as we said before, there is a substantial difference between the models without and with contracts from the point of view of the strategic incentives of agents. In fact, the hospital-optimal rule does not admit obvious manipulations in the context without contracts, but it does in the context with contracts, even in the simplest one-to-one case.

Finally, an additional open problem arises by interchanging the roles between doctors and hospitals and considering the manipulability of rules by hospitals. Even in the college admission model, no stable rule exists that is not manipulable for colleges (see Roth, 1985). Therefore, in this simple context, the consideration of non-obvious manipulable matching rules looks like an interesting problem.

(i) For all $Z \in A(\mathbf{X}_h)$ such that $\mid Z \mid > q$, we have $\emptyset \succ_h Z$.
(ii) For every pair of contracts $x, y \in \mathbf{X}_h \cup \emptyset$, and every $Z \subseteq \mathbf{X}_h \setminus \{x, y\}$ with $\mid Z \mid < q$, whenever $Z \cup \{x\}$ and $Z \cup \{y\}$ are in $A(\mathbf{X}_h)$, we have

$$Z \cup \{x\} \succ_h Z \cup \{y\} \text{ if and only if } \{x\} \succ_h \{y\}.$$

References

ARRIBILLAGA, R. P. AND A. G. BONIFACIO (2022): "Obvious manipulations of tops-only voting rules," *arXiv preprint arXiv:2210.11627*.

AZIZ, H. AND A. LAM (2021): "Obvious manipulability of voting rules," in *International Conference on Algorithmic Decision Theory*, Springer, 179–193.

CHEN, P., M. EGESDAL, M. PYCIA, AND M. B. YENMEZ (2016): "Median stable matchings in two-sided markets," *Games and Economic Behavior*, 97, 64–69.

——— (2021): "Quantile stable mechanisms," *Games*, 12, 43.

FERNANDEZ, M. A. (2020): "Deferred acceptance and regret-free truth-telling," Working Paper.

GALE, D. AND L. SHAPLEY (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly*, 69, 9–15.

HATFIELD, J. AND P. MILGROM (2005): "Matching with contracts," *American Economic Review*, 95, 913–935.

MARTÍNEZ, R., J. MASSÓ, A. NEME, AND J. OVIEDO (2004): "On group strategy-proof mechanisms for a many-to-one matching model," *International Journal of Game Theory*, 33, 115–128.

ORTEGA, J. AND E. SEGAL-HALEVI (2022): "Obvious manipulations in cake-cutting," *Social Choice and Welfare*, 1–20.

PEPA RISMA, E. (2015): "Binary operations and lattice structure for a model of matching with contracts," *Mathematical Social Sciences*, 73, 6–12.

PSOMAS, A. AND P. VERMA (2022): "Fair and efficient allocations without obvious manipulations," *arXiv preprint arXiv:2206.11143*.

ROTH, A. (1985): "The college admissions problem is not equivalent to the marriage problem," *Journal of Economic Theory*, 36, 277–288.

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

Troyan, P., D. Delacrétaz, and A. Kloosterman (2020): "Essentially stable matchings," *Games and Economic Behavior*, 120, 370–390.

Troyan, P. and T. Morrill (2020): "Obvious manipulations," *Journal of Economic Theory*, 185, 104970.