Soc Choice Welf (2013) 41:701–720
DOI 10.1007/s00355-012-0703-1

ORIGINAL PAPER

Games with capacity manipulation: incentives and Nash equilibria

Antonio Romero-Medina · Matteo Triossi

Received: 20 August 2011 / Accepted: 29 September 2012 / Published online: 20 October 2012

Abstract Studying the interactions between preference and capacity manipulation in matching markets, we prove that acyclicity is a necessary and sufficient condition that guarantees the stability of a Nash equilibrium and the strategy-proofness of truthful capacity revelation under the hospital-optimal and intern-optimal stable rules. We then introduce generalized games of manipulation in which hospitals move first and state their capacities, and interns are subsequently assigned to hospitals using a sequential mechanism. In this setting, we first consider stable revelation mechanisms and introduce conditions guaranteeing the stability of the outcome. Next, we prove that every stable non-revelation mechanism leads to unstable allocations, unless restrictions on the preferences of the agents are introduced.

1 Introduction

The literature has studied preference and capacity manipulation separately and has thus overlooked the interaction between the two. However, many hiring and admission procedures allow firms, hospitals and schools to state the number of their vacant positions before candidates are assigned. This creates a possibility of capacity manipulation before the matching process. Furthermore, ex-ante manipulation does not prevent agents from misrepresenting their preferences during the matching process. In

A. Romero-Medina
Departmento de Economía, Universidad Carlos III de Madrid, Calle Madrid 126,
28903 Getafe-Madrid, Spain
e-mail: aromero@eco.uc3m.es

M. Triossi (✉)
Centro de Economía Aplicada, Departamento de Ingeniería Industrial, Universidad de Chile,
Avenida Republica 701, Santiago, Chile
e-mail: mtriossi@dii.uchile.cl

this paper, we present a many-to-one matching model that allows for both capacity and preference manipulation. Our objective is to understand whether a mechanism that includes a capacity reporting stage can implement stable allocations.

Indeed, there exists a widespread opinion that markets producing stable outcomes are more successful than those that do not produce such outcomes (see Roth and Sotomayor 1990; Roth 2002). We begin our analysis by isolating the strategic options at work in our settings. As a preliminary step, we concentrate solely on capacity manipulation. We focus on the Nash equilibria (NE) of capacity reporting games. First, we provide an equivalence result, that is, the NE of capacity reporting games are stable if and only if the stable rule used is immune to capacity manipulation. Second, we provide conditions under which capacity reporting games yield stable matchings at the NE. For this reason, we introduce the concept that the agents' preferences are acyclical. A cycle in the preferences of hospitals (interns) occurs when there is an alternating list of hospitals and interns "on a circle" such that every hospital (intern) prefers the intern (hospital) on its clockwise side to the intern (hospital) on its counterclockwise side and finds both acceptable. We say that preferences are acyclical if there are no cycles. In addition, we say that a group of agents form a simultaneous cycle if they form a cycle both in the preferences of interns and hospitals. Acyclicity holds, in particular, when the preferences of the agents on one side of the market are aligned. We prove that an absence of simultaneous cycles in the preferences of the agents guarantees the stability of the NE of capacity reporting games when any stable rule is used.

In addition, acyclicity is the minimal condition guaranteeing the stability of the NE when the hospital-optimal stable rule is employed. However, acyclicity is not necessary for the stability of NE outcomes under the intern-optimal stable rule. Thus, the intern-optimal stable rule is less prone to capacity manipulation than the hospital-optimal stable rule. We prove that the capacity reporting game can produce unstable NE if and only if the preferences of the hospitals satisfy a complex cycle condition. First, the preferences of the hospitals must be non-monotonic in population. Then, the cycles in the preferences of hospitals must be linked in a particular way. These findings extend the results of Konishi and Ünver (2006) and are related to the work of Kesten (2012).

Third, we proceed to study what we call generalized games of manipulation (GGM). GGM are two-stage extensive form games. In the first stage, each hospital states its capacity. In the second stage, the agents play a general assignment game. We do not specify a particular assignment game, but we consider two classes of mechanisms: revelation stable and non-revelation stable mechanisms.

In stable revelation mechanisms, agents are asked to submit their preferences, and stable matching is then implemented. This type of mechanism has been successfully used in practice (see, for instance, Roth and Sotomayor 1990; Roth 2002) but can be manipulated thorough the misrepresentation of both preferences (see Dubins and Freedman 1981) and capacities (Sönmez 1997).[1] prove that the iterated elimination of weakly dominated strategies produces a stable matching if the intern-optimal stable

[1] The under-reporting of capacities was a source of major concern in the school choice program in NYC before it was redesigned (see Abdulkadiroğlu et al. 2005).

rule is employed. Additionally, when the preferences of the hospitals are known, any stable rule produces stable allocations if the preferences of the agents do not have simultaneous cycles.

A stable non-revelation mechanism is any sequential game of complete information such that the interaction of agents leads to stable allocations with respect to the stated capacities (some examples of non-revelation stable mechanisms are presented in Kara and Sönmez 1997; Alcalde and Romero-Medina 2000; Sotomayor 2003). We show that there is no family of such games that implements stable matchings at every Subgame Perfect Equilibrium (*SPE*). However, if only acyclical preferences are allowed, any non-revelation mechanism implements stable allocations.

1.1 Related literature

The issue of preference manipulation has been widely discussed in the literature. Roth and Sotomayor (1990) present detailed references. Additionally, most of the mechanisms that scholars, such as Kara and Sönmez (1997) and Alcalde and Romero-Medina (2000), have introduced to implement stable allocations in matching markets do not include a capacity reporting stage.

Capacity manipulation has been studied in isolation as well. Sönmez (1997) demostrates that every stable revelation mechanism is prone to manipulation via capacities. Konishi and Ünver (2006) present the conditions under which capacity revelation games have pure-strategy NE , and show that under the assumption of common preferences, truthful capacity reporting is a dominant strategy for colleges. Mumcu and Saglam (2009) consider sequential capacity allocation under an assumption of common preferences. Kesten (2012) studies capacity manipulation of the intern-optimal stable rule and the top-trading cycle rule in school admission problems. Kesten (2012) result proves that if a particular acyclicity condition holds, the intern-optimal stable matching cannot be manipulated via capacities. Finally, Ehlers (2010) relates capacity manipulation to two forms of preference manipulation.[2] To our knowledge, the only paper that considers both capacity and preference manipulation is that of Kojima and Pathak (2008), and they find that the intern-optimal stable matching leaves little room for manipulation in large markets.

The structure of this paper is as follows. Section 2 presents the model, Sect. 3 studies capacity manipulation, and Sect. 4 extends our analysis to generalized games of manipulation. Finally, Sect. 5 concludes the paper. The proofs are presented in the Appendix.

2 The model

There are two disjoint sets of agents, a set of interns $I = (i_1, \ldots, i_n)$ and a set of hospitals $H = (h_1, \ldots, h_m)$. Generic agents from the two sets are denoted, respectively, as i and h, whereas a generic agent is denoted by $x \in H \cup I$. Hospitals hire a set of

[2] See also Kojima (2007).

interns, and interns train in no more than one hospital. Each hospital has a capacity $q_h \geq 1$, which denotes the maximum number of interns that hospital h can accept. Each intern $i \in I$ has a complete, transitive and strict preference ordering P_i over the set of hospitals $H \cup \{i\}$. Let R_i be the weak preference relation associated with P_i. Each hospital $h \in H$ has a complete, strict and transitive preference ordering P_h over the set of interns $I \cup \{h\}$. Similarly, R_h denotes the weak preference relation associated with P_h. Let $P_I = \left(P_{i_1}, \ldots, P_{i_n}\right)$ be the preference profile of interns over hospitals and let $P_H = \left(P_{h_1}, \ldots, P_{h_m}\right)$ be the preference profile of hospitals over subsets of interns. The quadruple (H, I, q, P), where $P = (P_H, P_I)$ and $q = (q_1, \ldots, q_m)$ is a **hospital-intern market**. The problem consists of matching hospitals with subsets of interns, allowing for the possibility that some agents remain unmatched.

Let $I' \subseteq I$ be a subset of interns. The best group of interns for hospital h among those belonging to I' is called the **choice set from** I' and is denoted by $Ch_h(I', P_h)$ or $Ch_f(I')$ when no ambiguity is possible. Formally, $Ch_h(I', P_h) = \arg\max_{P_h} \left\{I'' : I'' \subseteq I'\right\}$. Let $i \in I$ be an intern. If $\varnothing P_h i$, hospital h prefers not to employ any intern rather than employing i. In this case, i is **unacceptable to** h.[3] Otherwise, i is **acceptable to** h. $A(h)$ denotes the set of interns who are individually acceptable to h. Similarly, for every intern $i \in I$ P_i is a strict preference order defined on $H \cup \{i\}$. Any hospital h such that $i P_i h$ is **unacceptable to** i. Otherwise, h is **acceptable to** i. $A(i)$ denotes the set of hospitals that are acceptable to i.

We assume that the preferences of the hospitals over sets of interns are responsive with respect to hospitals' preferences over individual interns. A hospital h has responsive preferences if, for any two assignments that differ in only one intern, it prefers the assignment containing the most preferred intern. Formally, P_h is **responsive** if for all $I' \subset I$ and for all interns $i, i' \in I$: (1) $I' \cup \{i\} P_h I' \cup \left\{i'\right\} \Leftrightarrow i P_h i'$ and (2) $I' \cup \{i\} P_h I' \Leftrightarrow i \in A(h)$. We say that hospitals' preferences satisfy **strong monotonicity in population** if every hospital h prefers a group of acceptable interns of larger cardinality to sets of acceptable interns of smaller cardinality. Formally, if for all h, for all $J, K \subset A(h)$ $|J| > |K| \Rightarrow J P_h K$.[4] A **matching** on (H, I, q, P) is a function $\mu : H \cup I \to 2^I \cup H$ such that, for every $(h, i) \in H \times I$: (1) $\mu(h) \in 2^I$, (2) $\mu(i) \in H \cup \{i\}$, (3) $\mu(i) = h \Leftrightarrow i \in \mu(h)$, and (4) $|\mu(h)| \leq q_h$. Let $\mathscr{M}_q$ be the set of matchings on (H, I, q, P). In other words, a matching is an assignment of interns to hospitals such that no intern is hired by more than one hospital and no hospital hires more interns than indicated by its capacity.

When there is no ambiguity, we use P_H and P_I to denote the following binary relations within the set of matchings: for every μ, ν matchings, let $\mu P_H \nu$ if and only if $\mu(h) R_h \nu(h)$ for all $h \in H$ and $\mu(h) P_h \nu(h)$ for at least one h. Let $\mu P_I \nu$ if and only if $\mu(i) R_i \nu(i)$ for all $i \in I$ and $\mu(i) P_i \nu(i)$ for at least one i. Analogously, we write $\mu P_h \nu$ and $\mu P_i \nu$ if $\mu(h) P_h \nu(h)$ and $\mu(i) P_i \nu(i)$, respectively.

A matching μ is **individually rational** if (1) $\mu(h) \subseteq A(h)$ for all $h \in H$, and (2) $\mu(i) \in A(i)$ for all $i \in I$. In other words, a matching is individually rational if each hospital is assigned acceptable interns and every intern prefers to join her

[3] For all $i, i' \in I$ $i P_h i'$, $i P_h \varnothing$ and $\varnothing P_h i$ denote $\{i\} P_h \left\{i'\right\}$, $\{i\} P_h \varnothing$ and $\varnothing P_h \{i\}$, respectively.

[4] The symbol $|X|$ denotes the cardinality of the set X.

assigned hospital rather than stay unemployed. A matching μ is **blocked by the pair** $(h, i) \in H \times I$ if (1) $hP_i\mu(i)$ and (2) $i \in Ch_h(\mu(h) \cup \{i\})$. A matching μ is **stable in** (H, I, q, P) if it is individually rational and no pair blocks it. Therefore, a hospital-intern pair (h, i) blocks a matching μ if an intern i prefers joining a hospital h over her match or not being matched at all and hospital h prefers i to one of its interns or prefers to leave a position vacant. Otherwise, μ is **unstable**. $\Gamma(H, I, q, P)$ denotes the **stable set**, the set of matchings that are stable in market (H, I, q, P). If the hospitals have responsive preferences, the stable set is not empty. There is a stable matching, which is the **hospital-optimal** stable matching that is (weakly) preferred to any other stable matching by every hospital. Another stable matching, the **intern-optimal** stable matching, is (weakly) preferred to any other stable matching by every intern. The **hospital-optimal deferred acceptance algorithm** (Gale and Shapley 1962) generates the hospital-optimal stable matching of (H, I, q, P), and the **intern-optimal deferred acceptance algorithm** generates the intern-optimal stable matching of (H, I, q, P). The hospital-optimal and the intern-optimal stable matchings of (H, I, q, P) are denoted by $\varphi^H(H, I, q, P)$ and $\varphi^I(H, I, q, P)$, respectively. When there is no ambiguity, $\varphi^H(q)$ and $\varphi^I(q)$ are used rather than $\varphi^H(H, I, q, P)$ and $\varphi^I(H, I, q, P)$, respectively. Finally, we denote by $\varphi(q)$ any stable matching of market (H, I, q, P), and we call the function φ a **stable rule**.

Let φ be a stable rule. In a capacity reporting game, each hospital h simultaneously reports a capacity q_h, and the outcome is determined according to φ. Interns are passive players, and information is complete. The **capacity reporting game** induced by φ is a normal form game of complete information. The set of players is H, and the strategy space of hospital h is $\mathcal{Q}(q_h) = \{1, \ldots, q_h\}$(see also Hurwicz et al. 1995). The outcome function is φ. The preferences of hospitals over outcomes are generated by their preferences over the subsets of interns. Finally, a mechanism or rule is manipulable via capacities if there is a hospital that is strictly better off by under-reporting its capacity. Formally, the mechanism φ is manipulable by capacities at (q, P) if there exists $h \in H$ and $q'_h < q_h$ such that $\varphi\left(q'_h, q_{-h}\right) P_h \varphi(q)$. Given a profile of preferences P and a mechanism φ, we will say that φ is **capacity-proof** if stating the true capacities is a weakly dominant strategy under $\varphi(P, \cdot)$.

3 A look at Nash equilibria

In this section, we concentrate on the stability of NE outcomes of capacity reporting games. The objective is to provide the necessary and sufficient conditions that guarantee the existence and the stability of pure strategy NE.

Konishi and Ünver (2006) devote their attention to discovering sufficient conditions for the existence of pure strategy NE in capacity reporting games. They also prove that under the assumption of common preferences, stating the true capacities is a dominant strategy for hospitals.

Our first result links the stability of NE outcomes and capacity manipulation.

Lemma 1 *Let $V \in \{H, I\}$. Let q be a NE of the capacity revelation game induced by φ^V at (H, I, q^*, P). If h belongs to a pair blocking $\varphi^V(q)$ in (H, I, q^*, P), then $\varphi^V(q) P_h \varphi^V\left(q^*_h, q_{-h}\right)$.*

Lemma 1 shows that if a NE produces an unstable matching, then any hospital belonging to some blocking pair is strictly better off by manipulating its capacity. We employ this result throughout the paper.

3.1 The hospital-optimal rule

The literature on capacity reporting games has devoted attention to the property of strong monotonicity. Every counterexample in Konishi and Ünver (2006) and in Sönmez (1997) uses preferences that are not strongly monotonic. Strong monotonicity is intuitively linked to capacity manipulation. However, it is neither necessary nor sufficient for the stability of NE outcomes, as the following example demonstrates.

Example 1 Consider the following 2×2 problem. The preferences of the hospital are strongly monotonic such that $P_{h_1} : \{i_1, i_2\}, \{i_1\}, \{i_2\}$ and $P_{h_2} : \{i_1, i_2\}, \{i_2\}, \{i_1\}$. The preferences of the interns are $P_{i_1} : h_2, h_1$ and $P_{i_2} : h_1, h_2$. When the capacities are $(2, 2)$, $(1, 2)$ or $(2, 1)$, the unique stable matching is

$$\mu_1 = \begin{pmatrix} h_1 & h_2 \\ \{i_2\} & \{i_1\} \end{pmatrix}.$$

where $\begin{matrix} h_1 \\ \{i_2\} \end{matrix}$ denotes that $\mu_1(h_1) = i_2$. When the capacities are $(1, 1)$ the matching μ_1 is the intern-optimal stable matching. The hospital-optimal stable matching is:

$$\mu_2 = \begin{pmatrix} h_1 & h_2 \\ \{i_1\} & \{i_2\} \end{pmatrix}.$$

When the capacities are $(2, 2)$ the capacity revelation game induced by φ^H has two NE $(1, 1)$ and $(2, 2)$. The former yields μ_2 as an outcome, which is blocked by the pair (h_1, i_2). The latter yields μ_1 as an outcome.

When the hospitals state their true capacities, the interns receive offers from both hospitals, along the deferred acceptance algorithm. Each intern can choose her favorite hospital, and every hospital ends up hiring its least-preferred intern. However, both hospitals would be willing to switch their interns because there is a "cycle" in their preferences: $i_1 P_{h_1} i_2 P_{h_2} i_1$. This can be accomplished if both hospitals understate their capacity. In this way, each hospital only makes an offer to its favorite intern. Every intern accepts her unique offer and each hospital ends up hiring its favorite intern. Notice that this possibility arises because there is also a "cycle" in the preferences of the interns, which moves in the opposite direction of the cycle for the preferences of the hospitals: $h_2 P_{i_1} h_1 P_{i_2} h_2$.

The findings of Example 1 are intrinsic to capacity manipulation. It is the presence of simultaneous cycles of preferences that allows for the possibility of capacity manipulation under the hospital-optimal rule.

In general, a cycle in the preferences of the hospitals arises when there is a list of hospitals and interns alternating "on a circle" such that every hospital in the cycle prefers the intern on its clockwise side to the intern on its counterclockwise side but finds both acceptable. We present this concept formally in the following definitions.

Definition 1 A **hospitals' cycle (of length** $T+1$**)** is given by $h_0, \ldots, h_T$ with $h_l \neq h_{l+1}$ for $i = 0, \ldots, T$ and distinct $i_0, i_1, \ldots, i_T$ such that

1. $i_0 P_{h_0} i_T P_{h_T} i_{T-1} \ldots i_1 P_{h_1} i_0$,
2. for every l, $i_l \in A(h_l) \cap A(h_{l+1})$.[5]

The preferences of the hospitals are **acyclical** if they have no cycles of any length.

Assume that a cycle exists. If every i_l is initially assigned to h_{l+1}, every hospital is willing to exchange its assigned intern with its successor. If the preferences are acyclical, in particular, there are no cycles of length 2. Thus, each pair of hospitals has the same preferences over the set of mutually acceptable interns. Therefore, the notion of acyclicity generalizes the notion of *common preferences* presented by Konishi and Ünver (2006).

The notion of a cycle in the preferences of interns' preferences is specular.

Definition 2 An **interns' cycle (of length** $T+1$**)** is given by $h_0, \ldots, h_T$ and $i_0, i_1, \ldots, i_T$ such that

1. $h_0 P_{i_T} h_T P_{i_{T-1}} h_{T-1} \ldots s h_1 P_{i_0} h_0$,
2. for every l, $h_l \in A(i_{l-1}) \cap A(i_l)$.

The preferences of the interns are **acyclical** if there are no cycles of any length.

A simultaneous cycle arises when there is a list of hospitals and interns alternating "on a circle" such that every hospital (intern) prefers the intern (hospital) on its clockwise side to the intern (hospital) on its counterclockwise side but finds both acceptable. Formally:

Definition 3 A **simultaneous cycle** is given by hospitals $h_0, \ldots, h_T$ and interns $i_0, i_1, \ldots, i_T$ such that

1. $i_T P_{h_T} i_{T-1} P_{h_{T-1} i_{T-2}} \ldots i_o P_{h_0} i_T$,
2. $h_0 P_{i_T} h_T P_{i_{T-1}} h_{T-1} \ldots h_1 P_{i_0} h_0$,
3. for every l, $i_l \in A(h_l) \cap A(h_{l+1})$,
4. for every l, $h_l \in A(i_{l-1}) \cap A(i_l)$.

A simultaneous cycle naturally defines two "partial-matchings" μ_1 and μ_2 where $\mu_1(i_t) = h_t$ and $\mu_2(i_t) = h_{t+1}$. Every hospital in the cycle prefers μ_1, and every intern in the cycle prefers μ_2. The intuition developed in Example 1 helps to state the following lemma.

Lemma 2 *Let* $V \in \{H, I\}$. *If* $\varphi^V(q) P_h \varphi^V\left(q_h^*, q_{-h}\right)$ *for some* h *and some* $q_h < q_h^*$, *then there exists a simultaneous cycle.*

From Lemmas 1 and 2, it follows that if no simultaneous cycle exists, stating the true capacities is a dominant strategy for hospitals in both the hospital-optimal and the intern-optimal stable matchings. From Proposition 1 in Romero-Medina and Triossi (2012), it follows that this result extends to any stable rule.

[5] From now on, indices are considered modulo $T+1$.

Proposition 1 *Assume that no simultaneous cycle exists and let φ be any stable rule. Then,*

(1) φ is capacity-proof.
(2) For each q, the capacity revelation games induced by φ have a unique NE, that is, the unique stable matching of (H, I, q, P).

Notice that requiring a profile of preferences not to have simultaneous cycles is much less demanding than requiring acyclicity in the preferences of the hospitals or of the interns. Assume for instance that hospital h prefers intern i_k to intern i_l and hospital h' prefers intern i_l to intern i_k. Assume that i_k and i_l are acceptable to both h_i and h_j. We have a cycle of length 2. However, if i_k and i_l rank h and h' in the same way we do not have a simultaneous cycle. More precisely, let H_{kl} be the set of hospitals that prefer intern i_k to intern i_l and find both acceptable. If there are no simultaneous cycles, then the preferences of i_k and i_l must coincide on all pairs of hospitals $\left(h, h'\right) \in H_{kl} \times H_{lk}$. In particular, the result holds when either the preferences of the hospitals or the preferences of the interns are acyclical, and thereby generalizes Theorems 6 and 7 in Konishi and Ünver (2006). Actually, acyclicity is the weakest condition that guarantees that stating the true capacities is a dominant strategy and that every NE yields a stable matching under the hospital-optimal stable rule.

Proposition 2 *Assume that the preferences of the hospitals (interns) have a cycle. Then, there exists a preference profile for the interns (hospitals) and a vector of capacities q such that the capacity reporting game induced by φ^H yields an unstable matching at equilibrium at (H, I, q, P).*

Thus, from Propositions 1 and 2 and Lemma 1 follows the following equivalence result.

Corollary 1 *Let P_H and I be given. Then, $\varphi(P, \cdot)$ is capacity-proof for any P_I if and only if P_H does not contain any cycle.*

It follows that the hospital-optimal stable matching is manipulable via capacities under relatively weak conditions. Indeed, assume that all interns (hospital) are acceptable to every hospital (intern). In this case, assuming acyclicity is equivalent to the assumption that all hospitals (interns) have the same preferences for individual interns (for hospitals) (see Romero-Medina and Triossi 2012).

3.2 The intern-optimal rule

The intern-optimal stable matching makes stating their true preferences a dominant strategy for interns. Furthermore, Kojima and Pathak (2008) find that the intern-optimal stable matching leaves little room for manipulation in large markets. According toPathak and Sönmez (2009) the intern-optimal stable matching is strongly more manipulable via colleges preferences than the hospital-optimal stable matching. Nevertheless, several matching procedures have been redesigned to use intern-optimal stable matching. Examples of this include the *NRMP* and the school allocation method currently used in Boston.

In the case of manipulation via capacities, the evidence is inconclusive. Roth and Peranson (1999) observed little evidence of differential manipulability via capacities between the initial $NRMP$ and the intern-optimal version of the same algorithm. We find that the game induced by the intern-optimal stable matching is more resistant to capacity manipulation.

First, to include capacity manipulation in the intern-optimal stable matching, at least three interns are needed. Consider, for instance, a matching market with only two interns and assume that at least one hospital has a capacity of two. If the two interns are assigned to one hospital, this hospital cannot benefit from rejecting one of them because the preferences are responsive. If the interns are assigned to two different hospitals, reducing capacities does not affect the outcome of the game.

There is a second and more important difference between the manipulability of the hospital-optimal and intern-optimal stable matchings. Under the hospital-optimal rule, a hospital that understates capacities refrains from granting admission to some interns in the deferred acceptance algorithm. In this way, it prevents potential cycles of rejections of hospitals by interns. Under intern-optimal stable matching, the situation is different. By understating capacities a hospital generates a chain of rejections of interns by hospitals. Therefore, such a hospital might receive more applications from interns, but it will be able to fill fewer positions. As we will later prove, a hospital under the intern-optimal rule needs non-monotonic preferences to profit from capacity manipulation. Notice that if there are only two interns, the capacity revelation game induced by φ^I yields the intern-optimal stable matching as a NE outcome, in contrast to the case of φ^H. Example 2 provides the basic intuition that explains how the intern-optimal stable rule can result in unstable matchings.

Example 2 Let $I=\{i_1, i_2, i_3, i_4\}$, $H = \{h_1, h_2\}$. Let P_{h_1} be such that P_{h_1} :$\{i_1, i_2, i_3\}$, $\{i_1, i_2\}$, $\{i_1, i_3\}$, $\{i_1\}$, $\{i_2, i_3\}$, $\{i_2\}$, $\{i_3\}$, $\{i_4\}$, and let P_{h_2} be strongly monotonic in population according to the following preference over individual interns P_{h_2}:$\{i_4\}$, $\{i_3\}$, $\{i_2\}$, $\{i_1\}$. Let P_{i_1} : h_2, h_1 P_{i_2} : h_1, h_2 P_{i_3} : h_1, h_2, and P_{i_4} :h_2, h_1. When the capacity is $(2, 2)$, the intern-optimal stable matching is

$$\mu_1 = \begin{pmatrix} h_1 & h_2 \\ \{i_2, i_3\} & \{i_1, i_4\} \end{pmatrix}.$$

When the capacity is $(1, 2)$, the intern-optimal stable matching is

$$\mu_2 = \begin{pmatrix} h_1 & h_2 & \emptyset \\ \{i_1\} & \{i_3, i_4\} & \{i_2\} \end{pmatrix}.$$

When the capacity is $(2, 2)$, the unique NE under the intern-optimal rule is $(1, 2)$, which yields an unstable matching, μ_2.

In Example 2, if h_1 states its true capacity, it only receives applications from i_2 and i_3 and it never receives an application from i_1 under the intern-optimal deferred acceptance algorithm. If h_1 understates its capacity, it rejects the application from i_3 in the first stage of the deferred acceptance algorithm. In the second stage of the deferred acceptance algorithm, i_3 applies to h_2 and induces the rejection of i_1 by h_2. Finally, h_1 receives an application from i_1 and rejects i_2. The non-monotonicity of

h_1's preferences is necessary to generate the instability. The cycle at h_1 makes the chain of rejections possible.

Assume that hospital h has capacity q_h and fills the q_h^{th} position at stage k of the deferred acceptance algorithm for the first time. Let I_h be the set of interns employed at h at stage $k-1$. Stating capacity $q_h - 1$ can be profitable to h only if some intern i filling the q_h^{th} position applies to hospital h' and induces a chain of rejections such that some interns must apply to and be accepted by h. In this situation, h ends up with a new set of interns I'_h with at most $q_h - 1$ interns. If capacity manipulation is profitable, then $I'_h P_h(I_h \cup \{i\})$. Therefore, the preferences of h must not be strictly monotonic, as $|I_h \cup \{i\}| > \left|I'_h\right|$.

To describe the appropriate chains of rejections, we show that a new notion of cycles is necessary.

Definition 4 A **generalized cycle (of length** $T+1$**) at** h is given by a cycle in hospital's preferences $h = h_0, \ldots, h_T\, i_0, i_1, \ldots, i_T$ and by i_{-1} such that: $i_0 P_{h_0} i_{-1} P_{h_0} i_T$.

Notice that in Example 2, there is a generalized cycle at h_1: $i_1 P_{h_1} \{i_2, i_3\}\ i_1 P_{h_1} i_3 P_{h_2} i_2 P_{h_2} i_1$. If every hospital finds all interns to be acceptable, any generalized cycle can be reduced to a generalized cycle of length 2 (see Ergin 2002). Assume that a generalized cycle of length 2 at h exists. Let h_0 be matched with two interns i_{-1} *and* i_1. and let h_1 be matched with i_0. Assume also that $i_0 P_{h_0} \{i_{-1}, i_1\}$. Hospital h_0 would be willing to exchange its two interns for i_0 only, and h_1 would accept the proposal (potentially only hiring i_1). In general, non-monotonicity in the preferences of the hospitals and the generalized cycles must be connected in a particular way for capacity manipulation to be profitable under the intern-optimal rule.

Definition 5 A **non-monotonic cycle** at h is given by $M, M' \subseteq I$, with $|M| < \left|M'\right|$ such that

(1) $M P_h M'$.
(2) Let $M \setminus M' = \left\{i^1, \ldots, i^s\right\}$. For $k = 1, \ldots, s$ there is a generalized cycle at h, $h_0^k, \ldots, h_{T^k}^k, i_{-1}^k, i_0^k, i_1^k, \ldots, i_{T^k}^k$, $T^k \geq 1$ such that $i^k = i_0^k$ and $i_{-1}^k, i_{T^k}^k \in M' \setminus M$.
(3) For $k \neq k'$, $i_l^k \neq i_{l'}^{k'}$ for all $l = 0, \ldots T^k$, $l' = 0, \ldots, T^{k'}$.

The definition of a non-monotonic cycle is simple but demanding. It links non-monotonicity with cycles of rejection. It requires that (1) h prefers some set of interns containing fewer elements, M, to a set of interns containing more elements, M', and (2) any intern who belongs to the set with more interns but not to the one with fewer interns must be the starting point of a generalized cycle at h, for which the last intern of the cycle and i_{-1}^k belong to the larger set (M') but not to the smaller set (M); and (3) all the cycles in (2) must be disconnected.

The main result of this section weakens the requirements of Proposition 1: the intern-optimal stable matching is non-manipulable via capacities under relatively weak conditions.

Proposition 3 *Assume that no non-monotonic cycle exists. Then,*

(1) φ^I *is capacity-proof, and*

(2) for each q, the capacity revelation game induced by φ^I yields the intern-optimal stable matching of (H, I, q, P) at every NE.

If the preferences of the hospitals satisfy strong monotonicity in population (see definition in Sect. 2), no non-monotonic cycle exists and Proposition 3 implies and extends Theorem 5 in Konishi and Ünver (2006), as we formally state in the following Corollary.

Corollary 2 *Assume that one of the following conditions holds: the preferences of the hospitals satisfy strong monotonicity in population, there is no cycle of a length larger than 2 in the preferences of the hospitals, and there are no generalized cycles. Then, φ^I is capacity-proof and the game yields the intern-optimal stable matching at every NE.*

The absence of non-monotonic cycles is the minimal condition required to prevent capacity manipulation. If a non-monotonic cycle exists, there is a preference profile for the interns and a vector of capacities q such that the capacity reporting game yields an unstable matching in equilibrium. Additionally, if the preferences of the interns have a cycle of length at least 3, there exists a preference profile for the hospitals and a vector of capacities q such that the capacity reporting game yields an unstable matching in equilibrium. The same applies to the preferences profile for those interns with cycles of lengths less than 3. The following proposition shows which hospitals might benefit from capacity manipulation.

Proposition 4 *Assume that there exists a non-monotonic cycle at h or that the preferences of the interns have a cycle length at least 3. Then,*

(1) there is a preferences profile for the interns and a vector of capacities q such that the capacity reporting game induced by φ^I yields an unstable matching at equilibrium, and
(2) there is a preferences profile for the interns and a vector of capacities q such that hospital h can manipulate φ^I at (q, P).

Notice that Theorem 1 in Kesten (2012) shows that given a vector of capacity and preference (q, P), the intern-optimal stable matching cannot be manipulated via capacities if and only if (q, P) has no cycles. In Kesten (2012) (see also Ergin 2002), a priority structure contains a cycle if the following two conditions are satisfied: (1) There is a generalized cycle of length 2 $h_0, h_1\, i_0, i_1, i_{-1}$. (2) There exist disjoint sets of interns $N_{h_0}\, N_{h_1}$ such that $i\, P_{h_0}\, i_{-1}$ for all $i \in N_{h_0}\, i\, P_{h_1}\, i_0$ for all $i \in N_{h_1}$ $\left|N_{h_0}\right| = q_{h_0} - 1$ and $\left|N_{h_1}\right| = q_{h_1} - 1$. This definition of a cycle imposes a restriction on capacities that is absent in ours. Furthermore, given a cycle of length 2, there always exists a capacity vector such that condition (2) is satisfied with $N_{h_0} = N_{h_1} = \emptyset$ and $q_{h_0} = q_{h_1} = 1$. Finally, our condition for a non-monotonic cycle in Definition 5 is more restrictive. It requires the existence of non-monotonic preferences and, at least, a generalized cycle.[6]

[6] The example that proves the sufficiency of Kesten's condition for capacity manipulation includes a non-monotonic cycle.

4 Generalized games of manipulation

In most real life mechanisms, the strategic possibilities of agents go beyond capacity manipulation. For example, after hospitals have revealed their capacities, interns are assigned to hospitals according to stated preferences (for instance, in the $NRMP$, the Boston and New York mechanisms). The game that follows the capacity revelation stage has been modeled in different ways in the literature (see Alcalde and Romero-Medina 2000; Abdulkadiroğlu et al. 2005; Sotomayor 2008). At this stage, both hospitals and interns can manipulate the outcome by misrepresenting their preferences. We now define a class of games that allows for the manipulation of both capacities and preferences.

Definition 6 Let (H, I, q, P) be a hospital-intern market. A **generalized game of manipulation** is $\{G_{q'}\}_{q' \leq q}$ where $G_{q'} = (H, I, M_{q'}, g_{q'})$ is a game form. The set of players is $H \cup I$, the strategy space is $M_{q'} = \prod_{i \in I} M_{q',i} \times \prod_{h \in H} M_{q',h}$, and the outcome function is $g_{q'} : M_{q'} \to \mathscr{M}_{q'}$.

For every $q' \leq q$ $G_{q'}$ describes the game played by the agents following the revelation of a vector of capacities q'.

In the remainder of this section, we consider games in which hospitals simultaneously reveal a capacity q in the first stage and agents play the game $G_q = (H, I, M_q, g_q)$ in the second stage. We explore both revelation and non-revelation GGM.

4.1 Revelation games

We assume that the game played after the capacity revelation stage is a revelation game induced by a stable rule φ. Formally, $M_{q,x} = \mathscr{P}_x$ for all $x \in H \cup I$, $g_q(m) \in \Gamma(H, I, q, P)$ for all q. Such a generalized game of manipulation will be called preference-capacity manipulation game.

It is well known that no stable capacity revelation game makes the revelation of both every agent's preferences and capacities a dominant strategy. Therefore, the concept of a dominant strategy is too demanding for this framework. However, when the intern-optimal stable matching is used, stating true preferences is always a dominant strategy for interns. From Proposition 3, we also know that if the interns strategy P_I is a vector of acyclic preferences,[7] then stating true capacities is a dominant strategy for hospitals. In addition, the following result holds.

Proposition 5 *Assume that the preferences of the interns are acyclical. When the intern-optimal stable rule is used, the unique outcome that survives the iterated elimination of weakly dominated strategies in the preference-capacity manipulation game is the intern-optimal stable matching.*

An analogous result does not hold when the hospital-optimal stable rule is employed because truth-telling is not a dominant strategy for any agent.

[7] Notice that acyclicity implies non-monotonic cycles

However, there are situations in which the preferences of the hospitals can be taken as given. This situation is due, for instance, to institutional constraints. In this case, we can consider $M_{q,h} = \{P_h\}$ for all $h \in H$ $M_{q,i} = \mathscr{P}_i$ for all $i \in I$, and $g_q(m) \in \Gamma(H, I, q, P)$ for all q. Sotomayor (2008) shows that, when capacities are known, the game induced by the hospital-optimal rule implements the stable set in NE. However, this is not enough to prevent capacity manipulation. Only the assumption of acyclicity prevents the implementation of unstable allocations.

Proposition 6 *Let $V \in \{H, I\}$ and let $g_q(m) = \varphi^V(P, q)$ for all q. If the preferences of either interns or hospitals have no simultaneous cycles, the preference-capacity manipulation games induced by φ^V yield the unique stable matching of (H, I, q, P) as a SPE outcome.*

Proposition 6 follows from Sotomayor (2008) and Proposition 1.

4.2 Non-revelation games

In this section we consider capacity manipulation in non-revelation games, that are games where the strategy space of each agent does not necessarily coincide with her type space. Kara and Sönmez (1997) prove that the stable set is implementable in NE through a non-revelation game. Alcalde and Romero-Medina (2000), Sotomayor (2003), and Romero-Medina and Triossi (2010) present extensive form games capable of implementing the stable set and the intern-optimal stable matching in *SPE*.

For the remainder of the section, we assume that every $G_{q'}$ is an extensive form game. Let $SPE\left(G_{q'}, q', P\right)$ be the set of *SPE* outcomes of $G_{q'}$ when the capacity-preference vector is $\left(q', P\right)$. We assume that $SPE\left(G_{q'}, q', P\right) \neq \emptyset$ for all q' and that all such *SPE* outcomes are stable with respect to the stated capacities, which are $\emptyset \subsetneqq SPE\left(G_{q'}, q', P\right) \subseteq \Gamma(H, I, q', P)$ for all q'.[8] We call the family $\left\{G_{q'}\right\}_{q'}$, **stable**.

Even if the family $\left\{G_{q'}\right\}_{q'}$ is well behaved, adding a capacity manipulation stage does not guarantee that the resulting GGM produces stable matching in every *SPE*. In fact, the negative result is even stronger.

Proposition 7 *Assume that there are at least two hospitals and three interns. There is no family of stable non-revelation mechanisms $\left\{G_{q'}\right\}_{q'}$ such that the associated generalized game of manipulation yields stable SPE for all q.*

Proof The proof is by means of an example, based on Sönmez (1997).

Let $H \supseteq \{h_1, h_2\}$ and let $I \supseteq \{i_1, i_2, i_3\}$. Let $P_{h_1} : \{i_1, i_2, i_3\}, \{i_1, i_2\}, \{i_1, i_3\}, \{i_1\}, \{i_2, i_3\}, \{i_2\}, \{i_3\}$, and let $P_{h_2} : \{i_1, i_2, i_3\}, \{i_2, i_3\}, \{i_1, i_3\}, \{i_3\}, \{i_1, i_2\}, \{i_2\}, \{i_1\}$. Let $P_{i_1} = h_2, h_1$ $P_{i_2} = h_1, h_2$, and $P_{i_3} = h_1, h_2$. Finally, let $q_1 = q_2 = 2$ $q_1' = q_2' = 1$ be the possible capacities.

Assume that $q_{h_l} = 1$ for all $l \geq 3$. Let P_H be such that i_j $j = 1, 2, 3$ is not acceptable to h_l $l > 2$ such that i_j $j > 3$ is not acceptable to h_1 or to h_2 and such that

[8] While restrictive, this condition is nonetheless necessary for the GGM to yield stable allocations.

each i_j $j > 3$ is acceptable to at most one hospital. Let μ be the unique stable matching of the market $\left(H \setminus \{h_1, h_2\}, I \setminus \{i_1, i_2, i_3\}, q_{H\setminus\{h_1,h_2\}}, P_{H\setminus\{h_1,h_2\}}, P_{I\setminus\{i_1,i_2,i_3\}}\right)$. Let $\mu^0 = \begin{pmatrix} h_1 & h_2 & \emptyset \\ \{i_1\} & \{i_3\} & \{i_2\} \end{pmatrix}$, $\mu^1 = \begin{pmatrix} h_1 & h_2 \\ \{i_2\} & \{i_1,i_3\} \end{pmatrix}$, $\mu^2 = \begin{pmatrix} h_1 & h_2 \\ \{i_1\} & \{i_2,i_3\} \end{pmatrix}$, $\mu^3 = \begin{pmatrix} h_1 & h_2 \\ \{i_1,i_2\} & \{i_3\} \end{pmatrix}$, and $\mu^4 = \begin{pmatrix} h_1 & h_2 \\ \{i_2,i_3\} & \{i_1\} \end{pmatrix}$. Then:

$\Gamma(1, 1, q_{H\setminus\{h_1,h_2\}}) = \{(\mu^0, \mu)\}$, $\Gamma(1, 2, q_{H\setminus\{h_1,h_2\}}) = \{(\mu^1, \mu), (\mu^2, \mu)\}$,
$\Gamma(2, 1, q_{H\setminus\{h_1,h_2\}}) = \{(\mu^3, \mu), (\mu^4, \mu)\}$, $\Gamma(2, 2, q_{H\setminus\{h_1,h_2\}}) = \{(\mu^4, \mu)\}$.

We prove that for every family $\{G_{q'}\}_{q'}$ of stable mechanisms, the generalized game of manipulation induced by $\{G_{q'}\}_{q'}$ yields an unstable matching at some *SPE* when the true capacity vector is $(2, 2, q_{H\setminus\{h_1,h_2\}})$.

Assume by contradiction that there is a family $\{G_{q'}\}_{q'}$ of stable mechanisms such that the generalized game of manipulation induced by $\{G_{q'}\}_{q'}$ yields a selection of the stable set in *SPE* for every q. When both capacities are equal to 2 the *SPE* outcome is (μ^4, μ). There are two possibilities: either the *SPE* yielding (μ^4, μ) includes hospitals' h_1 and h_2 true capacities or it does not.

From subgame perfection, it follows that when both hospitals have capacity 2 (μ^4, μ) must be the unique NE outcome of one of the following games or no such games can have a pure strategy NE (without loss of generality, we disregard the moves of hospitals h_l, for $l \geq 3$:

(1)
$h_1\backslash h_2$	1	2
1	i_1, i_3	$i_2, \{i_1,i_3\}$
2	$\{i_2, i_3\}, i_1$	$\{i_2, i_3\}, i_1$

(2)
$h_1\backslash h_2$	1	2
1	i_1, i_3	$i_1, \{i_2, i_3\}$
2	$\{i_2, i_3\}, i_1$	$\{i_2, i_3\}, i_1$

(3)
$h_1\backslash h_2$	1	2
1	i_1, i_3	$i_2, \{i_1,i_3\}$
2	$\{i_1, i_2\}, i_3$	$\{i_2, i_3\}, i_1$

(4)
$h_1\backslash h_2$	1	2
1	i_1, i_3	$i_2, \{i_1,i_3\}$
2	$\{i_1, i_2\}, i_3$	$\{i_2, i_3\}, i_1$

where the table above presents the outcomes at matching μ^4 as a result of the capacities declared by h_1 and h_2. For example, $\mu^4(h_1 \mid (q_{h_1}, q_{h_2}) = (1, 1)) = i_1$, $\mu^4(h_2 \mid (q_{h_1}, q_{h_2}) = (1, 2)) = \{i_1,i_3\}$ and so on. Games (1) and (2) have $(1, 2)$ as NE. Games (3) and (4) have $(2, 1)$ as NE. None of the NE yields μ^4, thus yielding a contradiction. □

However, if there are no simultaneous cycles, then any such mechanisms implement the stable allocations.

Proposition 8 *Assume that the family of non-revelation mechanisms* $\{G_q\}_{q'\leq q}$ *is stable. Assume that the preferences of the agents P have no simultaneous cycles. Then every SPE of the generalized game of capacity manipulation induced by* $\{G_q\}_{q'\leq q}$ *yields the unique stable matching of* (H, I, q, P).

5 Conclusions

In this paper, we study the interaction between preference and capacity manipulation in many-to-one matching markets. This interaction, which has been largely overlooked

in the literature, is relevant in determining the likelihood of finding stable allocations in these markets. We first provide the necessary and sufficient conditions that guarantee the stability of NE and the strategy-proofness of truthful capacity revelation under the hospital-optimal and the intern-optimal stable rules. It turns out that the hospital-optimal rule is more prone to capacity manipulation than the intern-optimal rule. This result is in line with that of Kojima and Pathak (2008), who show how the intern-optimal rule leaves little room for manipulation in large markets. Second, we study generalized games of manipulation. A GGM is a multi-stage game in which hospitals first state their capacities and then interns are assigned to hospitals using a sequential mechanism. In the GGM, the agents develop the full extent of their strategic capabilities in a setting in which both capacity and preference manipulation are allowed. In this setting, we first present an impossibility result: none of the games can implement stable allocations in a general domain. However, if we restrict the preference domain, implementation becomes feasible. We show that the absence of simultaneous cycles guarantees the stability of NE outcomes when the preferences of hospitals are known, i.e., in a stable revelation mechanism. Furthermore, in the case of stable non-revelation mechanisms, we find that there is no possibility of implementing stable matching, unless preferences are acyclical.

The previous results in GGM provide insight as to the reasons why capacity manipulation may hinder the implementability of stable matching in some markets. First, the choice of the rule to be implemented is determinant because the hospital-optimal rule favors capacity manipulation. Moreover, the consequences of the previous choice differ depending on whether the GGM is designed with a revelation or a non-revelation mechanism.

Acknowledgments We are grateful to Alejandro Neme and Jorge Oviedo for valuable comments. Both authors acknowledge financial support from MEC ECON2008/027038 and MEC ECO2011/25330. Triossi acknowledges financial support from Fondecyt under project Nos. 11080132 and 1120974.

Appendix

Proof of Lemma 1 Let q be a NE when the capacity vector is q^* and let $\mu = \varphi^V(q)$ be the matching outcome. Assume μ is unstable in (H, I, q^*, P). Let $(h, j) \in H \times I$ be a hospital-intern pair blocking μ and set $\mu^* = \varphi^V\left(q_h^*, q_{-h}\right)$.

We next prove that $\mu P_h \mu^*$. We already know that $\mu(h) R_h \mu^*(h)$ because q is a NE. First, notice that $q_h < q_h^*$ and $|\mu(h)| = q_h$, otherwise (h, j) would block μ in (H, I, q, P). Consider the related one-to-one matching market. Let h_c' denote a copy of hospital $h' \in H$ in that market. From Proposition 2 in Gale and Sotomayor 1985a it follows that $\mu^* R_I \mu$ and $\mu R_{h_c'} \mu^*$ for every $h' \neq h$ and $\mu R_{h_c} \mu^*$ for every h_c such that $\mu(h_c) \neq h_c$. Furthermore, $\mu R_h \mu^*$ because q is a NE and $\mu \neq \mu^*$ because μ is unstable in $\left(H, I, \left(q_h^*, q_{-h}\right), P\right)$. Thus, $\mu P_H \mu^*$ and $\mu^* P_I \mu$. Finally, $\mu P_h \mu^*$, otherwise (h, j) would block μ in (H, I, q, P). □

Proof of Lemma 2 Let q be a vector of capacities. Let $h \in H$. Let $q_h < q_h^*$ and let q_{-h} be the vector of capacities for the other hospitals. Set $\mu = \varphi^V(q)$ and set $\mu^* = \varphi^V\left(q_h^*, q_{-h}\right)$. We prove that if $\mu P_h \mu^*$, then a simultaneous cycle exists. Proposition

2 in Gale and Sotomayor 1985a (applied to the related one-to-one matching market) implies that $\mu^* P_I \mu$ and $\mu P_H \mu^*$. More precisely, it implies that $i P_{h'} j$ for all h' such that $\mu(h') \neq \mu^*(h')$, for all $i \in \mu(h') \setminus \mu^*(h')$ and for all $j \in \mu^*(h') \setminus \mu(h')$. Set $I' = \{i : \mu^* P_i \mu\} \neq \emptyset$. Let $h_0 \in \mu(I')$, then $\mu P_{h_0} \mu^*$ and set $i_0 = max_{P_{h_0}} \mu(h_0) \setminus \mu^*(h_0)$ $i_0 \in I'$. For all $l \geq 1$, set $h_{l+1} = \mu^*(i_l)$ if $h_{l+1} \neq h_t$ for every $t < l+1$ and set $h_{l+1} = h_l$ otherwise. Observe that $h_0 \neq h_1$. Let $i_l = max_{P_{h_{l-1}}} \mu(h_{l-1}) \setminus (\mu^*(h_{l-1}) \cup \{i_0, \ldots, i_{l-1}\})$ if $\mu^*(h_{l-1}) \cup \{i_1, \ldots, i_{l-1}\} \not\supseteq \mu(h_{l-1})$, and set $i_{l+1}=i_l$ otherwise. The sequence is stationary because I' is finite. Let $\bar{l}$ be the minimal number $l \geq 1$ such that $h_l = h_{l+1}$. Let k be such that $h_k = h_{\bar{l}}$. Set $j_l = i_{l+k}$ and $r_l = h_{l+k}$ for every $l \leq \bar{l} - k$. The sequence satisfies $\mu(j_l) = h_l = \mu^*(j_{l-1})$ for $1 \leq l \leq \bar{l} - k - 1$, and $\mu^*(j_{\bar{l}-k}) = r_0 = \mu(j_0)$. We have: (1) $j_l P_{r_l} j_{l+1}$ for all $1 \leq l \leq \bar{l} - k$ and $j_0 P_{r_0} j_{\bar{l}-k}$; (2) $r_{l+1} P_{j_l} r_l$ for all $0 \leq l \leq \bar{l} - k - 1$ and $r_0 P_{j_{\bar{l}-k}} j_{\bar{l}-k}$. Thus, $h_0, \ldots, h_k, r_0, \ldots, r_k$ constitute a simultaneous cycle. □

Proof of Proposition 1 (1) From Lemma 2, φ^H cannot be manipulated through capacities. From Proposition 1 in Romero-Medina and Triossi (2012) it follows that if P has no simultaneous cycles, the set of stable matchings is a singleton for every q, then $\varphi^H(P, q) = \varphi(P, q)$ for every stable rule φ, for every q. It follows that no stable rule φ can be manipulates through capacities.

(2) From (1) a NE yielding a stable matching exists. From Lemma 1 and (1) the game does not yield unstable matchings at equilibrium. The rest of the claim follows from Proposition 1 in Romero-Medina and Triossi (2012), which shows that if there are no simultaneous cycles the set of stable matchings is a singleton. □

Proof of Proposition 2 Assume that there is a hospitals' cycle. Let $h_0, \ldots, h_T$ and $i_0, i_1, \ldots, i_T$ be defined as in Definition 1. We define a preference profile for the interns as follows. Let $h_{l+1} P_{i_l} h_l$ and $A(i_l) = \{h_l, h_{l+1}\}$ for $l = 0, \ldots, T$. Let $P_{I \setminus \{i_0, \ldots, i_T\}}$ be any vector of preferences. Consider the market $(H \setminus \{h_0, \ldots, h_T\}, I \setminus \{i_0, \ldots, i_T\}, q_{-\{h_0, \ldots, h_T\}}, P_{H \setminus \{h_0, \ldots, h_T\}}, P_{I' \setminus \{i_0, \ldots, i_T\}})$ and let μ' be the hospital-optimal stable matching. Let $P_{I \setminus \{i_0, \ldots, i_T\}}$ such that $A(i) = \mu(i)$ for every $i \in I$. When $q_{h_l} = 2$ for $l = 0, \ldots, T$, the market (H, I, q, P) has a unique stable matching: $\mu(i) = \mu'(i)$ for every $i \in I' \setminus \{i_0, \ldots, i_T\}$ and $\mu(i_l) = h_{l+1}$, for $l = 0, \ldots, T$. It is easy to see that when $q = (2, \ldots, 2, q_{-\{h_0, \ldots, h_T\}})$, the message $(1, \ldots, 1, q_{-\{h_0, \ldots, h_T\}})$ is a NE. The matching outcome is μ^*, where $\mu^*(i) = \mu'(i)$ for every $i \in I' \setminus \{i_0, \ldots, i_T\}$ $i \neq i_1, i_2$ $\mu^*(i_l) = h_{l+1}$, for $l = 0, \ldots, T$. The matching μ^* is blocked by (h_2, i_1). The proof of the remainder of the claim is identical and thus omitted. □

Proof of Proposition 3 Claim (1). Let $h \in H$. Let $q_h < q_h^*$ and let q_{-h} be a vector of capacities for hospitals other than h. Set $\mu = \varphi^I(q)$ and $\mu^* = \varphi^I(q_h^*, q_{-h})$. We prove that if $\mu P_h \mu^*$, a non-monotonic cycle exists. Proposition 2 in Gale and Sotomayor 1985b, applied to the related one-to-one market, implies that for every $h' \in H$ and $i, j \in I$ such that $i \in \mu(h') \setminus \mu^*(h')$ $j \in \mu^*(h') \setminus \mu(h')$ we have $i P_{h'} j$. From $\mu P_h \mu^*$ it follows that $\mu P_H \mu^*$. Proposition 2 in Gale and Sotomayor 1985b also implies that $\mu^* P_I^* \mu$.

There is no loss of generality in assuming that $\mu^*(i)$ is i's favorite hospital, for every $i \in I$, because $\mu P_H \mu^*$ and $\mu^* P_I \mu$. Consider the deferred acceptance algorithm where interns apply and the capacity vector is q. Let i be the first intern rejected by $\mu^*(i) = h'$. When i is rejected, hospital h' has all its $q_{h'}$ positions filled; hence i is rejected in favor of an intern in $\mu^*(h')$. It follows that $|\mu(h')| < |\mu^*(h')|$ and $h = h'$, thus the preferences of h are not monotonic.

Set $M = \mu(h)$ and $M' = \mu^*(h)$. Let $M \setminus M' = \{i^1, \ldots, i^s\}$ and $M' \setminus M = \{j^1, \ldots, j^q\}$. Set $r = |M'| - |M|$. It has been assumed that $\mu^*(i)$ is i's favorite hospital. Remember that $M P_h M'$. Consider the deferred acceptance algorithm where interns apply to hospitals and the capacity vector is q, which leads to μ. For every $i \in I$, intern i applies to $\mu^*(i)$ in the first stage of the deferred acceptance algorithm leading to μ. It must be the case that exactly r interns are rejected by h in the first stage of the deferred acceptance algorithm.

The remainder of the proof of Claim (1) is divided into two parts, where we find the elements of the non-monotonic cycles at h that appear in Definition 5, separately

(a) First, we find i^t_{-1}, as in Definition 5, using the following algorithm.

Step 1. Consider i^1. Let $d_0 > 1$ be the stage of the deferred acceptance algorithm leading to μ where i^1 has been accepted by h, and let $w_1 \in I$ be an intern that has been rejected by h in favor of i^1. If $w_1 \in M'$, stop and set $i^1_{-1} = w_1$, otherwise at step d_1 $1 < d_1 < d_0$ w_1 has been accepted and an intern w_2 has been rejected in favor of w_1. For all $k \geq 2$, if $w_k \in M'$, stop and set $i^1_{-1} = w_k$, otherwise at step d_k, $1 < d_k < d_{k-1}$ w_k has been accepted by h and an intern w_{k+1} has been rejected by h in favor of w_k. The sequence eventually stops at a $w_{K^1} \in M'$ who has been rejected by h in a step $d_{K^1} > 1$ of the deferred acceptance algorithm.[9] Set $i^1_{-1} = w_{K^1}$ and $W^1 = \{w_1, \ldots, w_{K^1}\}$. Notice that i^1_{-1} belongs to $M' \setminus M$. There is no loss of generality in assuming that $i^1_{-1} = j^1$. We have $i^1 P_h j^1$.

Step t. $2 \leq t \leq s$. Let $d_0 > 1$ be the stage of the deferred acceptance algorithm leading to μ where i^t has been accepted by h and let $w_1 \in I \setminus \bigcup_{l=1}^{t-1} W^l$ be an intern that has been rejected by h in favor of i^t. This is possible because if a number of interns are accepted by a college $\bar{h}$ at the same stage $t > 1$ of the deferred acceptance algorithm, the same number of interns who were previously employed at $\bar{h}$ are rejected. If $w_1 \in M'$, stop and set $i^t_{-1} = w_1$, otherwise at step d_1 $1 < d_1 < d_0$ w_1 has been accepted and an intern $w_2 \in I \setminus \bigcup_{l=1}^{t-1} W^l$ has been rejected in favor of w_1. For all $k \geq 2$, if $w_k \in M'$, stop and set $i^t_{-1} = w_k$, otherwise at step d_k, $1 < d_k < d_{k-1}$ w_k has been accepted by h and an intern w_{k+1} has been rejected by h in favor of w_k. The sequence eventually stops at some $w_{K^t} \in M'$ who has been rejected by h in a step $d_{K^t} > 1$ of the deferred acceptance algorithm.[10] Set $i^t_{-1} = w_{K^t}$ and set $W^t = \{w_1, \ldots, w_{K^t}\}$. Notice that i^t_{-1} belongs to $M' \setminus M$. There is no loss of generality in assuming that $i^t_{-1} = j^t$. We have $i^t P_h j^t$.

By construction $i^t_{-1} \neq i^l_{-1}$ for $l \neq t$.

(b) Next, for every $k = 1, \ldots, s$ we find the $h^k_0, \ldots, h^k_{T^k}, i^k_0, i^k_1, \ldots, i^k_{T^k}$ from Definition 5 and conclude.

[9] Every intern in the sequence is rejected because of the arrival of an application from another intern.

[10] See footnote 9.

For $k = 1, \ldots, s$ set $i_0^k = i^k$.

Step 1.k. Let i_1^k be the intern in favor of which i_0^k has been rejected by $\mu^*\left(i_0^k\right) = h_1^k$.

Step $t.k$ $t \geq 2$. At a stage d_t of the deferred acceptance algorithm leading to $\mu\, i_t^k$ has been rejected by $h_{p+1}^k = \mu^*\left(i_t^k\right) \neq h_t^k$ in favor of an intern $i_{t+1}^k \notin \mu^*\left(h_t^k\right)$. If $h_t^k = h_l^{k'}$ and $d_t^k = d_l^{k'}$ for some $k' < k$ and for some $l h_t^k$ has received at least $\left|\left\{k' < k : h_t^k = h_l^{k'} \text{for some } l \text{ and } d_t^k = d_l^{k'}\right\}\right| + 1$ applications that are better than i_t^k. Hence we can choose a i_t^k that is different from every other $i_l^{k'}$ $0 \leq k' < k$. We have $h_{t+1}^k = \mu^*\left(i_t^k\right)$ for all k $i_t^k P_{h_{t+1}^k} i_{t+1}^k$ and $h_{t+1}^k P_{i_t^k} h_t^k$.[11] The sequence stops at a T^k where $h_{T^k}^k = h$ rejects some interns in the first stage of the algorithm. By (a), $i_0^k P_h i_{-1}^k P_h i_{T^k}^k$. Therefore, there is a hospital $h \in H$ and M, M' subsets of interns, that satisfy $|M| < \left|M'\right|$ $M P_h M'$ and generalized cycles $h_0^k, \ldots, h_{T^k}^k, i_{-1}^k, i_0^k, i_1^k, \ldots, i_{T^k}^k$ with $T^k \geq 1$ such that $h = h_0^k$ and $i^k = i_0^k$ where $i_{-1}^k, i_{T^k}^k \in M' \setminus M$ and $i_{-1}^k \neq i_{-1}^{k'}$, for $k \neq k'$. Therefore, there is a non-monotonic cycle at h.

Claim (2). By Claim (1) there exists a NE that yields a stable matching. By Lemma 1 there are no unstable equilibria; hence every equilibrium outcome is stable. By contradiction, assume that the outcome is not the intern-optimal stable matching. It must be the case that some hospital has misrepresented its true capacity. Let q be a NE of the game and $q^* \geq q$ be the true capacity vector. Set $\mu = \varphi^I(q)$ and $\mu^* = \varphi^I(q^*)$. From Claim (1) μ is stable in (H, I, q^*, P), so $\mu P_H \mu^*$ and $\mu^* P_I \mu$. There is no loss of generality in assuming that $\mu^*(i)$ is intern i's favorite hospital. The matching μ is obtained through the intern-optimal deferred acceptance algorithm. It must be the case that at least one i is rejected by $\mu^*(i) = h$ in the first stage of the deferred acceptance algorithm. Every intern applies to her hospital under μ^* at this stage because h has misrepresented its true capacity. Hence h has fewer interns under μ than under μ^*. This yields a contradiction because both matchings are stable in (H, I, q^*, P). □

Proof of Proposition 4 Assume that there is a non-monotonic cycle at h. Using the notation from Definition 5, let $I' = \left\{i_{T^k}^1 : h_l^k = h\right\} \cap M' \cup \left\{i_{-1}^1, \ldots, i_{-1}^s\right\}$. Set $M^* = M' \cap M \cup I'$. Notice that $\left|M'\right| > |M|$ and $M P_h M'$. Set the preferences of the interns as follows. Let $A\left(i_l^k\right) = \left\{h_l^k, h_{l+1}^k\right\}$ and $h_{l+1}^k P_{h_k} h_l^k$ for all k and all l. Let $A(i) = \{h\}$ if $i \in M' \cap M$. For all other interns, let $A(i) = \{h(i)\}$ for a hospital $h(i) \notin \left\{h_l^k : k = 1, \ldots, s; l = 1, \ldots, T^k\right\}$. Let $q_{h_0} = q_h = \left|M^*\right|$ and $q = 1$ for all k, l such that $h_l^k \neq h$. Set all other capacities arbitrarily. We have $\varphi^I(q) = M^*$. From the property of the non-monotonic cycle at h, we know that $\varphi^I\left(q_h', q_{-h}\right) R_h M P_h M^*$. Let q_h' be h's best response to q_{-h}. We have $q_h' < q_h$. It is easy to see that $\left(q_h', q_{-h}\right)$ is a NE at (H, I, q, P). It yields a matching that is unstable because in any stable matching of (H, I, q, P) h is matched to $|M^*| > q_h'$ interns. □

Proof of Proposition 5 When the intern-optimal rule is used the revelation of true preferences is a dominant strategy for interns. From Proposition 3 we have $\varphi^I(q_h, q_{-h}, P)\, R_h \varphi^I\left(q_h', q_{-h}, P\right)$ for all q_h', q_h such that $q_h' \leq q_h$ and for all h. Thus,

[11] Because i_p^k first applies to h_{p+1}^k in the deferred acceptance algorithm.

to complete the proof of the claim it suffices to show that $\varphi^I(q, P_H, P_I) R_h \varphi^I(q, P'_h, P_{-h}, P_I)$ for all q, and P'_h as well as for all h if the preferences of the interns are acyclical. But this follows from Lemma 3 in Romero-Medina and Triossi (2012). □

Proof of Proposition 6 The claim follows from Theorem 1 on Sotomayor (2008) (pp. 631–632) and Proposition 1. □

Proof of Proposition 8 The claim follows from Proposition 1. □

References

Abdulkadiroğlu A, Pathak PA, Roth AE (2005) The New York City high school match. Am Econ Rev 95:364–367

Alcalde J, Romero-Medina A (2000) Simple mechanisms to implement the core of college admissions problems. Games Econ Behav 31(2):294–302

Dubins LE, Freedman DA (1981) Machiavelli and the Gale-Shapley algorithm. Am Math Mon 88:485–494

Ehlers L (2010) Manipulation via capacities revisited. Games Econ Behav 69:302–311

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

Gale D, Shapley LS (1962) College admissions and the stability of marriage. Am Math Mon 69:9–15

Gale D, Sotomayor M (1985a) Some remarks on the stable marriage problem. Discrete Appl Math 11: 223–232

Gale D, Sotomayor M (1985b) Ms Machiavelli and the stable matching problem. Am Math Mon 92:261–268

Hurwicz L, Maskin E, Postlewaite A (1995) Feasible Nash implementation of social choice rules when the designer does not know endowments or production sets. In: Ledyard JOThe economics of informational decentralization: complexity, efficiency and stability. Kluwer Academic Publishers, Amsterdam

Kara T, Sönmez T (1997) Implementation of college admission rules. Econ Theory 9:197–218

Kesten O (2012) On two kinds of manipulation for school choice problems. Econ Theory. doi:10.1007/s00199-011-0618-6

Kojima F (2007) When can manipulations be avoided in two-sided matching markets?—maximal domain results. B.E. J Theor Econ 7(1) (Contributions), Article 32

Kojima F, Pathak PA (2008) Incentives and stability in large two sided matching markets. Am Econ Rev 99(3):608–627

Konishi H, Ünver MU (2006) Games of capacity manipulation in hospital-intern markets. Soc Choice Welf 27:3–24

Mumcu A, Saglam I (2009) Games of capacity allocation in many-to-one matching with an aftermarket. Soc Choice Welf 33:383–403

Pathak PA, Sönmez T (2009) Comparing mechanisms by their vulnerability to manipulation. WP, MIT and Boston College, Cambridge

Romero-Medina A, Triossi M (2010) Non-revelation mechanisms in many-to-one markets. Available at SSRN: http://ssrn.com/abstract=1675222

Romero-Medina A, Triossi M (2012) Acyclicity and singleton cores in matching markets. Econ Lott (forthcoming)

Roth AE (2002) The economist as engineer: game theory, experimentation as tools for design economics. Econometrica 70:1341–1378

Roth AE, Peranson E (1999) The redesign of the matching market for American physicians: some engineering aspects of economic design. Am Econ Rev 89:748–780

Roth AE, Sotomayor M (1990) Two-sided matching: a study in game theoretic modeling and analysis. Cambridge University Press, Cambridge

Sönmez T (1997) Manipulation via capacities in two-sided matching markets. J Econ Theory 77:197–204

Sotomayor M (2003) Reaching the core of the marriage market through a non-revelation matching mechanism. Int J Game Theory 32:241–251
Sotomayor M (2008) The stability of the equilibrium outcomes in the admission games induced by stable matching rules. Int J Game Theory 36:621–640