Soc Choice Welf (2012) 39:23–33
DOI 10.1007/s00355-011-0549-y

ORIGINAL PAPER

Manipulability in matching markets: conflict and coincidence of interests

Itai Ashlagi · Flip Klijn

Received: 16 June 2010 / Accepted: 23 March 2011 / Published online: 10 April 2011

Abstract We study comparative statics of manipulations by women in the men-proposing deferred acceptance mechanism in the two-sided one-to-one marriage market. We prove that if a group of women weakly successfully manipulates or employs truncation strategies, then all other women weakly benefit and all men are weakly harmed. We show that these results do not appropriately generalize to the many-to-one college admissions model.

1 Introduction

We study the effect of strategic agents on non-strategic agents in two-sided matching markets. Consider the marriage market introduced by Gale and Shapley (1962) where the two (finite) sides of the market are "men" and "women," each agent having preferences over the other side of the market and the prospect of being alone. An outcome for a marriage market is a matching in which each agent either marries an agent from the other side of the market or remains single. A key property for a matching is stability. A matching is stable if each agent has an acceptable match and there is no pair of a man and a woman who like each other better than their current matches. Using their deferred acceptance algorithm, Gale and Shapley (1962) constructively proved that there exists a stable matching for each profile of preferences. Moreover, Knuth (1976) showed that the set of stable matchings is a distributive lattice with respect to the preferences of the agents. An important consequence is that on the set of stable

I. Ashlagi
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

F. Klijn (✉)
Institute for Economic Analysis (CSIC), Campus UAB, 08193 Bellaterra, Barcelona, Spain
e-mail: flip.klijn@iae.csic.es

matchings each side of the market has common interests that are in conflict with those of the other side.[1]

In this article, we show that the conflict and coincidence of interests extend to the effects of manipulations in the direct-revelation game based on the deferred acceptance algorithm.[2] Consider the direct-revelation mechanism induced by the men-proposing deferred acceptance algorithm. It is in the best interest of each man to report his true preferences (Dubins and Freedman 1981; Roth 1982), but women typically have incentives to misreport their true preferences. Concerning her strategic options, a woman can restrict her strategy space to truncation strategies, which are the strategies obtained by removing a tail of men (i.e., some least preferred men) from her (true) ordered list of acceptable men. More precisely, for any (general) manipulation by a woman, there is a truncation strategy which is at least as good (Roth and Vande Vate 1991). We show that under the men-proposing deferred acceptance mechanism,

- any weakly successful group manipulation[3] by women is weakly beneficial to all other women and weakly harmful to all men (Proposition 3.1), and
- truncating preferences by some women is weakly beneficial to all other women and weakly harmful to all men (Proposition 3.2).

Finally, we consider extending our results to the many-to-one college admissions model where students have to be assigned to colleges (with possibly multiple seats). A minor adaptation of the proof of Proposition 3.2 shows that under the student-proposing deferred acceptance mechanism, any truncation of preferences by some colleges is weakly beneficial to the other colleges and weakly harmful to all students. However, Kojima and Pathak (2009) showed that under the student-proposing deferred acceptance mechanism, truncation strategies typically do not exhaust the strategic options of the colleges. They proved that so-called dropping strategies constitute a class of exhaustive strategies. A dropping strategy of a college is obtained by removing some students from its (true) ordered lists of acceptable students (i.e., not necessarily a tail of least preferred students). We show that neither of our results extends to the college admissions model in an appropriate way: there are dropping strategies and successful manipulations that strictly harm some other college and strictly benefit some student. We do obtain a positive result for the college admissions model if we strengthen the definition of weakly successful manipulation (Proposition 4.1).

Our results complement work by Crawford (1991), Ma (2010), and Roth (1984b). Crawford (1991) studied general many-to-one matching markets and investigated the effect of the entrance of an agent on the welfare of the other agents. When restricted to college admissions markets, his result is the particular case of our result in which a college submits an empty truncation strategy. For the game induced by the men-proposing deferred acceptance mechanism in marriage markets, Roth (1984b) showed that the equilibria in which the men play their weakly dominant strategy of truth telling yield stable matchings. In fact, Ma (2010) proved that each such equilibrium yields

[1] See also Roth (1984a, 1985b) for further results on polarization of interests in two-sided markets.

[2] For the important role of the deferred acceptance algorithm in both matching theory and many real-life applications we refer to Roth (2008).

[3] That is, none of the manipulating agents is strictly worse off.

the women-optimal stable matching if women play truncation strategies. For the game induced by the student-proposing deferred acceptance mechanisms in college admissions markets, Ma (2010) showed that each equilibrium in which students play their weakly dominant strategy of truth telling and colleges play truncation strategies yields an unstable matching or the college-optimal stable matching.

To obtain a concise exposition and to avoid unnecessary notation, we recall in Sect. 2 the college admissions model with general quotas. In Sect. 3, we present our results when all quotas are equal to one (the special case of the marriage model). Finally, in Sect. 4 we present our findings for the case with general quotas.

2 College admissions

There are two finite and disjoint sets of agents: a set S of students and a set C of colleges. We denote a generic student, college, and agent by s, c, and i, respectively. For each college c, there is a fixed quota q_c that represents the number of positions it offers.[4]

Each student s has a complete, transitive, and strict preference relation over the colleges and the prospect of being unmatched. Hence, student s's preferences can be represented by a strict ordering P_s of the elements in $C \cup s$. For $c, c' \in C \cup s$, we write $c\ P_s\ c'$ if student s strictly prefers c to c' ($c \neq c'$), and $c\ R_s\ c'$ if s likes c at least as well as c' ($c\ P_s\ c'$ or $c = c'$). If $c \in C$ such that $c P_s s$, then we call c an acceptable college for student s. Let $P_S = (P_s)_{s \in S}$.

Each college c has a complete, transitive, and strict preference relation over the *individual* students and the prospect of being unmatched. Hence, college c's preferences over individual students can be represented by a strict ordering P_c of the elements in $S \cup \{\emptyset\}$. For $s, s' \in S \cup c$, we write $s\ P_c\ s'$ if college c strictly prefers s to s' ($s \neq s'$), and $s\ R_c\ s'$ if c likes s at least as well as s' ($s\ P_c\ s'$ or $s = s'$). If $s \in S$ such that $s P_c \emptyset$, then we call s an acceptable student for college c. Let $P_C = (P_c)_{c \in C}$.

With some abuse of notation we also represent a student s's preferences P_s as an ordered list of the elements in $C \cup s$. For instance, $P_s = c_3\, c_2\, s\, c_1\, \dots\, c_4$ indicates that s prefers c_3 to c_2 and he prefers remaining single to any other college. Similarly, college c's preferences over individual students can be represented as an ordered list P_c of the elements in $S \cup \{\emptyset\}$. We often omit the unacceptable agents from agent i's ordered list P_i.

A set of students $S' \subseteq S$ is feasible for college c if $|S'| \leq q_c$. Each college c has a complete and transitive preference relation over feasible sets of students, which can be represented by a weak ordering P_c^* of the elements in $\mathcal{P}(S, q_c) \equiv \{S' \subseteq S : |S'| \leq q_c\}$. For $S', S'' \in \mathcal{P}(S, q_c)$, we write $S'\ P_c^*\ S''$ if college c strictly prefers S' to S'' ($S' \neq S''$), and $S'\ R_c^*\ S''$ if c likes S' at least as well as S''.

We will assume that for each college c, P_c^* is a responsive extension of P_c,[5,6] i.e., for all $S' \in \mathcal{P}(S, q_c)$,

---

[4] The marriage model is the special case of one-to-one (two-sided) matching where for all $c \in C$, $q_c = 1$.

[5] Sometimes we also say that R_c^* is a responsive extension of R_c.

[6] See Roth (1985a) and Roth and Sotomayor (1989) for a discussion of this assumption.

(r1) if $s \notin S'$ and $|S'| < q_c$, then $(S' \cup s) P_c^* S'$ if and only if $s P_c \emptyset$ and
(r2) if $s \notin S'$ and $t \in S'$, then $\big((S' \backslash t) \cup s\big) P_c^* S'$ if and only if $s P_c t$.

By responsiveness, P_c^* coincides with P_c on the set of individual students. Note also that P_c allows for multiple responsive extensions.

A college admissions market is a triple (S, C, P), where $P = (P_S, P_C^*)$. A matching for college admissions market (S, C, P) is a function μ on the set $S \cup C$ such that

(m1) each student is either matched to exactly one college or unmatched, i.e., for all $s \in S$, either $\mu(s) \in C$ or $\mu(s) = s$,
(m2) each college is matched to a feasible set of students, i.e., for all $c \in C$, $\mu(c) \in \mathcal{P}(S, q_c)$, and
(m3) a student is matched to a college if and only if the college is matched to the student, i.e., for all $s \in S$ and $c \in C$, $\mu(s) = c$ if and only if $s \in \mu(c)$.

Given matching μ, we call $\mu(s)$ student s's match and $\mu(c)$ college c's match. We often denote a matching as a list of the colleges to which the students are matched.

A key property of matchings is stability. First, we impose a voluntary participation condition. A matching μ is individually rational if neither a student nor a college would be better off by breaking a current match, i.e., if $\mu(s) = c$, then $c P_s s$ and $\mu(c) P_c^* (\mu(c) \backslash s)$. By responsiveness of P_c, the latter requirement can be replaced by $s P_c \emptyset$. Thus alternatively, a matching μ is individually rational if any student and any college that are matched to one another are mutually acceptable. Second, if a student s and a college c are not matched to one another at a matching μ but the student would prefer to be matched to the college and the college would prefer to either add the student or replace another student by student s, then we would expect this mutually beneficial adjustment to be carried out. Formally, a pair (s, c), $s \notin \mu(c)$, is a blocking pair if $c P_s \mu(s)$ and (a) $[\,|\mu(c)| < q_c$ and $s P_c \emptyset\,]$ or (b) [there exists $t \in \mu(c)$ such that $s P_c t$].[7] A matching is stable if it is individually rational and there are no blocking pairs. Since stability does not depend on the particular responsive extensions of the colleges' preferences over individual students, a college admissions market is henceforth a triple $(S, C, P = (P_S, P_C))$, or simply P.

The deferred acceptance (DA) algorithm (Gale and Shapley 1962) yields a stable matching. Let $Q = (Q_S, Q_C)$ be a preference profile (of preferences over individual agents). The student-proposing DA algorithm applied to Q, denoted by DA(Q) for short, finds a matching through the following steps.[8]

Step 1: Each student s proposes to the college that is ranked first in Q_s (if there is no such college then s remains single). Each college c considers its proposers and tentatively assigns its q_c positions to these students one at a time following the preferences Q_c. All other proposers are rejected.

Step k, $k \geq 2$: Each student s that is rejected in Step $k-1$ proposes to the next college in his list Q_s (if there is no such college then s remains single). Each

[7] Recall that by responsiveness (a) implies $(\mu(c) \cup s) P_c^* \mu(c)$ and (b) implies $((\mu(c) \backslash t) \cup s) P_c^* \mu(c)$.

[8] Note that the DA algorithm does not depend on the particular responsive extensions.

college c considers the students it has been holding together with its new proposers and tentatively assigns its q_c positions to these students one at a time following the preferences Q_c. All other proposers are rejected.

The algorithm stops when no student is rejected. Then, all tentative matches become final. With some abuse of notation, let $\mu(Q)$ denote the matching. For $i \in S \cup C$, let $\mu(Q, i)$ denote the match of agent i at $\mu(Q)$. Gale and Shapley (1962) proved that matching $\mu(Q)$ is stable with respect to preference profile Q. In fact, from Roth and Sotomayor (1990, Theorem 5.31) it follows that $\mu(Q)$ is the best (worst) stable matching for the students (colleges) with respect to preference profile Q. Roth (1985a, Theorem 5*) proved that under the direct-revelation mechanism induced by μ it is a weakly dominant strategy for the students to reveal their true preferences. Therefore, we will assume that students are truthful and that colleges are the only strategic agents. In fact, there is no stable mechanism that makes it a weakly dominant strategy for each college to report its true preferences (Roth 1985a, Proposition 2).

Let P be a college admissions market. A *(group) manipulation* by a group of colleges C' is a strategy-profile $P'_{C'} = (P'_c)_{c \in C'}$. If $|C'| = 1$, then $P'_{C'}$ is an *individual manipulation*. A manipulation is *weakly successful* if for all $c \in C'$, $\mu(P', c)$ R^*_c $\mu(P, c)$ where $P' = (P'_{C'}, P_{-C'})$. A manipulation is *successful* if for all $c \in C'$, $\mu(P', c) R^*_c$ $\mu(P, c)$ and for some $c' \in C'$, $\mu(P', c')$ $P^*_{c'}$ $\mu(P, c')$.

The following well-known result states that students and colleges have opposite interests whenever a manipulation leads to a stable matching.

Lemma 2.1 *Under the student-proposing DA mechanism, a group manipulation by some colleges C' is weakly beneficial to all colleges (for all responsive extensions) and weakly harmful to all students if the induced matching is stable. If the matching is not stable then each blocking pair contains a college from C'.*

Proof Let $P'_{C'}$ be a group manipulation and let $P' = (P'_{C'}, P_{-C'})$. By assumption, $\mu(P')$ is stable for the market P. Hence, by student-optimality of $\mu(P)$, all colleges weakly prefer $\mu(P')$ to $\mu(P)$ and all students weakly prefer $\mu(P)$ to $\mu(P')$. The second statement follows from the observation that $\mu(P')$ is stable for P' and that for each pair (s, c) with $c \notin C'$, $P_s = P'_s$ and $P_c = P'_c$. □

The following example illustrates that a manipulation may lead to an unstable matching, even if the manipulating colleges are strictly better off at the new matching.

Example 1 (A successful manipulation that induces an unstable matching.) Consider the matching market with 3 students, 3 colleges, and preferences P given by the columns in the table below. For instance, $c_3\ P_{s_1}\ c_1\ P_{s_1}\ c_2\ P_{s_1}\ s_1$.

Students			Colleges		
s_1	s_2	s_3	c_1	c_2	c_3
$\boxed{c_3}$	$\boxed{c_2}$	$\boxed{c_1}$	s_1	$\mathbf{s_1}$	$\mathbf{s_3}$
c_1	$\mathbf{c_1}$	$\mathbf{c_3}$	$\mathbf{s_2}$	$\boxed{s_2}$	$\boxed{s_1}$
$\mathbf{c_2}$	c_3	c_2	$\boxed{s_3}$	s_3	s_2

One easily verifies that $\mu(P) = (s_3, s_2, s_1)$—the boxed matching in the table. Suppose that college c_1 submits the list $P'_{c_1} = s_2$. Then, $\mu(P') = (s_2, s_1, s_3)$—the boldfaced matching in the table. Note P'_{c_1} is a successful manipulation since $\mu(P', c_1) = s_2 \; P_{c_1} \; s_3 = \mu(P, c_1)$. However, $\mu(P')$ is not stable with respect to the true preferences P (the unique blocking pair is (s_1, c_1)). ◇

In Example 1, all colleges that do not manipulate weakly benefit and all students are weakly harmed. Since the resulting matching is not stable this observation does not follow from Lemma 2.1. Nevertheless, we will prove in the next section that the observed opposed interests are a feature of two important classes of manipulations if all colleges' quotas are equal to one.

3 All quotas equal to one

In this section, we focus on the marriage model, i.e., the special case of college admissions where each college has one position. Hence, *we will assume throughout this section that for all $c \in C$, $q_c = 1$*. Before turning to our results, we first provide the formal definition of an important class of manipulations.

A truncation strategy (Roth and Vande Vate 1991) of a college c is a strategy (or equivalently, an ordered list) P'_c obtained from P_c by making a tail of acceptable students unacceptable.[9] Formally, P'_c is a *truncation strategy* if for all $s, s' \in S$, (a) [if $s \; R'_c \; s' R'_c \; \emptyset$ then $s \; R_c \; s' R_c \; \emptyset$], and (b) [if $s \; P'_c \; \emptyset$ and $s' \; P_c \; s$ then $s' \; P'_c \; \emptyset$].

Note that not every weakly successful, individual manipulation is a truncation strategy (see, for instance, Example 1). Likewise, not every truncation strategy is a weakly successful manipulation. For instance, an empty truncation strategy leaves the college unmatched. However, truncation strategies are exhaustive in the sense that any individual manipulation can be replicated or improved upon by some truncation strategy (Roth and Vande Vate 1991, Theorem 2).[10]

We next show that the opposed interests observed in Example 1 are a feature of the marriage model as long as the group manipulations are weakly successful or consist of truncation strategies. To state and prove our results, we introduce the following additional notation. For every integer $k \geq 1$, let $S(Q, c, k)$ be the set of students that will have proposed to college c by step k under DA(Q), i.e., in some step $l \in \{1, \ldots, k\}$ of DA(Q). Let $S(Q, c)$ be the set of students that will have proposed to c by the last step of DA(Q), i.e., $S(Q, c) = \cup_k S(Q, c, k)$.

The proofs of our results are similar to Crawford (1991). We include them since for marriage markets the arguments are shorter and more transparent.

Proposition 3.1 *Under the student-proposing DA mechanism, any weakly successful group manipulation by colleges is weakly beneficial to the other colleges and weakly harmful to all students.*

---

[9] Truncation strategies have been observed in practice. See, for instance, Mongell and Roth (1991).

[10] This fact crucially depends on the assumption that each college has only one position. See also the discussion in Sect. 4.

Proof Let $P'_{C'}$ be a weakly successful manipulation of a group of colleges C' and let $P' = (P'_{C'}, P_{-C'})$. It is sufficient to show that for each college c and each step k, $S(P, c, k) \subseteq S(P', c)$. For $k = 1$, the inclusion is obvious since at step 1 of DA(P) and DA(P') each student proposes to exactly the same college.

Assume that the inclusion holds for k. We will show that the inclusion also holds for $k+1$. Let $s \in S(P, c, k+1)$. If $s \in S(P, c, k)$, then by induction, $s \in S(P', c)$. Therefore, assume $s \in S(P, c, k+1) \setminus S(P, c, k)$. Then, in DA$(P)$, student s proposed to c at step $k+1$ but not at step k. Therefore, s was rejected by some college $\bar{c} \neq c$ at step k of DA(P). By the induction hypothesis, $s \in S(P, \bar{c}, k) \subseteq S(P', \bar{c})$. If $\bar{c} \notin C'$ then $\bar{c}$ will also reject s in DA(P') since $P'_{\bar{c}} = P_{\bar{c}}$. If $\bar{c} \in C'$ then $\mu(P', \bar{c})\ R_{\bar{c}}\ \mu(P, \bar{c})\ P_{\bar{c}}\ s$, which implies that in the last step of DA(P') college $\bar{c}$ strictly prefers its match to s (according to its true preferences). Therefore, in either case $\bar{c}$ will also eventually reject s in DA(P'). Since s makes his proposals in the same order in DA(P) and DA(P'), he will have proposed to c by the last step of DA(P'). Hence, $s \in S(P', c)$. □

For marriage markets, the next proposition generalizes the results of Crawford (1991) from an *individual empty* truncation strategy to *arbitrary group* truncation strategies. It also shows that we can replace the weakly successful group manipulations in Proposition 3.1 by (possibly unsuccessful) truncation strategies.

Proposition 3.2 *Under the student-proposing DA mechanism, any group manipulation by colleges that consists of truncation strategies is weakly beneficial to the other colleges and weakly harmful to all students.*

Proof The proof is (almost) identical to that of Proposition 3.1: One only needs to replace the line "If $\bar{c} \in C'$ then ... (according to its true preferences)." by the following.

If $\bar{c} \in C'$, then $\bar{c}$ will also have rejected s by the last step of DA(P') since $P'_{\bar{c}}$ is a truncation strategy obtained from $P_{\bar{c}}$. □

4 General quotas

In this section, we consider extending our results to college admissions markets in which colleges have multiple positions, i.e., for all $c \in C$, $q_c \geq 1$.

A minor adaptation of the proof of Proposition 3.2 shows that under the student-proposing DA mechanism, any group manipulation by colleges that consists of truncation strategies is weakly beneficial to the other colleges (for all responsive extensions) and weakly harmful to all students. However, Kojima and Pathak (2009) showed that under the student-proposing deferred acceptance mechanism, truncation strategies typically do *not* exhaust the strategic options of the colleges. More precisely, they presented a many-to-one market in which for some college there is a strategy such that any truncation strategy yields a strictly worse match. They also proved that so-called dropping strategies do constitute a class of exhaustive strategies. A dropping strategy of a college is obtained by removing some students from its (true) ordered lists of acceptable students (i.e., not necessarily a tail of least preferred students). Formally, for a college c with preferences P_c over individual students, P'_c is a *dropping strategy* if for all students s, s', $[s\ R'_c\ s'\ R'_c\ \emptyset$ implies $s\ R_c\ s'\ R_c\ \emptyset]$.

Therefore, a possible appropriate extension of Proposition 3.2 to college admissions would involve dropping strategies rather than truncation strategies. The next example, however, shows that neither of our results in Sect. 3 extends to the college admissions model with general quotas in an appropriate way: there are markets in which a (successful) dropping strategy strictly harms some other college and strictly benefits some student.

Example 2 (Propositions 3.1 and 3.2 do not appropriately generalize to college admissions with general quotas.) Consider the following matching market with students s_1, s_2, s_3, and s_4, and colleges c_1 and c_2. Each college has two seats. The preferences P over individual agents are given by the columns in the table below. We assume that the colleges' responsive extensions $P^*_{c_1}$ and $P^*_{c_2}$ are such that $\{s_1, s_4\} P^*_{c_1} \{s_2, s_3\}$ and $\{s_1, s_4\} P^*_{c_2} \{s_2, s_3\}$.[11]

Students				Colleges	
s_1	s_2	s_3	s_4	c_1	c_2
$\boxed{c_2}$	$\boxed{c_1}$	$\boxed{c_1}$	$\boldsymbol{c_1}$	$\boldsymbol{s_1}$	$\boxed{s_4}$
$\boldsymbol{c_1}$	$\boldsymbol{c_2}$	$\boldsymbol{c_2}$	$\boxed{c_2}$	$\boxed{s_2}$	$\boldsymbol{s_2}$
				$\boxed{s_3}$	$\boldsymbol{s_3}$
				$\boldsymbol{s_4}$	$\boxed{s_1}$

One easily verifies that $\mu(P) = (c_2, c_1, c_1, c_2)$—the boxed matching in the table. Suppose that college c_1 submits the dropping strategy $P'_{c_1} = s_1, s_4$. Then, $\mu(P') = (c_1, c_2, c_2, c_1)$—the boldfaced matching in the table. Note that P'_{c_1} is a successful dropping strategy since college c_1 prefers $\{s_1, s_4\}$ to $\{s_2, s_3\}$. Since college c_2 is strictly worse off and student s_4 is strictly better off under $\mu(P')$ it follows that Propositions 3.1 and 3.2 do not appropriately generalize to college admissions with general quotas. ◇

Remark 1 In fact, using the many-to-one market in Example 2 one can construct a marriage market in which (unsuccessful) dropping strategies of two women make another woman strictly worse off and some man strictly better off.[12] This shows that Proposition 3.2 can also not be extended to the class of dropping strategies in one-to-one markets. For two reasons we do not provide further details and present Example 2 instead. First, the class of dropping strategies contains the strictly smaller class of truncation strategies, which is already exhaustive for one-to-one markets. Second, the market in Example 2 shows not only the impossibility of appropriately generalizing Proposition 3.2 but also the impossibility of generalizing Proposition 3.1. ◇

We note that Example 2 uncovers another difference between marriage markets and college admissions and adds to those already identified in Roth (1985a).

[11] Note that preferring {1, 4} to {2, 3} is compatible with responsiveness.

[12] We thank Bettina Klaus for pointing this out.

A positive result for the college admissions model with general quotas can be obtained through a change in what we mean by a "successful" manipulation. A manipulation $P'_{C'} = (P'_c)_{c \in C'}$ is *clearly weakly successful* if for all $c \in C'$ and all responsive extensions P^*_c of P_c, $\mu(P', c)\ R^*_c\ \mu(P, c)$.[13]

Proposition 4.1 *Under the student-proposing DA mechanism, any clearly weakly successful group manipulation by colleges is weakly beneficial to the other colleges (for all responsive extensions) and weakly harmful to all students.*

To prove Proposition 4.1, we need the following additional notation and result. Let $\bar{c} \in C$. A particular (additive) responsive extension of $P_{\bar{c}}$ can be obtained in the following way. Let $u : S \cup \{\emptyset\} \to \mathbb{R}$ be a function such that for all $s, t \in S \cup \{\emptyset\}$, $u(s) > u(t)$ if and only if $s\ P_{\bar{c}}\ t$. Define a function $\tilde{u} : \mathcal{P}(S, q_{\bar{c}}) \to \mathbb{R}$ by

$$\tilde{u}(S'') := \sum_{s \in \overline{S''}} u(s) \quad \text{for all } S'' \in \mathcal{P}(S, q_{\bar{c}}), \tag{1}$$

where $\overline{S''}$ is the multiset obtained from S'' by adding $q_{\bar{c}} - |S''|$ copies of $\emptyset$. Clearly, $\tilde{u}$ induces a weak ordering $P^{\tilde{u}}_{\bar{c}}$ of the elements in $\mathcal{P}(S, q_{\bar{c}})$ through

$$S'\ R^{\tilde{u}}_{\bar{c}}\ S'' \Leftrightarrow \tilde{u}(S') \geq \tilde{u}(S'') \quad \text{for all } S', S'' \in \mathcal{P}(S, q_{\bar{c}}). \tag{2}$$

One readily verifies the following result, the proof of which we have omitted.

Lemma 4.2 *$R^{\tilde{u}}_{\bar{c}}$ is a responsive extension of $R_{\bar{c}}$.*

Note that not all responsive extensions are additive.[14]

Proof of Proposition 4.1: Let C' be the group of colleges that manipulate. We will prove a stronger result. More precisely, we will only need that for all colleges $\bar{c} \in C'$ and all *additive* responsive extensions $R^*_{\bar{c}}$ of $R_{\bar{c}}$, $\mu(P', \bar{c})\ R^*_{\bar{c}}\ \mu(P, \bar{c})$.

The proof runs again similarly to that of Proposition 3.1. We only have to replace the line "If $\bar{c} \in C'$ then ... (according to its true preferences)." in Proposition 3.1 by the following.

Let $\bar{c} \in C'$. Let $T = \mu(P, \bar{c}) = \{t_1, \ldots, t_l\}$ and $T' = \mu(P', \bar{c}) = \{t'_1, \ldots, t'_r\}$. Expand the sets T and T' through the addition of $q_{\bar{c}} - |T|$ and $q_{\bar{c}} - |T'|$ copies of $\emptyset$ to obtain the multisets $\overline{T} = \{t_1, \ldots, t_l, t_{l+1}, \ldots, t_{q_{\bar{c}}}\}$ and $\overline{T'} = \{t'_1, \ldots, t'_r, t'_{r+1}, \ldots, t'_{q_{\bar{c}}}\}$, respectively. Without loss of generality, assume that

$$t_1\ R_{\bar{c}}\ t_2\ R_{\bar{c}} \cdots R_{\bar{c}}\ t_{q_{\bar{c}}-1}\ R_{\bar{c}}\ t_{q_{\bar{c}}} \quad \text{and}$$
$$t'_1\ R_{\bar{c}}\ t'_2\ R_{\bar{c}} \cdots R_{\bar{c}}\ t'_{q_{\bar{c}}-1}\ R_{\bar{c}}\ t'_{q_{\bar{c}}}.$$

Claim: For all $i = 1, \ldots, q_{\bar{c}}$, $t'_i\ R_{\bar{c}}\ t_i$.

[13] We thank the associate editor for providing this suggestion.

[14] For instance, if $P_{\bar{c}} = s_1 s_2 s_3 s_4 s_5 \emptyset$, then there is a responsive extension $P^*_{\bar{c}}$ with $s_1 s_3 s_5\ P^*_{\bar{c}}\ s_1 s_2 s_4\ P^*_{\bar{c}}\ s_2 s_4\ P^*_{\bar{c}}\ s_3 s_5$. One easily verifies that $P^*_{\bar{c}}$ is not additive, i.e., there is no $\tilde{u}$ such that $R^*_{\bar{c}} = R^{\tilde{u}}_{\bar{c}}$.

Proof of the Claim: Suppose the Claim is not true. Let n be the smallest index for which $t_n \, P_{\bar{c}} \, t'_n$. Then, for all $i = 1, \ldots, n-1$, $t'_i \, R_{\bar{c}} \, t_i$. Let $u : S \cup \{\emptyset\} \to \mathbb{R}$ be a function such that the following conditions are satisfied:

- for all $s, t \in S \cup \{\emptyset\}$, $u(s) > u(t)$ if and only if $s \, P_{\bar{c}} \, t$;
- $u(t_1), \ldots, u(t_n) \in [q_{\bar{c}} + 1, q_{\bar{c}} + 2]$;
- $u(t'_1), \ldots, u(t'_{n-1}) \in [q_{\bar{c}} + 1, q_{\bar{c}} + 2]$;
- $u(t_n) \leq 1$;
- for all $s \in S \cup \{\emptyset\}$, $u(s) \geq 0$.

For the function $\tilde{u}$ defined through (1), we find that

$$
\begin{aligned}
\tilde{u}(T) &= \sum_{t \in \overline{T}} u(t) \\
&\geq n \cdot (q_{\bar{c}} + 1) + (q_{\bar{c}} - n) \cdot 0 \\
&= n(q_{\bar{c}} + 1) \\
&> (n-1)(q_{\bar{c}} + 1) + q_{\bar{c}} \\
&= (n-1) \cdot (q_{\bar{c}} + 2) + (q_{\bar{c}} - (n-1)) \cdot 1 \\
&\geq \sum_{t' \in \overline{T'}} u(t') = \tilde{u}(T').
\end{aligned}
$$

Therefore, from (2), it follows that for the responsive extension $R_{\bar{c}}^{\tilde{u}}$ (Lemma 4.2) of $R_{\bar{c}}$, we have

$$\mu(P, \bar{c}) = T \; P_{\bar{c}}^{\tilde{u}} \; T' = \mu(P', \bar{c}),$$

contradicting the assumption that P' is a clearly weakly successful group manipulation by colleges $C' \ni \bar{c}$. □

Since s is rejected by $\bar{c}$ in DA(P) we have that (a) s is not acceptable for $\bar{c}$ at P, i.e., $\emptyset \, P_{\bar{c}} \, s$, or (b) for all $t \in \mu(P, \bar{c})$, $t \, P_{\bar{c}} \, s$ and $|\mu(P, \bar{c})| = q_{\bar{c}}$.

Suppose (a). Let $t' \in \mu(P', \bar{c})$. By the Claim, there is some $t \in \mu(P, \bar{c}) \cup \{\emptyset\}$ with $t' \, R_{\bar{c}} \, t$. Since $\mu(P)$ is an individually rational matching for P, $t \, R_{\bar{c}} \, \emptyset$. Hence, $t' \, R_{\bar{c}} \, \emptyset \, P_{\bar{c}} \, s$. So, $s \notin \mu(P', \bar{c})$.

Suppose (b). Since $\mu(P)$ is an individually rational matching for P and $R_{\bar{c}}$ is a linear order, for all $t \in \mu(P, \bar{c})$, $t \, P_{\bar{c}} \, \emptyset$. Then, from $|\mu(P, \bar{c})| = q_{\bar{c}}$ and the Claim it follows that $|\mu(P', \bar{c})| = q_{\bar{c}}$. Let $t' \in \mu(P', \bar{c})$. By the Claim and $|\mu(P, \bar{c})| = |\mu(P', \bar{c})| = q_{\bar{c}}$, there is some $t \in \mu(P, \bar{c})$ with $t' \, R_{\bar{c}} \, t \, P_{\bar{c}} \, s$. Therefore, $s \notin \mu(P', \bar{c})$.

Therefore, in both (a) and (b), $s \notin \mu(P', \bar{c})$. Hence, $\bar{c}$ will also have rejected s by the last step of DA(P'). □

Acknowledgments We thank an associate editor, a referee, Bettina Klaus, and Alexandru Nichifor for useful suggestions and detailed comments. The first draft of this article was written while F. Klijn was visiting Harvard Business School. He gratefully acknowledges a research fellowship from HBS for academic year 2009–2010 and support from Plan Nacional I+D+i (ECO2008–04784), the Barcelona Graduate School of Economics, the Government of Catalonia, and the Consolider-Ingenio 2010 (CSD2006–00016) program.

References

Crawford VP (1991) Comparative statics in matching markets. J Econ Theory 54(1):389–400

Dubins LE, Freedman DA (1981) Machiavelli and the Gale-Shapley algorithm. American Mathematical Monthly 88(7):485–494

Gale D, Shapley LS (1962) College admissions and the stability of marriage. Am Math Monthly 69(1):9–15

Knuth DE (1976) Mariages stables. Les Presses de l'Université de Montréal, Montréal

Kojima F, Pathak PA (2009) Incentives and stability in large two-sided matching markets. Am Econ Rev 99(3):608–627

Ma J (2010) The Singleton core in the college admissions problem and its application to the National Resident Matching Program (NRMP). Games Econ Behav 69(1):150–164

Mongell S, Roth AE (1991) Sorority rush as a two-sided matching mechanism. Am Econ Rev 81(3): 441–464

Roth AE (1982) The economics of matching: stability and incentives. Math Oper Res 7(4):617–628

Roth AE (1984a) Stability and polarization of interests in job matching. Econometrica 52(1):47–58

Roth AE (1984b) Mispresentation and stability in the marriage problem. J Econ Theory 34(2):383–387

Roth AE (1985a) The college admission problem is not equivalent to the marriage problem. J Econ Theory 36(2):277–288

Roth AE (1985b) Conflict and coincidence of interest in job matching: some new results and open questions. Math Oper Res 10(3):379–389

Roth AE (2008) Deferred acceptance algorithms: history, theory, practice, and open questions. Int J Game Theory 36(3):537–569

Roth AE, Sotomayor MAO (1989) The college admissions problem revisited. Econometrica 57(3):559–570

Roth AE, Sotomayor MAO (1990) Two-sided matching: a study in game-theoretic modeling and analysis, Econometric Society Monograph Series. Cambridge University Press, New York

Roth AE, Vande Vate JH (1991) Incentives in two-sided matching with random stable mechanisms. Econ Theory 1(1):31–44