Games and Economic Behavior 75 (2012) 685–693

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

School choice: Impossibilities for affirmative action ☆

Fuhito Kojima

Department of Economics, Stanford University, United States

ARTICLE INFO

Article history:
Received 21 January 2011
Available online 16 March 2012

JEL classification:
C70
D61

Keywords:
Matching
Stability
School choice
Affirmative action
Deferred acceptance
Top trading cycles

ABSTRACT

This paper investigates the welfare effects of affirmative action policies in school choice. We show that affirmative action policies can have perverse consequences. Specifically, we demonstrate that there are market situations in which affirmative action policies inevitably hurt every minority student – the purported beneficiaries – under any stable matching mechanism. Furthermore, we show that another famous mechanism, the top trading cycles mechanism, suffers from the same drawback.

1. Introduction

Racial desegregation is a long-standing problem in American society as well as many parts of the world. In the United States, affirmative action policies have been playing an important role in achieving this goal. Individuals in minority groups often receive preferential treatment in employment and school admission decisions.

Affirmative action policies have been widely used in public education although they have also received various criticisms.[1] Although recent Court decisions have prohibited the explicit use of "racial tie breakers" in public school admission to achieve racial integration,[2] indirect, or "color-blind," affirmative action policies are still regularly employed.[3]

We study affirmative action policies in the context of the school choice problem as analyzed by Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu (2005).[4] Specifically, we investigate the consequences of adopting affirmative action policies on student welfare. Our main finding is that an affirmative action policy can have a perverse effect on student

☆ I am grateful to Toshiji Kawagoe, Tayfun Sönmez, Satoru Takahashi, Yosuke Yasuda, and especially Yusuke Narita, Parag Pathak, Al Roth, and two anonymous referees for discussion and comments. I thank Neil Prasad and Pete Troyan for their excellent research assistances.

E-mail address: fuhitokojima1979@gmail.com.

[1] Criticisms include, among others, the claim that affirmative action is a reverse discrimination and doubts about its effectiveness as desegregation measure.

[2] In *Parents Involved in Community Schools v. Seattle School District No. 1*, the Supreme Court prohibited the use of race as admission criterion, with Chief Justice Roberts famously declaring "The way to stop discrimination on the basis of race is to stop discrimination on the basis of race."

[3] Districts across the country are still using criteria that are highly correlated with race such as socio-economic status. In Cambridge, for instance, free or reduced price lunch eligibility is used (see Cambridge Public Schools' website, http://www.cpsd.us/FRC/Reg_inst.cfm, for detail). In San Francisco, a neighborhood is used to "promote diversity" (Board of Education, San Francisco United School District, http://rpnorton.files.wordpress.com/2011/05/draft-ms-assignment-05312011.pdf).

[4] See Roth (1991) for a related model in the context of British labor market for doctors, and Ergin and Sönmez (2006) who study the Boston mechanism in school choice. See Roth and Sotomayor (1990), Roth (2008), and Sönmez and Ünver (2011) for more general surveys of the literature.

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2012.03.003

welfare, including minority students, who are the purported beneficiaries. More specifically, we establish impossibility theorems stating that there are situations where *affirmative action policies inevitably hurt every minority student* under any stable matching mechanism, such as those used in New York City and Boston. Moreover, this misfortune is unavoidable under alternative ways to implement the affirmative action policy: The impossibility results hold both when certain school seats are reserved for minority students (by imposing type-specific quotas on majority students) and when minority students are given preferential treatment in receiving priority in schools. Furthermore, similar impossibility results hold under another popular mechanism, the top trading cycles (TTC) mechanism. These findings suggest that caution should be exercised in the use of affirmative action policies even if helping minority students is deemed desirable by society.

On the other hand, we also find that imposing affirmative action does not necessarily hurt majority students. In fact, there are cases in which a stronger affirmative action policy makes everyone better off, *including every majority student*. This observation appears to be contrary to popular beliefs and suggests that caution may be needed when assessing the cost of affirmative action policies as well as its benefit.

The analytical approach of this paper follows the tradition of impossibility studies in the matching literature. Roth (1982) shows that there exists no stable and strategy-proof mechanism. Sönmez (1997, 1999) studies two forms of manipulations – manipulations via capacities and via pre-arranged matches, respectively – and shows that immunity to either of these manipulations is incompatible with stability. While the analysis of the current paper is new to our knowledge, the formulations of our impossibility theorems are inspired by these studies.

The rest of the paper proceeds as follows. Section 2 sets up the model. Section 3 introduces the affirmative action policies and the associated desiderata. Section 4 presents our main results. Section 5 concludes.

2. Model

A market is tuple $G = (S, C, (\succsim_i)_{i \in S \cup C}, (\mathbf{q}_c)_{c \in C})$. S and C are finite and disjoint sets of students and schools. For each student $s \in S$, $\succsim_s$ is a preference relation over C and being unmatched (being unmatched is denoted by $\emptyset$). We assume that preferences are strict. We write $c' \succ_s c$ if and only if $c' \succsim_s c$ but not $c \succsim_s c'$. For each school c, $\succsim_c$ is a priority order over the set of students. We assume that priority orders are strict. We write $s' \succ_c s$ if and only if $s' \succsim_c s$ but not $s \succsim_c s'$. If $c \succ_s \emptyset$, then c is said to be **acceptable** to s. The set of students are partitioned to two subsets; the set S^M of **majority students** and S^m of **minority students**.[5] For each $c \in C$, $\mathbf{q}_c = (q_c, q_c^M)$ is the **capacity** of c: The first component q_c represents the total capacity of school c, while the second component q_c^M represents the type-specific capacity for majority students.

A **matching** μ is a mapping from $C \cup S$ to $C \cup S \cup \{\emptyset\}$ such that

(1) $\mu(s) \in C \cup \{\emptyset\}$,
(2) For any $s \in S$ and $c \in C$, $\mu(s) = c$ if and only if $s \in \mu(c)$,
(3) $\mu(c) \subseteq S$ and $|\mu(c)| \leqslant q_c$ for all $c \in C$,
(4) $|\mu(c) \cap S^M| \leqslant q_c^M$ for all $c \in C$.

All conditions except for (4) are standard in the literature. Condition (4) requires that the number of majority students matched to each school c is at most its type-specific capacity q_c^M.

A matching μ is **stable** if

(1) $\mu(s) \succsim_s \emptyset$ for each $s \in S$, and
(2) if $c \succ_s \mu(s)$, then either
  (a) $|\mu(c)| = q_c$ and $s' \succ_c s$ for all $s' \in \mu(c)$, or
  (b) $s \in S^M$, $|\mu(c) \cap S^M| = q_c^M$, and $s' \succ_c s$ for all $s' \in \mu(c) \cap S^M$.

All conditions except for (2b) are standard. Condition (2b) describes a case in which a potential blocking is not realized because of a type-specific capacity constraint for the majority students: Student s wants to be matched with school c, but she is a majority student and the seats for majority students are filled by students who have higher priority than s at c.

This model is a special case of the "controlled school choice" analyzed by Abdulkadiroğlu and Sönmez (2003). They present a more general model, where there are an arbitrary finite number of student types and there is a type specific capacity for each type of students. In this paper we focus on a very simple situation in which there are only two types of students S^M and S^m. We chose our simple model for expositional simplicity only, and this modeling choice is without loss of generality: Since our results are impossibility theorems, all claims that hold in our simple environment hold in their more general model.

[5] Although we use words such as "majority" and "minority," we do not necessarily assume that there are more majority students than minority students. Such an assumption changes none of the results of this paper.