

# Lotteries in student assignment: An equivalence result

PARAG A. PATHAK

Department of Economics, Massachusetts Institute of Technology

JAY SETHURAMAN

Department of Industrial Engineering and Operations Research, Columbia University

This paper formally examines two competing methods of conducting a lottery to assign students to schools, motivated by the design of the centralized high school student assignment system in New York City. The main result of the paper is that single and multiple lottery mechanisms are equivalent for the problem of allocating students to schools in which students have strict preferences and the schools are indifferent. In proving this result, a new approach is introduced that simplifies and unifies all the known equivalence results in the house allocation literature. Along the way, two new mechanisms—Partitioned Random Priority and Partitioned Random Endowment—are introduced for the house allocation problem. These mechanisms generalize widely studied mechanisms for the house allocation problem and may be appropriate for the many-to-one setting such as the school choice problem.

KEYWORDS. Matching, random assignment.

JEL CLASSIFICATION. C78, D63.

## 1. INTRODUCTION

A large number of school districts around the world employ centralized mechanisms to assign students to schools.<sup>1</sup> The design of the current student assignment mechanism

---

Parag A. Pathak: [ppathak@mit.edu](mailto:ppathak@mit.edu)

Jay Sethuraman: [jay@ieor.columbia.edu](mailto:jay@ieor.columbia.edu)

This paper is a combination of working papers by Pathak (2008) and Sethuraman (2009). We are grateful to Debraj Ray for encouraging us to collaborate. We are grateful to various seminar participants and anonymous referees for comments. Pathak is indebted to Susan Athey, Drew Fudenberg, Alvin E. Roth, and Tayfun Sönmez. Sethuraman is grateful to Thiam Lee and Hervé Moulin for many insightful discussions during the course of this research: It was an early conversation with Thiam Lee that led to a formulation of the PRP mechanism and Hervé Moulin suggested thinking about the PRE mechanism. We thank Stergios Athanassoglou and Yeon-Koo Che for detailed comments on an earlier draft of the paper. Pathak's research was supported by the National Science Foundation under Grant SES-0924555 and Sethuraman's research was supported by the National Science Foundation under Grant CMMI-0916453.

<sup>1</sup>See Abdulkadiroğlu and Sönmez (2003) for descriptions of several systems in the United States, Balinski and Sönmez (1999) for a description of Turkish college admissions, Chiu and Weng (2009) for a description of the Chinese college admissions system, and Lavy (2010) for details on the public school choice plan in Tel Aviv, Israel. In England, centralized mechanisms are present in many urban areas (see, e.g., Coldron et al. 2008).

Copyright © 2011 Parag A. Pathak and Jay Sethuraman. Licensed under the [Creative Commons Attribution-NonCommercial License 3.0](http://creativecommons.org/licenses/by-nc/3.0/). Available at <http://econtheory.org>.

DOI: 10.3982/TE816

in New York City (NYC) is inspired by ideas from the mechanism design literature on the allocation of indivisible goods (e.g., Roth and Sotomayor 1990 and Abdulkadiroğlu and Sönmez 2003).<sup>2</sup> The mechanism in NYC consists of a main round, followed by a supplementary round that is meant to assign applicants who are unassigned after the main round. The assignment process in the main round takes several factors into account. These include student priorities at each school based on criteria such as geographic proximity and performance on standardized test scores, and has been studied as a two-sided matching problem (Abdulkadiroğlu et al. 2005a, 2009). In contrast, the assignment process in the supplementary round is quite simple: all unassigned applicants are invited to rank order high schools with unfilled seats; all students at this stage have the *same* priority to attend any school. The supplementary round involves over 8,000 students, a little more than 10% of applicants, who are unassigned after the main round.

The supplementary round can be modeled as a setting in which agents rank heterogeneous “objects” for which they have equal claims. This raises the following question: if many students rank the same school as their top choice, which of these students should be assigned to that school? A lottery mechanism is a way to achieve fairness in the assignment process. In a *single* lottery mechanism, a single random ordering of the agents is drawn; any ties (at *any* school) are broken in favor of the student whose lottery number is lower. A natural alternative is a *multiple* lottery mechanism, which allows each school to conduct its own lottery. The actual assignment is made by the core mechanism applied to the preference profiles of the students and the priority profiles of the schools (Abdulkadiroğlu and Sönmez 2003).

During the course of the design of the new assignment mechanism, policymakers from the NYC Department of Education (DOE) believed that a mechanism based on a single lottery is less equitable than lotteries at each school. After deciding on two rounds, they were still concerned with the question of how to resolve indifferences for each round. After community forums, one DOE official summarized the discussion:<sup>3</sup>

Although students might not get their first choices, they were considered separately for each program. There was a rank order established and each student had an equal chance to be selected. [...] If we want to give each child a shot at each program, the only way to accomplish this is to run a new random. [...] I cannot see how the children at the end of the line are not disenfranchised totally if only one run takes place. I believe that one line will not be acceptable to parents. When I answered questions about this at training sessions, (it did come up!) people reacted that the only fair approach was to do multiple runs.

This quotation motivates our comparison of single and multiple lotteries for student assignment problems. The main result, in Section 3, is a proof of the equivalence of the

<sup>2</sup>A large theoretical literature on school choice problems has developed following Abdulkadiroğlu and Sönmez (2003). See Abdulkadiroğlu et al. (2005a, 2005b, 2009), Erdil and Ergin (2008), Ergin and Sönmez (2006), Kesten (2010), and Pathak and Sönmez (2008).

<sup>3</sup>See Pathak (2007) for additional statements by DOE policymakers about the perceived fairness of multiple lotteries relative to single lotteries.

single and multiple lottery mechanisms in a setting where each school may have multiple seats and each student has strict preferences over schools. To prove this result, we assume that schools can admit only one student, but instead introduce a *new* mechanism that we call *Partitioned Random Priority* (PRP). Under the PRP mechanism, there is an arbitrary partition  $S_1, S_2, \dots, S_k$  of the schools: the schools within each  $S_i$  use a *common* lottery and distinct  $S_i$ 's use an independent lottery.<sup>4</sup> We show that the distribution of assignments under the PRP mechanism is the same, *regardless of the partition of the schools*.

We prove a similar equivalence result for another new mechanism that we introduce called *Partitioned Random Endowment* (PRE), which can be thought of as a generalization of both the single lottery mechanism and the Random Endowment (RE) mechanism. In the RE mechanism, each agent is endowed with a distinct house, each of the  $n!$  endowments being equally likely; the resulting allocation is the core of the associated economy, computing using the top trading cycles (TTC) procedure defined by Shapley and Scarf (1974). In PRE, there is a partition  $S_1, S_2, \dots, S_k$  of the schools; the students are partitioned—in a particular way that is described in Section 2.2—into sets  $N_1, N_2, \dots, N_k$  such that  $N_i$  has exactly  $|S_i|$  students, and every school in  $S_i$  has a common ranking of the students in  $N_i$ . Thus, if each school is in a partition by itself, we recover the RE mechanism; if all the schools form a single partition, we recover the single lottery mechanism. Our equivalence result thus subsumes the earlier equivalence result, due to Abdulkadiroğlu and Sönmez (1998) and Knuth (1996), of the RE and single lottery mechanisms.

Finally, in contrast to the proofs of earlier equivalence results, our proof technique is elementary, inductive, and relies on basic properties of the cycle structure of permutations. We illustrate the usefulness of this approach by providing, in Section 4, an analysis of the house allocation model with existing tenants, where some agents are endowed with a house. In this context, there is a natural generalization of the RE and RP mechanisms. Sönmez and Ünver (2005) illustrate the relationship between the two; our techniques allow us to provide a simple alternative proof of their result. We also establish a new equivalence result for this model.

## 2. THE MODEL

Let  $N$  denote the set of students and  $S$  denote the set of schools, and suppose  $|N| = |S| = n$ . Each student has a strict preference ordering of the schools and wishes to attend exactly one school. Each school wishes to admit exactly one student, but does not care which one it admits. Because the school preferences exhibit extreme indifference, we can view this problem as a *house allocation* problem in which the agents are students and the houses are schools. This analogy is possible only because the schools are treated as objects that do not care about their assignment. Not surprisingly, some of the mechanisms proposed for the school choice problem have been inspired by the house allocation literature (Abdulkadiroğlu and Sönmez 2003, Erdil and Ergin 2008, Pathak 2008, Abdulkadiroğlu et al. 2009).

<sup>4</sup>If each school is in a partition by itself, we recover the multiple lottery mechanism; if all the schools belong to a single partition, we recover the single lottery mechanism.

### 2.1 Existing mechanisms

An important ingredient in house allocation mechanisms is the precise manner in which *conflicts* between agents are resolved: if many agents rank the same house as their most preferred house, which one of them gets it? If houses had strict preferences (as would be the case in a *two-sided* matching problem), such conflicts can be resolved based on the preference ordering of the house in question. As houses do not have strict preferences, a natural idea is to generate a *priority list* of agents for each house that will be used to resolve conflicts. The various mechanisms proposed for the house allocation problem differ only in how these priority orderings are generated (equivalently, how such conflicts are resolved). We review the prominent mechanisms before describing the two new mechanisms we propose.

**RANDOM ENDOWMENT (RE).** Each agent is given top priority at a distinct house, with each of the  $n!$  possibilities being equally likely. For a given priority list, the resulting allocation is the (unique) core of the associated trading game in which the top-priority agent at any house is said to “own” that house. This allocation can be found using the TTC algorithm: construct a graph with one node for each agent; there is an arc from node  $i$  to node  $j$  if agent  $j$  owns agent  $i$ ’s most preferred house (note that  $j$  could be  $i$ , in which case there will be a self-loop). This graph must have a cycle as every node has out-degree 1 and the graph is finite. Each agent in the cycle is assigned the house owned by the agent he points to and all these nodes are deleted. The computation is repeated on the residual problem, which involves only the remaining agents and the remaining houses, each of which is necessarily owned by a distinct remaining agent.

**RANDOM PRIORITY (RP).** The agents are ordered randomly, with each of the  $n!$  orderings being equally likely. Every house uses this ordering as its priority list. For any given priority list, the resulting allocation is found as before by the TTC algorithm, with the modification that each house is owned by the agent who appears first in its priority list. Note that, with this new definition, an agent may own multiple houses at a time; moreover, a house owned by some agent may be owned by a different agent later on. Neither one of these aspects is present in the RE mechanism. Finally, this mechanism is equivalent to the more familiar description of the RP mechanism: the agents are ordered uniformly at random, make their choices according to this ordering, and pick their most preferred house that is still available when it is their turn.

**INDEPENDENT RANDOM PRIORITY (IRP).** Each house draws a random ordering of the agents, independently of the other houses, with each of the  $n!$  orderings being equally likely. Thus the priority profile of the houses is equally likely to be any one of the  $(n!)^n$  possibilities. For any given priority profile, the resulting allocation is found as before by the TTC algorithm.

The three mechanisms introduced so far—RE, RP, and IRP—all have very different descriptions. It is therefore remarkable that all of these mechanisms are equivalent. The RP and RE assignments are obtained by running the TTC algorithm on  $n!$  profiles.

That they find the same distribution over assignments was shown by [Abdulkadiroğlu and Sönmez \(1998\)](#) and [Knuth \(1996\)](#), independently. Both papers construct a bijective proof—with any given ordering of the agents (RP), they associate a distinct endowment (RE) such that the resulting final assignment is identical. In contrast, the IRP assignment is found by running the TTC algorithm on  $(n!)^n$  profiles. It is also possible to prove the equivalence of RP and IRP using the following approach: associate with any given ordering of the students a *set* of  $(n!)^{n-1}$  priority profiles of schools such that the resulting final assignment is identical, and such that the sets of priority profiles associated with distinct student orderings are disjoint (see [Pathak 2008](#) for these details).

## 2.2 New mechanisms

The mechanisms discussed so far were all for one-to-one allocation problems: each agent wishes to be assigned a house and each house can be assigned to exactly one agent. The assignment problems arising in school choice, however, are typically many-to-one: a school can admit many students, whereas each student still wishes to be assigned to one school. We can, of course, formulate an equivalent one-to-one version of the problem: if a school can admit 13 students, we can make 13 copies of that school and have each student rank all the copies of this school in an arbitrary (but fixed) way. There is no difficulty in extending the RP mechanism. But the IRP and the RE mechanisms are somewhat unnatural in the equivalent one-to-one formulation. Ideally, we would like the IRP mechanism to allow the priority list of each *school* to be generated independently. In the equivalent one-to-one formulation, IRP allows each “seat” to generate its own priority list, so that different copies of the same school may have different priority lists. A similar problem plagues the RE mechanism. To remedy this difficulty, we introduce the following two new mechanisms for the school choice problem. (To emphasize the application, we switch to students and schools instead of agents and houses.)

**PARTITIONED RANDOM PRIORITY (PRP).** Fix an arbitrary partition of the schools  $S$  into sets  $S_1, S_2, \dots, S_k$ . School priority lists are determined in the following way: For each  $i$ , the schools in  $S_i$  have identical priority lists, which are generated independently and uniformly. Thus, each of the  $(n!)^k$  priority profiles is equally likely. The distribution of assignments under the PRP mechanism can be found by running the TTC algorithm on each of the  $(n!)^k$  possible priority profiles. When each school is in a partition by itself, we recover the IRP mechanism. When all the schools are in one partition, we recover the RP mechanism.

**PARTITIONED RANDOM ENDOWMENT (PRE).** Fix an arbitrary partition of the schools  $S$  into sets  $S_1, S_2, \dots, S_k$ . School priority lists are determined in the following way: the students are arranged uniformly at random, all  $n!$  orderings being equally likely. From this random ordering, the first  $|S_1|$  students form the top  $|S_1|$  choices of each of the schools in  $S_1$ ; the next  $|S_2|$  students form the top  $|S_2|$  choices of each of the schools in  $S_2$ , etc. Again, the distribution of assignments under the PRE mechanism can be found by running the TTC algorithm on each of the  $n!$  possible priority profiles. When each school is in a partition by itself, we recover the RE mechanism. When all the schools are in one partition, we recover the RP mechanism.

Our main result (in [Section 3](#)) is that, for both mechanisms, the distribution of assignments is the same for *any partition* of the schools. In particular, the IRP and PRE mechanisms are equivalent to the RP mechanism. Thus our main result generalizes the earlier results of [Abdulkadiroğlu and Sönmez \(1998\)](#) and [Knuth \(1996\)](#).

### 3. THE EQUIVALENCE RESULT

Let  $S_1, S_2, \dots, S_k$  be a partition of the schools. What does it mean to generate a random preference profile with respect to this partition? We start with the PRP mechanism. Imagine  $n$  cards—one for each student—arranged in a sequence next to each school. The cards are all face down, and the sequences are identical for all the schools in each  $S_i$ . Consider now the PRE mechanism. As in the PRP mechanism, all the schools within a given  $S_i$  have an identical priority list. The difference is that in the PRE mechanism, the schools in  $S_i$  have an identical ranking of  $|S_i|$  students and these students are not in the “priority” list of any of the other schools. That is, the students are partitioned so that exactly  $|S_i|$  students appear in the priority list of the schools in  $S_i$  and they are ranked identically by the schools in  $S_i$ , with each of the  $|S_i|!$  rankings being equally likely.

When can the TTC algorithm find a given cycle  $C = (i_1, i_2, \dots, i_m)$ ? The TTC algorithm can find  $C$  only if, for every  $1 \leq \ell \leq m$ , student  $i_{\ell+1}$  has the top priority at student  $i_\ell$ 's most preferred school (where  $i_{m+1}$  is to be read as  $i_1$ ). In particular, the  $m$  students in  $C$  should all have distinct most preferred schools. Suppose the TTC algorithm finds cycle  $C$  in the given priority profile. We can delete the students in  $C$  along with their most preferred schools. We can also remove the cards bearing the name of any student in  $C$  from the remaining  $n - m$  schools, as these students have no further role to play in the TTC algorithm. We call the instance thus obtained the *residual instance* after eliminating cycle  $C$ . Let  $S'_1, S'_2, \dots, S'_k$  be the partition of the remaining  $n - m$  schools where  $S'_i$  contains those schools in  $S_i$  that are still in the residual instance. (Note that each  $S_i$  can lose at most one school. Also, if some  $S'_j$  is empty, it can be deleted, so the number of nonempty classes in the partition for the residual instance may be strictly smaller than  $k$ .) The critical observation is that—for both PRP and PRE mechanisms—this residual instance is *equally likely to be any of the possible profiles* with the partition  $S'_1, S'_2, \dots, S'_k$ . This follows from the observation that eliminating cycle  $C$  does not convey any information about the priority lists of any of the remaining schools with respect to the remaining students.

It is clear that *cycles* play an important role in both mechanisms. Moreover, the cycles that *could* be found as a first cycle are the same for both PRP and PRE, so the discussion that follows applies to both mechanisms. To make these notions precise, we introduce some notation. Recall that  $S_1, S_2, \dots, S_k$  is the given partition of the schools. For  $i = 1, 2, \dots, k$ , let  $N_i$  be the set of students whose most preferred school is in  $S_i$ . From the earlier discussion, any cycle that can be found as a first cycle can have at most one student from each  $N_i$ . Let  $\mathcal{C}$  be the collection of all such cycles. We use  $C$  to denote a generic member of  $\mathcal{C}$  and let  $\overline{C}$  be the *set* of students involved in the cycle  $C$ . Note that it is possible for  $C, C' \in \mathcal{C}$ ,  $C \neq C'$ , and  $\overline{C} = \overline{C'}$ . Given two cycles  $C, C' \in \mathcal{C}$ , we say that  $C$  and  $C'$  *conflict* if for some  $i$ ,  $N_i \cap \overline{C}$  and  $N_i \cap \overline{C'}$  are both nonempty. For any  $\ell \geq 1$ ,  $\mathcal{C}^\ell$

consists of products of  $\ell$  cycles of  $\mathcal{C}$  such that no pair of cycles conflicts. Note that  $\mathcal{C}^1$  is just  $\mathcal{C}$ , and  $\mathcal{C}^\ell$  consists of products of  $\ell$  distinct cycles of  $\mathcal{C}$ , all of which can be simultaneously present in some priority profile (and each of which can potentially be uncovered as a first cycle). We illustrate these definitions in the following example.

EXAMPLE 1. Consider the following instance with 5 students and 5 schools, each with one seat. The strict student preferences are

| 1        | 2        | 3        | 4        | 5        |
|----------|----------|----------|----------|----------|
| <i>a</i> | <i>a</i> | <i>b</i> | <i>d</i> | <i>c</i> |
| <i>b</i> | <i>d</i> | <i>a</i> | <i>b</i> | <i>e</i> |
| <i>c</i> | <i>c</i> | <i>d</i> | <i>a</i> | <i>a</i> |
| <i>d</i> | <i>b</i> | <i>c</i> | <i>c</i> | <i>b</i> |
| <i>e</i> | <i>e</i> | <i>e</i> | <i>e</i> | <i>d</i> |

Suppose the given partition of the schools is  $S_1 = \{a, d\}$ ,  $S_2 = \{b, e\}$ , and  $S_3 = \{c\}$ . Then  $N_1 = \{1, 2, 4\}$ ,  $N_2 = \{3\}$ , and  $N_3 = \{5\}$ . In this case

$$\mathcal{C} = \{(1), (2), (3), (4), (5), (13), (23), (43), (15), (25), (45), (35), (135), (153), (235), (253), (435), (453)\}.$$

Note that cycle (3) can be uncovered only if  $b$  has 3 as its top-priority student; cycle (25) can be uncovered only if  $c$  has 2 as its top-priority student and  $a$  has 5 as its top-priority student; finally (435) can be uncovered only if schools  $d$ ,  $b$ , and  $c$  have students 3, 5, and 4 as their top-priority students, respectively. We distinguish the cycle (453) from (435) because (453) can be uncovered only if schools  $d$ ,  $b$ , and  $c$  have students 5, 3, and 4 as their top-priority choices, respectively. (Thus the top-priority students of  $d$  and  $b$  are switched in these two cycles.) We turn now to  $\mathcal{C}^2$ . By definition,  $\mathcal{C}^2$  consists of products of two distinct cycles of  $\mathcal{C}$  that can be simultaneously present in a priority profile. Thus

$$\mathcal{C}^2 = \{(1)(3), (1)(5), (2)(3), (2)(5), (4)(3), (4)(5), (3)(5), (13)(5), (23)(5), (43)(5), (15)(3), (25)(3), (45)(3), (35)(1), (35)(2), (35)(4)\}.$$

Note that the product of cycles (1)(4) is *not* in  $\mathcal{C}^2$  even though the individual cycles (1) and (4) are both in  $\mathcal{C}$ . This is because it is not possible for both agents 1 and 4 to be part of a cycle in the first iteration (as their most preferred schools both belong to  $S_1$ ). Similarly, it is easy to check that

$$\mathcal{C}^3 = \{(1)(3)(5), (2)(3)(5), (4)(3)(5)\}$$

and  $\mathcal{C}^\ell = \emptyset$  for  $\ell > 3$ . ◇

Before stating the main result, we prove the following lemma on the cycle structure of permutations. This is a well known result, but we present a proof for the sake of completeness.

LEMMA 1. *Let  $c(m, \ell)$  be the number of permutations of  $m$  elements that can be written as a product of  $\ell$  cycles.<sup>5</sup> Then, for any  $m > 1$ ,*

$$\sum_{\ell=1}^m (-1)^{\ell-1} c(m, \ell) = 0.$$

PROOF. For any  $m > 1$ ,

$$c(m, \ell) = c(m-1, \ell-1) + (m-1)c(m-1, \ell). \quad (1)$$

The left-hand side is, by definition, the number of permutations of  $m$  elements that can be written as a product of  $\ell$  cycles. The right-hand side counts the same set of permutations in a different way: either the element  $m$  is in a cycle by itself, in which case the rest of the elements have to be written as a product of  $\ell-1$  cycles, or the element  $m$  is in a cycle with some other element(s), in which case the remaining elements have to be written as a product of  $\ell$  cycles, and element  $m$  can be inserted into this product in  $(m-1)$  ways. We use the recurrence relation (1) to prove the lemma by induction.

It is clear that  $c(2, 1) = c(2, 2) = 1$ , so the lemma is true when  $m = 2$ . For any  $m > 2$ , using the recurrence relation (1), we have

$$\sum_{\ell=1}^m (-1)^{\ell-1} c(m, \ell) = \sum_{\ell=1}^m (-1)^{\ell-1} c(m-1, \ell-1) + \sum_{\ell=1}^m (-1)^{\ell-1} (m-1)c(m-1, \ell).$$

Noting that  $c(m-1, 0) = c(m-1, m) = 0$ , we find that the first of these terms is the alternating series

$$-c(m-1, 1) + c(m-1, 2) - c(m-1, 3) + \cdots + (-1)^{m-1} c(m-1, m-1),$$

whereas the second is the alternating series

$$(m-1)(c(m-1, 1) - c(m-1, 2) + c(m-1, 3) - \cdots + (-1)^{m-2} c(m-1, m-1)).$$

Adding these two terms, we get

$$(m-2) \sum_{\ell=1}^{m-1} (-1)^{\ell-1} c(m-1, \ell),$$

which is zero by the induction hypothesis.  $\square$

We are now ready to state the main result.

THEOREM 2. *The distribution of assignments determined by the RP mechanism is identical to the distribution of assignments under the PRP mechanism, for any partition of the schools.*

<sup>5</sup>The numbers  $c(m, \ell)$  are known as the unsigned Stirling numbers of the first kind; see Stanley (1997) for further details.

PROOF. The result is clearly true when there is one student and one school. Consider an instance with  $n$  students and  $n$  schools, and suppose the result holds for all smaller problems. Fix a partition  $S_1, S_2, \dots, S_k$  of the schools.

Let  $\text{RP}(X)$  denote the outcome of the RP mechanism when the agents in  $N \setminus X$  are removed along with their most preferred schools. Since each agent is equally likely to be the first agent in the uniform random ordering chosen by the RP mechanism, we can conclude that

$$\text{RP}(N) = \sum_{i=1}^n \frac{1}{n} \{Y_i + \text{RP}(N \setminus \{i\})\}, \quad (2)$$

where  $Y_i$  indicates the assignment of student  $i$  to his most preferred school.

Now we turn to the PRP mechanism. As discussed earlier, the PRP mechanism works by finding a cycle (that is necessarily in  $\mathcal{C}$  by definition). Let  $P(C)$  be the probability that cycle  $C$  can be uncovered (i.e., cycle  $C$  is present in a priority profile that is drawn uniformly at random from all possible priority profiles) and let  $Y_{\bar{C}}$  indicate the assignment in which each student in  $\bar{C}$  is allocated his most preferred school. Thus, analogous to (2), we may be tempted to write

$$\text{PRP}(N) = \sum_{C \in \mathcal{C}} P(C) \{Y_{\bar{C}} + \text{PRP}(N \setminus \bar{C})\}.$$

This, however, is incorrect: it is possible for cycles  $C$  and  $C'$  both to be present in a given profile, in which case that profile is counted once for  $C$  and once for  $C'$  in the expression above. The profiles that are overcounted are precisely those in which a cycle in  $\mathcal{C}^2$  is present,<sup>6</sup> but subtracting these would undercount profiles that contain three distinct cycles from  $\mathcal{C}$ : these are counted thrice, once for each cycle, but subtracted thrice as well, once for each pair of cycles, so we need to add back the profiles that contain a cycle in  $\mathcal{C}^3$ . Taking this argument to its logical conclusion, we find

$$\text{PRP}(N) = \sum_{\ell=1}^k \sum_{C \in \mathcal{C}^\ell} (-1)^{\ell-1} P(C) \{Y_{\bar{C}} + \text{PRP}(N \setminus \bar{C})\}, \quad (3)$$

where  $Y_{\bar{C}}$  and  $\mathcal{C}^\ell$  are as defined earlier.

Consider any subset  $\bar{X}$  of students and let  $|\bar{X}| = m$ . Suppose there is a cycle  $C$  in  $\mathcal{C}^\ell$  such that  $\bar{C} = \bar{X}$ . Observe that  $P(C) = 1/n^m$ —for any agent in  $\bar{X}$ , the cycle  $C$  assigns his most preferred school to a particular agent (also in  $\bar{X}$ ) and this occurs with probability  $1/n$ , independently, for each of the  $m$  agents in  $\bar{X}$ . If, alternatively, there is no cycle in  $\mathcal{C}^\ell$  involving the agents in  $\bar{X}$ , we can take  $P(\bar{X}) = 0$ . In either case, we can write  $P(\bar{X})$  instead of  $P(C)$ , as this probability depends only on the number of agents in  $C$ . Furthermore, the term inside the braces in the right-hand side of (3) depends only on  $\bar{X}$ , not on

<sup>6</sup>To avoid cumbersome notation, we use  $C$  to denote a generic element in  $\mathcal{C}^\ell$  for any  $\ell \geq 1$ . Any element  $C \in \mathcal{C}^2$  is a *product* of two cycles, not a simple cycle.

the actual cycle involving the agents in  $\bar{X}$ . Using these observations, we can rewrite (3) as

$$\text{PRP}(N) = \sum_{\bar{X}: \bar{X} \subseteq N} (Y_{\bar{X}} + \text{PRP}(N \setminus \bar{X})) P(\bar{X}) \left\{ \sum_{\ell=1}^k \sum_{\substack{C \in \mathcal{C}^\ell \\ \bar{C} = \bar{X}}} (-1)^{\ell-1} \right\}. \quad (4)$$

We now focus on the term inside the braces in (4). As  $\bar{X}$  has only  $m$  students, it is clear that for any  $\ell > m$ , no cycle  $C \in \mathcal{C}^\ell$  can have  $\bar{C} = \bar{X}$ . This means that the upper limit of the first summation in that term can be changed from  $k$  to  $m$ . Therefore, we have

$$\sum_{\ell=1}^k \sum_{\substack{C \in \mathcal{C}^\ell \\ \bar{C} = \bar{X}}} (-1)^{\ell-1} = \sum_{\ell=1}^m \sum_{\substack{C \in \mathcal{C}^\ell \\ \bar{C} = \bar{X}}} (-1)^{\ell-1} = \sum_{\ell=1}^m (-1)^{\ell-1} c(m, \ell) = \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{if } m > 1, \end{cases}$$

where the last expression follows from the definition of  $c(m, \ell)$  and [Lemma 1](#). Thus the only terms that survive in the right-hand side of (4) are those for which  $|\bar{X}| = 1$ . In this case, however,  $P(\bar{X}) = 1/n$  and (4) simplifies to

$$\text{PRP}(N) = \sum_{i=1}^n \frac{1}{n} \{Y_i + \text{PRP}(N \setminus \{i\})\}.$$

By the induction hypothesis,  $\text{PRP}(N \setminus \{i\}) = \text{RP}(N \setminus \{i\})$  regardless of the partition of the schools (in the problem without student  $i$ ). We see that the expressions that determine  $\text{PRP}(N)$  and  $\text{RP}(N)$  are identical, establishing the result.  $\square$

An important consequence of [Theorem 2](#) is an equivalence result for school choice in which schools can admit more than one student.

**THEOREM 3.** *Consider a school choice problem with  $n$  students and  $k$  schools. Suppose school  $i$  can admit  $q_i$  students, and suppose  $\sum_i q_i = n$ . The distribution of assignments obtained by the TTC algorithm is the same, whether a common lottery is used for all the schools or whether each school uses its own independent lottery.*

**PROOF.** Make  $q_i$  copies of school  $i$  and replace all occurrences of school  $i$  in the student preference lists by an arbitrary, but consistent ordering of all its copies; this results in an equivalent one-to-one matching problem. Place all the copies of a given school into one partition, so that the partition structure of the “schools” for the PRP mechanism is precisely  $S_1, S_2, \dots, S_k$ , where  $S_i$  consists of all copies of school  $i$ . The RP mechanism finds the outcome when a common ordering is used by all the schools; the PRP mechanism finds the outcome in which each school uses an independent ordering of the students. The result now follows from [Theorem 2](#).  $\square$

We state and prove a parallel result for the PRE mechanism.

**THEOREM 4.** *The distribution of assignments determined by the RP mechanism is identical to the distribution of assignments under the PRE mechanism, for any partition of the schools.*

**PROOF.** The proof is identical to the proof of [Theorem 2](#), the only change being the expression for  $P(C)$ . For the PRP mechanism, we saw that  $P(C) = 1/n^m$  whenever  $C$  is a cycle involving  $m$  students; for the PRE mechanism, however,

$$P(C) = \frac{1}{n} \frac{1}{n-1} \cdots \frac{1}{n-m+1}.$$

This is so because the most preferred schools of the students involved in the cycle  $C$  all are required to have distinct top-priority students. The exact expression for  $P(C)$  does not matter whenever  $C$  has more than one student because of the cancellation of terms (by [Lemma 1](#)). When  $C$  has exactly one student or, equivalently, when  $m = 1$ ,  $P(C) = 1/n$  as in the proof of [Theorem 2](#) and the result follows.  $\square$

This theorem has a number of implications and extensions. First, [Theorem 4](#) provides an alternative proof of the equivalence of RP and RE mechanisms: this alternative proof relies on induction and counting instead of the original bijective argument due to [Abdulkadiroğlu and Sönmez \(1998\)](#) and [Knuth \(1996\)](#). A similar comment applies to the alternative equivalence proof of RP and IRP, implicit in [Theorem 2](#).

Moreover, the discussion preceding [Lemma 1](#) suggests a simple equivalence proof of the RE and IRP mechanisms. Recall that in the IRP mechanism, each school is in a partition by itself and generates its priority list uniformly at random, independently of the other schools. As the sets  $C^\ell$  are identical for both RE and IRP for any  $\ell \geq 1$ , we may conclude that the expressions—like (3)—for the RE and IRP mechanisms are identical, making [Lemma 1](#) unnecessary. Note, however, that [Lemma 1](#) is still needed: although RE and IRP are very similar,  $P(C)$  is different for these two mechanisms, so the recursive expressions are different.

Furthermore, we defined the PRP and PRE mechanisms with respect to a fixed partition of the schools. However, the equivalence proof of [Theorem 4](#) applies more broadly. Consider a model in which there is an initial partition of the schools and the top-priority student for each partition is drawn as in the PRP (or PRE) mechanism. The outcome of the TTC algorithm in this step determines the partition of the *remaining* schools in the next step, at which point the top priority student for each partition is determined again. Thus, the partition at stage  $t$  of the algorithm could depend on the entire history of allocations determined in stages  $1, 2, \dots, t-1$ . This and closely related models are discussed in [Pápai \(2000\)](#), [Chen \(2007\)](#), and [Svensson and Larsen \(2005\)](#). [Theorem 4](#) establishes the equivalence of all of these mechanisms to the RP mechanism.

Finally, consider the following model for generating priority lists for the schools. Suppose we are given  $(N_1, N_2, \dots, N_n)$  with each  $N_i \geq 1$ . Consider a box in which there are  $N_i$  balls of color  $i$ . A ball is drawn at random from this box, and if the drawn ball is of color  $i$ , then  $i$  gets top priority, all balls of color  $i$  are removed from the box, and the procedure is repeated. Note that the probability of agent  $i$  having a higher priority than

agent  $j$  is  $N_i/(N_i + N_j)$ , regardless of which other agents are present. (If the  $N_i$  are identically 1, we get the uniform distribution.) It is a simple matter to check that if all priority lists are drawn from this distribution, then the uniform lottery mechanism is equivalent to the multiple lottery mechanism.

#### 4. HOUSE ALLOCATION WITH EXISTING TENANTS

These ideas can be used to obtain a simpler proof of another equivalence result, due to [Sönmez and Ünver \(2005\)](#), on house allocation with existing tenants. In an instance of this problem there are  $n$  agents and  $n$  houses;  $n - r$  of the agents are existing tenants, each of whom already occupies a distinct house; the remaining  $r$  agents and the  $r$  unoccupied houses constitute the new arrivals to the problem. There is a natural generalization of the RP and RE mechanisms to this setting. First we describe the RP mechanism: order the agents uniformly at random and let agents make their choices in this order, except that whenever an agent  $j$  demands a house already occupied by an agent  $j'$ , then  $j'$  gets promoted to just before  $j$  in the ordering. In this case,  $j'$  is invited to make his choice and  $j$  picks only after  $j'$  is assigned a house (that may possibly be the one that he already occupies). For the RE mechanism: endow the *new* houses uniformly at random to the *new* agents, each existing tenant is endowed the house he occupies, and apply the TTC algorithm to this trading game. It is not hard to see that, as defined, the RE mechanism is *not* equivalent to the RP mechanism. This is because the new agents are treated preferentially in the RE mechanism: if  $j'$  is an existing tenant who happens to occupy a house that is the last choice for all the agents (including himself!),  $j'$  cannot be assigned any other house under the RE mechanism, but he could get other houses under RP. [Sönmez and Ünver \(2005\)](#) modify the RP mechanism in the following way: the new agents are ordered uniformly at random and the existing tenants appear *after* the new agents in an *arbitrary* order.<sup>7</sup> They show that this modified RP mechanism is equivalent to the RE mechanism. We prove this result next. (To emphasize the distinction in context, we use the terms “agents” and “houses” in this section instead of students and schools.)

**THEOREM 5.** *For the house allocation mechanism with existing tenants, the RE mechanism (which assigns the new houses randomly to the new agents) is equivalent to the RP mechanism in which all the new agents are ordered uniformly at random before any of the existing tenants.*

**PROOF.** We prove the result by induction on the number of existing tenants. If there are no existing tenants, the result follows from [Theorem 4](#).

Suppose there are  $\ell$  existing tenants. We may assume that initially there is no cycle involving the existing tenants alone: that is, there is no set of existing tenants  $i_1, i_2, \dots, i_m$  such that the most preferred house of agent  $i_j$  is the one occupied by  $i_{j+1}$  (where  $i_{m+1}$  is read as  $i_1$ ). If there is one, observe that the RE and RP mechanisms will *always* assign

<sup>7</sup>We could, of course, look at uniform random orderings of the existing tenants as well, but as it turns out their exact ordering does not matter as long as they appear *after* the new applicants. This will be apparent from the proof.

these houses to these agents such that each agent in this set gets their most preferred house; we can then delete these agents and their respective houses, and the result follows by induction (as we have an instance with fewer existing tenants).

Suppose there are  $n$  agents and  $n$  houses in all, and suppose there are  $r$  new applicants, who, without loss of generality, we assume to be agents  $1, 2, \dots, r$ . Then, in the RP mechanism, each of the  $r$  new agents is equally likely to be the first. Suppose (new) agent  $i$  is first. As we assume that there is no cycle involving the existing tenants alone, observe that agent  $i$  will necessarily receive his first choice—either  $i$ 's first choice is an unoccupied house or  $i$ 's first choice is an occupied house in which case the owner of that house will ask for his best choice; this latter house is either unoccupied (in which case the chain stops) or is occupied by another existing tenant, who will now ask for *his* best choice, etc. By our assumption, this chain will have to end with an existing tenant asking for an unoccupied house. The agents involved in this chain, including agent  $i$ , will all get their most preferred house. Let  $C_i$  be the set of agents involved in the chain started by agent  $i$ . Note that  $C_i$  contains  $i$  and does not contain any other new applicant. The relevant recursion for the RP mechanism is

$$\text{RP}(N) = \sum_{i=1}^r \frac{1}{r} \{Y_{C_i} + \text{RP}(N \setminus \{C_i\})\},$$

where  $Y_{C_i}$  indicates the assignment in which the agents in  $C_i$  get their most preferred houses.

Let  $\mathcal{C}$  be the set of cycles that can be uncovered as a first cycle by the RE mechanism. As before, we think of cycles as involving just the agents as it is understood that the houses involved are all the most preferred houses of the agents in the cycle. Let  $\mathcal{C}^\ell$  be defined analogously as well, for  $\ell = 1, 2, \dots$ , with  $\mathcal{C}^1 = \mathcal{C}$ . Finally, note that the existing tenants are always endowed with the houses they occupy. Clearly,

$$\text{RE}(N) = \sum_{\ell=1}^r \sum_{C \in \mathcal{C}^\ell} (-1)^{\ell-1} P(C) \{Y_{\overline{C}} + \text{RE}(N \setminus \overline{C})\}, \quad (5)$$

where  $Y_{\overline{C}}$  and  $\mathcal{C}^\ell$  are as defined before.

Consider an arbitrary set  $\overline{X} \subseteq N$  of agents. Suppose there is a cycle  $C$  in  $\mathcal{C}^\ell$  such that  $\overline{C} = \overline{X}$ . As the endowment of the occupied houses is fixed, it is clear that for any agent  $i \in C$ , if  $i$ 's most preferred house is an occupied house, then  $i$  *must* be followed by the existing tenant occupying that house. So the only agents in  $C$  for whom a successor is not already determined are those whose most preferred house is *new*. Suppose there are exactly  $m$  of these agents in  $C$ . Then observe that

$$P(C) = \frac{1}{r} \frac{1}{r-1} \cdots \frac{1}{r-m+1}.$$

This is because the new houses that each of these agents desire must each be endowed to a distinct new applicant—there are  $r$  choices for the first new applicant,  $r-1$  for the second, and so on. As before, if there is no cycle in  $\mathcal{C}^\ell$  involving the agents in  $\overline{X}$ , we can

take  $P(\bar{X}) = 0$ . In either case, we can write  $P(\bar{X})$  instead of  $P(C)$ , as the probability does not depend on the cycle, but only on the set of agents involved in the cycle (in fact, it depends only on  $m$  and  $r$ ). Furthermore, the term inside the braces in the right-hand side of (5) depends only on  $\bar{X}$ , not on the cycle  $X$  itself. Using these observations, we can rewrite (5) as

$$\text{RE}(N) = \sum_{\bar{X}: \bar{X} \subseteq N} (Y_{\bar{X}} + \text{RE}(N \setminus \bar{X})) P(\bar{X}) \left\{ \sum_{\ell=1}^r \sum_{\substack{C \in \mathcal{C}^\ell \\ \bar{C} = \bar{X}}} (-1)^{\ell-1} \right\}. \quad (6)$$

We now focus on the term inside the braces in (6). As  $|\bar{X}|$  has only  $m$  agents whose most preferred house is new and as each cycle in  $\mathcal{C}$  has at least one of these agents, it is clear that for any  $\ell > m$ , no cycle  $C \in \mathcal{C}^\ell$  can have  $\bar{C} = \bar{X}$ . This means that the upper limit of the second summation can be changed from  $r$  to  $m$ . Therefore, we have

$$\sum_{\ell=1}^r \sum_{\substack{C \in \mathcal{C}^\ell \\ \bar{C} = \bar{X}}} (-1)^{\ell-1} = \sum_{\ell=1}^m \sum_{\substack{C \in \mathcal{C}^\ell \\ \bar{C} = \bar{X}}} (-1)^{\ell-1} = \sum_{\ell=1}^m (-1)^{\ell-1} c(m, \ell) = \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{if } m > 1, \end{cases}$$

where the last expression follows from the definition of  $c(m, \ell)$  and [Lemma 1](#). Thus the only terms that survive in the right-hand side of (6) are those for which there is exactly one agent whose most preferred house is new. This also implies that there is exactly one new applicant in the cycle—as the new houses are all endowed to the new applicants, the number of new houses in any cycle should equal the number of new applicants in the cycle. In this case, however,  $P(C) = 1/r$ , and (5) simplifies to

$$\text{RE}(N) = \sum_{i=1}^r \frac{1}{r} \{Y_{C_i} + \text{RE}(N \setminus \{C_i\})\}.$$

By the induction hypothesis,  $\text{RE}(N \setminus \{C_i\}) = \text{RP}(N \setminus \{C_i\})$ . We see that the expressions that determine  $\text{RE}(N)$  and  $\text{RP}(N)$  are identical, establishing the result.  $\square$

The equivalence of RP and IRP mechanisms can be generalized to this setting as well. The RP mechanism was defined earlier in this section, but the following equivalent definition will be convenient: first a common priority ordering of the agents is drawn for all houses. The priority ordering for each new house is this common priority ordering; for each of the occupied houses, the existing tenant is moved to the top of the priority ordering, leaving the relative ordering of the others the same. The TTC algorithm applied to the given preference profile along with this (random) priority profile finds the same assignment as the one found by the RP mechanism described by [Abdulkadiroğlu and Sönmez \(1998\)](#). The PRP mechanism works in a similar manner, except that priorities are drawn uniformly and independently for each house: for each new house, the priority orderings are drawn uniformly at random from the set of all permutations of all agents; for each occupied house, the owner of that house appears first in the ordering, and a random priority ordering of the remaining agents is drawn uniformly. The proof of the

equivalence of these two mechanisms is very similar to that of [Theorem 2](#) and is omitted. As in the statement of that theorem, the results extend to situations in which houses are partitioned arbitrarily, with the caveat that whenever multiple occupied houses belong to the same block in the partition, the owners of these occupied houses are given top priority at the houses they occupy (the relative ordering of the other agents will be the same for all the houses in the partition).

More generally, the philosophical discussion on multiple versus single randomization and the perceptions of policymakers that motivate the analysis here may also be relevant in other distinct contexts, where randomization is applied for fairness reasons. In the model analyzed here, students do not know their lottery number before submitting their preferences, but they are modeled as knowing their preferences or type. It may be possible to examine models where a randomization determines a student's type following a Harsanyi veil of ignorance thought experiment. From a welfare perspective, this lottery over student identities may not be equivalent to a lottery over outcomes, a point made by [Grant et al. \(2010\)](#). Another issue raised from this case study is that even though these two mechanisms are equivalent, policymakers may prefer to conduct multiple lotteries because they are perceived as being fairer on procedural grounds. These issues are likely an important constraint in practice, and developing models of how mechanisms are perceived is an important area for future work.

#### REFERENCES

- Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2005a), "The New York City high school match." *American Economic Review (Papers and Proceedings)*, 95, 364–367. [2]
- Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2009), "Strategy-proofness versus efficiency in matching with indifference: Redesigning the NYC high school match." *American Economic Review*, 99, 1954–1978. [2, 3]
- Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2005b), "The Boston public school match." *American Economic Review (Papers and Proceedings)*, 95, 368–371. [2]
- Abdulkadiroğlu, Atila and Tayfun Sönmez (1998), "Random serial dictatorship and the core from random endowments in house allocation problems." *Econometrica*, 66, 689–701. [3, 5, 6, 11, 14]
- Abdulkadiroğlu, Atila and Tayfun Sönmez (2003), "School choice: A mechanism design approach." *American Economic Review*, 93, 729–747. [1, 2, 3]
- Balinski, Michel and Tayfun Sönmez (1999), "A tale of two mechanisms: Student placement." *Journal of Economic Theory*, 84, 73–94. [1]
- Chen, Bo (2007), "The random endowment and inheritance rules for the house allocation problem." Unpublished paper, Department of Economics, University of Wisconsin. [11]

Chiu, Y. Stephen and Weiwei Weng (2009), “Endogenous preferential treatment in centralized admissions.” *RAND Journal of Economics*, 40, 258–282. [1]

Coldron, John, Emily Tanner, Steven Finch, Lucy Shipton, Claire Wolstenholme, Ben Willis, Sean Demack, and Bernadette Stiell (2008), “Secondary school admissions.” Research Report DCSF-RR020, U.K. Department for Children, Schools and Families. [1]

Erdil, Aytek and Haluk Ergin (2008), “What’s the matter with tie-breaking? Improving efficiency in school choice.” *American Economic Review*, 98, 669–689. [2, 3]

Ergin, Haluk and Tayfun Sönmez (2006), “Games of school choice under the Boston mechanism.” *Journal of Public Economics*, 90, 215–237. [2]

Grant, Simon, Atsushi Kajii, Ben Polak, and Zvi Safra (2010), “Generalized utilitarianism and Harsanyi’s impartial observer theorem.” *Econometrica*, 78, 1939–1971. [15]

Kesten, Onur (2010), “School choice with consent.” *Quarterly Journal of Economics*, 125, 1297–1348. [2]

Knuth, Donald E. (1996), “An exact analysis of stable allocation.” *Journal of Algorithms*, 20, 431–442. [3, 5, 6, 11]

Lavy, Victor (2010), “Effects of free choice among public schools.” *Review of Economic Studies*, 77, 1164–1191. [1]

Pápai, Szilvia (2000), “Strategyproof assignment by hierarchical exchange.” *Econometrica*, 68, 1403–1433. [11]

Pathak, Parag A. (2007), *Essays on Real-Life Allocation Problems*. Ph.D. thesis, Harvard University. [2]

Pathak, Parag A. (2008), “Lotteries in student assignment: The equivalence of queueing and a market-based approach.” Unpublished paper, Department of Economics, MIT. [1, 3, 5]

Pathak, Parag A. and Tayfun Sönmez (2008), “Leveling the playing field: Sincere and sophisticated players in the Boston mechanism.” *American Economic Review*, 98, 1636–1652. [2]

Roth, Alvin E. and Marilda A. Sotomayor (1990), *Two-Sided Matching*. Cambridge University Press, Cambridge. [2]

Sethuraman, Jay (2009), “An equivalence result in school choice.” Unpublished paper, Department of Industrial Engineering and Operations Research, Columbia University. [1]

Shapley, Lloyd S. and Herbert E. Scarf (1974), “On cores and indivisibility.” *Journal of Mathematical Economics*, 1, 23–37. [3]

Sönmez, Tayfun and M. Utku Ünver (2005), “House allocation with existing tenants: An equivalence.” *Games and Economic Behavior*, 52, 153–185. [3, 12]

Stanley, Richard P. (1997), *Enumerative Combinatorics*, Volume 1. Cambridge University Press, Cambridge. [8]

Svensson, Lars-Gunnar and Bo Larsen (2005), “Strategy-proofness, core, and sequential trade.” *Review of Economic Design*, 9, 167–190. [11]

---

Submitted 2009-10-2. Final version accepted 2010-9-15. Available online 2010-3-1.