IMPLEMENTING RANDOM ASSIGNMENTS: A GENERALIZATION OF THE BIRKHOFF-VON NEUMANN THEOREM

ERIC BUDISH, YEON-KOO CHE, FUHITO KOJIMA, AND PAUL MILGROM

ABSTRACT. The Birkhoff-von Neumann Theorem shows that any bistochastic matrix can be written as a convex combination of permutation matrices. In particular, in a setting where n objects must be assigned to n agents, one object per agent, any random assignment matrix can be resolved into a deterministic assignment in accordance with the specified probability matrix. We generalize the theorem to accommodate a complex set of constraints encountered in many real-life market design problems. Specifically, the theorem can be extended to any environment in which the set of constraints can be partitioned into two hierarchies. Further, we show that this bihierarchy structure constitutes a maximal domain for the theorem, and we provide a constructive algorithm for implementing a random assignment under bihierarchical constraints. We provide several applications, including (i) single-unit random assignment, such as school choice; (ii) multi-unit random assignment, such as course allocation and fair division; and (iii) two-sided matching problems, such as the scheduling of inter-league sports matchups. The same method also finds applications beyond economics, generalizing previous results on the minimize makespan problem in the computer science literature.

KEYWORDS: Birkhoff-von Neumann Theorem, Market Design, Random Assignment, Probabilistic Serial, Utility Guarantee, Makespan, Maximal Domain, Fair Allocation, Santa Claus Problem, Optimal Assignment, Assignment Auction.

1. INTRODUCTION

Suppose a social planner wishes to allocate a set of indivisible objects amongst a set of agents. A natural method is to use some sort of an auction, but in a wide variety of allocation problems the use of monetary transfers is either impractical, undesirable, or illegal. Examples range from the allocation of slots in public schools, to the assignment of

Budish: Department of Economics, Harvard University and Harvard Business School, Boston, MA 02163, ebudish@hbs.edu. Che: Department of Economics, Columbia University, 420 West 118th Street, 1016 IAB New York, NY 10027, yeonkooche@gmail.com. Kojima: Cowles Foundation, Yale University, 30 Hillhouse Avenue, New Haven, CT 06510, fuhitokojima1979@gmail.com. Milgrom: Department of Economics, Stanford University, CA 94305, paul@milgrom.net. We are grateful to Hervé Moulin and Al Roth for helpful comments. We especially thank Jay Sethuraman who pointed out close connections between our analysis and integer programming literature.

tasks within an organization, to the allocation of course seats or dormitory rooms at universities, to the allocation of organs amongst patients needing transplants. Randomness plays a key role in the design of allocation procedures in such settings, both in theory and practice. The reason is that, with indivisible objects and no possibility of transfering utility amongst agents using money, ex-post assignments can be very unfair. Randomness can help to "even things out".

This paper answers a fundamental question in non-transferable utility (NTU) market design: *what random assignments can be implemented (and how)*. To motivate the conceptual difficulty, consider a random assignment in which three agents $\{1, 2, 3\}$ are assigned to three objects $\{a, b, c\}$ according to the following matrix:

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$

where entry P_{ia} represents the probability that agent i is assigned to object a. If we resolve each agent's lottery independently, we might allocate some object to two agents while some other object goes unallocated, violating feasibility. Instead, the random assignment should be resolved by choosing the assignments $(1 \to a, 2 \to b, 3 \to c)$ and $(1 \to b, 2 \to c, 3 \to a)$ with equal probabilities. Note for example that the event of agent 3 receiving object c is perfectly correlated with the event of agent 1 receiving object a. The precise method for correlating one agent's allocation to another's is not always obvious from the random assignment itself, and whether a method can be found to implement *any* arbitrary random assignment is unclear. The celebrated Birkhoff-von Neumann theorem (Birkhoff (1946); von Neumann (1953)) provides an answer for this simple setting. It shows that any bistochastic matrix[1] can be expressed as a convex combination of permutation matrices, i.e., bistochastic matrices each containing only zeros or ones as entries. This ensures that any random assignment can be implemented in the single-unit assignment setting, in which the number of agents equals the number of objects, and agents have unit demand. The theorem has proved useful in two important NTU market designs, namely the pseudo-market mechanism by Hylland and Zeckhauser (1979) and more recently the probabilistic serial mechanism by Bogomolnaia and Moulin (2001).

In real-world assignment problems, however, there are many features and constraints that are not allowed for in the Birkhoff-von Neumann theorem. (Indeed, despite their

[1]A square matrix is bistochastic if (i) all entries are between 0 and 1 inclusive, and (ii) all rows and columns sum to one.

theoretical appeal, we are not aware of either the pseudo-market mechanism or the probabilistic serial mechanism ever being used in practice). Two useful generalizations are straightforward to handle. First, in many allocation problems it is acceptable for some agents or objects to remain unassigned; for instance, in the (public) school-choice problem, some students may exit the system, opting for a private school. Second, some agents might consume multiple units; for instance, in the course-allocation problem, students seek a schedule consisting of multiple courses. The first generalization can be accomodated easily by introducing a "null" object; the second is only a bit more complicated.

Other constraints are more challenging to deal with. One difficult problem occurs when different groups of agents must be treated differently in the assignment. For instance, schools may wish to seek balance between the genders, or may require that the number of students of a certain racial, ethnic, geographic or income group may not exceed and/or fall short of some target quota. Another difficult problem occurs when the supply of objects is not exogenously fixed but rather produced endogenously according to some technological constraint. This feature arises when a public school authority wishes to install multiple school programs in one building and the relative sizes of these programs can be changed, but the total number of students in these programs is constrained by the building size. In multi-unit assignment problems such as course allocation, agents might have constraints on the kinds of sets of objects they are able to consume. For instance, a student might have a preference to take no more than a certain number of courses in a particular subject matter, or be required to take at least a certain number.

These are just a few well-known examples of constraints that are not readily accommodated by the Birkhoff-von Neumann Theorem. There are many other similar examples. It is thus important to understand to what extent the theorem generalizes beyond the classical single-unit assignment setting.

We generalize the Birkhoff-von Neumann theorem in two directions. First, we allow for the number of agents to differ from the number of objects, and for the supply of objects and agents' demands for them to take any integer amount, positive or negative. (A negative integer is interpreted as supply by that agent of the object in question. Accordingly, the notion of a random assignment is generalized to allow for any real (positive or negative) values for the entries of the associated matrix.

The second and non-trivial direction is in terms of the kinds of constraints we accommodate. In principle, it is possible to define constraints on any subset of entries in the matrix. The Birkhoff-von Neumann theorem allows for just two kinds of constraints: constraints on whole rows of the matrix (the number of objects an agent consumes), and constraints on whole columns of the matrix (the number of times an object is allocated). We allow for constraints to be placed on arbitrary subsets of the matrix, and we allow for both floor and ceiling constraints, not necessarily equal to each other as in the Birkhoff-von Neumann theorem. What we require is that the set of constraint sets can be partitioned into two hierarchies; that is, for any pair of sets in the same hierarchy, either they are disjoint or one set is a subset of the other. Under this **bihierarchy** structure, we show that any generalized random assignment matrix can be expressed as a convex combination of integer-valued matrices satisfying all the constraints on the sets in the two hierarchies. This means that as long as the subsets of entries that are subject to quota constraints form a bihierarchical structure and the quota ceilings/floors are integer-valued, any random assignment satisfying these constraints can be implemented by conducting a lottery over non-random assignments each of which satisfies all the constraints.

The above extension is the most general statement to our knowledge, and accommodates all the kinds of real-world constraints we discussed above. Moreover, we provide a constructive algorithm for implementing random assignments under bihierarchical constraints, and we show a maximal domain result which indicates that, subject to some technical conditions, ours is the most general statement possible. That is, if the desired constraints do not form a bihierarchy, then there exists a random assignment that satisfies all the constraints, but for which it is impossible to implement that random assignment by a lottery over outcomes each of which satisfies the constraints. In that sense, the bihierarchical structure implicit in the Birkhoff-von Neumann theorem was not only sufficient but necessary for the development of the random assignment mechanisms described above.

We hope that our generalization of the Birkhoff-von Neummann theorem will find a wide variety of applications in NTU mechanism design (and that the maximal domain result will facilitate the understanding of what kinds of random mechanisms are *not* possible). In this paper we illustrate two distinct ways to use the hierarchical structure to yield new design possibilities.

First, we use our method to extend a specific random-assignment mechanism, that of Bogomolnaia and Moulin (2001), to more realistic settings. Many indivisible resources such as school slots, housing, offices, etc., are commonly allocated by a procedure called the random serial dictatorship (RSD). Serial dictatorship is a deterministic mechanism in which agents choose objects one at a time according to an exogenously specified priority order (i.e., by serial number). Random serial dictatorship is the same, except the priority order is chosen randomly; that is, RSD is essentially a deterministic mechanism except that which agent occupies which role in the society is randomized. In an influential recent paper, Bogomolnaia and Moulin (2001) propose a mechanism called probabilistic serial (PS) that directly produces a random assignment in response to agents' reports of their preferences. PS satisfies a stronger notion of efficiency than does RSD, and otherwise the two mechanisms' properties are similar.[2] Yet, we do not observe the use of PS in practice. One possibility for why is that PS is not able to handle complex features of real-life scenarios like the ones mentioned above. RSD is more flexible; for instance, RSD easily accommodates quota constraints on groups of agents.

We extend the probabilistic serial mechanism in three steps. First, we show how to adapt the Bogomolnaia and Moulin (2001) "eating" algorithm to handle bihierarchical constraints. Second, we use our main theorem to show that the resulting random assignment can be implemented. And, finally, we show that the resulting random allocating continues to have the desirable efficiency properties of the original.

In our second application, we show how our method can be used to implement *any* random assignment in a way that guarantees that agents' utilities are always "close" – specifically, within at most the value of a single object – to the (expected) utility associated with the random assignment. This result is of course vacuous in the context of single-unit assignment. But it can be quite powerful in the context of multi-unit resource allocation problems such as course allocation, task assignment, and the fair division of estates.

One attractive method for solving such problems is to treat objects as "divisible" and solve for an optimal fractional (random) assignment. The difficulty is that there are many ways to resolve a given random assignment, some of which could entail an outcome quite different from the original random assignment. For instance, suppose in the context of course allocation that under an attractive divisible-goods allocation procedure some student receives a fractional assignment in which he has a one-half probability of obtaining each of twenty courses, ten of which are "good" and ten of which are "bad". The only constraint is that he receive ten courses overall. One way to resolve this lottery would be to give the agent a one-half chance of obtaining all ten good courses and a one-half chance of obtaining all ten bad courses. This resolution exposes the agent to substantial risk. What we would do is supplement the actual constraints of the problem with a hierarchical set of artificial quota constraints in a way that bounds the extent to which each agent's utility can vary over different resolutions of the (artificially constrained) random assignment. In the context of this simple example, we provide this "utility guarantee" by adding

[2]One difference is that RSD is strictly strategyproof in any size market, whereas PS is strategyproof only in an ordinal sense; namely, no agent under that mechanism can do strictly better by lying in the sense of first-order stochastic dominance (Bogomolnaia and Moulin 2001). In large finite markets PS becomes strictly strategyproof.(Kojima and Manea 2008).

constraints on the number of good courses the agent receives (here, floor = ceiling = 5) and the number of bad courses the agent receives. We add different constraints for different agents in a way that depends on their preferences, and use the generalized BvN theorem to ensure that the resulting artificially constrained random assignment can be implemented.

Our framework lends itself to extension to the two-sided matching setting in which both sides of the market are agents. This can be done by interpreting the object side as another set of agents. The utility guarantee can then be readily applied to a two-sided matching problem: starting with any random matching, we can find an exact matching that always gives the agents on both sides realized utilities that are similar to those they expect from the random matching. This method can be used to design a fair schedule of inter-league matchups in sports scheduling, or a fair speed-dating mechanism.

2. Setup

Let $\mathbb{Z}$ be the set of integers. We consider tuples $\langle N, O, \mathcal{H}\rangle$ where:

- N is the set of agents where $|N| \geq 2$,
- O is the set of objects where $|O| \geq 2$,
- $\mathcal{H} \subset 2^{N\times O}$ is a set of subsets from $N \times O$ that forms two hierarchies, that is, there exist $\mathcal{H}_N$ and $\mathcal{H}_O$ such that
  - $\mathcal{H} = \mathcal{H}_N \cup \mathcal{H}_O$ and $\mathcal{H}_N \cap \mathcal{H}_O = \emptyset$, that is, $\mathcal{H}_N$ and $\mathcal{H}_O$ partition $\mathcal{H}$,
  - for each $S \in \mathcal{H}$, $S \subseteq N \times O$,
  - if $S, S' \in \mathcal{H}_N$, then $S \subset S'$ or $S' \subset S$ or $S \cap S' = \emptyset$.
  - if $S, S' \in \mathcal{H}_O$, then $S \subset S'$ or $S' \subset S$ or $S \cap S' = \emptyset$.

  We say $\mathcal{H}$ is **bihierarchical** if $\mathcal{H}$ satisfies these conditions, and the partitions $\mathcal{H}_N$ and $\mathcal{H}_O$ are hierarchical. In many applications $\mathcal{H}_N$ and $\mathcal{H}_O$ include, respectively, sets of the form $\{i\} \times O$ and $N \times \{a\}$ where $i \in N$, $a \in O$. These sets represent constraints imposed on each agent and object, thus the mnemonic notation $\mathcal{H}_N$ and $\mathcal{H}_O$. However, at this point we do not impose such a restriction, and we will state the restriction whenever applicable.

We also consider input of the form $\langle N, O, \mathcal{H}, (\underline{q}_S, \overline{q}_S)_{S\in\mathcal{H}}\rangle$ where $(\underline{q}_S, \overline{q}_S)_{S\in\mathcal{H}}$ is the set of quota constraints for each set in the hierarchies. We call $\underline{q}_S$ the **floor constraint** and $\overline{q}_S$ the **ceiling constraint** for S. For each $S \in \mathcal{H}$, we assume $\underline{q}_S$ and $\overline{q}_S$ are integers except that we allow $\underline{q}_S = -\infty$ and $\overline{q}_S = +\infty$. We simply write $q = (\underline{q}_S, \overline{q}_S)_{S\in\mathcal{H}}$ when there is no confusion. The tuple $\mathcal{E} = \langle N, O, \mathcal{H}, q\rangle$ is called an **environment**.

A (generalized) **random assignment** is a $|N| \times |O|$ matrix, $\mathbf{P} = [P_{ia}]$ where $P_{ia} \in (-\infty, \infty)$ for all $i \in N, a \in O$. A **deterministic assignment** is a random assignment $\mathbf{P}$ whose entries are integers. Note that we allow for assigning more than one unit of a good and even for assigning a negative amount of a good. Receiving a negative amount of a good corresponds to supplying that good in that amount.

Given environment $\mathcal{E} = \langle N, O, \mathcal{H}, q \rangle$, P is said to be **feasible** if

$$\underline{q}_S \leq P_S \leq \overline{q}_S, \text{ for all } \quad S \in \mathcal{H}, \tag{1}$$

where we define

$$P_S := \sum_{(i,a) \in S} P_{ia},$$

for any random assignment P and $S \in \mathcal{H}$. We denote by $\mathcal{P}_{\mathcal{E}}$ the set of random assignments that are feasible in $\mathcal{E}$.

2.1. **Examples.**

2.1.1. *The Simple Environment.* Consider the environment $\langle N, O, \mathcal{H}, q \rangle$ where

$$\begin{aligned}
&\mathcal{H}_N = \{\{(i,a)\} | (i,a) \in N \times O\} \cup \{\{i\} \times O | i \in N\}, \\
&\mathcal{H}_O = \{N \times \{a\} | a \in O\}, \\
&\underline{q}_{\{(i,a)\}} = 0, \overline{q}_{\{(i,a)\}} = 1, \quad \text{ for all } (i,a) \in N \times O, \\
&\underline{q}_S = \overline{q}_S = 1, \quad \text{ for all } S \in \mathcal{H} \setminus \{\{(i,a)\} | (i,a) \in N \times O\}.
\end{aligned}$$

This is an environment in which each agent receives exactly one object and each object is allocated to exactly one agent (note that $|N| = |O|$ is implied), and no other constraints are imposed. We denote this environment by $\mathcal{E}_{BvN}$, where the subscript stands for Birkhoff and von Neumann, who studied this environment extensively and showed Corollary 1 presented below.

A random assignment feasible in $\mathcal{E}_{BvN}$ is called a **bistochastic matrix** (alt., doubly stochastic matrix). Formally, P is a bistochastic matrix if

(1) $P_{ia} \geq 0$ for all $i \in N$ and $a \in O$,
(2) $P_{\{i\} \times O} = 1$ for all $i \in N$, and
(3) $P_{N \times \{a\}} = 1$ for all $a \in O$.

2.1.2. *Flexible Capacity.* The current framework can accommodate situations where objects can be produced in a flexible manner subject to certain constraints. An organization may wish to divide its workforce or resources among a few projects based on the skill set and preferences of its employees. In the context of school choice, different types of objects may represent different programs, say within a school specializing in different disciplines and activities, only a subset of which may be chosen depending on the preferences of incoming students. The multi-program design problem in school choice mentioned in the introduction also fits in this framework; a public school authority chooses the relative sizes of several education programs subject to the constraint that the total number of enrollments in those programs does not exceed a certain exogenous quota. Such a situation can be represented by a hierarchy $\mathcal{H}_O$ containing sets of the form $S = N \times O'$ with $|O'| \geq 2$. The ceiling $\overline{q}_S$ then describes total capacities that can be allocated within O'. Further, the hierarchical structure means that flexible production can be nested; e.g., a subset of programs may be chosen and, within each chosen program, a subset of subprograms may be chosen, and so on.

2.1.3. *Group-specific Quotas.* The mechanism designer may need to treat different groups of agents differently in assignment. For example, affirmative action or a desire by a school for diversity may entail the use of quotas for applicants of different ethnic, racial or economic profiles. Such a practice is called "controlled choice" and is used in many school districts in the United States. Some other forms of constraints are mathematically similar. For example, a subset of schools in New York City (the so-called Educational Option programs) require balanced distributions of test score: Namely, 16 percent of the seats should be allocated to students who were rated top performers in a standardized English Language Arts exam, 68 percent to middle performers, and 16 percent to lower performers (Abdulkadiroğlu, Pathak, and Roth 2005).[3] Quotas may be based on the residence of applicants as well: The school choice program set to begin in 2010 in Seoul, Korea, limits the percentage of seats allocated to the applicants from outside the district to 20 percent,[4] and a number of school choice programs in Japan have similar quotas based on residential areas as well.

All such constraints can be easily incorporated by $\mathcal{H}_O$ containing sets of the form $N' \times \{a\}$ for $a \in O$ and $N' \subsetneq N$. The quota $\overline{q}_{N' \times \{a\}}$ then determines the maximum

[3]An exact implementation can be slightly different from the quotas analyzed by most of the literature cited above. See Kojima (2008a) for further discussion on this point.

[4]See "Students' High School Choice in Seoul Outlined," Digital Chosun Ilbo, October 16, 2008 (http://english.chosun.com/w21data/html/news/200810/200810160016.html).

number of agents school a can admit from group N'. We can accommodate multiple groups for each a as long as they do not overlap with each other. Note that we can also accommodate hierarchies of constraints: for instance, a school system can require that a school admit at most 50 students from district one, at most 50 students from district two, and at most 80 students from either district one or two.

2.1.4. *Course Allocation with Complex Constraints.* Consider the course allocation problem. The school administrator wants to allocate seats in courses to students. Each course has its own quota. Each student can enroll in more than one course but she does not wish to receive more than one seat in any course for obvious reasons. Moreover, she may have constraints imposed by the system, or based on her own preferences, such that she cannot (or does not want to) take more than a certain number of courses from a subset of the courses. For example, a student might not be allowed to take two courses that meet during the same time slot. Or, a student might prefer to take at most two courses on finance, at most three courses on marketing, and at most four courses on finance or marketing in total.

Such restrictions can be modeled in our framework. Assume that $\mathcal{H}_N$ contains sets of the form $\{(i,a)\}$ with $i \in N, a \in O$ and $\{i\} \times O$, $i \in N$. By setting $\overline{q}_{\{(i,a)\}} = 1$ and $\overline{q}_{\{i\}\times O} > 1$, we can assure that student i can obtain up to $\overline{q}_{\{i\}\times O} > 1$ courses but for each course a, he will get at most one seat. Letting F and M be finance courses and marketing courses, if $\mathcal{H}_N$ contains $\{i\} \times F$, $\{i\} \times M$ and $\{i\} \times F \cup M$, then we can express the constraints "student i can take at most $\overline{q}_{\{i\}\times F}$ courses in finance, $\overline{q}_{\{i\}\times M}$ courses in marketing, and $\overline{q}_{\{i\}\times(F\cup M)}$ in finance and marketing combined." Scheduling constraints are handled similarly.[5]

2.1.5. *Incompatibility of Goods and Agents.* Suppose that the social planner cannot (or does not want to) assign object a to agent i. Such a restriction appears in school choice, for example. Some students may not be eligible to apply to some schools since the student has not fulfilled academic requirements or because the school is not allowed to admit students from the district in which that student lives, for example. Such restrictions also arise in the context of organ transplantation.[6] In that context, N represents patients and O represents organs, and some organs cannot be transplanted into some agents because

[5]While very flexible, there are some limitations to the kinds of constraints that can be accommodated without violating bihierarchy. The course-allocation procedure proposed in Budish (2008) accommodates arbitrary constraints. See Section 5 for further discussion.

[6]For economic analysis related to organ donation, see Roth, Sönmez, and Ünver (2004) and references therein.

of incompatibility of blood type and other biological traits. In a different context, if N represents machines and O represents jobs to be processed, then certain jobs may not be compatible to certain machines due to operating system differences or other technical reasons. Such a constraint can be incorporated by letting $\mathcal{H}$ contain $\{(i,a)\}$ and $\underline{q}_{\{(i,a)\}} = \overline{q}_{\{(i,a)\}} = 0$.

2.1.6. *Interleague Play Design.* Some professional sports, most notably Major League Baseball (MLB) and the National Football League (NFL), have two separate leagues. In MLB, teams in the American League (AL) and National League (NL) had traditionally played against teams only within their own league during the regular season, but play across the AL and NL, called interleague play, was introduced in 1997.[7] Unlike the intraleague games, the number of interleague games is relatively small, and this can make the indivisibility problem particularly difficult to deal with in designing the matchups. For example, suppose there are two leagues, N and O, each with 9 teams. Suppose each team must play 15 games against teams in the other league. There are some matchup constraints: Each team in N has a geographic rival in O, and they must play twice. For fairness reasons, teams in each league must face opponents in the other league of similar difficulty. Specifically, one could order the teams of each league in (the descending order of) winning percentage, say $N = \{i_1, ..., i_9\}$ and $O = \{a_1, ..., a_9\}$, and then require each team to play at least 4 games with top 3 teams, 4 games with middle 3 teams and 4 games with bottom 3 teams of the other league. It is not difficult to see that the resulting constraint sets form a bihierarchy.

3. The Rounding Theorem

Theorem 1 (Rounding Theorem). *Given any $\langle N, O, \mathcal{H}\rangle$ and P, if $\mathcal{H}$ is a bihierarchy, then there exist $P^1, P^2, \ldots, P^K$ and $\lambda^1, \lambda^2, \ldots, \lambda^K$ such that*

(1) $P = \sum_{k=1}^{K} \lambda^k P^k$,
(2) $\lambda^k \in (0,1]$ *for all* k *and* $\sum_{k=1}^{K} \lambda^k = 1$,
(3) $P^k_S \in \{\lfloor P_S \rfloor, \lceil P_S \rceil\}$ *for all* $k \in \{1, \ldots, K\}$ *and* $S \in \mathcal{H}$.[8]

The Theorem shows that any random assignment P can be decomposed into matrices where the sum of the entries within each element of the bihierarchy is rounded up or down to the nearest integer. In practice we want resolution of uncertainty so that every

[7]See "Interleague play", Wikipedia (http://en.wikipedia.org/wiki/Interleagueplay).

[8]For any $x \in \mathbb{R}$, $\lfloor x \rfloor$ and $\lceil x \rceil$ are the largest integer no larger than x and the smallest integer no smaller than x, respectively.

assignment of every single pair is integer valued, and this can be achieved for free by adding singleton sets to one of the hierarchies. The proof of Theorem 1 is in the appendix. We also provide in the appendix a constructive algorithm for implementing the theorem.

Corollary 1 (Rounding Theorem for Stochastic Assignments)**.** *Consider an environment* $\langle N, O, \mathcal{H}, (\underline{q}_S, \overline{q}_S)_{S\in\mathcal{H}}\rangle$*. Assume* $\mathcal{H}$ *is a bihierarchy and* P *is a feasible random assignment, so* $\underline{q}_S \leq P_S \leq \overline{q}_S$ *for all* $S \in \mathcal{H}$*. Then there exist* $P^1, P^2, \ldots, P^K$ *and* $\lambda^1, \lambda^2, \ldots, \lambda^K$ *such that*

(1) $\underline{q}_S \leq P^k_S \leq \overline{q}_S$ *for all* k, $\quad S \in \mathcal{H}$,
(2) $P = \sum_{k=1}^{K} \lambda^k P^k$,
(3) $\lambda^k \in (0,1]$ *for all* k *and* $\sum_{k=1}^{K} \lambda^k = 1$.

Corollary 1 is a generalization of the well-known Birkhoff-von Neumann Theorem. Recall that a feasible random assignment matrix in $\mathcal{E}_{BvN}$ is called a bistochastic matrix and that a permutation matrix is a bistochastic matrix in which every entry is zero or one.

Corollary 2 (Birkhoff (1946); von Neumann (1953))**.** *Suppose that* P *is a bistochastic matrix. Then* P *can be written as a convex combination of permutation matrices.*

Proof. Consider the pair of bihierarchy $\mathcal{H} = \mathcal{H}_N \cup \mathcal{H}_O$ and constraints $(\underline{q}_S, \overline{q}_S)_{S\in\mathcal{H}}$ defined as

$$\mathcal{H}_N = \{\{(i,a)\} | (i,a) \in N \times O\} \cup \{\{i\} \times O | i \in N\},$$
$$\mathcal{H}_O = \{N \times \{a\} | a \in O\},$$
$$\underline{q}_{\{(i,a)\}} = 0, \overline{q}_{\{(i,a)\}} = 1, \quad \text{for all } (i,a) \in N \times O,$$
$$\underline{q}_S = \overline{q}_S = 1, \quad \text{for all } S \in \mathcal{H} \setminus \{\{(i,a)\} | (i,a) \in N \times O\}.$$

Applying Corollary 1 to this environment, we obtain the conclusion of Corollary 2. □

Moreover, it is easy to see that existing generalizations of the Birkhoff-von Neumann Theorem (Corollary 2) known in the mathematics literature, such as Watkins and Merris (1974), Lewandowski, Liu, and Liu (1986) and de Werra (1984), are special cases of Corollary 1.

It turns out that Theorem 1 is essentially the most general proposition possible. More precisely, we present the following "necessity" or "maximal domain" result.

Theorem 2 (Maximal Domain)**.** *Fix* $N, O, \mathcal{H}$ *and suppose that* $\mathcal{H}$ *contains all the sets of the form* $\{i\} \times O$ *and* $N \times \{a\}$*. If* $\mathcal{H}$ *is not bihierarchical, then there exists* P *such that there exists no set of scalars and matrices which satisfy conditions (1)–(3) of Theorem 1.*

Corollary 3. *Fix $N, O, \mathcal{H}$ and suppose that $\mathcal{H}$ contains all the sets of the form $\{i\} \times O$ and $N \times \{a\}$. If $\mathcal{H}$ is not bihierarchical, then there exist P and $(\underline{q}_S, \overline{q}_S)_{S \in \mathcal{H}}$ such that P satisfies $\underline{q}_S \leq P_S \leq \overline{q}_S$ for all $S \in \mathcal{H}$ but there exists no decomposition of P into a convex combination of matrices satisfying conditions (1)–(3) of Corollary 1.*

Note that the restriction that $\mathcal{H}$ contains all the sets of the form $\{i\} \times O$ and $N \times \{a\}$ is natural in any bilateral matching setting.

4. Application: Probabilistic Serial Mechanism with Complex Constraints

In this section, we consider an environment $\mathcal{E} = \langle N, O, \mathcal{H}, \{\underline{q}_S, \overline{q}_S\}_{S \in \mathcal{H}} \rangle$, where the bihierarchy $\mathcal{H} = \mathcal{H}_N \cup \mathcal{H}_O$ has the following structure:

$$\mathcal{H}_N = \{\{(i,a)\} | (i,a) \in N \times O\} \cup \{\{i\} \times O | i \in N\}$$

and

$$N \times \{a\} \in \mathcal{H}_O, \forall a \in O,$$

with constraints such that $\underline{q}_{\{i\} \times O} = \overline{q}_{\{i\} \times O} = 1$ for all $i \in N$ and $\underline{q}_S = 0$ for all $S \in \mathcal{H} \setminus \{\{i\} \times O | i \in N\}$. This describes a problem of assigning objects to agents, who each demand exactly one unit. This problem arises in a wide variety of situations, ranging from placement of students to schools, doctors to hospitals, advertisements to publishers, computer processors to jobs, etc. The maximum quota for each object a, $\overline{q}_{N \times \{a\}}$, can be arbitrary, unlike in $\mathcal{E}_{BvN}$. It is understood also that O contains a **null object** $\o$ with unlimited supply, that is, $\overline{q}_S = +\infty$ for any $S \in \mathcal{H}_O$ with $(N \times \{\o\}) \cap S \neq \emptyset$.

4.1. Applications.

As motivated earlier, the bihierarchy structure in this section accommodates a range of realistic situations and constraints that a mechanism designer faces. First, the objects may be produced endogenously based on the (reported) preferences of the agents. For instance, the objects may be a set of possible goods that can be produced, and the mechanism may select which array of goods will be produced based on the demand of the buyers. In the context of school choice, different types of objects may represent different programs, say within a school specializing in different disciplines and activities, only a subset of which may be chosen depending on the preferences of incoming students. Or school districts may run multiple school programs in the same building (this is often the case in public schools in NYC, for example) and relative sizes of these programs can be changed but the sum of students in these programs is constrained by the building size. Such situations are captured by a hierarchy $\mathcal{H}_O$ containing sets of the form $S = N \times O'$ with $|O'| \geq 2$. The ceiling $\overline{q}_S$ then describes total capacities that can be allocated within O'. Further, the hierarchical structure means that flexible production can be nested; e.g., a subset of programs may be chosen and, within each chosen program, a subset of subprograms may be chosen, and so on.

Another common situation confronting a mechanism designer is a need to treat different groups of agents differently in assignment. For example, affirmative action or a desire by a school for diversity may entail quotas for applicants of different ethnic, racial or economic profiles. Such a practice is called "controlled choice" and is employed in many school districts in the United States. Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu (2005) analyze assignment mechanisms under such constraints. Although the controlled choice plans face some opposition,[9] there are other forms of constraints that have similar structures. For example, a subset of schools in New York City (the so-called Educational Option programs) require balanced distributions of test scores: 16 percent of the seats should be allocated to students who were rated top performers in a standardized exam, 68 percent to middle performers, and 16 percent to lower performers (Abdulkadiroğlu, Pathak, and Roth 2005).[10] Quotas may be based on residence as well: In Seoul, Korea, a concern over traffic congestions caused by commuting students — a form of externality the students or their parents may not fully internalize — led the school system to limit the number of seats for students from distant areas, in their upcoming school choice program (to start in year 2010).[11] Similar quotas based on residential areas are observed in a number of school choice programs in Japan as well.[12] All such constraints (they can be affirmative action or test score distribution or neighborhood quotas) can be easily incorporated by $\mathcal{H}_O$ containing sets of the form $N' \times \{a\}$ for $a \in O$ and $N' \subsetneq N$. The quota $\overline{q}_{N' \times \{a\}}$ then determines the maximum number of agents school a can admit from group N'. We can accommodate multiple groups for each a as long as they do not overlap

[9]In 2007, U.S. Supreme court ruled against Seattle's and Louisville's use of race as an admission criterion in their school choice plans with a 5-4 vote (New York Times 06/29/2007).

[10]The exact implementation of affirmative action constraints and other constraints is often slightly different from the quotas analyzed in this section, but the current definition serves as a useful benchmark. Indeed, most of the literature such as those cited above adopt the same definition as ours. See Kojima (2008a) for further discussion on this point.

[11]We are not aware of school districts in the U.S. that use residential-area based quotas as modeled in this paper. In Boston, for instance, students living in a neighborhood of a school are given priority to that school over students outside of the schools' neighborhood (Abdulkadiroğlu, Pathak, Roth, and Sönmez 2005). Boston's constraint is based on a similar idea but formally different from our quota-based approach. We suspect that there may be school districts in the U.S. that use quota-based approach investigated here.

[12]Edogawa and Shinagawa districts in Tokyo, for example.

with each other. We can even accommodate hierarchies of constraints: For example, a constraint that a school admits at most 50 students from district one, at most 50 students from district two, and at most 80 students from either district one or two.

4.2. **The Probabilistic Serial Mechanism.** To formally study the problem of assignment, we need to introduce the preferences of agents. Each agent $i \in N$ has a **strict preference** $\succ_i$ over O. We write $a \succeq_i b$ if either $a \succ_i b$ or $a = b$ holds. When N and O are fixed, we write $\succ$ for $(\succ_i)_{i \in N}$, $\succ_{N'}$ for $(\succ_i)_{i \in N'}$ where $N' \subset N$. A quadruple $\Gamma = (N, O, \mathcal{H}, \{\underline{q}_S, \overline{q}_S\}_{S \in \mathcal{H}}, (\succ_i)_{i \in N})$ then defines a **random assignment problem**. Recall that a feasible random assignment is a matrix $P = [P_{ia}]_{i \in N, a \in O}$ satisfying $\underline{q}_S \leq \sum_{(i,a) \in S} P_{ia} \leq \bar{q}_S$ for each $S \in \mathcal{H}$. Recall that $\mathcal{P}_{\mathcal{E}}$ denotes the set of all random assignments feasible in $\mathcal{E} = \langle N, O, \mathcal{H}, q \rangle$. A special case of random assignment is a **deterministic assignment**, represented by a matrix $P \in \mathcal{P}_{\mathcal{E}}$ with $P_{ia} \in \{0, 1\}$ for each $(i, a) \in N \times O$.

A solution used in many applications is **random priority** (Bogomolnaia and Moulin 2001), also called **random serial dictatorship** (Abdulkadiroğlu and Sönmez 1999), originally studied under environment $\mathcal{E}_{BvN}$. In the current setup, we define random priority as follows: (i) randomly order agents with equal probability, and (ii) the first agent obtains her favorite object, the second agent obtains her favorite object among the remaining objects, and so on, *as long as* allocating the good to agents so far is consistent with all ceiling constraints. The mechanism is **strategy-proof**, that is, reporting true preferences is a dominant strategy for every agent. However, the random priority may result in loss of efficiency (Bogomolnaia and Moulin 2001). To see how the loss of efficiency may occur in our context, consider the following example, adapted from Bogomolnaia and Moulin (2001).

Example 1 (Random priority may result in a suboptimal production plan)**.** Let $N = \{1, 2, 3, 4\}$, $O = \{a, b, c, \emptyset\}$, $\mathcal{H}_O = [\bigcup_{a' \in O} \{N \times \{a'\}\}] \cup \{N \times \{a, b, c\}\}$. Assume $\overline{q}_{N \times \{a\}} = \overline{q}_{N \times \{b\}} = \overline{q}_{N \times \{c\}} = 1, \overline{q}_{N \times \{a,b,c\}} = 2$. This is a situation in which each good has individual quota of one, and furthermore only two out of three goods can actually be produced.

Let

$$\succ_1: a, b, \emptyset,$$

$$\succ_2: a, b, \emptyset,$$

$$\succ_3: c, b, \emptyset,$$

$$\succ_4: c, b, \emptyset,$$

where the notation means that agent one prefers a to b to $\emptyset$, and so on. Under the random priority mechanism, the assignment

$$RP = \begin{pmatrix} 5/12 & 1/12 & 0 & 1/2 \\ 5/12 & 1/12 & 0 & 1/2 \\ 0 & 1/12 & 5/12 & 1/2 \\ 0 & 1/12 & 5/12 & 1/2 \end{pmatrix},$$

will be obtained.[13] The following random assignment is preferred by everyone.

$$P' = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

One notable feature of the random priority mechanism is that, under this mechanism, good b is produced although everyone prefers some other good, a or c, to be produced. Good b is produced either when agent 1 and 2 get highest priorities or when agent 3 and 4 get highest priorities. All agents will be made better off if the social planner can first decide to produce a and c and then allocate the goods: It is easy to see that random assignment P' results if the random priority is conducted after the production plan is fixed to producing a and c. Of course, such a production plan is inefficient if agents prefer b to other goods. Thus a good mechanism may be one that simultaneously decides production of goods as well as the allocation of them, based on preference information reported by agents.

We begin by introducing the efficiency concept in our setup, called ordinal efficiency. A random assignment P **ordinally dominates** another random assignment $P' \in \mathcal{P}_{\mathcal{E}}$ **at** $\succ$ if for each agent i the lottery P_i first-order stochastically dominates the lottery P'_i, that is,

$$\sum_{b \succeq_i a} P_{ib} \geq \sum_{b \succeq_i a} P'_{ib} \ \forall i \in N, \forall a \in O,$$

[13] Given $N = \{1,2,3,4\}$, $O = \{a,b,c,\emptyset\}$ and random assignment P, we write

$$P = \begin{pmatrix} P_{1a} & P_{1b} & P_{1c} & P_{1\emptyset} \\ P_{2a} & P_{2b} & P_{2c} & P_{2\emptyset} \\ P_{3a} & P_{3b} & P_{3c} & P_{3\emptyset} \\ P_{4a} & P_{4b} & P_{4c} & P_{4\emptyset} \end{pmatrix}.$$

Similar notation will be used elsewhere in this paper as well.

with strict inequality for some i, a (we say that P_i weakly stochastically dominates P'_i when either P_i stochastically dominates P'_i or $P_i = P'_i$). If P ordinally dominates P' at $\succ$, then every agent i prefers P_i to P'_i according to any expected utility function with utility index consistent with $\succ_i$. The random assignment $P \in \mathcal{P}_{\mathcal{E}}$ is **ordinally efficient at** $\succ$ if it is not ordinally dominated at $\succ$ by any other random assignment in $\mathcal{P}_{\mathcal{E}}$. Note that our model allows for a complex set of constraints, so the current notion has the flavor of "constrained efficiency" in that the efficiency is defined within the set of assignments satisfying the constraints.

As with Bogomolnaia and Moulin (2001), a different characterization of ordinal efficiency proves useful. To this end, we first define the **minimal constraint set containing** (i, a):

$$\nu(i,a) := \bigcap_{S \in \mathcal{H}(i,a)} S,$$

if the set $\mathcal{H}(i,a) := \{S \in \mathcal{H}_O : (i,a) \in S, \sum_{(j,b)\in S} P_{jb} = \overline{q}_S\}$ is nonempty. If $\mathcal{H}(i,a) = \emptyset$ (or equivalently $\sum_{(j,b)\in S} P_{jb} < \overline{q}_S$ for all $S \in \mathcal{H}_O$ containing (i,a)), then we let $\nu(i,a) = N \times O$.

We next define the following binary relations on $N \times O$ given (Γ, P) as follows:[14]

$$(j,b) \rhd_1 (i,a) \iff i = j, b \succ_i a, \text{ and } P_{ia} > 0,$$

$$(j,b) \rhd_2 (i,a) \iff \nu(j,b) \subseteq \nu(i,a).$$

We then say

$$(j,b) \rhd (i,a) \iff (j,b) \rhd_1 (i,a) \text{ or } (j,b) \rhd_2 (i,a).$$

We say a binary relation $\rhd$ is **strongly cyclic** if there exists a finite cycle $(i_0, a_0) \rhd (i_1, a_1) \rhd \cdots \rhd (i_k, a_k) \rhd (i_0, a_0)$ such that $\rhd = \rhd_1$ for at least one relation. We next provide a characterization of ordinal efficiency.

Proposition 1. *Random assignment $P \in \mathcal{P}_{\mathcal{E}}$ is ordinally efficient if and only if $\rhd$ is not strongly cyclic given (Γ, P).*[15]

A remark is in order. In the environment $\mathcal{E}_{BvN}$, Bogomolnaia and Moulin (2001) define the binary relation $\rhd$ over the set of objects where $b \rhd a$ if there is an agent i such that $b \succ_i b$ and $P_{ia} > 0$. Bogomolnaia and Moulin show that, in $\mathcal{E}_{BvN}$, a random assignment is

[14]Given that $\mathcal{H}_O$ has a hierarchical structure,

$$(j,b) \rhd_2 (i,a) \iff (j,b) \in S \text{ for any } S \in \mathcal{H}_O \text{ such that } (i,a) \in S, P_S = \overline{q}_S.$$

[15]In Kojima and Manea (2008), ordinal efficiency is characterized by two conditions, acyclicity and non-wastefulness. We do not need non-wastefulness as a separate axiom in our current formulation since a "wasteful" random assignment (in their sense) contains a strong cycle as defined here.

ordinally efficient if and only if $\triangleright$ is acyclic. Our contribution over their characterization is that we expand the domain over which the binary relation is defined to the set of agent-good pairs, in order to capture the complexity that results from a more general environment than $\mathcal{E}_{BvN}$.

Now we introduce the **probabilistic serial** mechanism, which is an adaptation of the mechanism proposed by Bogomolnaia and Moulin to our setting. The idea is to regard each object as a divisible object of "probability shares." Each agent "eats" the best available object with speed one at every time $t \in [0,1]$. The resulting profile of shares of objects eaten by agents by time 1 obviously corresponds to a random assignment matrix, which we call the **probabilistic serial random assignment**.

Before giving a formal definition, note that we will need to modify the definition of the algorithm from the version of Bogomolnaia and Moulin (2001). First, we will specify availability of goods with respect to both agents and objects in order to accommodate complex constraints such as affirmative action. Second, we need to keep track of multiple constraints for each pair of agent-good pair (i,a) during the algorithm, since there are potentially multiple constraints that would make the consumption of the good a by the agent i no longer feasible.

Formally, the probabilistic serial mechanism is defined through the following **symmetric simultaneous eating algorithm**, or the eating algorithm for short. For any $(i,a) \in S \subseteq N \times O$, let

$$\chi(i,a,S) = \begin{cases} 1 & \text{if } (i,a) \in S \text{ and } a \succeq_i b \text{ for any } b \text{ with } (i,b) \in S, \\ 0 & \text{otherwise,} \end{cases}$$

be the indicator function that a is the most preferred object for i among objects b such that (i,b) is listed in S.

Given a preference profile $\succ$, the eating algorithm is defined by the following sequence of steps. Let $S^0 = N \times O, t^0 = 0$, and $P^0_{ia} = 0$ for every $i \in N$ and $a \in O$. Given $S^0, t^0, [P^0_{ia}]_{i \in N, a \in O}, \dots, S^{v-1}, t^{v-1}, [P^{v-1}_{ia}]_{i \in N, a \in O}$, for any $(i,a) \in S^{v-1}$ define

$$t^v(i,a) = \min_{S \in \mathcal{H}_O : (i,a) \in S} \sup \left\{ t \in [0,1] | \sum_{(j,b) \in S} [P^{v-1}_{jb} + \chi(j,b,S^{v-1})(t - t^{v-1})] < \overline{q}_S \right\}, \tag{2}$$

$$t^v = \min_{(i,a) \in S^{v-1}} t^v(i,a), \tag{3}$$

$$S^v = S^{v-1} \setminus \{(i,a) \in S^{v-1} | t^v(i,a) = t^v\}, \tag{4}$$

$$P^v_{ia} = P^{v-1}_{ia} + \chi(i,a,S^{v-1})(t^v - t^{v-1}). \tag{5}$$

Since $N \times O$ is a finite set, there exists $\bar{v}$ such that $t^{\bar{v}} = 1$. We define $PS(\succ) := P^{\bar{v}}$ to be the probabilistic serial random assignment for the preference profile $\succ$.

Bogomolnaia and Moulin (2001) show that the probabilistic serial mechanism results in an ordinally efficient random assignment in their simplified setting $\mathcal{E}_{BvN}$. Their proof can be adapted to our setting using Proposition 1, although the proof is somewhat more involved because of the constraints that are not present in their setting.

Proposition 2. *For any preference profile $\succ$, the probabilistic serial random assignment $PS(\succ)$ is ordinally efficient at $\succ$.*

Bogomolnaia and Moulin (2001) also show that the probabilistic serial mechanism is fair in a specific sense in their simple setting. Formally, a random assignment P is said to be **envy-free at** $\succ$ if P_i weakly first-order stochastically dominates P_j with respect to $\succ_i$ for every $j \in N$. It turns out that the probabilistic serial random assignment may not be envy-free in our environment. To see this point, consider a random assignment problem in which $N = \{1, 2, 3\}$, $O = \{a, \emptyset\}$, $\mathcal{H}_O = \{\{1, 2\} \times \{a\}, N \times \{a\}\}$, $\bar{q}_{\{1,2\}\times\{a\}} = 1$, $\bar{q}_{N\times\{a\}} = 2$, and $a \succ_i \emptyset$ for every $i \in N$. In this problem it is easy to see that

$$PS(\succ) = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \\ 1 & 0 \end{pmatrix}.$$

The probabilistic serial random assignment $PS(\succ)$ is not envy-free since $PS_3(\succ)$ is not weakly stochastically dominated by $PS_1(\succ)$ with respect to $\succ_1$ (indeed, $PS_3(\succ)$ stochastically dominates $PS_1(\succ)$ in this example). However, existence of envy may not immediately imply that the allocation is unfair. To see this point, note that it is infeasible to assign a to agent 1 with higher probability simply by moving probability share of a from agent 3 to agent 1, because there is a constraint on $\{1, 2\} \times \{a\}$. In that sense the envy is based on a desire of agent 1 that cannot be feasibly accommodated. Motivated by this observation, we introduce the following concept. Random assignment $P \in \mathcal{P}_{\mathcal{E}}$ is **feasible envy-free at** $\succ$ if there is no i and j such that $P_j \neq P_i$, P_j is not first-order stochastically dominated by P_i at $\succ_i$ and an assignment Q defined by

$$Q_{ka} = \begin{cases} P_{ja} & \text{if} \quad k = i, \\ 0 & \text{if} \quad k = j, \\ P_{ka} & \text{otherwise}, \end{cases} \tag{6}$$

is in $\mathcal{P}_{\mathcal{E}}$.[16] In the above example, $PS(\succ)$ is feasible envy-free. This property turns out to hold generally, as shown below.

Proposition 3. *For any preference profile $\succ$, the probabilistic serial random assignment $PS(\succ)$ is feasible envy-free at $\succ$.*

Ordinal efficiency and feasible envy-freeness are not satisfied by random priority. Indeed, random priority violates both properties even in the simplest setting $\mathcal{E}_{BvN}$ (Bogomolnaia and Moulin 2001).

Unfortunately the mechanism is not strategy-proof, that is, an agent is sometimes made better off misstating her preferences. However, Bogomolnaia and Moulin (2001) show that the probabilistic serial mechanism is **weakly strategy-proof**, that is, an agent cannot misstate his preferences and obtain a random assignment that stochastically dominates the one obtained under truth-telling. Formally, we claim that the probabilistic serial mechanism is weakly strategy-proof, that is, there exist no $\succ$, $i \in N$ and $\succ'_i$ such that $PS_i(\succ'_i, \succ_{-i})$ stochastically dominates $PS_i(\succ)$ at $\succ$ in our more general environment.[17]

Proposition 4. *The probabilistic serial mechanism is weakly strategy-proof.*

Proof. The proof is an adaptation of Proposition 1 of Bogomolnaia and Moulin (2001) and we omit the proof. □

5. Application: Multi-Unit Assignment

This section considers multi-unit resource allocation problems in which monetary transfers are prohibited. Examples include the assignment of schedules of courses to students, the assignment of tasks within an organization, the division of sentimental objects amongst a set of heirs, and the allocation of access to jointly-owned scientific resources.

This is well known to be a difficult problem. Papai (2001) shows that sequential dictatorships (considered unrealistic for many applications) are the only deterministic mechanisms that are nonbossy, strategy-proof and Pareto optimal. Ehlers and Klaus (2003) and

[16] Alternatively, one could define P to allow for feasible envy if exchanging P_i and P_j between i and j is feasible and P_j is not weakly stochastically dominated by P_i at i. This alternative definition is weaker than our current definition. Thus, by Proposition 3, the probabilistic serial mechanism is feasible envy-free with this alternative definition.

[17] Kojima and Manea (2008) show that truthtelling becomes a dominant strategy for a sufficiently large market under the probabilistic serial mechanism in a simpler environment than the current one. Showing such a claim in a more general environment is beyond the scope of this paper, but we conjecture that the argument readily extends.

Hatfield (2008) show similarly pessimistic results even if agents have more specific preferences. Kojima (2008b) shows that the difficulty is present even if random assignment mechanisms are allowed: There exists no mechanism that is ordinally efficient, envy-free and weakly strategy-proof. Sönmez and Ünver (2008) discuss theoretical difficulties with one mechanism used commonly in practice for allocating course seats at business schools in the United States. Budish and Cantillon (2008) document empirically that a different mechanism used in practice (at Harvard Business School) is heavily manipulated by students in a manner consistent with theory, harming welfare, and yet nevertheless is "less bad" than the random serial dictatorship on measures of both welfare and fairness.

Some recent progress has been made by Budish (2008). He proposes a mechanism in which students are randomly assigned approximately equal budgets of an artificial currency, which they use to purchase sure bundles of courses at an approximate competitive equilibrium price vector. That is, the randomness in the Budish (2008) mechanism is analogous to the randomness in the random serial dictatorship mechanism, in that, after an initial random allocation of resources (budgets or priority), the mechanism is essentially deterministic. (In fact the two mechanisms coincide for the case of single-unit demand).

Here, we provide a method that may facilitate development of multi-unit assignment mechanisms that directly allocate a random assignment. (That is, mechanisms whose randomness is analogous to that in PS, not RSD). Specifically, for a given random assignment prescribed by the social planner,[18] we show how to implement the random assignment in a manner that limits the variation in realized payoffs.

To motivate, suppose that two agents are to divide four objects a, b, c, d, listed in descending order of their (common) preferences. A fair fractional assignment would be for each agent to receive half of each object. One way to implement this fractional assignment is to give any two randomly chosen objects to one agent and give the remaining two to the other. This could entail a highly unfair outcome, however, in which one agent gets the two best objects, a and b, and the other gets the two worst c and d.

Our method avoids such outcomes. The idea is to supplement the actual constraints of the problem with a set of "artificial" hierarchical constraints, in a manner that limits

[18]We do not specify how the social planner prescribes the random assignment here, since our analysis holds for any given random assignment. One possibility is a multi-unit generalization of pseudo-market mechanism by Hylland and Zeckhauser (1979). Another approach may be a generalization of the probabilistic mechanism (Bogomolnaia and Moulin 2001), but it may not be a sensible mechanism in the current setting since the mechanism will have good incentive properties (as analyzed by Bogomolnaia and Moulin (2001), Kojima and Manea (2008) and Che and Kojima (2008)) only when agents have unit demand.

variation in utilities. Specifically, if there are k objects, for each agent we create k additional constraints, one each for his j most preferred objects for $j = 1 \dots k$. The floor and ceiling constraints are based on the total probabilities for each set prescribed by the social planner. In this simple example, we would add constraints, for each agent, that they get between 0 and 1 units of a, exactly 1 unit of $\{a, b\}$, between 1 and 2 units of $\{a, b, c\}$, and exactly 2 units of $\{a, b, c, d\}$.

5.1. **The Utility Guarantee.**

Theorem 3 (Utility Guarantee)**.** *Consider an input $\langle N, O, P, \mathcal{H}\rangle$ as in Theorem 1 for which, additionally,*

$$\mathcal{H}_N = \{\{(i, a)\}|(i, a) \in N \times O\} \cup \{\{i\} \times O|i \in N\}.$$

Suppose that there is a set of values $(v_{ia})_{(i,a)\in N\times O}$ such that, for all i, agent i's expected utility from a stochastic assignment P is $\sum_{a\in O} P_{ia}v_{ia}$. (Negative v's are interpreted as costs). Then, for any P, there exists a decomposition of P that satisfies all of the conditions of Theorem 1, and also:

$$\sum_a P'_{ia}v_{ia} - \sum_a P''_{ia}v_{ia} \in [-\overline{v}_i, \overline{v}_i], \tag{7}$$

$$\sum_a P'_{ia}v_{ia} \in \left[\sum_a P_{ia}v_{ia} - \overline{v}_i, \sum_a P_{ia}v_{ia} + \overline{v}_i\right], \tag{8}$$

for each i and each P' and P'' in the convex combination, where $\overline{v}_i = \max\{v_{ia}|a \in O, P_{ia} > 0\}$.

The Theorem enables the social planner to implement the lottery while guaranteeing that the variations in realized utilities are bounded by the valuation of a single object. The conclusion is most useful when each agent obtains a large number of objects, since in such a case the valuation of one object will often be relatively small compared to the utility of the bundle of goods as a whole. Then Theorem 3 implies that the utility variation among realized deterministic assignments may be relatively small.

We include the proof in the body of the text because it illustrates the versatility of hierarchical constraints.

Proof. Assume $P_{ia} > 0$ for all i, a (extension to cases when $P_{ia} = 0$ for some i, a is straightforward). For each $i \in N$, let $(a_i^1, a_i^2, \dots, a_i^{|O|})$ be a sequence of objects ordered in the decreasing order of i's preferences so that $v_{ia_i^1} \geq v_{ia_i^2} \geq \ldots, v_{ia_i^{|O|}}$. Define the class of sets $\mathcal{H}' = \mathcal{H}'_N \cup \mathcal{H}'_O$ by

$$\mathcal{H}'_N = \mathcal{H}_N \cup \bigcup_{\substack{i \in N, \\ k \in \{1,\ldots,|O|\}}} \{i\} \times \{a_i^1, \ldots, a_i^k\},$$

$$\mathcal{H}'_O = \mathcal{H}_O.$$

By inspection, $\mathcal{H}'$ is a bihierarchy. Therefore, by Theorem 1, there exists a convex decomposition such that

$$\sum_{(i,a)\in S} P'_{ia}, \sum_{(i,a)\in S} P''_{ia} \in \left\{ \left\lfloor \sum_{(i,a)\in S} P_{ia} \right\rfloor, \left\lceil \sum_{(i,a)\in S} P_{ia} \right\rceil \right\} \quad \text{for all } S \in \mathcal{H}', \tag{9}$$

for any integer-valued matrices P' and P'' that are part of the decomposition. In particular, property (9) holds for each $\{(i,a)\} \in \mathcal{H}'_N$ and $\{i\} \times \{a_i^1, \ldots, a_i^k\} \in \mathcal{H}'_N$. This means that

- For any i and k, $P'_{ia_i^k} - P''_{ia_i^k} \in \{-1, 0, 1\}$, and
- If $\sum_{l=1}^k P'_{ia_i^l} - \sum_{l=1}^k P''_{ia_i^l} = 1$, then $P'_{ia_i^{k+1}} - P''_{ia_i^{k+1}} \in \{-1, 0\}$.

Therefore the sequence $(P'_{ia_i^k} - P''_{ia_i^k})_{k\in\{1,\ldots,|O|\}}$ takes values $\{-1,0,1\}$ and alternates the sign whenever it takes a nonzero value. This fact and the assumption that $v_{ia_i^1} \geq v_{ia_i^2} \geq \ldots, v_{ia_i^{|O|}}$ imply

$$\sum_{a\in O} P'_{ia} v_{ia} - \sum_{a\in O} P''_{ia} v_{ia} = \sum_{k=1}^{|O|} (P'_{ia_i^k} - P''_{ia_i^k}) v_{ia_i^k}, \tag{10}$$

is between $-\overline{v}_i$ and $\overline{v}_i$. Thus we obtain property (7). Property (8) follows immediately from property (7). □

5.2. **The Pseudo-Market Approach to Fair Division.** To be added.

5.3. **The Maximin Approach to Fair Division.** Consider the following problem. The social planner has a number of indivisible objects O to be allocated to agents N. Utility of agents is additive in objects up to a fixed quota (the quota can be infinite), so utility of agent i from random assignment P is $\sum_{a\in O} v_{ia} P_{ia}$. The social planner wants to maximize the utility of the worst-off agent. This is sometimes called the Santa Claus problem: Santa Claus wants to give presents to children in such a way that the least fortunate child is as happy as possible given the fixed set of presents he has.

Formally, consider the social planner's problem:

$$
\begin{aligned}
\text{maximize } \omega \text{ subject to} & \\
P_{ia} \in \mathbb{N} \quad & \text{for all } i \in N, a \in O, \\
P_S \leq \overline{q}_S, \quad & \text{for all } S \in \mathcal{H}_O, \\
\sum_{a \in O} P_{ia} \leq \overline{q}_{\{i\} \times O}, \quad & \text{for all } i \in N, \\
\omega \leq \sum_{a \in O} P_{ia} v_{ia} \quad & \text{for all } i \in N.
\end{aligned}
\tag{11}
$$

This problem is known to be computationally difficult. Thus in practice, the social planner may need to use a mechanism that is easier to implement. On the other hand, she wants to attain the objective at least approximately. To attain these conflicting goals, consider the following two-stage algorithm. In the first stage, solve the following linear programming problem:

$$
\begin{aligned}
\text{maximize } \omega \text{ subject to} & \\
P_{ia} \in [0, \infty) \quad & \text{for all } i \in N, a \in O, \\
P_S \leq \overline{q}_S, \quad & \text{for all } S \in \mathcal{H}_O, \\
\sum_{a \in O} P_{ia} \leq \overline{q}_{\{i\} \times O}, \quad & \text{for all } i \in N, \\
\omega \leq \sum_{a \in O} P_{ia} v_{ia} \quad & \text{for all } i \in N.
\end{aligned}
\tag{12}
$$

Since the problem relaxes the integrality of the first constraint, the solution may be infeasible. On the other hand the optimal solution of this problem is easy to compute since it is a simple linear programming problem. In the second stage, given the optimal solution of the linear programming problem (12), round the solution into an integer-valued solution, making the assignment a feasible solution in problem (11). The cost of doing so is, of course, that the social welfare typically decreases when the social planner modifies the optimal fractional solution into an integral one. However, the following claim guarantees that the loss of efficiency can be bounded.

Corollary 4. *Let P^* be an optimal solution of the above linear programming problem (12) with optimal value ω^*. Then there exists an integer solution P' of the integer programming problem (11), with value $\omega' \geq \omega^* - \max_{i \in N} \overline{v}_i$, where $\overline{v}_i = \max\{v_{ia} | a \in O, P_{ia} > 0\}$. In* *particular, $\omega' \geq \omega^{**} - \max_{i \in N} \overline{v}_i$ where w^{**} is the optimal value of the original integer programming problem (11).*

Proof. Let P^* be the solution of the linear programming problem (12), with optimal value ω^*. By Theorem 3, there exists P' that is integer-valued and satisfies

$$\sum_a P'_{ia} v_{ia} \geq \sum_a P^*_{ia} v_{ia} - \overline{v}_i, \tag{13}$$

for each i. Since $\sum_a P^*_{ia} v_{ia} \geq \omega^*$ for each i by construction, inequality (13) implies that $\sum_a P'_{ia} v_{ia} \geq \omega^* - \overline{v}_i$, implying $\omega' \geq \omega^* - \max_{i \in N} \overline{v}_i$. Finally, $w^* \geq w^{**}$ since w^* is the optimal value of a less constrained problem (12) than problem (11). Thus we have $\omega' \geq \omega^{**} - \max_{i \in N} \overline{v}_i$, completing the proof. □

Corollary 4 generalizes Bezáková and Dani (2005), who proposed a similar two-stage algorithm when the environment is the simple $\mathcal{E}_{BvN}$. While Corollary 4 can handle more complex situations, we acknowledge that Corollary 4 is only a mild extension. On the other hand we emphasize the methodological innovation. Corollary 4 is shown by Theorem 3, which in turn is a direct consequence of our rounding result Theorem 1. Our contribution here is that apparently dissimilar results such as utility guarantee and fair division mechanisms can be derived from one fundamental result.

5.4. **Scheduling Jobs on Parallel Machines: Minimize Makespan Problem.** Our approach has can be applied to the so-called "minimize makespan problem" studied widely in computer science.

Consider the following setting, slightly generalizing Lenstra, Shmoys, and Tardos (1990). There is a set N of parallel machines and a set A of independent jobs. Each job needs to be assigned to one of the machines. The job is indivisible, that is, each job needs to be assigned to one machine in its entirety (or equivalently, it is prohibitively costly to process part of a job in one machine and process remaining parts in others). The processing of job a on machine i takes time c_{ia}. The machines are parallel and jobs are independent, that is, more than one machines can process jobs simultaneously and any job can be processed irrespective of whether other jobs are already completed. The **makespan** of the assignment of jobs to machines is the time needed to finish all jobs. The objective is to find a schedule that minimizes the makespan.

Let $J_i(t)$ denote the set of jobs that require at most time t when processed by machine i, and let $M_a(t)$ denote the set of machines that can process job a in no more than time t. Consider the relaxed problem in which random assignments are allowed, and let P be a fractional assignment of jobs to machines where each machine i finishes processing jobs by deadline d_i (a fractional solution is often easy to find because linear programming techniques are applicable). We will show the following slight generalization of the rounding theorem of Lenstra, Shmoys, and Tardos (1990).

Corollary 5 (Theorem 1 of Lenstra, Shmoys, and Tardos (1990))**.** *Let* $c = (c_{ia})_{(i,a)\in N\times O} \in \mathbb{R}_+^{|N|\times|O|}$, $d = (d_a)_{a\in O} \in \mathbb{R}_+^{|O|}$ *and* $t \in \mathbb{R}_+$ *be given. If there is a feasible solution* P *to the (in)equalities,*

$$\sum_{i\in M_a(t)} P_{ia} = 1, \qquad \text{for } a \in O,$$

$$\sum_{a\in J_i(t)} P_{ia}c_{ia} \leq d_i, \qquad \text{for } a \in O,$$

$$P_{ia} \geq 0, \qquad \text{for } a \in J_i(t), i \in N,$$

then there is an integer solution P' *to the following set of conditions,*

$$\sum_{i\in M_a(t)} P'_{ia} = 1, \qquad \text{for } a \in O, \tag{14}$$

$$\sum_{a\in J_i(t)} P'_{ia}c_{ia} \leq d_i + t, \qquad \text{for } a \in O,$$

$$P'_{ia} \in \{0,1\}, \qquad \text{for } a \in J_i(t), i \in N.$$

Proof. By Theorem 3, there exists P' that is integer-valued and satisfies (14) and

$$\sum_a P'_{ia}c_{ia} \leq \sum_a P_{ia}c_{ia} + \bar{c}_i.$$

By definition of $M_a(t)$ and the assumption, $P_{ia} > 0$ implies $c_{ia} \leq t$. This completes the proof. □

The result implies that there exists a feasible integer solution whose makespan is within time t of the optimal (infeasible) fractional solution, where t is the time of the single slowest job processed in the fractional solution. Since the optimal fractional solution is weakly better than the optimal feasible integer solution, we have a method that finds an integer solution that is "close" to the true optimum. Some generalizations of the minimize makespan problem, such as Theorem 2.1 of Shmoys and Tardos (1993), are also corollaries of Theorem 3, via a logic similar to the proof of Corollary 5.

5.5. **Two-Sided Matching.** Our approach can be applied to the two-sided matching environment. In this section, both N and O are sets of agents. We allow for many-to-many matching, that is, some agents in N can be matched with more than one agent in O each and vice versa.

Theorem 4. *Consider a problem as in Theorem 1 for which, additionally,*

$$\mathcal{H}_N = \{\{(i,a)\}|(i,a) \in N \times O\} \cup \{\{i\} \times O | i \in N\},$$
$$\mathcal{H}_O = \{N \times \{a\} | a \in O\}.$$

Suppose that there are sets of values $(v_{ia})_{(i,a)\in N\times O}$ *and* $(w_{ia})_{(i,a)\in N\times O}$ *such that, for each agent* $i \in N$ *(respectively agent* $a \in O$*), her expected utility from a random assignment* P *is* $\sum_{a\in O} P_{ia}v_{ia}$ *(respectively* $\sum_{i\in N} P_{ia}w_{ia}$*). Then, for any* P*, there exists a decomposition of* P *that satisfies all of the conditions of Theorem 1, and also:*

$$\sum_a P'_{ia}v_{ia} - \sum_a P''_{ia}v_{ia} \in [-\overline{v}_i, \overline{v}_i] \tag{15}$$

$$\sum_a P'_{ia}v_{ia} \in \left[\sum_a P_{ia}v_{ia} - \overline{v}_i, \sum_a P_{ia}v_{ia} + \overline{v}_i\right], \tag{16}$$

$$\sum_i P'_{ia}w_{ia} - \sum_i P''_{ia}w_{ia} \in [-\overline{w}_a, \overline{w}_a] \tag{17}$$

$$\sum_i P'_{ia}w_{ia} \in \left[\sum_i P_{ia}w_{ia} - \overline{w}_a, \sum_i P_{ia}w_{ia} + \overline{w}_a\right], \tag{18}$$

for each i, a *and each* P' *and* P'' *being part of the convex decomposition, where* $\overline{v}_i = \max\{v_{ia}|a \in O, P_{ia} > 0\}$ *and* $\overline{w}_a = \max\{w_{ia}|i \in N, P_{ia} > 0\}$.

Proof. The proof is a straightforward adaptation of the proof of Theorem 3 and hence is omitted. ☐

Let us suggest one possible application. There are two leagues of sports teams N and O, say the American League and National League in professional baseball, and the planner wants to schedule interleague play. The planner wants to ensure that the strength of opponents that teams in a league play against is as equalized as possible among teams in the same league. For that goal, the planner could first give a uniform probability for each match: That will give one specific random assignment in which any pair of teams in the same league is treated equally. Then, using Theorem 4, the planner finds a feasible match, that is, a deterministic assignment matrix, in which differences in strength of opponents is bounded by one game with the strongest opponent in the other league, no matter how many games are scheduled for each team.

We note that transforming this feasible match into a specific schedule - i.e., not only how often does Team A play Team B, but *when* - is considerably more complicated. For example, the problem involves scheduling both intraleague and interleague matches simultaneously, dealing with geographical constraints and so forth. We do not claim that our method can be directly used to such complicated situations. Rather, our point here is to suggest a possibility that our analysis may be a useful first step to solve some problems that have not been considered to be related to questions such as school choice or fair allocation. See Nemhauser and Trick (1998) for further discussion of sport scheduling.

6. Optimal Assignment with Monetary Transfers

The next application considers the optimal assignment problem with monetary transfers, and its generalizations. In a recent paper, Milgrom (2008) considers the "assignment auction," a multi-object auction for close substitutes, building on the optimal assignment literature by Koopmans and Beckman (1957) and Shapley and Shubik (1972). Buyers are assumed to have a valuation function that is additive in objects except that there are (possibly) hierarchical ceiling constraints on subsets of objects. He further describes the assignment exchange model, in which there are multiple buyers and sellers.

Assume there is a set of $\overline{n}$ buyers and $\overline{m}$ sellers where each buyer is indexed by n and each seller by m. Each buyer and seller is assumed to be able to submit multiple "bids" indexed by i and j respectively, where buyer n has $\bar{i}_n$ bids and seller m has $\bar{j}_m$ bids. For each pair (n, i), there is a ceiling constraint. Moreover, there is a set of additional constraints. Constraints have a hierarchical form $\mathcal{G}$. Formally, the problem is written as

$$
\text{(19)} \qquad \begin{aligned}
\text{maximize } \sum_{k=1}^{\bar{k}} \sum_{n=1}^{\overline{n}} \sum_{i=1}^{\bar{i}_n} v_{nik} x_{nik} - \sum_{k=1}^{\bar{k}} \sum_{m=1}^{\overline{m}} \sum_{j=1}^{\bar{j}_m} c_{mjk} y_{mjk} \text{ subject to} \\
\sum_{n,i} x_{nik} - \sum_{m,j} y_{mjk} \leq 0 \text{ for all product versions } k, \\
\sum_{(n,i,k)\in S} x_{nik} \leq z_S \text{ for all } S \in \mathcal{G}, \\
\sum_{k} y_{mjk} \leq q_{mj} \text{ for all offers } m, j, \\
x_{nik}, y_{mjk} \geq 0 \text{ for all } m, n, i, j, k.
\end{aligned}
$$

In this environment, we derive Theorem 7 of Milgrom (2008) using our framework.

Corollary 6 (Theorem 7 of Milgrom (2008))**.** *If z_S and q_{mj} are positive integers for all m, j, S, then there is an integer optimal solution to problem (19).*

Proof. Let P^* be an optimal solution of Problem (19), which is in general a random assignment (the existence of such a solution is clear). Let N be all the possible agent-bid pairs, and N_b and N_s be set of possible buyer-bid pairs and seller-bid pairs respectively. Then, consider the following bihierarchies:

$$\mathcal{H}_N = \mathcal{G} \cup \{\{(n,i,k)\} | (n,i) \in N, k \in O\} \cup \{\{(m,j)\} \times O | (m,j) \in N_s\},$$
$$\mathcal{H}_O = \{N \times \{a\} | a \in O\}.$$

Let $P_{nik} = x_{nik}$ and $P_{mjk} = -y_{mjk}$ for all m, j, k, and $v_{mjk} = c_{mjk}$ if $(m,j) \in N_s$. Then the constraints in (19) can be rewritten as

$$\sum_{(n,i,k) \in N \times \{k\}} P_{nik} \leq 0 \text{ for all product versions } k, \tag{20}$$
$$\sum_{(n,i,k) \in S} P_{nik} \leq z_S \text{ for all } S \in \mathcal{G},$$
$$-q_{mj} \leq \sum_{(m,j,k) \in \{(m,j)\} \times O} P_{mjk} \text{ for all offers } (m,j) \in N_s,$$
$$P_{nik} \geq 0 \text{ for all } n, i, k,$$
$$P_{mjk} \leq 0 \text{ for all } m, j, k.$$

P^* satisfies the above constraints represented by (20). Since $\mathcal{G}$ is a hierarchy, it is easy to see that $\mathcal{H}$ defined above is a bihierarchy. Constraints (20) satisfy the conditions for the conclusion of Corollary 1 to apply: the sum of entries in sets in $\mathcal{H}$ is constrained by weak inequalities where constraints are integers or ∞ or $-\infty$. Therefore, by Corollary 1, there is a decomposition of P^* for which the sum of entries of each set in $\mathcal{H}$ is an integer. Since $\{(n,i,k)\} \in \mathcal{H}$ for each $(n,i) \in N, k \in O$ by definition of $\mathcal{H}$, this in particular means that each matrix that is part of the decomposition has all its entries being integers. Since the objective function is linear, we conclude that there is an integer solution to Problem (19), completing the proof. □

More generally, Theorem 1 can be used to derive a conclusion as in Milgrom (2008) for slightly more general environments, namely under any bihierarchical constraint structure.

7. Generalization

Throughout the paper we have focused on random assignment of objects (or agents) to agents. Implicit in the description of the model was the assumption that there are two kinds of entities and the social planner considers matchings between them. However, some of our results can be extended to a more general environment as described below.

Let X be a finite set and $\mathcal{H}$ be a collection of subsets of X. We call a pair $\mathcal{X} = (X, \mathcal{H})$ **hypergraph**. Let $f : 2^X \to \mathbb{R}$ be a **weight function** satisfying additivity: $f(S) =$

$\sum_{x\in S} f(x)$ for each $S \subset X$. Let $\mathcal{F}$ be the set of all such weight functions. We say the hypergraph $\mathcal{X}$ is **BvN decomposable** if, for each $f \in \mathcal{F}$, there exist $f^1, ..., f^k$ each in $\mathcal{F}$ such that

$$f = \sum_{j=1}^{k} \alpha_j f^j,$$

where $\alpha^j > 0, j = 1, ..., k$, and $\sum_{j=1}^{k} \alpha_j = 1$, and for each $j = 1, ..., k$,

(i) $f^j(S) \in \mathbb{Z}$, $\forall S \in \mathcal{H}$;

(ii) $\lfloor f(S) \rfloor \le f^j(S) \le \lceil f(S) \rceil$, $\forall S \in \mathcal{H}$.

We say that a hypergraph $\mathcal{X}$ forms a **bihierarchy** if there are $\mathcal{H}^1$ and $\mathcal{H}^2$ such that $\mathcal{H}^1 \cup \mathcal{H}^2 = \mathcal{H}$, $\mathcal{H}^1 \cap \mathcal{H}^2 = \emptyset$ and $\mathcal{H}^i$, $i = 1, 2$ is a **hierarchy**: if $S, S' \in \mathcal{H}^i$, then either $S \cap S' = \emptyset$, or $S \subset S'$ or $S' \subset S$.

It is useful to define a dual of a hypergraph. Given a hypergraph $\mathcal{X} = (X, \mathcal{H})$, its dual is $\mathcal{X}^T = (\mathcal{H}, X)$. A bihierarchy can be defined for its dual. To this end, for each $x \in X$, let $\mathcal{S}(x) := \{S \in \mathcal{H} | x \in S\}$ be the collection of sets in $\mathcal{H}$ each containing x. We say the dual of $\mathcal{X}$ forms a **bihierarchy** if there are X_1 and X_2 such that $X_1 \cup X_2 = X$, $X_1 \cap X_2 = \emptyset$ and X_i, $i = 1, 2$, is a **dual hierarchy**: if $x, x' \in \mathcal{H}^i$, then either $\mathcal{S}(x) \cap \mathcal{S}(x') = \emptyset$, or $\mathcal{S}(x) \subset \mathcal{S}(x')$ or $\mathcal{S}(x') \subset \mathcal{S}(x)$.

We can define these concepts via matrices. Enumerating $X = \{x_1, ..., x_n\}$ and $\mathcal{H} = \{S_1, ..., S_m\}$, the hypergraph is represented by an $n \times m$ $\{0, 1\}$ incidence matrix $A = [a_{ij}]$ such that $a_{ij} = 1_{\{x_i \in S_j\}}$. The incidence matrix of the dual is A^T, the transpose of A.

Theorem 5. *A hypergraph $\mathcal{X} = (X, \mathcal{H})$ is BvN decomposable if either it forms a bihiearchy or its dual forms a bihierarchy.*

Example 2. *Consider $X = \{a, b, c, d, e, f\}$, and*

$$\mathcal{H} = \{\{a, d\}, \{a, e\}, \{a, f\}, \{b, d\}, \{b, e\}, \{b, f\}, \{c, d\}, \{c, e\}, \{c, f\}\}.$$

The hypergraph $\mathcal{X} = (X, \mathcal{H})$ is in fact a bipartite graph in this case. Even though it does not form a bihierarchy, its dual forms a bihierarchy. In fact, its dual is assignment between three agents and three objects, with only row and column constraints.

Example 3. *Consider $X = \{a, b, c, d, e, f, \alpha, \beta, \delta, \epsilon\}$, and*

$$\mathcal{H} = \{\{a, d, \alpha, \delta\}, \{a, e, \alpha, \epsilon\}, \{a, f\}, \{b, d, \beta, \delta\}, \{b, e, \beta, \epsilon\}, \{b, f\}, \{c, d\}, \{c, e\}, \{c, f\}\}.$$

The hypergraph $\mathcal{X} = (X, \mathcal{H})$ again does not form a bihierarchy, but its dual forms a bihierarchy. In fact, its dual is the 3 by 3 matching, with row and column constraints, and two subrow and two subcolumn constraints.

8. Conclusion

We generalize the Birkhoff-von Neumann theorem so that the implementation of lotteries is possible whenever the set of constraints can be partitioned into two hierarchies. Thus, given any random assignment satisfying constraints in two hierarchies, the assignment can be realized by using a lottery over outcomes each of which satisfies all the constraints. Moreover, we provide a maximal domain result, that indicates that the bi-hierarchical structure is necessary (subject to a technical condition) to guarantee that a random assignment can always be implemented by lotteries over feasible outcomes. We presented several applications, including (i) random assignment mechanisms (especially the probabilistic serial mechanism) under complex constraints, (ii) utility guarantee for problems with multi-unit demand, (iii) fair division, (iv) the minimize-makespan problem, (v) two-sided matching, and (vi) optimal assignment.

As the basic result is applicable to a wide range of situations as exemplified in this paper, we envision that the result will prove useful in other economic and non-economic applications. Finding more applications is an interesting topic left for the future.

Appendix A. Proofs of Theorems 1 and 5

Since Theorem 1 is a special case of Theorem 5, we prove the latter.

A matrix is **totally unimodular** if the determinant of every square submatrix is 0, -1 or $+1$.

Lemma 1. *(Hoffman and Kruskal) If the matrix is totally unimodular, then the vertices of the polytope defined by linear integral constraints are integer valued.*

We can easily use one of the conditions to say that if the incidence matrix of a hypergraph is totally unimodular, then the hypergraph is BvN decomposable.

Lemma 2. *(Ghouila-Houri) A* $\{0,1\}$ *incidence matrix is totally unimodular if and only if each collection of its columns can be partitioned into red and blue columns such that for every row of that collection, the sum of nonzero entries in the red columns differs by at most one from the sum of the nonzero entries in the blue columns.*

Proof of Theorem 5. Suppose first $\mathcal{X}$ forms a bihiearchy, with $\mathcal{H}^1$ and $\mathcal{H}^2$ such that $\mathcal{H}^1 \cup \mathcal{H}^2 = \mathcal{H}$, $\mathcal{H}^1 \cap \mathcal{H}^2 = \emptyset$ and $\mathcal{H}^i$, $i = 1, 2$ is a hierarchy. Let A be the associated incidence matrix. Take any collection of columns of A, corresponding to a subcollection E of $\mathcal{H}$. We shall partition E into two sets, B and R. First, for each $i = 1, 2$, we partition $E \cap \mathcal{H}^i$ into nonempty sets $E^i_1, E^i_2, ..., E^i_{k^i}$ defined recursively as follows: Given $E^i_0 \equiv \emptyset$, for each

$j = 1, ...,$ we let

$$E^i_j := \{S \in (E \cap \mathcal{H}^i) \setminus (\cup_{j'=1}^{j-1} E^i_{j'}) \,|\, \nexists S' \in (E \cap \mathcal{H}^i) \setminus (\cup_{j'=1}^{j-1} E^i_{j'} \cup S) \text{ such that } S' \supset S\}.$$

(The non-emptiness requirment means that once all sets in $E \cap \mathcal{H}^i$ are accounted for, the recursive definition stops, which it does at a finite $j = k^i$.) Since $\mathcal{H}^i$ is a hierarchy, any two sets in E^i_j must be disjoint, for each $j = 1, ..., k^i$. Hence, any element of X can belong to at most one set in each E^i_j. Observe next for $j < l$, $\bigcup_{S \in E^i_l} S \subset \bigcup_{S \in E^i_j} S$. In other words, if an element of X belongs to a set in E^i_l, it must also belong to a set in E^i_j for each $j < l$.

We now define sets B and R that partition E:

$$B := \{S \in E | S \in E^i_j, i + j \text{ is an even number }\},$$

and

$$R := \{S \in E | S \in E^i_j, i + j \text{ is an odd number }\}.$$

We call the elements of B "blue" sets, and call the elements of R "red" sets.

Fix any $x \in X$. If x belongs to any set in $E \cap \mathcal{H}^1$, then it must belong to exactly one set $S^1_j \in E^1_j$, for each $j = 1, ..., l$ for some $l \leq k^1$. These sets alternate in colors in $j = 1, 2..,$ starting with blue: S^1_1 is blue, S^1_2 is red, S^1_3 is blue, and so forth. Hence, the number of blue sets in $E \cap \mathcal{H}^1$ containing x either equals or exceeds by one the number of red sets in $E \cap \mathcal{H}^1$ containing x. By the same reasoning, if x belongs to any set in $E \cap \mathcal{H}^2$, then it must belong to one set $S^2_j \in E^2_j$, for each $j = 1, ..., m$ for some $m \leq k^2$. These sets alternate in colors in $j = 1, 2..,$ starting with red: S^2_1 is red, S^2_2 is blue, S^2_3 is red, and so forth. Hence, the number of blue sets in $E \cap \mathcal{H}^1$ containing x is less by one than or equal to the number of red sets in $E \cap \mathcal{H}^1$ containing x. In sum, the number of blue sets in E containing x differs at most by one from the number of red sets in E containing x. It thus follows that an arbitrary submatrix of A is equitably bicolorable. Hence, A is totally unimodular. By Hoffman and Kruskal, for any $f \in \mathcal{F}$, the vertices of the set

$$\{f' \in \mathcal{F} | \lfloor f(S) \rfloor \leq f'(S) \leq \lceil f(S) \rceil, \forall S \in \mathcal{H}\}$$

are integer valued, so the hypergraph $\mathcal{X}$ is BvN decomposable.

We next consider the case where the dual of $\mathcal{X}$ forms a bihierarchy. To this end, consider a hypergraph $\mathcal{X}^* = (X^*, \mathcal{H}^*)$ such that $X^* = \mathcal{H}$ and $\mathcal{H}^* = X$. That is, X^* is a finite ground set whose elements share the same labels as the hyperedges in $\mathcal{H}$, and $\mathcal{H}^*$ is a collection of subsets of $\mathcal{H}^*$ that have the same labels as X. Assume that $S \in X^*$ is an element of $x \in \mathcal{H}^*$ in $\mathcal{X}^*$ if and only if x is an element of S in $\mathcal{X}$. The fact that the dual of $\mathcal{X}$ forms a bihierarchy means that (the primal of) $\mathcal{X}^*$ forms a bihierarchy. The argument made above then implies that the incidence matrix A^* associated with $\mathcal{X}^*$ is totally unimodular. Since this matrix coincides with the incidence matrix of the dual of $\mathcal{X}$, $A^* = A^T$. Since a transpose of a totally unimodular matrix is totally unimodular in general by definition, it follows that the incidence matrix A of $\mathcal{X}$ must be also totally unimodular. Hence, $\mathcal{X}$ is BvN decomposable. □

A.1. **Computable Algorithm for Bihierarchy.** For the case of bihierarchy, we provide the following algorithm to find a decomposition. Observe that the algorithm gives a constructive proof of the Theorem for the bihierarchy case. We say a set $S \subset X$ is **integral** [resp. **nonintegral**] (under f) if $f(S) \in \mathbb{Z}$ and an element $x \in X$ is **integral** [resp. **nonintegral**] (under f) if $f(x) \in \mathbb{Z}$ [resp. $f(x) \notin \mathbb{Z}$]. In case there is no confusion, we suppress the qualifier inside the parenthesis.

We define the **degree of integrality** of f with respect to $\mathcal{H}$:

$$\deg[f(\mathcal{H})] := \#\{S \in \mathcal{H} | f(S) \in \mathbb{Z}\} + \#\{x \in X | \{x\} \notin \mathcal{H}, f(x) \in \mathbb{Z}\}.$$

Lemma 3. *(Decomposition) Suppose a hypergraph $\mathcal{X} = (X, \mathcal{H})$ forms bihierarchy. Then, for any $f \in \mathcal{F}$, there exist f^1 and f^2, both in $\mathcal{F}$, and $\gamma \in (0,1)$ such that*
(i) $f = \gamma f^1 + (1-\gamma) f^2$:
(ii) $f^1(S), f^2(S) \in [\lfloor f(S) \rfloor, \lceil f(S) \rceil]$, $\forall S \in \mathcal{H}$.
(iii) $deg[f^i(\mathcal{H})] > deg[f(\mathcal{H})]$ for $i = 1, 2$.

Our algorithm consists of two parts: Fission algorithm and Decomposition algorithm.

□ **Fission Algorithm**

1. Within-Hierarchy Unnesting Phase

(1) Let $\mathcal{C}_0^i$, $i = 1, 2$, be the collection of all integral sets of $\mathcal{H}^i$ under f.
(2) **Step** $t = 1, \ldots$**:** Find sets S, S' in $\mathcal{C}_{t-1}^i$ such that $S \subset S'$.
  (a) If no such sets exist, *move to* the **Dividing Phase**.
  (b) If such sets exist, *remove S' from $\mathcal{C}_{t-1}^i$ and replace it by $S' \setminus S$, and call the resulting collection $\mathcal{C}_t^i$, and iterate to Step $t+1$.*
(3) The unnesting phase stops in finite interations for each $\mathcal{H}^i$, $i = 1, 2$. *Call the resulting collections $\mathcal{D}_0^1$ and $\mathcal{D}_0^2$, respectively, and move to* **Dividing Phase**.

2. Dividing Phase

(1) **Step** $t = 1, \ldots$. Find $S \in \mathcal{D}_{t-1}^1$ and $S' \in \mathcal{D}_{t-1}^2$ such that $S \setminus S'$, $S' \setminus S$ and $S \cap S'$ are all nonempty and integral.
  (a) If no such sets exist, *stop and move to the Cross-Hierarchy Unnesting Phase.*

(b) If such sets exist, then *remove them from* $\mathcal{D}^1_{t-1} \cup \mathcal{D}^2_{t-1}$ *and add* $S \setminus S'$ *and* $S \cap S'$ *to* $\mathcal{D}^1_{t-1}$*, and add* $S' \setminus S$ *to* $\mathcal{D}^2_{t-1}$*, and call them* $\mathcal{D}^1_t$ *and* $\mathcal{D}^2_t$*, respectively, and move to Step* $t+1$.

(2) This process ends in finite steps. Call the resulting collections, $\mathcal{G}^1_0$ and $\mathcal{G}^2_0$, respectively, and move the Cross-Hierarchy Unnesting Phase.

3. Cross-Hierarchy Unnesting Phase

(1) **Step** $t = 1, \ldots.$ Find $S \in \mathcal{G}^i_{t-1}$ and $S' \in \mathcal{G}^j_{t-1}$, $i = 1, 2, j = 3 - i$, such that $S \subset S'$.
   (a) If no such sets exist, *stop and move to the Decomposition Algorithm.*
   (b) If such sets exist, then remove S from $\mathcal{G}^i_{t-1}$ and S' from $\mathcal{G}^j_{t-1}$, and add S and $S' \setminus S$ to $\mathcal{G}^j_{t-1}$, and call the resulting collections $\mathcal{G}^i_t$ and $\mathcal{G}^j_t$, respectively, and iterate to Step $t+1$.

(2) This process ends in finite steps. Call the resulting collections, $\overline{\mathcal{H}}^1$ and $\overline{\mathcal{H}}^2$, respectively, and move to the Decomposition Algorithm.

We make several observations on $\overline{\mathcal{H}}^1$ and $\overline{\mathcal{H}}^2$:

□ **Observations**

(1) The sets in $\overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$ are integral under f. Conversely, if all sets in $\overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$ are integral under any $f' \in \mathcal{F}$, then any set in $\mathcal{H}$ that is integral under f is also integral under f'. This latter observation follows from the fact that, by the Fission Algorithm, any set in $\mathcal{H}$ that is integral under f can be expressed as a union of disjoint sets in $\overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$.

(2) The sets within each $\overline{\mathcal{H}}^i$, $i = 1, 2$, are disjoint: Any nesting within each hierarchy is eliminated by the end of the within-unnesting phases, and the dividing phase and cross-unnesting phases do not create any new nesting within each hierarchy.

(3) There are no two sets $S, S' \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$ such that $S \subset S'$: This follows from Observation (2) and the cross-unnesting phase.

(4) Each element of X could be in at most one set of each $\overline{\mathcal{H}}^i$: This follows from Observation (2).

(5) If there are $S, S' \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$ such that $S \setminus S'$, $S' \setminus S$ and $S \cap S'$ are all nonempty, then all these sets must be non-integral: This follows from the original bihierarchy structure and the dividing phase.

Let $\mathcal{N} = \{x \in X | \nexists S \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2 \text{ s.t. } x \in S\}$ be the elements of X not in any integral sets in $\mathcal{C}$.

□ **Decomposition Algorithm**

(1) **Step 0:**

(a) If $\mathcal{N}$ is non-empty, then *circle any non-integral element of* $\mathcal{N}$ *and proceed to* **Termination - Dead End.**

(b) If there does not exist any non-integral element in $\mathcal{N}$, then *move to* **Step 1.**

(2) **Step 1:**

(a) If no set in $\overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$ contains any non-integral element, then *stop the Decomposition Algorithm.* [Every element in X is then integral.]

(b) If there exists a non-integral element x_1 in some $S_1 \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$, then *circle it and proceed to* **Step 2.**

(3) **Step** $t = 2, ...$**:**

(a) In case $x_{t-1} \in S_{t-1}$ is circled, where $S_{t-1} \in \overline{\mathcal{H}}^i$, $i = 1, 2$:

(i) Find a non-integral element in $S_{t-1} \setminus S_{t-2}$ different from x_{t-1}. (Let $S_0 \equiv \emptyset$.) Such an element exists since S_{t-1} is integral by Observation (1), and since, by Observation (5), $S_{t-1} \setminus S_{t-2}$ is non-integral whenever $S_{t-1} \cap S_{t-2}$ is non-empty. If there is any such element that is an element of S_i for some $i \in \{1, \dots, t-2\}$, then *erase the markings (e.g., circle and square) of all* x_j*, for* $j = 1, ..., i-1$, and *proceed to* **Termination - Cycle** (note that $S_i \in \overline{\mathcal{H}}^{3-i}$ by observation (4)). Otherwise, choose one element $x_t \neq x_{t-1}$, $x_t \in S_{t-1} \setminus S_{t-2}$ arbitrarily and square it.

(ii) If no set other than S_{t-1} contains x_t, then *stop and proceed to* **Termination - Chain**.

(iii) Suppose another set, S_t, contains x_t. Then, $S_t \in \overline{\mathcal{H}}^{3-i}$ since, by Observation (2), all sets in $\overline{\mathcal{H}}^i$ are disjoint. Also, by the construction in Step 3(a)i above, $S_t \neq S_i$ for any $i = \{1, \dots, t-1\}$. *Proceed to* **Step** $t+1$.

(b) In case $x_{t-1} \in S_{t-1}$ is squared, where $S_{t-1} \in \overline{\mathcal{H}}^i$, $i = 1, 2$:

(i) Find a non-integral element in $S_{t-1} \setminus S_{t-2}$ different from x_{t-1}. (Let $S_0 \equiv \emptyset$.) Such an element exists since S_{t-1} is integral by Observation (1), and since, by Observation (5), $S_{t-1} \setminus S_{t-2}$ is non-integral whenever $S_{t-1} \cap S_{t-2}$ is non-empty. If there is any such element that is an element of S_i for some $i \in \{1, \dots, t-2\}$, then *erase the markings (e.g., circle and square) of all* x_j*, for* $j = 1, ..., i-1$, and *proceed to* **Termination - Cycle** (note that $S_i \in \overline{\mathcal{H}}^{3-i}$ by observation (4)). Otherwise, choose one element $x_t \neq x_{t-1}$, $x_t \in S_{t-1} \setminus S_{t-2}$ arbitrarily and circle it.

(ii) If no set other than S_{t-1} contains x_t, then *stop and proceed to* **Termination - Chain**.

(iii) Suppose another set, S_t, contains x_t. Then, $S_t \in \overline{\mathcal{H}}^{3-i}$ since, by Observation (2), all sets in $\overline{\mathcal{H}}^i$ are disjoint. Also, by the construction in Step 3(b)i above, $S_t \neq S_i$ for any $i = \{1, \ldots, t-1\}$. *Proceed to* **Step** $t+1$.

Since X is finite, we eventually reach the Termination Step.

(4) **Termination - Dead End**
 (a) Construct a mapping f^1 which is the same as f, except at the circled element x. $f^1(x)$ is obtained by raising $f(x)$ as high as possible without "crossing" any constraint in $\mathcal{H}$. This amount α is positive.
 (b) Construct a mapping f^2 which is the same as f, except at the circled element x. $f^2(x)$ is obtained by reducing $f(x)$ as low as possible without "crossing" any constraint in $\mathcal{H}$. The amount of reduction β is positive.
 (c) Set γ by $\gamma\alpha + (1-\gamma)(-\beta) = 0$, i.e., $\gamma = \frac{\beta}{\alpha+\beta}$.
 (d) The decomposition of f into $f = \gamma f^1 + (1-\gamma) f^2$ satisfies the requirements of the Lemma by construction.

(5) **Termination - Chain**
 (a) Observe that the sets in $\overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$ containing either circled or squared term does not form a cycle. By construction, each of these sets contains precisely one circled element and one squared element.
 (b) Construct a mapping f^1 in $\mathcal{F}$ which is the same as f, except at the circled elements, X_C, and at squared elements, X_S. For each $x \in X_C$, set $f^1(x) = f(x) + \alpha$, and for each $x \in X_S$, set $f^1(x) = f(x) - \alpha$, where $\alpha > 0$ is the largest number that still satisfies all constraints in $\mathcal{H}$. By construction, $f^1(S) = f(S)$ for all $S \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$. Observation (1) then ensures that $f^1(S) = f(S)$ for each integral set $S \in \mathcal{H}$.
 (c) Construct a mapping f^2 which is the same as f, except at the circled elements, X_C, and at squared elements, X_S. For each $x \in X_C$, set $f^2(x) = f(x) - \beta$, and for each $x \in X_S$, set $f^2(x) = f(x) + \beta$, where $\beta > 0$ is the largest number that still satisfies all constraints in $\mathcal{H}$. By construction, $f^2(S) = f(S)$ for all $S \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$. Observation (1) then ensures that $f^2(S) = f(S)$ for each integral set $S \in \mathcal{H}$.
 (d) Set γ by $\gamma\alpha + (1-\gamma)(-\beta) = 0$, i.e., $\gamma = \frac{\beta}{\alpha+\beta}$.
 (e) The decomposition of f into $f = \gamma f^1 + (1-\gamma) f^2$ satisfies the requirements of the Lemma by construction.

(6) **Termination - Cycle**

(a) Observe that the sets $S_t, ..., S_{t+k}$ form a cycle (recall that $t = 1$, unless the construction involves the erasing of markings). Since these sets altenate between $\overline{\mathcal{H}}^i$ and $\overline{\mathcal{H}}^{3-i}$, the order of the cycle must be even. Further, by construction, each of these sets contains precisely one circled element and one squared element.

(b) Construct a mapping f^1 which is the same as f, except at the circled elements, X_C, and at squared elements, X_S. For each $x \in X_C$, set $f^1(x) = f(x) + \alpha$, and for each $x \in X_S$, set $f^1(x) = f(x) - \alpha$, where $\alpha > 0$ is the largest number that still satisfies all constraints in $\mathcal{H}$. By construction, $f^1(S) = f(S)$ for all $S \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$. Observation (1) then ensures that $f^1(S) = f(S)$ for all integral set $S \in \mathcal{H}$.

(c) Construct a mapping f^2 which is the same as f, except at the circled elements, X_C, and at squared elements, X_S. For each $x \in X_C$, set $f^2(x) = f(x) - \beta$, and for each $x \in X_S$, set $f^2(x) = f(x) + \beta$, where $\beta > 0$ is the largest number that still satisfies all constraints in $\mathcal{H}$. By construction, $f^2(S) = f(S)$ for all $S \in \overline{\mathcal{H}}^1 \cup \overline{\mathcal{H}}^2$. Observation (1) then ensures that $f^2(S) = f(S)$ for all integral set $S \in \mathcal{H}$.

(d) Set γ by $\gamma\alpha + (1-\gamma)(-\beta) = 0$, i.e., $\gamma = \frac{\beta}{\alpha+\beta}$.

(e) The decomposition of f into $f = \gamma f^1 + (1-\gamma) f^2$ satisfies the requirements of the Lemma by construction.

Appendix B. Proof of Theorem 2

In order to prove the Theorem, we study several cases.

- Assume there is $S \in \mathcal{H}$ such that $S = N' \times O'$ where $2 \le |N'| < |N|$ and $2 \le |O'| < |O|$. Let $\{i, j\} \times \{a, b\} \subseteq S$, $k \notin N'$ and $c \notin O'$ (observe that such $i, j, k \in N$ and $a, b, c \in O$ exist by the assumption of this case). Consider the matrix P where

$$P_{ia} = P_{ic} = P_{jb} = P_{kb} = P_{kc} = 0.5,$$

and all other entries are 0. This matrix P satisfies all the above constraints. We can show the following claim.

Claim 1. *There exists no convex decomposition of P into matrices each of which that satisfies conditions in Theorem 1.*

Proof. Suppose the contrary. Then there exists an integral matrix X that is part of decomposition of P, with

$$X_{ia} = 1. \tag{21}$$

Then, since the entries in row i should sum up to at most one by the assumption, we have $X_{ic} = 0$. Since $P_{ic} + P_{kc} = 1$, this implies $X_{kc} = 1$. Since the entries in row k should sum up to at most one by the assumption, we have $X_{kb} = 0$. Since $P_{jb} + P_{kb} = 1$, this implies

$$X_{jb} = 1. \tag{22}$$

This is a contradiction, since (21) and (22) imply that $X_S \geq X_{ia} + X_{jb} = 2 > 1 = \lceil P_S \rceil$. □

- Assume there is $S \in \mathcal{H}$ such that, for some $i, j \in N$ and $a, b \in O$, we have $(i, a), (j, b) \in S$ with $i \neq j$ and $a \neq b$, and $(i, b) \notin S$. Consider the matrix P where

$$P_{ia} = P_{ib} = P_{jb} = 0.5,$$

and all other entries are 0. An argument analogous to the proof of Claim 1 shows that this matrix has no convex decomposition satisfying conditions in Theorem 1.

By the above arguments, it suffices to consider a case in which all constraint sets in $\mathcal{H}$ have one of the following forms.

(1) $\{i\} \times O'$ where $i \in N$ and $O' \subseteq O$,
(2) $N' \times O$ where $N' \subseteq N$,
(3) $N' \times \{a\}$ where $a \in O$ and $N' \subseteq N$,
(4) $N \times O'$ where $O' \subseteq O$.

Therefore it suffices to consider the following cases.

(1) Assume that there are $S', S'' \in \mathcal{H}$ such that $S' = \{i\} \times O'$ and $S'' = \{i\} \times O''$ for some $i \in N$ and some $O', O'' \subset O$, $S' \cap S'' \neq \emptyset$ and S' is neither a subset nor a superset of S''. Then we can find $a, b, c \in O$ such that $a \in O' \setminus O''$, $b \in O' \cap O''$ and $c \in O'' \setminus O'$. Fix $j \neq i$. Consider the matrix P where

$$P_{ia} = P_{ib} = P_{ic} = P_{ja} = P_{jc} = 0.5,$$

and all other entries are 0. An argument analogous to the proof of Claim 1 shows that this matrix has no convex decomposition satisfying conditions in Theorem 1.

(2) Assume that there are $S', S'' \in \mathcal{H}$ such that $S' = N' \times O$ and $S'' = N'' \times O$ for some $N', N'' \subset N$, $S' \cap S'' \neq \emptyset$ and S' is neither a subset nor a superset of S''.

In such a case, we can find $i, j, k \in N$ such that $i \in N' \setminus N''$, $j \in N' \cap N''$ and $k \in N'' \setminus N'$. Fix $a, b \in O$ and consider the matrix P where

$$P_{ib} = P_{ja} = P_{kb} = 0.5,$$

and all other entries are 0. An argument analogous to the proof of Claim 1 shows that this matrix has no convex decomposition satisfying conditions in Theorem 1.

(3) Assume that there are $S', S'' \in \mathcal{H}$ such that $S' = N' \times \{a\}$ and $S'' = N'' \times \{a\}$ for some $a \in O$ and some $N', N'' \subset N$, $S' \cap S'' \neq \emptyset$ and S' is neither a subset nor a superset of S''. This is a symmetric situation with Case 1, so an analogous argument as before goes through.

(4) Assume that there are $S', S'' \in \mathcal{H}$ such that $S' = N \times O'$ and $S'' = N \times O''$ for some $O', O'' \subset O$, $S' \cap S'' \neq \emptyset$ and S' is neither a subset nor a superset of S''. This is a symmetric situation with Case 2, so an analogous argument as before goes through.

Appendix C. Proof of Propositions 1, 2, and 3

Proof of Proposition 1. **"Only if" part.** First note that the following property holds.

Claim 2. *$\rhd_1$ and $\rhd_2$ are transitive, that is,*

$$(k, c) \rhd_1 (j, b), (j, b) \rhd_1 (i, a) \Rightarrow (k, c) \rhd_1 (i, a),$$
$$(k, c) \rhd_2 (j, b), (j, b) \rhd_2 (i, a) \Rightarrow (k, c) \rhd_2 (i, a).$$

Proof. Suppose $(k, c) \rhd_1 (j, b)$ and $(j, b) \rhd_1 (i, a)$. Then, by definition of $\rhd_1$, we have $i = j = k$ and (i) $c \succ_i b$ since $(k, c) \rhd_1 (i, b)$ and (ii) $b \succ_i a$ since $(j, b) \rhd_1 (i, a)$. Thus $c \succ_i a$. Since $(j, b) \rhd_1 (i, a)$, we have $P_{ia} > 0$. Therefore $(k, c) \rhd_1 (i, a)$ by definition of $\rhd_1$.

Suppose $(k, c) \rhd_2 (j, b)$ and $(j, b) \rhd_2 (i, a)$. Then $\nu(k, c) \subseteq \nu(j, b)$ and $\nu(j, b) \subseteq \nu(i, a)$ by property (2). Hence $\nu(k, c) \subseteq \nu(i, a)$ which is equivalent to $(k, c) \rhd_2 (i, a)$, completing the proof by property (2). □

To show the "only if" part of the Proposition, suppose $\rhd$ is strongly cyclic. By Claim 2, there exists a cycle of the form

$$(i_0, b_0) \rhd_1 (i_0, a_0) \rhd_2 (i_1, b_1) \rhd_1 (i_1, a_1) \rhd_2 (i_2, b_2) \rhd_1 (i_2, a_2) \rhd_2 \cdots \rhd_1 (i_k, a_k) \rhd_2 (i_0, b_0),$$

in which every pair (i, a) in the cycle appears exactly once except for (i_0, b_0) which appears exactly twice, namely in the beginning and in the end of the cycle. Then there exists $\delta > 0$

such that a matrix Q defined by

$$Q_{ia} = \begin{cases} P_{ia} + \delta & \text{if} \quad (i,a) \in \{(i_0,b_0),(i_1,b_1),\dots,(i_k,b_k)\}, \\ P_{ia} - \delta & \text{if} \quad (i,a) \in \{(i_0,a_0),(i_1,a_1),\dots,(i_k,a_k)\}, \\ P_{ia} & \text{otherwise}, \end{cases}$$

is in $\mathcal{P}_{\mathcal{E}}$. Since $\delta > 0$ and $b_l \succ_{i_l} a_l$ for every $l \in \{0,1,\dots,k\}$, Q ordinally dominates P. Therefore P is not ordinally efficient.

"If" part. Suppose P is ordinally inefficient. Then, there exists $Q \in \mathcal{P}_{\mathcal{E}}$ which ordinally dominates P. We then prove that $\rhd$, given (Γ, P), must be strongly cyclic.

(1) **Step 1: Initiate a cycle.**
   (a)

   Claim 3. *There exist* $(i_0,a_0),(i_1,a_1) \in N \times O$ *such that* $i_0 = i_1$, $P_{i_1a_1} < Q_{i_1a_1}$ *and* $(i_1,a_1) \rhd_1 (i_0,a_0)$ *given* (Γ, P).

   Proof. Since Q ordinally dominates P, there exists $(i_1,a_1) \in N \times O$ such that $Q_{i_1a_1} > P_{i_1a_1}$ and $Q_{i_1a} = P_{i_1a}$ for all $a \succ_{i_1} a_1$. So there exists $a_0 \prec_{i_1} a_1$ with $P_{i_1a_0} > Q_{i_1a_0} \geq 0$ since $P_{\{i_1\}\times N} = Q_{\{i_1\}\times N}$ by assumption. Hence, we have $(i_1,a_1) \rhd_1 (i_1,a_0) = (i_0,a_0)$ given (Γ, P). □

   (b) If $(i_0,a_0) \in \nu(i_1,a_1)$, then $(i_0,a_0) \rhd_2 (i_1,a_1) \rhd_1 (i_0,a_0)$, so we have a strong cycle and we are done.
   (c) Else, circle (i_1,a_1) and go to Step 2.

(2) **Step $t+1$ ($t \in \{1,2\dots\}$): Consider the following cases.**
   (a) Suppose (i_t,a_t) is circled.
      (i)

      Claim 4. *There exists* $(i_{t+1},a_{t+1}) \in \nu(i_t,a_t)$ *such that* $P_{i_{t+1}a_{t+1}} > Q_{i_{t+1}a_{t+1}}$. *Hence,* $(i_{t+1},a_{t+1}) \rhd_2 \nu(i_t,a_t)$.

      Proof. Note that $\nu(i_t,a_t) \subsetneq N \times O$ since if $\nu(i_t,a_t) = N \times O$, then there exists $(i_{t'},a_{t'})$ with $t' < t$ and $(i_{t'},a_{t'}) \in \nu(i_t,a_t)$, so we have terminated the algorithm. Thus we have $\sum_{(i,a)\in\nu(i_t,a_t)} P_{ia} = \overline{q}_{\nu(i_t,a_t)}$. Since $P_{i_ta_t} < Q_{i_ta_t}$, there exists $(i_{t+1},a_{t+1}) \in \nu(i_t,a_t)$ such that $P_{i_{t+1}a_{t+1}} > Q_{i_{t+1}a_{t+1}}$. □

      (ii) If $(i_{t'},a_{t'}) \in \nu(i_{t+1},a_{t+1})$ for $t' < t$, then we have a strong cycle, $(i_{t'},a_{t'}) \rhd (i_{t+1},a_{t+1}) \rhd \dots \rhd (i_{t'},a_{t'})$, and at least one $\rhd$ is $\rhd_1$, so we are done.

(iii) Else, square (i_{t+1}, a_{t+1}) and move to the next step.

(b) Case 2: Suppose (i_t, a_t) is squared.

(i)

Claim 5. *There exists* $(i_{t+1}, a_{t+1}) \in \nu(i_t, a_t)$ *such that* $i_{t+1} = i_t$, $P_{i_{t+1}a_{t+1}} < Q_{i_{t+1}a_{t+1}}$, *and* $(i_{t+1}, a_{t+1}) \rhd_1 \nu(i_t, a_t)$.

Proof. Since (i_t, a_t) is squared, by Claim 4, $P_{i_t a_t} > Q_{i_t a_t}$. Since Q ordinally dominates P, there must be $(i_{t+1}, a_{t+1}) \in \nu(i_t, a_t)$ with $i_{t+1} = i_t$ such that $P_{i_{t+1}a_{t+1}} < Q_{i_{t+1}a_{t+1}}$, and $a_{t+1} \succ_{i_t} a_t$. Since $P_{i_t a_t} > Q_{i_t a_t} \geq 0$, we thus have $(i_{t+1}, a_{t+1}) \rhd_1 \nu(i_t, a_t)$. □

(ii) If $(i_{t'}, a_{t'}) \in \nu(i_{t+1}, a_{t+1})$ for $t' \leq t$, then we have a strong cycle, $(i_{t'}, a_{t'}) \rhd (i_{t+1}, a_{t+1}) \rhd ... \rhd (i_{t'}, a_{t'})$, and at least one $\rhd$ is $\rhd_1$, so we are done.

(iii) Else, circle (i_{t+1}, a_{t+1}) and move to the next step.

The process must end in finite steps and, at the end we must have a strong cycle. □

Proof of Proposition 2. Although the proof is a relatively simple modification of Theorem 1 of Bogomolnaia and Moulin (2001), we present the proof for completeness. We prove the claim by contradiction. Suppose that $PS(\succ)$ is ordinally inefficient for some $\succ$. Then, by Proposition 1 and Claim 2 there exists a strong cycle

$$(i_0, b_0) \rhd_1 (i_0, a_0) \rhd_2 (i_1, b_1) \rhd_1 (i_1, a_1) \rhd_2 (i_2, b_2) \rhd_1 (i_2, a_2) \rhd_2 \cdots \rhd_1 (i_k, a_k) \rhd_2 (i_0, b_0),$$

in which every pair (i, a) appears exactly once except for (i_0, b_0) which appears exactly twice, namely in the beginning and the end of the cycle. Let v^l and w^l be the steps of the symmetric simultaneous eating algorithm at which (i_l, a_l) and (i_l, b_l) become unavailable, respectively (that is, $(i_l, a_l) \in S^{v_l - 1} \setminus S^{v_l}$ and $(i_l, a_l) \in S^{w_l - 1} \setminus S^{w_l}$.) Since $(i_l, b_l) \rhd_1 (i_l, a_l)$, by the definition of the algorithm we have $w^l < v^l$ for each $l \in \{0, 1, \ldots, k\}$. Also, by $(i_l, a_l) \rhd_2 (i_{l+1}, b_{l+1})$, we have $v^l \leq w^{l+1}$ for any $l = \{0, 1, \ldots, k\}$ (with notational convention $(i_{k+1}, a_{k+1}) = (i_0, a_0)$.) Combining these inequalities we obtain $w^0 < v^0 \leq w^1 < v^1 \leq \cdots \leq w^k < v^k \leq w^{k+1} = w^0$, a contradiction. □

Proof of Proposition 3. Let $P = PS(\succ)$. Fix $i \in N$ and let O be ordered in the decreasing order of $\succ_i$, that is, $a_1 \succ_i a_2 \succ_i \cdots \succ_i a_{|N|}$. Let v_1 be the step in which i stops receiving probability share of a_1. In that step we have $P_{ia_1} = P^{v_1}_{ia_1} = t^{v_1}$ and there is $S_1 \in \mathcal{H}_O$ such that $(i, a_1) \in S_1$ and $P^{v_1}_{S_1} = \overline{q}_{S_1}$. Suppose $P_{ja_1} > P_{ia_1}$ for some $j \in N$. Then we have $(j, a_1) \notin S_1$ since $P_{ja_1} \leq t^{v_1} = P_{ia_1}$ if $(j, a_1) \in S_1$ by definition of the algorithm. Also $S_1 = N_1 \times \{a_1\}$ for some $N_1 \subseteq N$ with $i \in N_1$ and $j \notin N_1$ since $(i, a_1) \in S_1$ and $(j, a_1) \notin S_1$. Let Q be defined as in (6). Then, since $i \in N_1$ and $j \notin N_1$,

$$\begin{aligned} Q_{S_1} &\geq \sum_{k \in N_1} P_{ka_1} - P_{ia_1} + P_{ja_1} \\ &> \sum_{k \in N_1} P_{ka_1} \\ &\geq \sum_{k \in N_1} P^{v_1}_{ka_1} \\ &= P^{v_1}_{S_1} = \overline{q}_{S_1}, \end{aligned}$$

which implies that $Q \notin \mathcal{P}_{\mathcal{E}}$.

Let $l \geq 2$ and v_l be the step in which i stops receiving probability share of a_l. In that step we have $\sum_{m=1}^{l} P_{ia_m} = \sum_{m=1}^{l} P^{v_l}_{ia_m} = t^{v_l}$ and there is $S_l \in \mathcal{H}_O$ such that $(i, a_l) \in S_l$ and $P^{v_l}_{S_l} = \overline{q}_{S_l}$. Suppose $\sum_{m=1}^{m'} P_{ja_m} \leq \sum_{m=1}^{m'} P_{ia_m}$ for all $m' \leq l-1$ and $\sum_{m=1}^{l} P_{ja_m} > \sum_{m=1}^{l} P_{ia_m}$ for some $j \in N$. Then we have $(j, a_l) \notin S_l$ since $\sum_{m=1}^{l} P_{ja_m} \leq t^{v_l} = \sum_{m=1}^{l} P_{ia_m}$ if $(j, a_l) \in S_l$ by definition of the algorithm. Also $S_l = N_l \times \{a_l\}$ for some $N_l \subseteq N$ with $i \in N_l$ and $j \notin N_l$ since $(i, a_l) \in S_l$ and $(j, a_l) \notin S_l$. Let Q be defined as in (6). Then, since $i \in N_l$ and $j \notin N_l$,

$$\begin{aligned} Q_{S_l} &\geq \sum_{k \in N_l} P_{ka_l} - P_{ia_l} + P_{ja_l} \\ &> \sum_{k \in N_l} P_{ka_l} \\ &\geq \sum_{k \in N_l} P^{v_l}_{ka_l} \\ &= P^{v_l}_{S_l} = \overline{q}_{S_l}, \end{aligned}$$

which implies that $Q \notin \mathcal{P}_{\mathcal{E}}$. By induction, we complete the proof. □

References

Abdulkadiroğlu, A. (2005): "College Admission with Affirmative Action," *International Journal of Game Theory*, 33, 535–549.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2005): "The New York City High School Match," *American Economic Review Papers and Proceedings*, 95, 364–367.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2005): "The Boston Public School Match," *American Economic Review Papers and Proceedings*, 95, 368–372.

Abdulkadiroğlu, A., and T. Sönmez (1999): "House Allocation with Existing Tenants," *Journal of Economic Theory*, 88, 233–260.

——— (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Bezáková, I., and V. Dani (2005): "Allocating Indivisible Goods," *ACM SIGecom Exchanges*, 5, 11–18.

Birkhoff, G. (1946): "Three Observations on Linear Algebra," *Revi. Univ. Nac. Tucuman, ser A*, 5, 147–151.

Bogomolnaia, A., and H. Moulin (2001): "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 100, 295–328.

Budish, E. (2008): "A New Solution to the Course-Allocation Problem: Approximate Competitive Equilibrium from Approximately Equal Incomes," Harvard University, Unpublished mimeo.

Budish, E., and E. Cantillon (2008): "Strategic Behavior in Multi-unit Assignment Problems: Theory and Evidence from Course Allocations," Unpublished mimeo.

Che, Y.-K., and F. Kojima (2008): "Asymptotic Equivalence of the Random Priority and Probabilistic Serial Mechanisms," Unpublished mimeo.

de Werra, D. (1984): "Preemptive Scheduling, Linear Programming and Network Flows," *SIAM Journal of Algebraic and Descrite Methods*, 5, 11–20.

Ehlers, L., and B. Klaus (2003): "Coalitional strategy-proof and resource-monotonic solutions for multiple assignment problems," *Social Choice and Welfare*, 21, 265–280.

Hatfield, J. W. (2008): "Strategy-Proof and Nonbossy Quota Allocations," Unpublished mimeo.

Hylland, A., and R. Zeckhauser (1979): "The Efficient Allocation of Individuals to Positions," *Journal of Political Economy*, 87, 293–314.

Kojima, F. (2008a): "Games of school choice under the Boston mechanism with general priority structures," *Social Choice and Welfare*, 31, 357–365.

——— (2008b): "Random Assignment of Multiple Indivisible Objects," forthcoming, *Mathematical Social Sciences*.

Kojima, F., and M. Manea (2008): "Strategy-proofness of the Probabilistic Serial Mechanism in Large Random Assignment Problems," mimeo.

Koopmans, T., and M. Beckman (1957): "Assignment Problems and the Location of Economic Activities," *Econometrica*, 25, 53–76.

Lenstra, J. K., D. B. Shmoys, and E. Tardos (1990): "Approximation Algorithms for Scheduling Unrelated Parallel Machines," *Mathematical Programming*, 46, 259–271.

Lewandowski, J. L., C. L. Liu, and J. W. S. Liu (1986): "An Algorithmic Proof of a Generalization of the Birkhoff-von Neumann Theorem," *Journal of Algorithms*, 7, 323–330.

Milgrom, P. R. (2008): "Assignment Auctions," Unpublished mimeo.

Nemhauser, G., and M. Trick (1998): "Scheduling a Major College Basketball Conference," *Operations Research*, 46, 1–8.

Papai, S. (2001): "Strategyproof and nonbossy multiple assignments," *Journal of Public Economic Theory*, 3, 257–271.

Roth, A. E., T. Sönmez, and U. Ünver (2004): "Kidney Exchange," *Quarterly Journal of Economics*, 119, 457–488.

Shapley, L., and M. Shubik (1972): "The Assignment Game I: The Core," *International Journal of Game Theory*, 1, 111–130.

Shmoys, D. B., and E. Tardos (1993): "An Approximation Algorithm for the Generalized Assignment Problem," *Mathematical Programming*, 62, 461–474.

Sönmez, T., and U. Ünver (2008): "Course Bidding at Business Schools," forthcoming, *International Economic Review*.

von Neumann, J. (1953): "A certain zero-sum two-person game equivalent to the optimal assignment problem," in *Contributions to the theory of games, Vol. 2*, ed. by H. W. Kuhn, and A. W. Tucker. Princeton University Press, Princeton, New Jersey.

Watkins, W., and R. Merris (1974): "Convex Sets of Doubly Stochastic Matrices," *Journal of Combinatorial Theory, Series A*, 16, 129–130.