



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2021

---

## **Partial strategyproofness: Relaxing strategyproofness for the random assignment problem**

Mennle, Timo ; Seuken, Sven

**Abstract:** We present partial strategyproofness, a new, relaxed notion of strategyproofness for studying the incentive properties of non-strategyproof assignment mechanisms. Informally, a mechanism is partially strategyproof if it makes truthful reporting a dominant strategy for those agents whose preference intensities differ sufficiently between any two objects. We demonstrate that partial strategyproofness is axiomatically motivated and yields a parametric measure for “how strategyproof” an assignment mechanism is. We apply this new concept to derive novel insights about the incentive properties of the probabilistic serial mechanism and different variants of the Boston mechanism.

DOI: <https://doi.org/10.1016/j.jet.2020.105144>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-205589>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Mennle, Timo; Seuken, Sven (2021). Partial strategyproofness: Relaxing strategyproofness for the random assignment problem. *Journal of Economic Theory*, 191:105144.

DOI: <https://doi.org/10.1016/j.jet.2020.105144>



# Partial strategyproofness: Relaxing strategyproofness for the random assignment problem <sup>☆</sup>

Timo Mennle <sup>\*</sup>, Sven Seuken

*University of Zurich, Switzerland*

Received 17 July 2020; accepted 1 November 2020

Available online 17 November 2020

---

## Abstract

We present *partial strategyproofness*, a new, relaxed notion of strategyproofness for studying the incentive properties of non-strategyproof assignment mechanisms. Informally, a mechanism is partially strategyproof if it makes truthful reporting a dominant strategy for those agents whose preference intensities differ sufficiently between any two objects. We demonstrate that partial strategyproofness is axiomatically motivated and yields a parametric measure for “how strategyproof” an assignment mechanism is. We apply this new concept to derive novel insights about the incentive properties of the probabilistic serial mechanism and different variants of the Boston mechanism.

© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*JEL classification:* C78; D47; D82

*Keywords:* Mechanism design; Random assignment; Matching; Strategyproofness; Probabilistic serial; Boston mechanism

---

---

<sup>☆</sup> We would like to thank (in alphabetical order) Daniel Abächerli, Atila Abdulkadiroğlu, Haris Aziz, Ivan Balbuzanov, Craig Boutilier, Eric Budish, Gabriel Carroll, Lars Ehlers, Katharina Huesmann, Bettina Klaus, Flip Klijn, Fuhito Kojima, Eric Maskin, Antonio Miralles, Bernardo Moreno, Thayer Morrill, David Parkes, Baharak Rastegari, Alvin Roth, Steffen Schuldenzucker, Arunava Sen, Tayfun Sönmez, William Thomson, and Utku Ünver, as well as various anonymous reviewers for their thoughtful comments and suggestions. We are also thankful for the feedback we received from participants at the following events: EC'13, UECE Lisbon Meetings '13, CIREQ Matching Conference '14, EC'14, SCW'14, IWGTS'14, ESWC'15, GAMES'16. Any errors remain our own. Some of the ideas presented in this paper were also described in a two-page abstract that was published in the conference proceedings of EC'14 (see (Mennle and Seuken, 2014)). This research was supported by the Hasler Foundation under grant #12078 and the SNSF (Swiss National Science Foundation) under grant #156836.

<sup>\*</sup> Corresponding author.

*E-mail addresses:* [mennle@ifi.uzh.ch](mailto:mennle@ifi.uzh.ch) (T. Mennle), [seuken@ifi.uzh.ch](mailto:seuken@ifi.uzh.ch) (S. Seuken).

## 1. Introduction

The assignment problem is concerned with the allocation of indivisible objects to self-interested agents who have private preferences over these objects. Monetary transfers are not permitted, which makes this problem different from auctions and other problems that involve payments. In practice, assignment problems often arise in situations that are of great importance to people's lives; for example, when assigning students to seats at public schools (Abdulkadiroğlu and Sönmez, 2003), graduates to entry level positions (Roth, 1984), or tenants to subsidized housing (Abdulkadiroğlu and Sönmez, 1998).

In this paper, we study *ordinal assignment mechanisms*, which are mechanisms for the assignment problem that take preference orders over objects as input. As mechanism designers, we care specifically about incentives for truthtelling under the mechanisms we design. A mechanism is *strategyproof* if truthtelling is a dominant strategy equilibrium. Participating in a strategyproof mechanism is simple for the agents because it eliminates the need to take the preferences or strategies of other agents into account. Strategyproofness thus yields a robust prediction of equilibrium behavior. These and other advantages explain the popularity of strategyproofness as an incentive concept (Pathak and Sönmez, 2008; Abdulkadiroğlu et al., 2009).

The advantages of strategyproofness, however, come at a cost: Zhou (1990) showed that, in the assignment problem, it is impossible to achieve the optimum with respect to incentives, efficiency, and fairness simultaneously.<sup>1</sup> This makes the assignment problem an interesting mechanism design challenge. For example, the random serial dictatorship mechanism is strategyproof and anonymous, but only *ex-post* efficient. The more demanding ordinal efficiency is achieved by the probabilistic serial mechanism, but any mechanism that achieves ordinal efficiency and symmetry cannot be strategyproof (Bogomolnaia and Moulin, 2001). Finally, rank efficiency, an even stronger efficiency concept, can be achieved with rank value mechanisms.<sup>2</sup> However, rank efficiency is incompatible with strategyproofness, even without additional fairness requirements (Featherstone, 2020). Obviously, strategyproofness is in conflict with many other desiderata, and mechanism designers are therefore interested in studying non-strategyproof mechanisms. This highlights the need for good tools to capture the incentive properties of non-strategyproof mechanisms and to analyze what trade-offs are possible between incentives for truthtelling and other desirable properties.

In practice, non-strategyproof mechanisms are ubiquitous. Examples include variants of the Boston mechanism that are used in many cities for the assignment of seats at public schools,<sup>3</sup> a rank-efficient mechanism for the assignment of teachers to positions through the Teach for America program (Featherstone, 2020), a mechanism for the assignment of on-campus housing at MIT that minimizes the number of unassigned units,<sup>4</sup> and the HBS draft mechanism for the assignment of course schedules at Harvard Business School (Budish and Cantillon, 2012). It is

<sup>1</sup> Specifically, Zhou (1990) showed that no (possibly random and possibly cardinal) assignment mechanism can satisfy strategyproofness, *ex-ante* efficiency, and symmetry.

<sup>2</sup> *Rank efficiency* (Featherstone, 2020) requires that an assignment's rank distribution (i.e., the number of assigned first, second, etc. choices) cannot be first-order stochastically dominated by the rank distribution of any other feasible assignment. This is strictly more demanding than ordinal efficiency.

<sup>3</sup> Variants of the non-strategyproof Boston mechanism are used in Minneapolis, Seattle, Lee County (Kojima and Ünver, 2014), San Diego, Freiburg (Germany) (Dur et al., 2018), and throughout the German state of Nordrhein Westfalen (Basteck et al., 2015).

<sup>4</sup> Source: MIT Division of Student Life, retrieved September 8, 2015: <https://housing.mit.edu/>, and personal communication.

therefore important to understand the incentive properties of these mechanisms beyond the fact that they are “not fully strategyproof.”

The incompatibility of strategyproofness with other desiderata in theory and the prevalence of non-strategyproof assignment mechanisms in practice explain why researchers have been calling for useful relaxations of strategyproofness (Budish, 2012).<sup>5</sup> In this paper, we introduce *partial strategyproofness*, a new, relaxed notion of strategyproofness, which is particularly suited to the analysis of assignment mechanisms. Next, we illustrate the definition of partial strategyproofness with a motivating example, and then we explain our main results and what they mean for mechanism design.

Consider a setting with three agents, conveniently named 1, 2, 3, and three objects,  $a, b, c$ , with unit capacity. Suppose that the agents’ preferences are

$$P_1 : a \succ b \succ c, \tag{1}$$

$$P_2 : b \succ a \succ c, \tag{2}$$

$$P_3 : b \succ c \succ a, \tag{3}$$

and that the non-strategyproof probabilistic serial mechanism (Bogomolnaia and Moulin, 2001) is used to assign the objects. If all agents report truthfully, then agent 1 receives  $a, b, c$  with probabilities  $3/4, 0, 1/4$  respectively. If agent 1 instead reports

$$P'_1 : b \succ a \succ c, \tag{4}$$

then these probabilities change to  $1/2, 1/3, 1/6$ . Observe that whether or not the misreport  $P'_1$  increases agent 1’s expected utility depends on how intensely it prefers  $a$  over  $b$ : If  $u_1(a)$  is close to  $u_1(b)$ , then agent 1 would benefit from the misreport  $P'_1$ . If  $u_1(a)$  is significantly larger than  $u_1(b)$ , then agent 1 would prefer to report truthfully. Specifically, agent 1 prefers to report truthfully if  $3/4u_1(a) \geq u_1(b)$  (assuming  $u_1(c) = 0$ ).

Our definition of partial strategyproofness generalizes the intuition from this motivating example: For some bound  $r$  in  $[0, 1]$ , we say that agent  $i$ ’s utility function  $u_i$  satisfies *uniformly relatively bounded indifference with respect to  $r$*  (URBI( $r$ )) if  $ru_i(a) \geq u_i(b)$  holds whenever  $i$  prefers  $a$  to  $b$  (after appropriate normalization). Fixing a setting (i.e., the number of agents and objects and their capacities), a mechanism is  *$r$ -partially strategyproof (in that setting)* if it makes truthful reporting a dominant strategy for any agent whose utility function satisfies URBI( $r$ ). For example, as we show in Section 5.1, the probabilistic serial mechanism has a *degree of strategyproofness* of exactly  $r = 3/4$  in the setting of the motivating example. Thus, from a market design perspective, we can now give honest and useful strategic advice to agents facing this mechanism in this setting: They are best off reporting their preferences truthfully as long as they value their second choice at least a factor  $3/4$  less than their first choice.

We argue that partial strategyproofness is a natural and useful way to think about the incentive properties of non-strategyproof assignment mechanisms, and we present two main arguments that support this claim:

**Partial strategyproofness has a compelling axiomatic motivation (Sections 3 and 4.1).** We first prove that full strategyproofness can be decomposed into three simple axioms. Each of the

<sup>5</sup> Prior work has already led to some useful concepts for this purpose, e.g., the comparison by *vulnerability to manipulation* (Pathak and Sönmez, 2013) or *strategyproofness in the large* (Azevedo and Budish, 2019). We describe their relation to partial strategyproofness in Appendix E and Section 4.4.

axioms restricts the way in which a mechanism can react when an agent swaps two consecutively ranked objects in its preference report (e.g., from  $P_i : a \succ b$  to  $P'_i : b \succ a$ ).

1. A mechanism is *swap monotonic* if either the swap makes it more likely that the agent receives  $b$  (the object that it claims to prefer), or the mechanism does not change the agent's assignment at all.
2. A mechanism is *upper invariant* if, by swapping  $a$  and  $b$ , the agent cannot affect its chances for objects that it prefers to  $a$ .
3. A mechanism is *lower invariant* if, by swapping  $a$  and  $b$ , the agent cannot affect its chances for objects that it likes less than  $b$ .

In words, swap monotonicity ensures that the mechanism responds to changes in the agent's preference reports by increasing the assignment for the object that the agent claims to prefer, upper invariance essentially means that the agent cannot benefit from truncating its preference list (Hashimoto et al., 2014), and lower invariance is the complement of upper invariance but for less-preferred objects. For our first main result, we show that strategyproofness can be decomposed into these three axioms: A mechanism is strategyproof if and only if it is swap monotonic, upper invariant, and lower invariant (Theorem 1). Intuitively, lower invariance is the least important of the three axioms (see Remark 2 for two formal arguments that support this intuition), and by dropping it, we arrive at the larger class of partially strategyproof mechanisms: We show that in a given setting, a mechanism is  $r$ -partially strategyproof for some  $r > 0$  if and only if it is swap monotonic and upper invariant (Theorem 2). Thus, partial strategyproofness describes the incentive properties of (non-strategyproof) mechanisms that satisfy the first two of the three axioms.

**Partial strategyproofness yields a parametric measure for “how strategyproof” an assignment mechanism is (Sections 4.2 and 4.3).** To quantify the extent to which a mechanism makes truthful reporting a dominant strategy, we consider the maximal value of  $r$  for which the mechanism is  $r$ -partially strategyproof. We call this value the *degree of strategyproofness*. Since  $r$ -partial strategyproofness is equivalent to strategyproofness on the restricted domain of utility functions that satisfy URBI( $r$ ), a lower degree of strategyproofness corresponds to strategyproofness on a smaller restricted domain. We prove maximality of the domain restriction by showing that a utility function satisfies URBI( $r$ ) if and only if truthful reporting is a dominant strategy under *all*  $r$ -partially strategyproof mechanisms for an agent with that utility function (Proposition 1). Thus, without additional information about a mechanism besides the fact that it is  $r$ -partially strategyproof, URBI( $r$ ) is the only set of utility functions for which truthful reporting is guaranteed to be a dominant strategy. In this sense, the degree of strategyproofness is a tight measure for incentive properties.

The degree of strategyproofness parametrizes a *spectrum of incentive properties*: Since all utility functions satisfy URBI(1), full strategyproofness is obviously the upper limit concept for  $r$ -partial strategyproofness as  $r$  approaches 1. Regarding the lower limit, *lexicographic dominance (LD)* (Cho, 2018) is the weakest among the common dominance notions,<sup>6</sup> and the corresponding *LD-strategyproofness* is therefore a minimal notion of strategyproofness for random assignment mechanisms. We prove that LD-strategyproofness is the lower limit concept for

<sup>6</sup> Lexicographic dominance is implied by *trivial, deterministic, sure thing*, and *first-order stochastic dominance*, but it is the only one among them under which all pairs of lotteries over objects are comparable (Aziz et al., 2013a).

$r$ -partial strategyproofness as  $r$  approaches 0 (Proposition 2). The degree of strategyproofness thus parametrizes the entire spectrum between full strategyproofness and LD-strategyproofness.

To complete the picture, we establish the formal relationships between partial strategyproofness and other relaxed notions of strategyproofness from prior work (Section 4.4). For the assignment problem, Bogomolnaia and Moulin (2001) used *weak SD-strategyproofness* to describe the incentive properties of the probabilistic serial mechanism, and Balbuzanov (2016) used *convex strategyproofness* to refine their results. We show that partial strategyproofness implies both notions but is not implied by either. *Approximate strategyproofness* (Carroll, 2013) and *strategyproof in the large* (Azevedo and Budish, 2019) are two additional notions that are useful in a wider range of mechanism design problems beyond the assignment problem. When focusing on the assignment problem, we show that partial strategyproofness implies approximate strategyproofness in a meaningful way and that convergence of a mechanism's degree of strategyproofness to 1 in large markets implies strategyproofness in the large. Thus, establishing that a mechanism is partially strategyproof immediately provides additional insights about its incentive properties in terms of these notions from prior work.

To illustrate the usefulness of the partial strategyproofness concept in mechanism design, we provide two applications (Section 5): First, it enables the strongest description of the incentive properties of the probabilistic serial mechanism in finite settings known to date (refining (Bogomolnaia and Moulin, 2001; Balbuzanov, 2016)) as well as in the limit as markets get large (refining (Kojima and Manea, 2010; Azevedo and Budish, 2019)). Second, partial strategyproofness yields a distinction between two common variants of the Boston mechanism for school choice: Under the classic *Boston mechanism (BM)* (Abdulkadiroğlu and Sönmez, 2003), all applications in round  $k$  go to the students'  $k$ th choices, but under the *adaptive Boston mechanism (ABM)* (Alcalde, 1996; Miralles, 2008; Harless, 2019; Dur, 2019), students apply to their respective *best available* school in each round. Intuitively, one would expect ABM to have better incentive properties than BM because strategic skipping of exhausted schools is not a useful heuristic for manipulation under ABM. However, formalizing this intuition is surprisingly challenging (Dur et al., 2018). Using our new partial strategyproofness concept, we can now show formally that ABM indeed has better incentive properties than BM: When priorities are sufficiently random,<sup>7</sup> ABM is partially strategyproof while BM is not.

Partial strategyproofness is relevant for both theory and practical applications beyond the two concrete examples discussed above. For theory, the simplicity of our axioms and our decomposition results enable new approaches for axiomatic analysis.<sup>8</sup> Furthermore, the parametric nature of partial strategyproofness enables a new quantitative analysis of the possible and necessary trade-offs between strategyproofness and other mechanism design objectives (see, e.g., (Mennle and Seuken, 2017b)). For practical applications, partial strategyproofness allows policy makers (e.g., school boards) to give honest and useful strategic advice to participating agents: An agent is best off reporting truthfully if its preference intensities differ sufficiently between any two objects; otherwise, the agent's potential gain from misreporting may be positive but it is quantifiably bounded (because of the formal relationship between partial and approximate strategyproofness).

<sup>7</sup> Concretely, *sufficiently random* means that priorities must support all single priority profiles. In Remark 5, we discuss an approach that allows an extension of partial strategyproofness to problems with coarse priorities.

<sup>8</sup> For example, subsequent work has already used our results to explore alternative combinations of the axioms (Chun and Yun, 2019; Noda, 2019; Mennle and Seuken, 2017e), to identify when truthful reporting is a strictly dominant strategy under the deferred acceptance mechanism (Fack et al., 2015), and to study ordinal Bayesian incentive compatibility in the random assignment problem (Dasgupta and Mishra, 2020).

Furthermore, partial strategyproofness can be helpful to assess the extent to which misreporting may be a concern (e.g., by enabling mechanism designers to estimate for how many agents misreporting is potentially beneficial). Therefore, we expect partial strategyproofness to become a useful addition to the mechanism designer’s toolbox for the analysis of non-strategyproof assignment mechanisms.

## 2. Preliminaries

### 2.1. Model

A *setting*  $(N, M, q)$  consists of a set of *agents*  $N$  with  $n = |N|$ , a set of *objects*  $M$  with  $m = |M| \geq 2$ , and *capacities*  $q = (q_1, \dots, q_m)$  (i.e.,  $q_j$  units of object  $j$  are available). We assume  $n \leq \sum_{j \in M} q_j$  (i.e., there are not more agents than the total number of units; otherwise we include a dummy object with capacity  $n$ ). Each agent  $i \in N$  has a strict *preference order*  $P_i$  over objects, where  $P_i : a \succ b$  indicates that agent  $i$  prefers object  $a$  to object  $b$ . Let  $\mathcal{P}$  be the set of all possible preference orders over  $M$ . A *preference profile*  $P = (P_i)_{i \in N} \in \mathcal{P}^N$  is a collection of preference orders of all agents, and  $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$  is a collection of preference orders of all agents except  $i$ . We assume that agents have von Neumann–Morgenstern *utility functions*, where agent  $i$ ’s utility function  $u_i : M \rightarrow \mathbb{R}^+$  is *consistent with*  $P_i$  (i.e.,  $u_i(a) > u_i(b)$  whenever  $P_i : a \succ b$ ), and  $U_{P_i}$  denotes the set of all utility functions that are consistent with  $P_i$ .

An *assignment* is represented by an  $n \times m$ -matrix  $x = (x_{i,j})_{i \in N, j \in M}$ . The value  $x_{i,j}$  is the probability that agent  $i$  gets object  $j$ . The  $i$ th row  $x_i = (x_{i,j})_{j \in M}$  of  $x$  is called the *assignment vector of  $i$*  (or  *$i$ ’s assignment*).  $x$  is *feasible* if no object is assigned beyond capacity (i.e.,  $\sum_{i \in N} x_{i,j} \leq q_j$  for all  $j \in M$ ) and each agent receives some object with certainty (i.e.,  $\sum_{j \in M} x_{i,j} = 1$  for all  $i \in N$  and  $x_{i,j} \geq 0$  for all  $i \in N, j \in M$ ). Unless explicitly stated otherwise, we only consider feasible assignments throughout this paper.  $x$  is called *deterministic* if  $x_{i,j} \in \{0, 1\}$  for all  $i \in N, j \in M$ . The Birkhoff–von Neumann Theorem and its extensions (Budish et al., 2013) ensure that, for any given assignment, we can find a lottery over deterministic assignments that implements the respective marginal probabilities. Finally, we denote by  $X$  and  $\Delta(X)$  the sets of all deterministic and random assignments, respectively.

A *mechanism* is a mapping  $f : \mathcal{P}^N \rightarrow \Delta(X)$  that selects an assignment based on a preference profile.  $f_i(P_i, P_{-i})$  denotes  $i$ ’s assignment when  $i$  reports  $P_i$  and the other agents report  $P_{-i}$ . The mechanism  $f$  is *deterministic* if it only selects deterministic assignments (i.e.,  $f : \mathcal{P}^N \rightarrow X$ ). Finally, given an agent  $i$  with a utility function  $u_i$  who reports  $P_i$  when the other agents report  $P_{-i}$ , that agent’s *expected utility* is  $\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] = \sum_{j \in M} u_i(j) f_{i,j}(P_i, P_{-i})$ .

### 2.2. Strategyproofness and auxiliary concepts

The most common incentive concept for the design of assignment mechanisms requires that agents maximize their expected utility by reporting their ordinal preferences truthfully, independent of the other agents’ preference reports.

**Definition 1.** A mechanism  $f$  is *expected-utility strategyproof* (*EU-strategyproof*) if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , all misreports  $P'_i \in \mathcal{P}$ , and all utility functions  $u_i \in U_{P_i}$ , we have  $\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] \geq \mathbb{E}_{f_i(P'_i, P_{-i})}[u_i]$ .

Alternatively, strategyproofness can be defined in terms of stochastic dominance: For a preference order  $P_i \in \mathcal{P}$  and two assignment vectors  $x_i$  and  $y_i$ , we say that  $x_i$  *stochastically dominates*  $y_i$  at  $P_i$  if, for all objects  $a \in M$ , we have

$$\sum_{j \in M \text{ with } P_i: j > a} x_{i,j} \geq \sum_{j \in M \text{ with } P_i: j > a} y_{i,j}. \tag{5}$$

In words,  $i$ 's chances of obtaining one of its top- $k$  choices are weakly higher under  $x_i$  than under  $y_i$ , for all ranks  $k$ . A mechanism  $f$  is *stochastic dominance strategyproof (SD-strategyproof)* if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in \mathcal{P}$ ,  $i$ 's assignment  $f_i(P_i, P_{-i})$  stochastically dominates  $f_i(P'_i, P_{-i})$  at  $P_i$ . SD-strategyproofness and EU-strategyproofness are equivalent (Erdil, 2014), and we therefore refer to this property as *(full) strategyproofness*.

Cho (2016) presented a substantially weaker incentive concept: Agents with lexicographic preferences prefer any (arbitrarily small) increase in the probability for a more-preferred object to any (arbitrarily large) increase in the probability for any less-preferred object; and the associated lexicographic dominance strategyproofness requires that such agents prefer to report their preferences truthfully.

**Definition 2.** For a preference order  $P_i \in \mathcal{P}$  and assignment vectors  $x_i, y_i$ , we say that  $x_i$  *lexicographically dominates*  $y_i$  at  $P_i$  if  $x_i = y_i$ , or  $x_{i,a} > y_{i,a}$  for some  $a \in M$  and  $x_{i,j} = y_{i,j}$  for all  $j \in U(a, P_i)$ . A mechanism  $f$  is *lexicographic dominance strategyproof (LD-strategyproof)* if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in \mathcal{P}$ ,  $f_i(P_i, P_{-i})$  lexicographically dominates  $f_i(P'_i, P_{-i})$  at  $P_i$ .

Finally, we introduce the auxiliary concepts of neighborhoods and contour sets: The *neighborhood of a preference order*  $P_i$  is the set of all preference orders that differ from  $P_i$  by a swap of two consecutively ranked objects, denoted  $N_{P_i}$  (e.g., the neighborhood of  $P_i : a > b > c > d$  contains  $P'_i : b > a > c > d$  but does not contain  $P''_i : c > a > b > d$ ). The *upper contour set of  $j$  at  $P_i$*  is the set of objects that  $i$  strictly prefers to  $j$ , denoted  $U(j, P_i)$ . Conversely, the *lower contour set of  $j$  at  $P_i$*  is the set of objects that  $i$  likes strictly less than  $j$ , denoted  $L(j, P_i)$ . For example, the upper contour set of  $b$  at  $P_i : a > b > c > d$  is  $U(b, P_i) = \{a\}$  and the lower contour set is  $L(b, P_i) = \{c, d\}$ .

### 3. A new decomposition of strategyproofness

In this section, we introduce three axioms and we show that strategyproofness decomposes into these axioms. Swapping two consecutively ranked objects in the true preference order (or equivalently, reporting a preference order from its neighborhood) is a basic kind of misreport. The axioms we define limit the ways in which a mechanism can change an agent's assignment when that agent uses such a basic misreport.

**Axiom 1 (Swap monotonicity).** A mechanism  $f$  is *swap monotonic* if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in N_{P_i}$  with  $P_i : a > b$  but  $P'_i : b > a$ , one of the following holds:

- either:  $f_i(P_i, P_{-i}) = f_i(P'_i, P_{-i})$ ,
- or:  $f_{i,b}(P'_i, P_{-i}) > f_{i,b}(P_i, P_{-i})$ .



In words, swap monotonicity requires that the mechanism reacts to the swap in a *direct* and *monotonic* way: If the swap that brings  $b$  forward affects the agent’s assignment at all, then at least its assignment for  $b$  must be affected directly. Moreover, this change must be monotonic in the sense that the agent’s assignment for  $b$  must increase when  $b$  is reportedly more preferred.

For deterministic mechanisms, swap monotonicity is equivalent to strategyproofness (Proposition 5 in Appendix D). For the more general class of random mechanisms, swap monotonicity is weaker than strategyproofness but prevents a certain obvious kind of manipulability: Consider a mechanism that assigns an agent’s reported first choice with probability  $1/3$  and its reported second choice with probability  $2/3$ . The agent is unambiguously better off by ranking its second choice first. Swap monotonicity precludes such opportunities for manipulation. Nevertheless, even swap monotonic mechanisms may be manipulable in a stochastic dominance sense, as the next example shows.

**Example 1.** Consider a setting with four objects  $a, b, c, d$  and a single agent  $i$  with preference order  $P_i : a \succ b \succ c \succ d$ . Suppose that reporting a preference for  $b$  over  $c$  leads to an assignment of  $(0, 1/2, 0, 1/2)$  for  $a, b, c, d$  respectively, and reporting a preference for  $c$  over  $b$  leads to  $(1/2, 0, 1/2, 0)$ . This is consistent with swap monotonicity; yet, the latter assignment stochastically dominates the former at  $P_i$ .

Note that the misreport in Example 1 affects the agent’s assignment for  $a$ , an object that the agent strictly prefers to both  $b$  and  $c$ , the objects that get swapped. Our next axiom precludes such effects.

**Axiom 2 (Upper invariance).** A mechanism  $f$  is *upper invariant* if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in N_{P_i}$  with  $P_i : a \succ b$  but  $P'_i : b \succ a$ , we have that  $f_{i,j}(P_i, P_{-i}) = f_{i,j}(P'_i, P_{-i})$  for all  $j \in U(a, P_i)$ .

Intuitively, upper invariance ensures that agents cannot influence their chances of obtaining more-preferred objects by changing the order of less-preferred objects. The axiom was originally introduced by Hashimoto et al. (2014) (who called it *weak invariance*). It was one of the central axioms in their characterization of the probabilistic serial mechanism. If an outside option is available and if the mechanism is individually rational, then upper invariance is equivalent to *truncation robustness* (i.e., no agent can benefit by ranking the outside option above acceptable objects). Many assignment mechanisms satisfy upper invariance, including random serial dictatorship, probabilistic serial, the Boston mechanism, deferred acceptance (for agents on the proposing side), top-trade cycles, and the HBS draft mechanism.

Finally, our third axiom is symmetric to upper invariance but restricts how swaps can affect the assignment for less-preferred objects.

**Axiom 3 (Lower invariance).** A mechanism  $f$  is *lower invariant* if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in N_{P_i}$  with  $P_i : a \succ b$  but  $P'_i : b \succ a$ , we have that  $f_{i,j}(P_i, P_{-i}) = f_{i,j}(P'_i, P_{-i})$  for all  $j \in L(b, P_i)$ .

In words, a mechanism is lower invariant if changing the order of two consecutively ranked objects does not affect the agents’ chances of obtaining any less-preferred objects. Examples of lower invariant mechanisms are strategyproof mechanisms, like random serial dictatorship

and top-trade cycles. Constrained serial dictatorship (Noda, 2019) is swap monotonic and lower invariant, and deferred acceptance is lower invariant for agents on both sides.

Each of the three axioms affects incentives by preventing misreports from being beneficial in particular ways: swap monotonicity forces mechanisms to change the assignment for the respective objects directly and in the right direction, upper invariance is essentially equivalent to truncation robustness, and lower invariance mirrors upper invariance but for less-preferred objects. In combination, they give rise to our first main result, the decomposition of strategyproofness into these axioms.

**Theorem 1** (*Decomposition of strategyproofness*). *A mechanism is strategyproof if and only if it is swap monotonic, upper invariant, and lower invariant.*

**Proof.** *Sufficiency* ( $\Rightarrow$ ). Assume towards contradiction that a mechanism  $f$  is strategyproof but not upper invariant. Then there exists an agent  $i \in N$ , a preference profile  $P = (P_i, P_{-i}) \in \mathcal{P}^N$ , a misreport  $P'_i \in N_{P_i}$  with  $P_i : a \succ b$  but  $P'_i : b \succ a$ , and an object  $j$  which  $i$  prefers strictly to  $a$  with  $f_{i,j}(P'_i, P_{-i}) \neq f_{i,j}(P_i, P_{-i})$ . Without loss of generality, let  $j$  be  $i$ 's most-preferred such object and  $f_{i,j}(P'_i, P_{-i}) > f_{i,j}(P_i, P_{-i})$  (otherwise, reverse the roles of  $P'_i$  and  $P_i$ ). Then  $f_i(P_i, P_{-i})$  does not stochastically dominate  $f_j(P'_i, P_{-i})$ , a contradiction to SD-strategyproofness. Lower invariance follows analogously, except that we take  $j$  to be the *least*-preferred object for which  $i$ 's assignment changes.

By upper and lower invariance, any swap of consecutively ranked objects (e.g., from  $P_i : a \succ b$  to  $P'_i : b \succ a$ ) leads to a re-distribution of probability between  $a$  and  $b$ . If reporting  $P'_i$  leads to a strictly higher assignment for  $a$ , then  $f_i(P'_i, P_{-i})$  strictly stochastically dominates  $f_i(P_i, P_{-i})$  at  $P_i$ , a contradiction. This implies swap monotonicity.

*Necessity* ( $\Leftarrow$ ). We invoke a result of Carroll (2012) that strategyproofness can be shown by verifying that no agent can benefit from swapping two consecutively ranked objects. Let  $f$  be a swap monotonic, upper invariant, and lower invariant mechanism, and consider an agent  $i \in N$ , a preference profile  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and a misreport  $P'_i \in N_{P_i}$  with  $P_i : a \succ b$  but  $P'_i : b \succ a$ . Observe that  $f_i(P_i, P_{-i})$  stochastically dominates  $f_i(P'_i, P_{-i})$  at  $P_i$ : By upper and lower invariance,  $i$ 's assignment for all objects remains constant under the misreport, except possibly for  $a$  and  $b$ ; and, by swap monotonicity,  $i$ 's assignments for  $a$  and  $b$  can only decrease and increase, respectively.  $\square$

Theorem 1 highlights that strategyproofness is quite restrictive: If an agent swaps two objects (e.g., from  $P_i : a \succ b$  to  $P'_i : b \succ a$ ), the only way in which a strategyproof mechanism can react is by increasing that agent's assignment for  $b$  and decreasing its assignment for  $a$  by the same amount. The appeal of our decomposition in Theorem 1 lies in the choice of axioms, which are simple and easily motivated. They make the decomposition particularly useful, for example when verifying strategyproofness of mechanisms or when encoding strategyproofness as constraints under the automated mechanism design paradigm (Sandholm, 2003).<sup>9</sup>

**Remark 1.** While we focus on strict preferences in the present paper, Theorem 1 can be extended to the case of weak preferences; see (Mennle and Seuken, 2017a).

<sup>9</sup> The decomposition has been used in subsequent work, e.g., to identify when truth-telling is a strictly dominant strategy under the deferred acceptance mechanism (Fack et al., 2015) and to study Bayesian incentive compatibility of ordinal mechanisms (Dasgupta and Mishra, 2020).

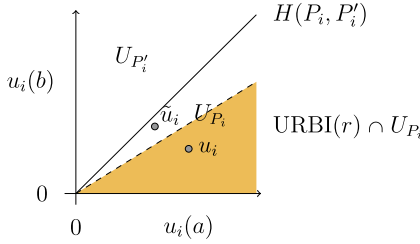


Fig. 1. Illustration of uniformly relatively bounded indifference.

### 4. Partial strategyproofness

Recall the motivating example from the introduction, where agent 1 was contemplating a misreport under the probabilistic serial mechanism.  $r = 3/4$  was the pivotal ratio between the agent’s (normalized) utility for its first and second choices which determined whether the misreport was beneficial or not. The following domain restriction generalizes the intuition from this basic example to agents whose preference intensities differ sufficiently between any two objects.

**Definition 3.** A utility function  $u_i$  satisfies *uniformly relatively bounded indifference with respect to*  $r \in [0, 1]$  ( $URBI(r)$ ) if, for all objects  $a, b \in M$  with  $u_i(a) > u_i(b)$ ,

$$r \left( u_i(a) - \min_{j \in M} u_i(j) \right) \geq u_i(b) - \min_{j \in M} u_i(j). \tag{6}$$

If  $\min_{j \in M} (u_i(j)) = 0$  (i.e.,  $i$  has utility 0 for its last choice), then Inequality (6) simplifies to

$$ru_i(a) \geq u_i(b). \tag{7}$$

In words,  $URBI(r)$  requires that agent  $i$  values  $b$  at least a factor  $r$  less than  $a$ . Consider Fig. 1, which depicts the utility space for a setting with only two objects  $a$  and  $b$ , such that every utility function corresponds to a point in this figure. This allows the following geometric interpretation: The  $URBI(r)$  condition means that  $i$ ’s utility function (e.g., the point labeled  $u_i$ ) cannot be arbitrarily close to the indifference hyperplane  $H(P_i, P'_i)$ , but it must lie within the shaded triangular area.  $r$  is the slope of the dashed line that bounds this area at the top. Any utility function in  $U_{P_i}$  that lies outside the shaded area (e.g., the point labeled  $\tilde{u}_i$ ) violates  $URBI(r)$ . For convenience, we also use  $URBI(r)$  to denote the *set of utility functions* that satisfy the constraints.

We are now ready to formally define our new incentive concept.

**Definition 4.** Given a setting  $(N, M, q)$  and a bound  $r \in [0, 1]$ , a mechanism  $f$  is  *$r$ -partially strategyproof in  $(N, M, q)$*  if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , all misreports  $P'_i \in \mathcal{P}$ , and all utility functions  $u_i \in U_{P_i} \cap URBI(r)$ , we have  $\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] \geq \mathbb{E}_{f_i(P'_i, P_{-i})}[u_i]$ .

In words, an  $r$ -partially strategyproof mechanism makes truthful reporting a dominant strategy at least for those agents whose utility functions satisfy  $URBI(r)$ . Note that the value of  $r$  depends on the setting. Since all utility functions satisfy  $URBI(1)$ , 1-partial strategyproofness is equivalent to full strategyproofness, and since the sets  $URBI(r)$  are nested (i.e.,  $URBI(r) \subseteq URBI(r')$  for  $r \leq r'$ ), the  $r$ -partial strategyproofness requirement is less demanding for smaller  $r$ .

### 4.1. A decomposition of partial strategyproofness

For our second main result, we prove that dropping lower invariance but still requiring swap monotonicity and upper invariance leads to partial strategyproofness.

**Theorem 2** (*Decomposition of partial strategyproofness*). *Given a setting  $(N, M, q)$ , a mechanism is  $r$ -partially strategyproof for some  $r > 0$  if and only if it is swap monotonic and upper invariant.*

**Proof.** Our three axioms swap monotonicity, upper invariance, and lower invariance are *local* constraints in the sense that they only restrict how mechanisms can react to swaps. Conversely, full and partial strategyproofness are *global* constraints because they restrict reactions to arbitrary report changes. In our proof of Theorem 1, we have used the local sufficiency result of Carroll (2012) to bridge the gap between local and global constraints. To bridge the analogous gap in the proof of Theorem 2, we use a local sufficiency result for lexicographic dominance strategyproofness by Cho (2016).

It is clear that swap monotonicity and upper invariance are equivalent to LD-adjacent strategyproofness (i.e., for all  $i \in N$ ,  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and  $P'_i \in N_{P_i}$ ,  $f_i(P_i, P_{-i})$  lexicographically dominates  $f_i(P'_i, P_{-i})$  at  $P_i$ ). By Theorem 1 of Cho (2016), swap monotonicity and upper invariance are thus equivalent to LD-strategyproofness. With this, we must show the following:

**Lemma 1.** *Given a setting  $(N, M, q)$ , a mechanism  $f$  is  $r$ -partially strategyproof for some  $r > 0$  if and only if  $f$  is LD-strategyproof.*

We can restrict attention to mechanisms under which at least one agent has a direct influence on its own assignment; formally, there exists an agent  $i \in N$ , a preference profile  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and a misreport  $P'_i \in \mathcal{P}$  such that  $f_i(P_i, P_{-i}) \neq f_i(P'_i, P_{-i})$ .<sup>10</sup> Let  $\delta$  be the smallest non-zero change in any agent’s assignment resulting from any change of report by that agent, defined by

$$\delta = \min \left\{ |f_{i,j}(P_i, P_{-i}) - f_{i,j}(P'_i, P_{-i})| \mid \begin{array}{l} i \in N, j \in M, (P_i, P_{-i}) \in \mathcal{P}^N, P'_i \in \mathcal{P}, \\ \text{s.t. } |f_{i,j}(P_i, P_{-i}) - f_{i,j}(P'_i, P_{-i})| > 0 \end{array} \right\}. \quad (8)$$

It is easy to see that  $\delta$  is well-defined and strictly positive for the mechanisms we consider.

*Necessity (partial strategyproofness  $\Leftarrow$  LD-strategyproofness).* For an agent  $i \in N$  and a preference profile  $(P_i, P_{-i}) \in \mathcal{P}^N$ , suppose that  $i$  considers misreport  $P'_i \in \mathcal{P}$ . If  $f_i(P_i, P_{-i}) \neq f_i(P'_i, P_{-i})$ , then LD-strategyproofness implies that there exists some object  $a$  for which  $i$ ’s assignment strictly decreases, and  $i$ ’s assignment for all more-preferred objects  $j \in U(a, P_i)$  remains unchanged. Since  $i$ ’s assignment for  $a$  changes, it must decrease by at least  $\delta$ . Let  $b$  be the object that  $i$  ranks directly below  $a$  in  $P_i$ . Then  $i$ ’s gain in expected utility from misreporting is greatest if, first,  $i$  receives  $a$  with probability  $\delta$  and its last choice with probability  $(1 - \delta)$  when being truthful, and second,  $i$  receives  $b$  with certainty when misreporting. The gain is therefore bounded from above by

$$u_i(b) - \left( \delta u_i(a) + (1 - \delta) \min_{j \in M} u_i(j) \right), \quad (9)$$

<sup>10</sup> Mechanisms that violate this property are trivially strategyproof for all agents and thus both LD-strategyproof and  $r$ -partially strategyproof for all  $r \in [0, 1]$ .

which is weakly negative if

$$\delta \left( u_i(a) - \min_{j \in M} u_i(j) \right) \geq u_i(b) - \min_{j \in M} u_i(j). \tag{10}$$

Inequality (10) holds for all utility functions in  $U_{P_i} \cap \text{URBI}(\delta)$ . Thus, if  $i$ 's utility function satisfies  $\text{URBI}(\delta)$ , then no misreport increases  $i$ 's expected utility, or equivalently,  $f$  is  $\delta$ -partially strategyproof.

*Sufficiency (partial strategyproofness  $\Rightarrow$  LD-strategyproofness).* Let  $f$  be  $r$ -partially strategyproof for some  $r > 0$ , and assume towards contradiction that  $f$  violates LD-strategyproofness. This is equivalent to saying that there exists an agent  $i \in N$ , a preference profile  $(P_i, P_{-i}) \in \mathcal{P}^N$ , a misreport  $P'_i \in \mathcal{P}$  with  $f_i(P_i, P_{-i}) \neq f_i(P'_i, P_{-i})$ , and an object  $a \in M$ , such that  $f_{i,a}(P_i, P_{-i}) < f_{i,a}(P'_i, P_{-i})$  but  $f_{i,j}(P_i, P_{-i}) = f_{i,j}(P'_i, P_{-i})$  for all  $j \in U(a, P_i)$ . Again, let  $b$  be the object that  $i$  ranks directly below  $a$  in  $P_i$ . Since  $i$ 's assignment for  $a$  increases, it must increase by at least  $\delta$ . Thus,  $i$ 's gain in expected utility is *smallest* if, first,  $i$  receives  $b$  with certainty when being truthful, and second,  $i$  receives  $a$  with probability  $\delta$  and its last choice with probability  $(1 - \delta)$  when misreporting. This makes

$$\left( \delta u_i(a) + (1 - \delta) \min_{j \in M} u_i(j) \right) - u_i(b) \tag{11}$$

a lower bound on  $i$ 's gain from misreporting. This bound is strictly positive if

$$\delta \left( u_i(a) - \min_{j \in M} u_i(j) \right) > u_i(b) - \min_{j \in M} u_i(j), \tag{12}$$

which holds for all utility functions in  $U_{P_i} \cap \text{URBI}(r)$  for  $r < \delta$ . Therefore,  $f$  cannot be  $r$ -partially strategyproof for any  $r > 0$ , a contradiction.  $\square$

Theorem 2 provides an axiomatic motivation for our definition of partial strategyproofness: The class of partially strategyproof mechanisms consists exactly of those mechanisms that are swap monotonic and upper invariant, but they may violate lower invariance.

**Remark 2.** Two main arguments suggest dropping lower invariance and keeping swap monotonicity and upper invariance as a sensible approach towards relaxing strategyproofness. First, on the positive side, upper invariance is essentially equivalent to robustness to manipulation by truncation (Hashimoto et al., 2014), and, for deterministic mechanisms, swap monotonicity is equivalent to strategyproofness (Proposition 5 in Appendix D). Second, on the negative side, dropping either swap monotonicity or upper invariance (instead of lower invariance) does not admit the construction of interesting ordinally efficient mechanisms: the probabilistic serial mechanism (Bogomolnaia and Moulin, 2001) is swap monotonic, upper invariant (but not lower invariant), and ordinally efficient, and it satisfies the additional axioms *symmetry*, *anonymity*, *neutrality*, and *non-bossiness*. In contrast, no mechanism can be upper invariant, lower invariant, ordinally efficient, and symmetric; and no mechanism can be swap monotonic, lower invariant, ordinally efficient, anonymous, neutral, and non-bossy.<sup>11</sup> This means that these popular combinations of mechanism design axioms become unattainable when dropping swap monotonicity or upper invariance (instead of lower invariance).

<sup>11</sup> See (Mennle and Seuken, 2017e) for formal definitions of these axioms and both impossibility results.

Despite these two arguments, whether an axiom is more or less “important” is also a question of taste. We have chosen to drop lower invariance in the present paper, but investigating the consequences of dropping swap monotonicity or upper invariance is definitely an interesting research question.<sup>12</sup>

4.2. Maximality of the URBI( $r$ ) domain restriction

In this section, we study how well the URBI( $r$ ) domain restriction captures the incentive properties of non-strategyproof assignment mechanisms. By definition,  $r$ -partial strategyproofness implies that the set URBI( $r$ ) must be contained in the set of utility functions for which truthful reporting is a dominant strategy. However, the two sets may not be exactly equal, as the following example shows.

**Example 2.** Consider a setting with 4 agents and 4 objects with unit capacity. In this setting, the adaptive Boston mechanism with priorities determined by a single uniform lottery ( $ABM^U$ , see Section 5.2) is  $1/3$ -partially strategyproof but not  $r$ -partially strategyproof for any  $r > 1/3$ . However, it is a simple (though tedious) exercise to verify that an agent with utility function  $u_i(a) = 6, u_i(b) = 2, u_i(c) = 1, u_i(d) = 0$  cannot benefit from misreporting, independent of the reports from the other agents. But  $u_i$  violates URBI( $1/3$ ), since  $(u_i(c) - \min_{j \in M} u_i(j)) / (u_i(b) - \min_{j \in M} u_i(j)) = 1/2 > 1/3$ . Thus, the set of utility functions for which  $ABM^U$  makes truthful reporting a dominant strategy is strictly larger than the set URBI( $1/3$ ).

Example 2 shows that, for some  $r$ -partially strategyproof mechanism, there may exist utility functions that violate URBI( $r$ ) but for which truthful reporting is nonetheless a dominant strategy. This raises the question whether URBI( $r$ ) is “too small” in the sense that it excludes some utility functions for which *all*  $r$ -partially strategyproof mechanisms make truthful reporting a dominant strategy. Our next proposition dispels this concern because it shows *maximality* of the URBI( $r$ ) domain restriction.

**Proposition 1 (Maximality).** Consider a setting  $(N, M, q)$  with  $m \geq 3$  objects, a bound  $r > 0$ , and an agent  $i$  with utility function  $u_i$ . Then  $u_i$  satisfies URBI( $r$ ) if and only if truthful reporting maximizes the expected utility for  $i$  under all  $r$ -partially strategyproof, anonymous mechanisms.

**Proof.** Towards contradiction, assume that a utility function  $\tilde{u}_i$  violates URBI( $r$ ). Then

$$\frac{\tilde{u}_i(b) - \min_{j \in M} \tilde{u}_i(j)}{\tilde{u}_i(a) - \min_{j \in M} \tilde{u}_i(j)} = \tilde{r} \tag{13}$$

for some  $a, b \in M$  and  $\tilde{r} > r$ . Observe that  $b$  cannot be  $i$ 's last choice because  $0 / (\tilde{u}_i(a) - \min_{j \in M} \tilde{u}_i(j)) \leq r$  is trivially satisfied. We construct a mechanism  $\tilde{f}$  that is  $r$ -partially strategyproof but manipulable for agent  $i$  with utility function  $\tilde{u}_i$ , setting the assignment for the distinguished agent  $i$  as follows: Fix parameters  $\delta_a, \delta_b \in [0, 1/m]$ ; then, for all  $P_{-i} \in \mathcal{P}^N$ , all preference orders  $\hat{P}_i \in \mathcal{P}$ , and all objects  $j \in M$ , let

<sup>12</sup> In subsequent work, Noda (2019) considered the combination of swap monotonicity and lower invariance; Chun and Yun (2019) considered the combination of upper and lower invariance, and they independently proved the incompatibility with ordinal efficiency and symmetry.

$$\tilde{f}_{i,j}(\hat{P}_i, P_{-i}) = \begin{cases} 1/m, & \text{if } \hat{P}_i : a \succ b, \\ 1/m + \delta_b \mathbb{1}_{\{j=b\}} - \delta_a \mathbb{1}_{\{j=a\}} + (\delta_a - \delta_b) \mathbb{1}_{\{j=d\}}, & \text{if } \hat{P}_i : b \succ a, \end{cases}$$

where  $d$  is the last choice under  $\hat{P}_i$ . For all other agents, distribute the remaining probabilities evenly. With parameters  $\delta_a, \delta_b \in [0, 1/m]$ , this mechanism is well-defined. Next, let  $\delta_b = 1/m$  and choose  $\delta_a \in [r/m, \tilde{r}/m]$ . It is straightforward to verify that  $\tilde{f}$  has the desired properties (Lemma 2 in Appendix A). To obtain an anonymous mechanism with the same properties, we randomly assign each agent to the role of the distinguished agent  $i$ . This yields a contradiction to the assumption that  $i$  finds truthful reporting a dominant strategy under any  $r$ -partially strategyproof, anonymous mechanism.

Sufficiency holds by definition of  $r$ -partial strategyproofness.  $\square$

Proposition 1 means that, for any utility function that violates  $URBI(r)$ , there exists some  $r$ -partially strategyproof mechanism under which truthful reporting is not a dominant strategy for an agent with that utility function. Thus, unless we are given additional structural information about the mechanism besides the fact that it is  $r$ -partially strategyproof (and possibly anonymous),  $URBI(r)$  is in fact the largest set of utility functions for which truthful reporting is guaranteed to be a dominant strategy.

### 4.3. The degree of strategyproofness

Partial strategyproofness induces a natural parameter to measure “how strategyproof” a non-strategyproof assignment mechanism is.

**Definition 5.** Given a setting  $(N, M, q)$  and a mechanism  $f$ , we define the *degree of strategyproofness of  $f$  (in the setting  $(N, M, q)$ )*<sup>13</sup> as

$$\rho_{(N,M,q)}(f) = \max \{r \in [0, 1] \mid f \text{ is } r\text{-partially strategyproof in } (N, M, q)\}. \tag{14}$$

Observe that, for  $0 \leq r < r' \leq 1$  we have  $URBI(r) \subset URBI(r')$  by construction. Thus, a lower degree of strategyproofness corresponds to a weaker guarantee. By maximality from Proposition 1, the degree of strategyproofness constitutes a meaningful measure for incentive properties: If the only known attributes of  $f$  are that it is swap monotonic and upper invariant and that  $r = \rho_{(N,M,q)}(f)$ , then truthful reporting is guaranteed to be a dominant strategy for all agents with utility functions in  $URBI(r)$ , but it is impossible to give such a guarantee for any utility function that violates  $URBI(r)$ .

The degree of strategyproofness can be used to compare two mechanisms by their incentive properties:  $\rho_{(N,M,q)}(f) > \rho_{(N,M,q)}(g)$  means that  $f$  makes truthful reporting a dominant strategy on a strictly larger  $URBI(r)$  domain restriction than  $g$  does.<sup>14</sup> In Section 5, we apply this comparison to differentiate three assignment mechanisms, namely the probabilistic serial mechanism and two variants of the Boston mechanism. Furthermore, in (Mennle and Seuken, 2017b) we have used the degree of strategyproofness to quantify trade-offs between strategyproofness

<sup>13</sup> To see that  $\rho_{(N,M,q)}$  is well-defined, observe that  $URBI(r)$  is topologically closed. Thus, a mechanism that is  $r'$ -partially strategyproof for all  $r' < r$  must also be  $r$ -partially strategyproof.

<sup>14</sup> Pathak and Sönmez (2013) proposed a concept to compare mechanisms by their *vulnerability to manipulation*. As Proposition 6 in Appendix E shows, this comparison is consistent with (but not equivalent to) the comparison of mechanisms by their degrees of strategyproofness.

and efficiency that are achievable via hybrid mechanisms (i.e., convex combinations of a strategyproof mechanism and a manipulable mechanism with better efficiency properties).

Observe that the degree of strategyproofness parametrizes a spectrum of incentive concepts, where the upper limit of this spectrum is full strategyproofness. Conversely, we would also like to understand what the *lower limit* is. To answer this question, we consider the weaker notion of lexicographic dominance strategyproofness (Definition 2). Recall that a mechanism is LD-strategyproof if truthful reporting is a dominant strategy for agents who prefer any (arbitrarily small) increase in the probability for a more-preferred object to any (arbitrarily large) increase in the probability for any less-preferred object. The next proposition formalizes the upper and lower limits for the partial strategyproofness concept.

**Proposition 2** (*Limit concepts*). *Given a setting  $(N, M, q)$  and an  $r > 0$ , let  $SP(N, M, q)$ ,  $r$ -PSP( $N, M, q$ ), and LD-SP( $N, M, q$ ) be the sets of mechanisms that are strategyproof,  $r$ -partially strategyproof, and LD-strategyproof in the setting  $(N, M, q)$ , respectively. Then*

$$SP(N, M, q) = \bigcap_{r < 1} r\text{-PSP}(N, M, q), \tag{15}$$

$$LD\text{-SP}(N, M, q) = \bigcup_{r > 0} r\text{-PSP}(N, M, q). \tag{16}$$

**Proof.** Towards Equality (15), observe that any utility function  $u_i$  consistent with a strict preference order  $P_i$  is also contained in URBI( $r$ ) for some  $r < 1$ . A mechanism that is  $r$ -partially strategyproof for all  $r < 1$  therefore makes truthful reporting a dominant strategy for all utility functions. This implies EU-strategyproofness.

Equality (16) follows directly from Lemma 1.  $\square$

In words, Proposition 2 shows that there is no gap “at the top” between the least manipulable partially strategyproof mechanisms and those that are fully strategyproof, and there is no gap “at the bottom” between the most manipulable partially strategyproof mechanisms and mechanisms that are merely LD-strategyproof. The degree of strategyproofness thus parametrizes the entire spectrum between full strategyproofness and the minimal LD-strategyproofness.

An important consequence of Proposition 2 is that, for a fixed setting, any LD-strategyproof mechanism is also  $r$ -partially strategyproof for some  $r > 0$ . On a technical level, this means that it suffices to verify LD-strategyproofness to show partial strategyproofness. We use this in Section 5 to prove partial strategyproofness of the probabilistic serial mechanism. Furthermore, describing a mechanism as being LD-strategyproof ignores the additional information about incentive properties captured by partial strategyproofness, i.e., the parametric nature of the set of utility functions for which truthful reporting is guaranteed to be a dominant strategy under that mechanism.

**Remark 3** (*Computability*). For any mechanism and any setting, the degree of strategyproofness is computable. To show this, we provide a simple algorithm (Algorithm 1 in Appendix F), which exploits an equivalent formulation of  $r$ -partial strategyproofness in terms of a finite system of inequalities.<sup>15</sup> Computability distinguishes partial strategyproofness from other concepts for

<sup>15</sup> Note that computability is not obvious because the definition of  $r$ -partial strategyproofness (Definition 4) involves infinitely many utility functions.



which computability has not been established, e.g., the *comparison of mechanisms by their vulnerability to manipulation* (Pathak and Sönmez, 2013), *strategyproofness in the large* (Azevedo and Budish, 2019), and *convex strategyproofness* (Balbuzanov, 2016). However, tractability is an issue; we can currently only compute the degree of strategyproofness for small settings because Algorithm 1 considers all (exponentially many) preference profiles. In Appendix F, we discuss approaches towards improving tractability.

#### 4.4. Relationships with other notions of strategyproofness

In this section, we explore the relationship between partial strategyproofness and other notions of strategyproofness. Proposition 2 already established full strategyproofness and LD-strategyproofness as the upper and lower limit concepts. In the following, we describe the relationships with weak SD-strategyproofness (Bogomolnaia and Moulin, 2001), convex strategyproofness (Balbuzanov, 2016), approximate strategyproofness (Carroll, 2013), and strategyproofness in the large (Azevedo and Budish, 2019).

In their seminal work, Bogomolnaia and Moulin (2001) used *weak SD-strategyproofness* to describe the incentive properties of the probabilistic serial mechanism.

**Definition 6.** A mechanism  $f$  is *weakly SD-strategyproof* if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in \mathcal{P}$ , the assignment vector  $f_i(P_i, P_{-i})$  stochastically dominates  $f_i(P'_i, P_{-i})$  at  $P_i$  whenever the two assignment vectors are comparable by stochastic dominance at  $P_i$ .

Equivalently, this concept can be formulated in terms of expected utilities: A mechanism is weakly SD-strategyproof if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in \mathcal{P}$ , there exists a utility function  $u_i \in U_{P_i}$  such that  $\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] \geq \mathbb{E}_{f_i(P'_i, P_{-i})}[u_i]$ . Observe that  $u_i$  can depend on  $i$ ,  $P_i$ ,  $P_{-i}$ , and  $P'_i$ , which makes this requirement very weak. The slightly stronger incentive concept of *convex strategyproofness* (Balbuzanov, 2016) arises if  $u_i$  may only depend on  $i$  and  $P_i$  but must be independent of  $P'_i$  and  $P_{-i}$ .

**Definition 7.** A mechanism  $f$  is *convex strategyproof* if, for all agents  $i \in N$  and all preference orders  $P_i \in \mathcal{P}$ , there exists a utility function  $u_i \in U_{P_i}$  such that, for all preferences  $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$  and all misreports  $P'_i \in \mathcal{P}$ , we have  $\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] \geq \mathbb{E}_{f_i(P'_i, P_{-i})}[u_i]$ .

Balbuzanov (2016) showed that the probabilistic serial mechanism is convex strategyproof and constructed a mechanism that is weakly SD-strategyproof but not convex strategyproof, which shows that the latter is a strictly stronger requirement.

While convex strategyproofness makes truthful reporting a dominant strategy for *some* agents, the different notion of *approximate strategyproofness* applies to *all* agents but only bounds their potential gain from misreporting by a small albeit positive amount. However, in ordinal domains, bounding these gains in a meaningful way is challenging because utilities are typically not comparable across agents. Nevertheless, one can formalize approximate strategyproofness for assignment mechanisms with the additional assumption that the agents' utility functions take values between 0 and 1 (Birrell and Pass, 2011; Carroll, 2013; Lee, 2015).

**Definition 8.** Given a setting  $(N, M, q)$  and a bound  $\varepsilon \in [0, 1]$ , a mechanism  $f$  is  $\varepsilon$ -*approximately strategyproof* if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ ,

all misreports  $P'_i \in \mathcal{P}$ , and all utility functions  $u_i \in U_{P_i}$  with  $u_i : M \rightarrow [0, 1]$ , we have  $\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] \geq \mathbb{E}_{f_i(P'_i, P_{-i})}[u_i] - \varepsilon$ .

Finally, Azevedo and Budish (2019) proposed *strategyproofness in the large*, a concept which formalizes the intuition that agents have a diminishing benefit from misreporting as markets get larger. To formalize the sense in which markets *get large*, we follow Kojima and Manea (2010) and consider a sequence of settings  $(N^n, M^n, q^n)_{n \geq 1}$  with a constant set of objects ( $M^n = M$ ), a growing agent population ( $|N^n| = n$ ), and growing capacities ( $\min_{j \in M} q_j^n \rightarrow \infty$  as  $n \rightarrow \infty$ ) that satisfy overall demand ( $\sum_{j \in M} q_j^n \geq n$ ).

**Definition 9.** For a finite set of utility functions  $\{u^1, \dots, u^K\}$  and a sequence of settings  $(N^n, M^n, q^n)_{n \geq 1}$  with the above properties, a mechanism is *strategyproof in the large (SP-L)* if, for all  $\varepsilon > 0$ , there exists  $n_0 \in \mathbb{N}$  such that, for all  $n \geq n_0$ , no agent with a utility function from  $\{u^1, \dots, u^K\}$  can gain more than  $\varepsilon$  by misreporting.<sup>16</sup>

**Theorem 3.** *The following hold:*

1. *In any fixed setting  $(N, M, q)$ ,  $r$ -partial strategyproofness for some  $r > 0$  implies convex strategyproofness and thus weak SD-strategyproofness. Conversely, if  $m = 2$ , then full, partial, convex, and weak SD-strategyproofness are equivalent; if  $m \geq 3$ , there exists a mechanism that is convex strategyproof (and thus weakly SD-strategyproof) but not  $r$ -partially strategyproof for any  $r > 0$ .*
2. *In any fixed setting  $(N, M, q)$ ,  $r$ -partial strategyproofness for some  $r > 0$  implies  $\varepsilon$ -approximate strategyproofness for some  $\varepsilon < 1$ ; and for all  $\varepsilon > 0$  there exists  $r < 1$  such that  $r$ -partial strategyproofness implies  $\varepsilon$ -approximate strategyproofness. Conversely, for all  $\varepsilon > 0$ , there exists an  $\varepsilon$ -approximately strategyproof mechanism that is not  $r$ -partially strategyproof for any  $r > 0$ .*
3. *Fix a finite set of utility functions  $\{u^1, \dots, u^K\}$  and a sequence  $(N^n, M^n, q^n)_{n \geq 1}$  with  $|N^n| = n$ ,  $M^n = M$ ,  $\sum_{j \in M} q_j^n \geq n$ , and  $\min_{j \in M} q_j^n \rightarrow \infty$  as  $n \rightarrow \infty$ . If the degree of strategyproofness of a mechanism  $f$  converges to 1 as  $n \rightarrow \infty$ , then  $f$  is strategyproof in the large. Conversely, there exists a mechanism that is strategyproof in the large but not  $r$ -partially strategyproof for any  $r > 0$  in any of the settings  $(N^n, M^n, q^n)_{n \geq 1}$ .*

**Proof.** *Statement 1.* Observe that  $r$ -partial strategyproofness implies convex strategyproofness because any utility function from  $\text{URBI}(r)$  can take the role of  $u_i$  in Definition 7, and convex strategyproofness implies weak SD-strategyproofness. Example 3 in Appendix B gives a mechanism that is convex strategyproof but violates upper invariance (and therefore partial strategyproofness).

*Statement 2.* Let  $f$  be  $r$ -partially strategyproof for some  $r > 0$ . The proof of Lemma 1 shows that any misreport under an  $r$ -partially strategyproof mechanism reduces the agent’s assignment

<sup>16</sup> The original definition of strategyproofness in the large weakens strategyproofness further by also taking an *interim perspective*, which assumes that agents are uncertain about the other agents’ reports. By omitting the interim perspective in our simplified Definition 9, we obtain a more demanding version of SP-L. In Theorem 3, we show in what sense partial strategyproofness implies this more demanding version and therefore also the original definition of SP-L.

for the most preferred object for which the assignment changes by at least  $\delta > 0$ . The mechanism is therefore  $(1 - \delta)$ -approximately strategyproof.

By Proposition 7 in Appendix F,  $r$ -partial strategyproofness of  $f$  can be equivalently expressed in terms of the following finite system of inequalities: for all  $i \in N$ ,  $(P_i, P_{-i}) \in \mathcal{P}^N$  (where  $P_i : j_1 \succ \dots \succ j_m$ ),  $P'_i \in \mathcal{P}$ , and  $K \in \{1, \dots, m\}$ , we have

$$\sum_{k=1}^K r^k f_{i,j_k}(P_i, P_{-i}) \geq \sum_{k=1}^K r^k f_{i,j_k}(P'_i, P_{-i}). \tag{17}$$

With  $r > 0$ , we obtain

$$\begin{aligned} \sum_{k=1}^K f_{i,j_k}(P'_i, P_{-i}) - f_{i,j_k}(P_i, P_{-i}) &\leq \sum_{k=1}^K (1 - r^k) (f_{i,j_k}(P'_i, P_{-i}) - f_{i,j_k}(P_i, P_{-i})) \\ &\leq \sum_{k=1}^K (1 - r^k) \leq \sum_{k=1}^m (1 - r^k). \end{aligned} \tag{18}$$

By choosing  $r$  close to 1, the last term in (18) can be made arbitrarily small. This implies a property that Liu and Pycia (2016) called  $\varepsilon$ -strategy-proofness, which is equivalent to  $\varepsilon$ -approximate strategyproofness (by Theorem 1 in (Mennle and Seuken, 2016)). To see why the converse does not hold, refer to our proof of Statement 3.

*Statement 3.* Observe that any utility function  $u_i \in U_{P_i}$  satisfies URBI( $r$ ) for some  $r < 1$ . Thus, we can choose  $\bar{r} < 1$  large enough such that  $u^k \in \text{URBI}(\bar{r})$  for all  $u^k \in \{u^1, \dots, u^K\}$ .  $f$  is  $\bar{r}$ -partially strategyproof in all sufficiently large markets because the degree of strategyproofness of  $f$  converges to 1 as markets get large (by assumption). Thus, in sufficiently large markets, truthful reporting is a dominant strategy for all agents with utility functions in  $\{u^1, \dots, u^K\}$ .

The converse does not hold: Let  $f$  be a mechanism under which an agent gets some fixed object with probability  $(n - 1)/n$  and its reported last choice with probability  $1/n$ . This mechanism is  $1/n$ -approximately strategyproof in the  $n$ th setting and therefore strategyproof in the large, but it is not even weakly SD-strategyproof in any setting.<sup>17</sup> □

Theorem 3 significantly expands the usefulness of the partial strategyproofness concept: First, given a partially strategyproof mechanism, Statement 1 shows that weak SD-strategyproofness and convex strategyproofness are implied. Thus, for instance, partial strategyproof of the probabilistic serial mechanism (see Proposition 3 in Section 5.1) tightens prior results of Bogomolnaia and Moulin (2001) and Balbuzanov (2016).

Second, by definition,  $r$ -partially strategyproof mechanisms make truthful reporting a dominant strategy for all agents whose utility functions satisfy URBI( $r$ ). The definition of  $r$ -partial strategyproofness does not rule out that agents with utility functions violating URBI( $r$ ) may benefit from misreporting. However, Statement 2 shows that for those agents, their potential gain from misreporting is bounded and that this bound is tighter for greater  $r$ .

<sup>17</sup> While  $f$  is a valid counterexample, it may appear artificial and inefficient. To address this concern, in Appendix B, we construct a mechanism that is ex-post efficient, anonymous, upper invariant, monotonic (i.e., ranking an object higher does not lead to a strictly lower assignment for that object), and  $\varepsilon$ -approximately strategyproof for a given  $\varepsilon < 1$ , yet fails to be  $r$ -partially strategyproof for any  $r > 0$ . This highlights that partial strategyproofness and approximate strategyproofness are fundamentally different concepts: the former requires that there is no benefit from misreporting for a specific subset of the agents, while the latter requires that the benefit from misreporting is bounded for all agents.

The connection between partial and approximate strategyproofness allows a refinement of the straightforward and honest strategic advice to agents who participate in an  $r$ -partially strategyproof mechanism: *An agent is best off reporting truthfully if its preference intensities differ sufficiently (i.e., the ratio between its normalized utilities for any two objects is at most  $r$ ); otherwise, the agent's potential gain from misreporting may be positive but it is quantifiably bounded (by  $\varepsilon$  times the utility difference between its first and last choice).* Ideally, the values of  $r$  and  $\varepsilon$  can be determined by the mechanism designer for a given mechanism in a given setting. We discuss computability of  $r$  in Appendix F; regarding computability of  $\varepsilon$  see (Mennle and Seuken, 2016). When the exact values of  $r$  and  $\varepsilon$  are not available, upper or lower bounds may present a useful second best (e.g., see (Abächerli, 2017) for an analytical lower bound on the degree of strategyproofness of the probabilistic serial mechanism).

Finally, Azevedo and Budish (2019) developed strategyproofness in the large to identify mechanisms for which incentives to misreport vanish in large markets. Convergence of a mechanism's degree of strategyproofness to 1 in large markets is an alternative way to formalize a similar idea. It is reassuring to know that, by Statement 3, both concepts are consistent in the sense that, if a mechanism's degree of strategyproofness converges to 1 in large markets, then it is also strategyproof in the large. We use this result in our proof of Corollary 1.

## 5. Applications

We now present two applications of our new partial strategyproofness concept. In Section 5.1, we prove partial strategyproofness of the probabilistic serial mechanism. This yields the most demanding description of the incentive properties of this mechanism known to date, both in any finite setting and in the limit as markets get large. In Section 5.2, we show how partial strategyproofness can be employed to distinguish two common variants of the Boston mechanisms by their incentive properties, a distinction that has remained elusive until now.

### 5.1. New insights about the probabilistic serial mechanism

The *probabilistic serial (PS)* mechanism (Bogomolnaia and Moulin, 2001) is one of the most well-studied mechanisms for the random assignment problem. It uses the *Simultaneous Eating* algorithm to determine an assignment: All agents begin by consuming probability shares of their respective most-preferred objects at equal speeds. Once an object's capacity has been completely consumed, the agents consuming this object move on to their respective next most-preferred objects and continue consuming shares of these. This process continues until all agents have collected a total of 1 in probability shares, and these shares constitute their final assignments.

The next proposition shows that PS is partially strategyproof.

**Proposition 3.** *Given a setting  $(N, M, q)$ , PS is  $r$ -partially strategyproof for some  $r > 0$ .*

**Proof.** PS is LD-strategyproof by Theorem 3 in (Cho, 2018). Therefore, in any fixed setting, it is  $r$ -partially strategyproof for some  $r > 0$  by Lemma 1.  $\square$

Regarding the incentive properties of PS, Bogomolnaia and Moulin (2001) already showed that it is weakly SD-strategyproof (but not strategyproof), and Balbuzanov (2016) strengthened their result by showing that it is convex strategyproof. Since partial strategyproofness is strictly

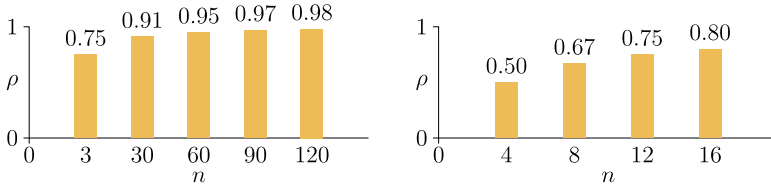


Fig. 2. Plot of  $\rho_{(N,M,q)}(PS)$  for  $m = 3$  (left) and  $m = 4$  (right) objects, for varying numbers of agents  $n$ , and evenly distributed capacities  $q_j = n/m$ .

stronger than both properties, Proposition 3 establishes the most demanding description of the incentive properties of PS in finite settings known to date.

To see what degrees of strategyproofness PS achieves, we have computed these values in various settings using Algorithm 1 from Appendix F. Fig. 2 shows the results. Observe that the values increase as the number of agents increases (while keeping the number of objects constant and evenly increasing capacities to satisfy demand). This suggests the hypothesis that the degree of strategyproofness of PS converges to 1 as the settings get large.<sup>18</sup> In subsequent work, Abächerli (2017) proved this hypothesis to be true. We use his convergence result to obtain an elegant, parametric proof of the following corollary.

**Corollary 1.** *PS is strategyproof in the large.*<sup>19</sup>

**Proof.** Let  $(N^n, M^n, q^n)_{n \geq 1}$  be a sequence of settings that get large as in Definition 9. Theorem 1 of Abächerli (2017) implies that  $\rho_{(N^n, M^n, q^n)}(PS) \rightarrow 1$  as  $n \rightarrow \infty$ . With this, Statement 3 of Theorem 3 implies that PS is strategyproofness in the large.  $\square$

The new insights about incentives under PS for both finite settings and in large markets highlight the usefulness of the partial strategyproofness concept. In particular, they refine our understanding by providing the structure of the sets of utility functions (via the URBI( $r$ ) domain restriction) for which truthful reporting is guaranteed to be a dominant strategy.

**Remark 4.** Our motivating example in the introduction considered PS in a setting with 3 agents and 3 objects with unit capacity. Recall that agent 1 could benefit from misreporting if  $3/4u_1(a) < u_1(b)$ . On the other hand, the degree of strategyproofness of PS in this setting is  $3/4$  (see Fig. 2), which implies that agent 1 is best off reporting truthfully if  $3/4u_1(a) \geq u_1(b)$ . Thus, in this setting, the set of utility functions for which PS makes truthful reporting a dominant strategy is *exactly* URBI( $3/4$ ). Similarly, in a setting with 4 agents and 4 objects with unit capacity, this set is *exactly* URBI( $1/2$ ). However, in a setting with 5 agents and 5 objects with unit capacity, an analogous statement fails to hold (see (Abächerli, 2017) for details on the settings with 4 and 5 agents).

<sup>18</sup> Note that, independent of our numerical results, Kojima and Manea (2010) presented a result about incentives under PS in large markets, which also suggests the same hypothesis but does not imply it.

<sup>19</sup> Azevedo and Budish (2019) first observed strategyproofness in the large for PS by invoking the result of Kojima and Manea (2010).

5.2. Comparing variants of the Boston mechanism

Many school districts around the world employ school choice mechanisms to assign students to seats at public schools. The *Boston mechanism (BM)* (Abdulkadiroğlu and Sönmez, 2003) is frequently used in the United States. Under BM, students apply to their respective first choices in the first round, and schools accept applicants by priority.<sup>20</sup> If a school receives more applications than it has seats, then it rejects applicants with the lowest priorities. Rejected students enter the second round, where they apply to their respective second choices. Again, schools accept additional applicants according to priority and up to capacity, and they reject the remaining applicants once all seats are filled. This process continues with third, fourth, etc. choices until all seats are taken or all students have been assigned.

One notable aspect of BM is that a student who has been rejected by her first choice may find that all seats at her second choice have been taken in the first round as well. Thus, when applying to her second choice in the second round, she effectively wastes one round where she could have competed for unfilled seats at other schools. This makes “skipping exhausted schools” an obvious heuristic for manipulation. A different variant, the *adaptive Boston mechanism (ABM)*, is more common in Europe (Dur et al., 2018).<sup>21</sup> Under ABM, students apply to their *best available choice* in every round. Skipping exhausted schools is therefore no longer necessary.

Intuitively, this makes ABM less susceptible to manipulation than BM. However, this difference is surprisingly challenging to formalize. Pathak and Sönmez (2013) proposed a natural concept for comparing mechanisms by their *vulnerability to manipulation*, but this concept fails to capture the difference between BM and ABM: For the case of strict and fixed priorities, it indicates equivalence, and for the case when priorities are determined by a single uniform lottery, the comparison is inconclusive (Dur et al., 2018). Interestingly, we can use partial strategyproofness to recover a meaningful distinction between BM and ABM.

To state this result formally, we require additional notation: A *priority order*  $\pi_j$  is a strict order over agents, where  $\pi_j : i > i'$  indicates that agent  $i$  has priority over agent  $i'$  at object  $j$ . We denote by  $\Pi$  the set of all possible priority orders. A *priority profile*  $\pi = (\pi_j)_{j \in M} \in \Pi^M$  is a collection of priority orders of all objects, and  $\pi$  is a *single priority profile* if  $\pi_j = \pi_{j'}$  for all  $j, j' \in M$ . A *priority distribution*  $\mathbb{P}$  is a probability distribution over priority profiles  $\Pi^M$ , and  $\mathbb{P}$  *supports all single priority profiles* if  $\mathbb{P}[\pi] > 0$  for all single priority profiles  $\pi$ . A *school choice mechanism*  $\varphi$  is a mapping  $\varphi : \mathcal{P}^N \times \Pi^M \rightarrow X$  that selects a deterministic assignment based on a preference profile and a priority profile. BM and ABM are examples of such school choice mechanisms. Finally, for a priority distribution  $\mathbb{P}$ , we define  $\varphi^{\mathbb{P}}(P) = \sum_{\pi \in \Pi} \varphi(P, \pi) \mathbb{P}[\pi]$ . Then  $f = \varphi^{\mathbb{P}}$  is a random assignment mechanism, and it captures the strategic situation of agents when priorities are drawn from  $\mathbb{P}$ .

With this, we can formalize the distinction between mechanisms  $BM^{\mathbb{P}}$  and  $ABM^{\mathbb{P}}$ .

**Proposition 4.** *Given a setting  $(N, M, q)$  and a priority distribution  $\mathbb{P}$  that supports all single priority orders,  $BM^{\mathbb{P}}$  and  $ABM^{\mathbb{P}}$  are upper invariant and  $ABM^{\mathbb{P}}$  is swap monotonic, but  $BM^{\mathbb{P}}$  violates swap monotonicity if  $m \geq 4$ ,  $n \geq 4$ , and  $\sum_{j \in M} q_j = n$ .*

<sup>20</sup> School choice mechanisms take into account priorities (Abdulkadiroğlu et al., 2006), which may depend on observable factors (e.g., grades, siblings, walk zones) and on lotteries (e.g., to break ties).

<sup>21</sup> See (Alcalde, 1996; Miralles, 2008; Harless, 2019; Mennle and Seuken, 2017d; Dur, 2019) for prior work that studied ABM.

**Proof.** For all priority profiles  $\pi \in \Pi^M$ , upper invariance of  $\text{BM}(\cdot, \pi)$  and  $\text{ABM}(\cdot, \pi)$  is obvious. Since  $\text{BM}^{\mathbb{P}}$  and  $\text{ABM}^{\mathbb{P}}$  are simply convex combinations of these mechanisms for different priority profiles, they inherit this property. Next, observe that  $\text{ABM}(\cdot, \pi)$  is *monotonic* (i.e., for all  $i \in N$ ,  $(P_i, P_{-i}) \in \mathcal{P}^N$ ,  $P'_i \in N_{P_i}$  with  $P_i : a \succ b$  but  $P'_i : b \succ a$ , we have  $\text{ABM}_{i,b}((P_i, P_{-i}), \pi) \leq \text{ABM}_{i,b}((P'_i, P_{-i}), \pi)$ ), and this property is again inherited by  $\text{ABM}^{\mathbb{P}}$ . Lemma 3 in Appendix C shows that  $\text{ABM}_i^{\mathbb{P}}(P_i, P_{-i}) \neq \text{ABM}_i^{\mathbb{P}}(P'_i, P_{-i})$  implies existence of a single priority profile  $\pi$  with  $\text{ABM}_{i,b}((P_i, P_{-i}), \pi) < \text{ABM}_{i,b}((P'_i, P_{-i}), \pi)$ . With this, monotonicity of  $\text{ABM}^{\mathbb{P}}$ , and the fact that  $\mathbb{P}$  supports all single priority profiles, we obtain swap monotonicity of  $\text{ABM}^{\mathbb{P}}$ . Example 5 in Appendix C shows that  $\text{BM}^{\mathbb{P}}$  violates swap monotonicity in the respective settings.  $\square$

Proposition 4 provides a formal justification for the intuition that “ABM has better incentive properties than BM.” If priorities are sufficiently random in the sense that all single priority profiles are possible, then  $\text{ABM}^{\mathbb{P}}$  is  $r$ -partially strategyproof for some  $r > 0$  while  $\text{BM}^{\mathbb{P}}$  has a degree of strategyproofness of 0.

**Remark 5.** Proposition 4 requires priorities to support all single priority profiles. Various examples from practice meet this requirement, including middle school assignment in Beijing (Lai et al., 2009), inter-district assignment to elementary schools in Estonia (Lauri et al., 2014), the supplementary round of high school assignment in New York City (Pathak and Sethuraman, 2011), and secondary school assignment in Berlin (Basteck et al., 2018).

When priorities are coarse (e.g., based on siblings or walk-zones) or even strict (e.g., based on GPA or entrance examinations), Proposition 4 cannot be used directly to establish partial strategyproofness of  $\text{ABM}^{\mathbb{P}}$ . To apply the partial strategyproofness concept in these cases, and more generally, to mechanisms with insufficient intrinsic randomness (and even deterministic ones), one can take an *interim perspective* by considering the additional randomness that arises from the agents’ uncertainty about the preference reports from the other agents (see Section 2.8 of Mennle (2016) for a formal treatment of this approach).<sup>22</sup>

To obtain some insights regarding what degrees of strategyproofness ABM achieves, we focus on the *single uniform priority distribution*  $\mathbb{U}$  (i.e., the distribution that selects all single priority profiles with equal probability). Thus, for  $\text{ABM}^{\mathbb{U}}$ , we have computed its degrees of strategyproofness in various settings.<sup>23</sup> The results are shown in Fig. 3. Observe that  $\rho_{(N,M,q)}(\text{ABM}^{\mathbb{U}})$  is significantly lower than  $\rho_{(N,M,q)}(\text{PS})$  (Fig. 2). Furthermore, it appears to be constant as the number of agents increases (while keeping the number of objects constant and evenly increasing capacities to satisfy demand). At least for these small settings, our numbers support the intuition that “PS has better incentive properties than  $\text{ABM}^{\mathbb{U}}$ .”<sup>24</sup>

<sup>22</sup> This interim perspective is central to the definition of strategyproofness in the large (Azevedo and Budish, 2019). In work subsequent to the present paper, Dasgupta and Mishra (2020) used an interim perspective to study ordinal Bayesian incentive compatibility of random assignment mechanisms.

<sup>23</sup> The fastest known algorithm to compute the assignments under  $\text{ABM}^{\mathbb{U}}$  has exponential run-time (in contrast to PS, for which it is polynomial). For this reason, we have only computed the degrees of strategyproofness of  $\text{ABM}^{\mathbb{U}}$  in settings with up to 9 agents and up to 4 objects (see Remark 3 and Appendix F for a discussion of computability and tractability).

<sup>24</sup> Note that we do not advocate the use of any mechanism solely on the basis of partial strategyproofness. Instead, such a decision must involve additional criteria like fairness and efficiency. Moreover, our results do not pertain to the controversy in the school choice literature that strategic reporting under the Boston mechanism can lead to ex-ante welfare gains (Abdulkadiroğlu et al., 2011).

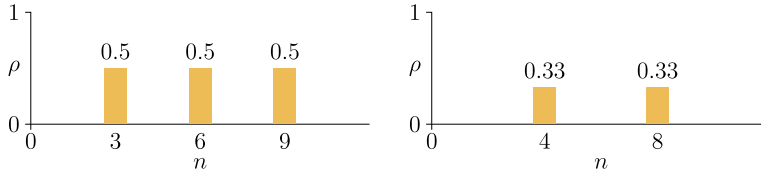


Fig. 3. Plot of  $\rho_{(N,M,q)}(\text{ABM}^U)$  for  $m = 3$  (left) and  $m = 4$  (right) objects, for varying numbers of agents  $n$ , and for  $q_j = n/m$ .

### 6. Conclusion

In this paper, we have introduced partial strategyproofness, a new concept to understand the incentive properties of non-strategyproof assignment mechanisms. This research is motivated by the restrictive impossibility results pertaining to strategyproofness in the assignment domain as well as the prevalence of non-strategyproof assignment mechanisms in practice.

In quasi-linear domains, such as auction problems, the monetary benefit from misreporting can serve as a quantifiable proxy for the extent to which agents care about strategizing. However, such a proxy is more challenging to define in ordinal domains, where cardinal preferences are typically not comparable across agents. The partial strategyproofness concept elegantly circumvents this problem: It exploits the observation that whether an agent can manipulate an assignment mechanism is often driven by how close that agent is to being *indifferent* between any two objects. The  $\text{URBI}(r)$  domain restriction separates agents by this criterion. By requiring good incentives only on the restricted domain, partial strategyproofness relaxes strategyproofness in a meaningful way and creates room for designing mechanisms that perform well on other dimensions.

The partial strategyproofness concept strikes a unique balance between two conflicting goals: It is strong enough to produce new insights yet weak enough to expand the mechanism design space. Regarding new insights, it allows us to provide honest and useful strategic advice to all agents: They are best off reporting their preferences truthfully if their preference intensities differ sufficiently between any two objects; otherwise, if they are close to being indifferent between some objects, then their potential gain from misreporting is quantifiably bounded in the sense of approximate strategyproofness. Regarding the expansion of the mechanism design space, we have demonstrated that partial strategyproofness can be applied to all mechanisms that satisfy the minimal requirement of lexicographic dominance strategyproofness. In particular, these include some of the most important assignment mechanisms, like probabilistic serial and variants of the Boston mechanism.

An important open question is what values of  $r$  are “high enough” to provide acceptable incentive guarantees. We consider this an important and interesting subject of future research. However, we do not believe that a universal answer to this question exists. Rather, the appropriate degree of strategyproofness will depend on the particular mechanism design problem at hand. For specific markets, it could be derived from survey data or revealed preference data (e.g., when preferences are expressed in terms of bidding points). Such research could be complemented by laboratory experiments that aim to identify the role of the degree of strategyproofness in human decisions to manipulate non-strategyproof assignment mechanisms.

Despite this open question, partial strategyproofness can already guide design decisions. For example, we can use it to compare non-strategyproof assignment mechanisms by their incentive properties (as we have illustrated by applying it to PS, ABM, and BM). In addition, the



parametric nature of partial strategyproofness enables new quantifiable trade-offs between incentives and other design desiderata. For example, in (Mennle and Seuken, 2017b), we have used it to identify the possible and necessary trade-offs between strategyproofness and efficiency that can be achieved via hybrid mechanisms. Going forward, we are confident that the partial strategyproofness concept will be a useful addition to the mechanism designer’s toolbox and that it will facilitate the study of non-strategyproof assignment mechanisms and the design of new ones.

**Appendix A. Proof of Lemma 2 for Proposition 1**

**Lemma 2.** *The mechanism  $\tilde{f}$  constructed in the proof of Proposition 1 is  $r$ -partially strategyproof and manipulable for agent  $i$  with utility function  $\tilde{u}_i$ .*

**Proof.** Observe that, under  $\tilde{f}$ ,  $i$ ’s assignment is independent of the other agents’ preference reports, so the mechanism is strategyproof for all other agents except  $i$ . Thus, we can guarantee  $r$ -partial strategyproofness of  $\tilde{f}$  by verifying that truthful reporting maximizes  $i$ ’s expected utility for any preference order  $P_i$  and any utility function  $u_i \in U_{P_i}$  that satisfies URBI( $r$ ).

If  $P_i : a \succ b$ , then  $i$ ’s assignment remains unchanged, unless  $i$  claims to prefer  $b$  to  $a$  (i.e.,  $\hat{P}_i : b \succ a$ ). If  $a$  is  $i$ ’s reported last choice under  $\hat{P}_i$ , then the change in  $i$ ’s expected utility from the misreport is  $-\delta_b u_i(a) + \delta_b u_i(b) \leq 0$ . If  $\hat{d} \neq a$  is  $i$ ’s reported last choice, then this change is  $-\delta_a u_i(a) + \delta_b u_i(b) + (\delta_a - \delta_b) u_i(\hat{d})$ . Since  $\delta_a < \delta_b$  by assumption, this value strictly increases if  $i$  ranks its true last choice  $d$  last. Thus,  $i$ ’s gain from misreporting is upper bounded by

$$-\delta_a u_i(a) + \delta_b u_i(b) + (\delta_a - \delta_b) u_i(d) \tag{19}$$

$$= -\delta_a \left( u_i(a) - \min_{j \in M} u_i(j) \right) + \delta_b \left( u_i(b) - \min_{j \in M} u_i(j) \right). \tag{20}$$

This bound is guaranteed to be weakly negative if  $\delta_a \geq r\delta_b$ , where we use the fact that  $u_i$  satisfies URBI( $r$ ).

Next, if  $P_i : b \succ a$ , then  $i$ ’s assignment from truthful reporting stochastically dominates any assignment that  $i$  can obtain by misreporting.

Finally, we need to verify that  $i$  finds a beneficial misreport if its utility function is  $\tilde{u}_i$ . Let  $P'_i$  be the same preference order as  $\tilde{P}_i$ , except that  $a$  and  $b$  trade positions. The change in  $i$ ’s expected utility from reporting  $P'_i$  is

$$-\delta_a \tilde{u}_i(a) + \delta_b \tilde{u}_i(b) + (\delta_a - \delta_b) \tilde{u}_i(d) \tag{21}$$

$$= -\delta_a \left( \tilde{u}_i(a) - \min_{j \in M} \tilde{u}_i(j) \right) + \delta_b \left( \tilde{u}_i(b) - \min_{j \in M} \tilde{u}_i(j) \right), \tag{22}$$

where  $d \neq b$  is  $i$ ’s true last choice. This change is strictly positive if  $\delta_a < \tilde{r}\delta_b$ .  $\square$

**Appendix B. Examples used in the proof of Theorem 3**

**Example 3** (*Convex strategyproofness does not imply upper invariance*). Consider a setting with one agent  $i$  and three objects  $a, b, c$  with unit capacity. Suppose that ranking  $b$  over  $c$  leads to an assignment of  $y_i = (0, 1/2, 1/2)$  for  $a, b, c$ , respectively, and ranking  $c$  over  $b$  leads to  $x_i = (1/4, 0, 3/4)$ . Without loss of generality, let  $u_i$  be such that the utility for the last choice is 0 and the utility for the second choice is 1. Then  $i$  can only benefit from misreporting in the following two cases:

1.  $P_i : a \succ b \succ c$  and  $u_i(a) > 2$ ,
2.  $P_i : c \succ b \succ a$  and  $u_i(c) \in (1, 2)$ .

The mechanism is therefore convex strategyproof but not upper invariant.

**Example 4** ( *$\epsilon$ -approximate strategyproofness does not imply  $r$ -partial strategyproofness*). For any  $\beta \in [0, 1]$ , let  $h^\beta$  be the mechanism that collects the agents' preference reports and then applies  $BM^\cup$  (the Boston mechanism with priorities determined by a single uniform lottery) with probability  $\beta$  and applies RSD (the random serial dictatorship mechanism with the order of dictators determined by a single uniform lottery) with probability  $(1 - \beta)$ . Ex-post efficiency of  $h^\beta$  follows from Theorem 2 of (Mennle and Seuken, 2017b) and the fact that  $BM^\cup$  and RSD are both ex-post efficient. Similarly,  $h^\beta$  inherits upper invariance, monotonicity, and anonymity from  $BM^\cup$  and RSD (for a proof of monotonicity of  $BM^\cup$ , see Claim 9 in Section 3.B.2 of (Mennle, 2016)). Next, observe that RSD is 0-approximately strategyproof and  $BM^\cup$  is trivially 1-approximately strategyproof. With this,  $h^\beta$  is  $\beta$ -approximately strategyproof by Theorem 2 of (Mennle and Seuken, 2016). Finally, Example 2 of (Mennle and Seuken, 2017b) demonstrates that  $h^\beta$  is not weakly SD-strategyproof in a setting with  $n = 6$  agents and  $m = 6$  objects with unit capacity, and  $h^\beta$  is therefore not  $r$ -partially strategyproof for any  $r > 0$  in this setting.

**Appendix C. Example and lemma used in the proof of Proposition 4**

**Example 5** (*Violation of swap monotonicity by BM*). Consider a setting with  $n \geq 4$  agents  $N = \{1, 2, 3, 4, \dots, n\}$ ,  $m \geq 4$  objects  $M = \{a, b, c, d, \dots\}$ , and capacities  $q_j \geq 1$  such that  $\sum_{j \in M} q_j = n$ . Let agent 1 have preference order  $P_1 : a \succ b \succ c \succ d \succ \dots$ , let agents 2 through  $q_a + q_c + q_d$  have preference order  $P : a \succ b \succ \dots$ , and for all objects  $j \in M \setminus \{a, b, d\}$ , let there be  $q_j$  agents with preference order  $P : j \succ \dots$ . Then  $BM_1^P(P_1, P_{-1}) = (x, y, 0, 1 - x - y, 0, \dots)$  for some  $x \in (0, 1)$ ,  $y \in (0, 1 - x)$  because  $\mathbb{P}$  supports all single priority profiles. However, if  $i$  reports  $P'_i : a \succ c \succ b \succ d \succ \dots$ , then  $BM_1^P(P'_i, P_{-1}) = (x, 0, 0, 1 - x, 0, \dots)$ . Observe that agent 1's assignment changes but its assignment for  $c$  remains unchanged, a contradiction to swap monotonicity.

**Lemma 3.** *Given a setting  $(N, M, q)$ , for all priority distributions  $\mathbb{P}$ , all  $i \in N$ , all  $(P_i, P_{-i}) \in \mathcal{P}^N$ , all  $P'_i \in N_{P_i}$  with  $P_i : a \succ b$  but  $P'_i : b \succ a$ , if*

$$ABM_i^P(P_i, P_{-i}) \neq ABM_i^P(P'_i, P_{-i}), \tag{23}$$

*then there exists a priority profile  $\pi^*$  such that*

$$ABM_{i,a}((P_i, P_{-i}), \pi^*) = ABM_{i,b}((P'_i, P_{-i}), \pi^*) = 1. \tag{24}$$

**Proof.** From  $ABM_i^P(P_i, P_{-i}) \neq ABM_i^P(P'_i, P_{-i})$  it follows that there exists a priority profile  $\pi^0$  with  $ABM_i((P_i, P_{-i}), \pi^0) \neq ABM_i((P'_i, P_{-i}), \pi^0)$ . Let  $ABM_{i,j}((P_i, P_{-i}), \pi^0) = 1$  and  $ABM_{i,j'}((P'_i, P_{-i}), \pi^0) = 1$ , i.e.,  $i$  gets  $j$  by reporting  $P_i$  and  $i$  gets  $j' \neq j$  by reporting  $P'_i$ , and let  $K$  be the round in which  $i$  applies to (or skips)  $a$  when reporting  $P_i$ .

We differentiate 6 cases, identify that the first 3 are impossible, and construct single the priority profile  $\pi^*$  for the other 3:

**$P_i : j \succ a$  or  $P_i : j' \succ a$ :** Then upper invariance implies  $j' = j$ , a contradiction.

$j=b$  or  $j'=a$ : Then monotonicity implies  $j' = j$ , a contradiction.

$P_i:b \succ j$  and  $P_i:b \succ j'$ : The fact that  $i$  gets neither  $a$  nor  $b$  when reporting  $P_i$  or  $P'_i$  means that both objects are exhausted by other agents at the end of round  $K$ .  $i$  thus applies to at most one of them and skips the other. Since  $i$  is rejected, the application process after round  $K$  is the same under both reports. This implies  $j = j'$ , a contradiction.

$j=a$  and  $j'=b$ . Then  $i$  receives  $a$  or  $b$  in round  $K$  under the different reports, respectively. Observe that the application process before this round is independent of whether  $i$  reports  $P_i$  or  $P'_i$ . For all  $k \in \{1, \dots, K - 1\}$ , let  $N^k$  be the agents who receive their object in round  $k$ , and let  $\pi_j^*$  be a priority order which, first, gives highest priority to all agents in  $N_1$ , then all agents in  $N_2$ , and so on, and second, ranks  $i$  directly after the agents in  $N_{K-1}$ . Then the single priority profile  $\pi^*$  where all objects have the same priority order  $\pi_j^*$  is a single priority profile with  $ABM_{i,a}((P_i, P_{-i}), \pi^*) = ABM_{i,b}((P'_i, P_{-i}), \pi^*) = 1$ .

$j=a$  and  $P_i:b \succ j'$ : If  $b$  is exhausted at the end of round  $K - 1$ , then  $i$  would skip  $b$  when reporting  $P'_i$  and still receive  $a$ . Thus, there is still capacity of  $b$  available at the beginning of round  $K$ . Let  $\pi^*$  be the same single priority profile as in the previous case. Then  $ABM_{i,a}((P_i, P_{-i}), \pi^*) = ABM_{i,b}((P'_i, P_{-i}), \pi^*) = 1$ .

$j'=b$  and  $P_i:b \succ j$ : This case is symmetric to the previous case but where the roles of  $a$  and  $b$  are inverted.

Thus, the single priority profile  $\pi^*$  exists in all cases that do not imply contradictions.  $\square$

#### Appendix D. Equivalence of swap monotonicity and strategyproofness for deterministic mechanisms

**Proposition 5.** *A deterministic mechanism  $f$  is strategyproof if and only if it is swap monotonic.*

**Proof.** Since deterministic mechanisms are just special cases of random mechanisms, Theorem 1 applies: A deterministic mechanism  $f$  is strategyproof if and only if it is swap monotonic, upper invariant, and lower invariant. Thus, strategyproofness implies swap monotonicity (i.e., *sufficiency* in Proposition 5). For *necessity*, observe that swap monotonicity implies upper and lower invariance for deterministic mechanisms: If a swap (say from  $P_i : a \succ b$  to  $P'_i : b \succ a$ ) affects an agent's assignment, then the assignment must change strictly for the two objects  $a$  and  $b$  that are swapped. But under a deterministic mechanism, this change can only be from 0 to 1 or from 1 to 0. The only possible changes are therefore the ones where an agent receives  $a$  with certainty if it reports  $P_i : a \succ b$  and receives  $b$  with certainty if she reports  $P'_i : b \succ a$ .  $\square$

#### Appendix E. Comparing mechanisms by vulnerability to manipulation and degree of strategyproofness

The next proposition shows that the comparison of mechanisms by their vulnerability to manipulation and by their degrees of strategyproofness is *consistent* but not equivalent.

**Proposition 6.** *For any setting  $(N, M, q)$  and mechanisms  $f, g$ , the following hold:*

1. *If  $g$  is strongly as manipulable as  $f$ , then  $\rho_{(N,M,q)}(f) \geq \rho_{(N,M,q)}(g)$ .*

2. If  $\rho_{(N,M,q)}(f) > \rho_{(N,M,q)}(g)$ , and if  $f$  and  $g$  are comparable by the strongly as manipulable as relation, then  $g$  is strongly as manipulable as  $f$ .

In Proposition 6, the strongly as manipulable as relation is extended to random assignment mechanisms as follows:

**Definition 10.** For a given setting  $(N, M, q)$  and two mechanisms  $f, g$ , we say that  $g$  is *strongly as manipulable as  $f$*  if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all utility functions  $u_i \in U_{P_i}$ , the following holds: If there exists a misreport  $P'_i \in \mathcal{P}$  such that

$$\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] < \mathbb{E}_{f_i(P'_i, P_{-i})}[u_i], \tag{25}$$

then there exists a (possibly different) misreport  $P''_i \in \mathcal{P}$  such that

$$\mathbb{E}_{g_i(P_i, P_{-i})}[u_i] < \mathbb{E}_{g_i(P''_i, P_{-i})}[u_i]. \tag{26}$$

In words,  $g$  is strongly as manipulable as  $f$  if any agent who can manipulate  $f$  in a given situation can also manipulate  $g$  in the same situation.

**Proof of Proposition 6. Statement 1.** Observe that, if  $f$  is strongly as manipulable as  $g$ , then any agent who can manipulate  $g$  also finds a manipulation to  $f$ . Thus, the set of utility functions on which  $g$  makes truthful reporting a dominant strategy cannot be larger than the set of utilities on which  $f$  does the same. This in turn implies  $\rho_{(N,M,q)}(f) \geq \rho_{(N,M,q)}(g)$ .

**Statement 2.** Observe that, if  $\rho_{(N,M,q)}(f) > \rho_{(N,M,q)}(g)$ , then there exists a utility function  $\tilde{u}$  in  $\text{URBI}(\rho_{(N,M,q)}(f))$ , which is not in  $\text{URBI}(\rho_{(N,M,q)}(g))$ , and for which  $g$  is manipulable, but  $f$  is not. Thus,  $f$  cannot be strongly as manipulable as  $g$ , but the converse is possible.  $\square$

### Appendix F. Discounted dominance and computability

In this section, we define a new dominance notion we call *r-discounted dominance* and prove that the induced incentive concept, *r-discounted dominance strategyproofness*, is equivalent to *r-partial strategyproofness*. We then show that this equivalence allows us to devise a simple algorithm for computing the degree of strategyproofness. We also discuss computational challenges and potential remedies.

#### F.1. Discounted dominance

The new dominance notion generalizes stochastic dominance but includes  $r$  as a discount factor.

**Definition 11.** For a bound  $r \in [0, 1]$ , a preference order  $P_i \in \mathcal{P}$  with  $P_i : j_1 \succ \dots \succ j_m$ , and assignment vectors  $x_i, y_i$ , we say that  $x_i$  *r-discounted dominates*  $y_i$  at  $P_i$  if, for all ranks  $K \in \{1, \dots, m\}$ , we have

$$\sum_{k=1}^K r^k x_{i,j_k} \geq \sum_{k=1}^K r^k y_{i,j_k}. \tag{27}$$

Observe that, for  $r = 1$ , this is precisely the same as stochastic dominance. However, for  $r < 1$ , the difference in the agent's assignment for the  $k$ th choice is discounted by the factor  $r^k$ . Analogous to stochastic dominance for SD-strategyproofness, we can use  $r$ -discounted dominance ( $r$ -DD) to define the corresponding incentive concept.

**Definition 12.** Given a setting  $(N, M, q)$  and a bound  $r \in (0, 1]$ , a mechanism  $f$  is  $r$ -DD-strategyproof if, for all agents  $i \in N$ , all preference profiles  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and all misreports  $P'_i \in \mathcal{P}$ ,  $f_i(P_i, P_{-i})$   $r$ -discounted dominates  $f_i(P'_i, P_{-i})$  at  $P_i$ .

The next proposition yields the equivalence to  $r$ -partial strategyproofness.

**Proposition 7.** Given a setting  $(N, M, q)$  and a bound  $r \in [0, 1]$ , a mechanism  $f$  is  $r$ -partially strategyproof if and only if it is  $r$ -DD-strategyproof.

**Proof.** Given the setting  $(N, M, q)$ , we fix an agent  $i \in N$ , a preference profile  $(P_i, P_{-i}) \in \mathcal{P}^N$ , and a misreport  $P'_i \in \mathcal{P}$ . The following claim establishes equivalence of the  $r$ -partial strategyproofness constraints and the  $r$ -DD-strategyproofness constraints for any such combination  $(i, (P_i, P_{-i}), P'_i)$  with  $x = f_i(P_i, P_{-i})$  and  $y = f_i(P'_i, P_{-i})$ .

**Claim 1.** Given a setting  $(N, M, q)$ , a preference order  $P_i \in \mathcal{P}$ , assignment vectors  $x, y$ , and a bound  $r \in [0, 1]$ , the following are equivalent:

- A. For all utility functions  $u_i \in U_{P_i} \cap \text{URBI}(r)$  we have  $\sum_{j \in M} u_i(j)x_j \geq \sum_{j \in M} u_i(j)y_j$ .
- B.  $x_i$   $r$ -discounted dominates  $y_i$  at  $P_i$ .

**Proof of Claim 1.** Sufficiency ( $B \Rightarrow A$ ). Let  $P_i : j_1 \succ \dots \succ j_m$ . Assume towards contradiction that Statement B holds, but that for some utility function  $u_i \in U_{P_i} \cap \text{URBI}(r)$ , we have

$$\sum_{l=1}^m u_i(j_l) (x_{j_l} - y_{j_l}) < 0. \tag{28}$$

Without loss of generality, we can assume  $\min_{j \in M} u_i(j) = 0$ . Let  $\delta_k = x_{j_k} - y_{j_k}$  and let

$$S(K) = \sum_{k=1}^K u_i(j_k)(x_{j_k} - y_{j_k}) = \sum_{k=1}^K u_i(j_k)\delta_k. \tag{29}$$

By assumption,  $S(m) = S(m - 1) < 0$  (see (28)), so there exists a smallest value  $K' \in \{1, \dots, m - 1\}$  such that  $S(K') < 0$ , but  $S(k) \geq 0$  for all  $k < K'$ . Using Horner's method, we rewrite the partial sum and get

$$S(K') = \sum_{k=1}^{K'} u_i(j_k)\delta_k = \left( \frac{S(K' - 1)}{u_i(j_{K'})} + \delta_{K'} \right) u_i(j_{K'}) \tag{30}$$

$$= \left( \left( \frac{S(K' - 2)}{u_i(j_{K'-1})} + \delta_{K'-1} \right) \frac{u_i(j_{K'-1})}{u_i(j_{K'})} + \delta_{K'} \right) u_i(j_{K'}) \tag{31}$$

$$= \left( \left( \dots \left( \delta_1 \frac{u_i(j_1)}{u_i(j_2)} + \delta_2 \right) \frac{u_i(j_2)}{u_i(j_3)} + \dots \right) \frac{u_i(j_{K'-1})}{u_i(j_{K'})} + \delta_{K'} \right) u_i(j_{K'}). \tag{32}$$

Since  $u_i$  satisfies URBI( $r$ ), the fraction  $\frac{u_i(j_{K'-1})}{u_i(j_{K'})}$  is bounded from below by  $1/r$ . But since  $S(K' - 1) \geq 0$  and  $u_i(j_{K'-1}) > 0$ , we must have that

$$\left( \frac{S(K' - 2)}{u_i(j_{K'-1})} + \delta_{K'-1} \right) = \frac{S(K' - 1)}{u_i(j_{K'-1})} \geq 0. \tag{33}$$

Therefore, when replacing  $\frac{u_i(j_{K-1})}{u_i(j_K)}$  by  $1/r$  in (32) we only make the term smaller. By the same argument, we can successively replace all the terms  $\frac{u_i(j_{k-1})}{u_i(j_k)}$  and obtain

$$0 > S(K') \geq \left( \left( \dots \left( \frac{\delta_1}{r} + \delta_2 \right) \frac{1}{r} + \dots \right) \frac{1}{r} + \delta_{K'} \right) u_i(j_{K'}) = \frac{u_i(j_{K'})}{r^{K'}} \sum_{k=1}^{K'} r^k \delta_k. \tag{34}$$

This contradicts  $r$ -discounted dominance of  $x$  over  $y$  at  $P_i$ , since

$$\sum_{k=1}^{K'} r^k \delta_k = \sum_{k=1}^{K'} r^k (x_{j_k} - y_{j_k}) \geq 0. \tag{35}$$

*Necessity (A  $\Rightarrow$  B).* Let  $P_i : j_1 > \dots > j_m$ . Assume towards contradiction that Statement A holds, but  $x$  does not  $r$ -discounted dominate  $y$  at  $P_i$ , i.e., for some  $K \in \{1, \dots, m\}$ , we have

$$\sum_{k=1}^K r^k x_{j_k} < \sum_{k=1}^K r^k y_{j_k}, \tag{36}$$

and let  $K$  be the smallest rank for which Inequality (36) is strict. Then the value

$$\Delta = \sum_{k=1}^K r^k (y_{j_k} - x_{j_k}), \tag{37}$$

is strictly positive. Let  $D \geq d > 0$ , and let  $u_i$  be the utility function defined by

$$u_i(j_k) = \begin{cases} Dr^k, & \text{if } k \leq K, \\ dr^k, & \text{if } K + 1 \leq k \leq m - 1, \\ 0, & k = m. \end{cases} \tag{38}$$

This utility function satisfies URBI( $r$ ). Furthermore, we have

$$\sum_{j \in M} u_i(j)x_j - \sum_{j \in M} u_i(j)y_j = \sum_{l=1}^m u(j_l) (x_{i,j_l} - y_{i,j_l}) \tag{39}$$

$$\begin{aligned} &= D \sum_{k=1}^K r^k (x_{j_k} - y_{j_k}) + d \sum_{k=K+1}^{m-1} r^k (x_{j_k} - y_{j_k}) \\ &\leq -D\Delta + d. \end{aligned} \tag{40}$$

For  $d < D\Delta$ ,  $\sum_{j \in M} u_i(j)x_j - \sum_{j \in M} u_i(j)y_j$  is strictly negative, a contradiction.  $\square$

This concludes the proof of Proposition 7.  $\square$

Proposition 7 generalizes the equivalence between EU-strategyproofness and SD-strategyproofness (Erdil, 2014). Moreover, it yields an alternative definition of  $r$ -partial strategyproofness in terms of discounted dominance. This shows that the partial strategyproofness concept integrates nicely into the landscape of existing incentive concepts, many of which are defined using dominance notions (e.g., SD-, weak SD-, LD-strategyproofness, and sure thing dominance strategyproofness (Aziz et al., 2013a)).

### F.2. Computability

The dominance interpretation from Proposition 7 enables a computational approach to partial strategyproofness: Recall that Definition 4 of  $r$ -partial strategyproofness imposes inequalities for all utility functions within the set  $URBI(r)$ . This set is infinite, which makes an algorithmic verification of  $r$ -partial strategyproofness infeasible via its original definition. However, by the equivalence from Proposition 7, it suffices to verify that all (finitely many) constraints for  $r$ -discounted dominance strategyproofness are satisfied (i.e., the Inequalities (27) from Definition 11). This yields Algorithm 1 to compute the degree of strategyproofness of any mechanism in any setting (provided that the mechanism  $f$  itself is computable).

---

**Algorithm 1:** Find degree of strategyproofness.

---

**Input:** Mechanism  $f$ , setting  $(N, M, q)$

**Output:** Degree of strategyproofness  $\rho_{(N, M, q)}(f)$

$r \leftarrow 1$

**for all preference profiles**  $P \in \mathcal{P}^N$  **do**

**for all agents**  $i \in N$  **do**

        let  $j_1, \dots, j_m \in M$  be objects such that  $P_i : j_1 \succ \dots \succ j_m$

**for all misreports**  $P'_i \in \mathcal{P}, P'_i \neq P_i$  **do**

**for all ranks**  $K \in \{1, \dots, m\}$  **do**

                define polynomial  $p(\cdot) : s \mapsto \sum_{k=1}^K s^k (f_{i, j_k}(P_i, P_{-i}) - f_{i, j_k}(P'_i, P_{-i}))$

$r' \leftarrow \max \{\bar{s} \in [0, 1] : p(s) \geq 0 \text{ for all } s \in [0, \bar{s}]\}$

$r \leftarrow \min(r, r')$

**end**

**end**

**end**

**end**

**return**  $r$

---

The algorithm begins by optimistically assuming  $r = 1$ . It then iterates over all possible combinations of preference profiles, agents, misreports, and ranks. For each combination, it checks whether the current  $r$  must be corrected downwards to avoid a violation of the respective constraint for discounted dominance from Definition 11. Specifically, we define the polynomial  $p(\cdot)$  to represent the respective constraint. Next, we determine the largest bound  $r'$  such that  $p(s) \geq 0$  for all  $s \in [0, r']$ , and we reduce  $r$  if  $r'$  is smaller than  $r$ .<sup>25</sup> This ensures that all discounted dominance constraints checked so far are satisfied for the current  $r$  but not for any larger values. Therefore, the value of  $r$  when the algorithm terminates is the degree of strategyproofness of the mechanism  $f$  in the setting  $(N, M, q)$ .

---

<sup>25</sup> Since  $p(0) = 0$  by construction,  $r'$  is well-defined. Determining  $r'$  requires an analysis of the roots of  $p(\cdot)$ . Finding roots of polynomials with rational coefficients is possible in polynomial time via the LLL algorithm (Lenstra et al., 1982).

We have used Algorithm 1 to compute the degrees of strategyproofness of PS and  $\text{ABM}^{\text{U}}$  in a number of small settings in Section 5. The limited setting sizes are owed to the exponential number of preference profiles and misreports over which the algorithm iterates. This issue is further exacerbated by the fact that computing the assignment matrix for even a single preference profile can be a computationally expensive problem for some mechanisms (e.g., Aziz et al. (2013b) proved that computing the assignment resulting from random serial dictatorship is  $\#P$ -hard, and for  $\text{ABM}^{\text{U}}$  the fastest known algorithm has exponential runtime). For these reasons, Algorithm 1 is only practical in small settings and tractability may be an issue in larger settings.

To address the problem of tractability, we observe that most applications come with additional restrictions, such as fairness requirements, admissible types of mechanisms, or limitations on preference reports. Exploiting these restrictions provides avenues for future research. For example, axioms like neutrality, symmetry, anonymity, or non-bossiness impose equality constraints. These constraints can be exploited to reduce the number of preference profiles that must be considered. Suppose that we restrict attention to anonymous mechanisms; then, for every preference profile we consider, we can skip all other preference profiles that are equivalent to the first profile up to a permutation of the agents.

A second approach towards improving tractability arises from local sufficiency. Even though full strategyproofness requires that none of the  $(m! - 1)$  misreports improve the agent's assignment, it actually suffices to verify this for the  $(m - 1)$  misreports from the neighborhood of the agent's true preference order (Carroll, 2012). The same is true for LD-strategyproofness, the lower limit concept of partial strategyproofness (Cho, 2016). In Mennle and Seuken (2017c), we have shown that  $r$ -local partial strategyproofness<sup>26</sup> implies  $r^2$ -partial strategyproofness and that this bound is tight (i.e.,  $r$ -local partial strategyproofness does not imply  $r^{2-\varepsilon}$ -partial strategyproofness for any  $\varepsilon > 0$ ). Verifying that a mechanism is  $r$ -locally partially strategyproof thus yields a lower bound  $r^2$  on its degree of strategyproofness but only requires checking  $(m - 1)$  misreports instead of  $(m! - 1)$ .

## References

- Abächerli, Daniel, 2017. A Parametric Proof for Strategyproofness in the Large of the Probabilistic Serial Mechanism. Master's thesis. University of Zurich.
- Abdulkadiroğlu, Atila, Roth, Alvin, Pathak, Parag, Sönmez, Tayfun, 2006. Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access. Working Paper.
- Abdulkadiroğlu, Atila, Sönmez, Tayfun, 1998. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica* 66 (3), 689–702.
- Abdulkadiroğlu, Atila, Sönmez, Tayfun, 2003. School choice: a mechanism design approach. *Am. Econ. Rev.* 93 (3), 729–747.
- Abdulkadiroğlu, Atila, Pathak, Parag, Roth, Alvin, 2009. Strategy-proofness versus efficiency in matching with indifference: redesigning the NYC high school match. *Am. Econ. Rev.* 99 (5), 1954–1978.
- Abdulkadiroğlu, Atila, Che, Yeon-Koo, Yasuda, Yosuke, 2011. Resolving conflicting preferences in school choice: the “Boston mechanism” reconsidered. *Am. Econ. Rev.* 101 (1), 399–410.
- Alcalde, Jose, 1996. Implementation of stable solutions to marriage problems. *J. Econ. Theory* 69 (1), 240–254.
- Azevedo, Eduardo, Budish, Eric, 2019. Strategy-proofness in the large. *Rev. Econ. Stud.* 86 (1), 81–116.
- Aziz, Haris, Brandt, Felix, Brill, Markus, 2013a. On the tradeoff between economic efficiency and strategy proofness in randomized social choice. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).

<sup>26</sup>  $r$ -local partial strategyproofness requires that, for each agent whose utility function satisfies  $\text{URBI}(r)$ , truthful reporting dominates all misreports from the *neighborhood* of its true preference order.



- Aziz, Haris, Brandt, Felix, Brill, Markus, 2013b. The computational complexity of random serial dictatorship. *Econ. Lett.* 121 (3), 341–345.
- Balbuzanov, Ivan, 2016. Convex strategyproofness with an application to the probabilistic serial mechanism. *Soc. Choice Welf.* 46 (3), 511–520.
- Basteck, Christian, Klaus, Bettina, Kübler, Dorothea, 2018. How Lotteries in School Choice Help to Level the Playing Field. Working Paper.
- Basteck, Christian, Huesmann, Katharina, Nax, Heinrich, 2015. Matching practices for secondary schools - Germany, MiP country profile 21. [www.matching-in-practice.eu](http://www.matching-in-practice.eu). (Accessed 26 June 2020).
- Birrell, Eleanor, Pass, Rafael, 2011. Approximately strategy-proof voting. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*.
- Bogomolnaia, Anna, Moulin, Hervé, 2001. A new solution to the random assignment problem. *J. Econ. Theory* 100 (2), 295–328.
- Budish, Eric, 2012. Matching “versus” mechanism design. *ACM SIGecom Exch.* 11 (2), 4–15.
- Budish, Eric, Cantillon, Estelle, 2012. The multi-unit assignment problem: theory and evidence from course allocation at Harvard. *Am. Econ. Rev.* 102 (5), 2237–2271.
- Budish, Eric, Che, Yeon-Koo, Kojima, Fuhito, Milgrom, Paul, 2013. Designing random allocation mechanisms: theory and applications. *Am. Econ. Rev.* 103 (2), 585–623.
- Carroll, Gabriel, 2012. When are local incentive constraints sufficient? *Econometrica* 80 (2), 661–686.
- Carroll, Garbiel, 2013. A Quantitative Approach to Incentives: Application to Voting Rules. Working Paper.
- Cho, Wonki, 2018. Probabilistic assignment: an extension approach. *Soc. Choice Welf.* 51 (1), 137–162.
- Cho, Wonki Jo, 2016. Incentive properties for ordinal mechanisms. *Games Econ. Behav.* 95, 168–177.
- Chun, Youngsub, Yun, Kiyong, 2019. Upper-contour strategy-proofness in the probabilistic assignment problem. *Soc. Choice Welf.* 54 (2), 667–687.
- Dasgupta, Sulagna, Mishra, Debasis, 2020. Ordinal Bayesian Incentive Compatibility in Random Assignment Model. Working Paper.
- Dur, Umut, 2019. The modified Boston mechanism. *Math. Soc. Sci.* 101, 31–40.
- Dur, Umut, Mennle, Timo, Seuken, Sven, 2018. First-choice maximal and first-choice stable school choice mechanisms. In: *Proceedings of the 19th ACM Conference on Economics and Computation (EC)*.
- Erdil, Aytek, 2014. Strategy-proof stochastic assignment. *J. Econ. Theory* 151 (1), 146–162.
- Fack, Gabrielle, Grenet, Julien, He, Yinghua, 2015. Beyond Truth-Telling: Preference Estimation with Centralized School Choice. Working Paper, earlier version of (Fack et al., 2019).
- Fack, Gabrielle, Grenet, Julien, He, Yinghua, 2019. Beyond truth-telling: preference estimation with centralized school choice and college admissions. *Am. Econ. Rev.* 109 (4).
- Featherstone, Clayton, 2020. Rank Efficiency: Modeling a Common Policymaker Objective. Working Paper.
- Harless, Patrick, 2019. Immediate Acceptance with or without Skips? Comparing School Assignment Procedures. Working Paper.
- Hashimoto, Tadashi, Hirata, Daisuke, Kesten, Onur, Kurino, Morimitsu, Ünver, Utku, 2014. Two axiomatic approaches to the probabilistic serial mechanism. *Theor. Econ.* 9 (1), 253–277.
- Kojima, Fuhito, Manea, Mihai, 2010. Incentives in the probabilistic serial mechanism. *J. Econ. Theory* 145 (1), 106–123.
- Kojima, Fuhito, Ünver, Utku, 2014. The Boston school-choice mechanism: an axiomatic approach. *Econ. Theory* 55 (3), 515–544.
- Lai, Fang, Sadoulet, Elisabeth, de Janvry, Alain, 2009. The adverse effects of parents’ school selection errors on academic achievement: evidence from the Beijing open enrollment program. *Econ. Educ. Rev.* 28 (4), 485–496.
- Lauri, Triin, Kaire, Pöder, André, Veski, 2014. Matching Practices for Elementary Schools - Estonia, MiP Country Profile 18. (Accessed 26 June 2020).
- Lee, David, 2015. Efficient, private, and epsilon-strategyproof elicitation of tournament voting rules. In: *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.
- Lenstra, Arjen, Lenstra, Hendrik, Lovász, László, 1982. Factoring polynomials with rational coefficients. *Math. Ann.* 261 (4), 515–534.
- Liu, Qingmin, Pycia, Marek, 2016. Ordinal Efficiency, Fairness, and Incentives in Large Markets. Working Paper.
- Mennle, Timo, 2016. Trade-offs between Strategyproofness and Efficiency of Ordinal Mechanisms. PhD diss. University of Zurich.
- Mennle, Timo, Seuken, Sven, 2014. An axiomatic approach to characterizing and relaxing strategyproofness of one-sided matching mechanisms [extended abstract]. In: *Proceedings of the 15th ACM Conference on Economics and Computation (EC)*.
- Mennle, Timo, Seuken, Sven, 2016. The Pareto frontier for random mechanisms [extended abstract]. In: *Proceedings of the 17th ACM Conference on Economics and Computation (EC)*.

- Mennle, Timo, Seuken, Sven, 2017a. An Axiomatic Decomposition of Strategyproofness for Ordinal Mechanisms with Indifferences. Research Note.
- Mennle, Timo, Seuken, Sven, 2017b. Hybrid Mechanisms: Trading Off Strategyproofness and Efficiency of Random Assignment Mechanisms. Working Paper.
- Mennle, Timo, Seuken, Sven, 2017c. Local Sufficiency for Partial Strategyproofness. Research Note.
- Mennle, Timo, Seuken, Sven, 2017d. Trade-offs in School Choice: Comparing Deferred Acceptance, the Naïve and the Adaptive Boston Mechanism. Working Paper.
- Mennle, Timo, Seuken, Sven, 2017e. Two New Impossibility Results for the Random Assignment Problem. Research Note.
- Miralles, Antonio, 2008. School Choice: the Case for the Boston Mechanism. Working Paper.
- Noda, Shunya, 2019. A Planner-Optimal Matching Mechanism and Its Incentive Compatibility in a Restricted Domain. Working Paper.
- Pathak, Parag, Sethuraman, Jay, 2011. Lotteries in student assignment: an equivalence result. *Theor. Econ.* 6 (1), 1–17.
- Pathak, Parag, Sönmez, Tayfun, 2008. Leveling the playing field: sincere and sophisticated players in the Boston mechanism. *Am. Econ. Rev.* 98 (4), 1636–1652.
- Pathak, Parag, Sönmez, Tayfun, 2013. School admissions reform in Chicago and England: comparing mechanisms by their vulnerability to manipulation. *Am. Econ. Rev.* 103 (1), 80–106.
- Roth, Alvin, 1984. The evolution of the labor market for medical interns and residents: a case study in game theory. *J. Polit. Econ.* 92 (6), 991–1016.
- Sandholm, Tuomas, 2003. Automated mechanism design: a new application area for search algorithms. In: *Proceedings of the International Conference on Principles and Practice of Constraint Programming (CP)*.
- Zhou, Lin, 1990. On a conjecture by Gale about one-sided matching problems. *J. Econ. Theory* 52 (1), 123–135.