Partial Strategyproofness: Relaxing Strategy-proofness for the Random Assignment Problem*

Timo Mennle
University of Zurich

Sven Seuken
University of Zurich

September 8, 2018

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 that it provides a unified perspective on incentives in the assignment domain. We then apply it to derive novel insights about the incentive properties of the Probabilistic Serial mechanism and different variants of the Boston mechanism.

Keywords: Mechanism Design, Ordinal Mechanisms, Random Assignment, Matching, Strategyproofness, Stochastic Dominance, Probabilistic Serial, Boston Mechanism

JEL: C78, D47, D82

*First version: January 13, 2014; University of Zurich, Switzerland. Email: {mennle, seuken}@ifi.uzh.ch. For updates see www.ifi.uzh.ch/ce/publications/PSP.pdf. 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, David Parkes, Baharak Rastegari, Alvin Roth, Steffen Schuldenzucker, Arunava Sen, Tayfun Sönmez, William Thomson, and Utku Ünver for helpful comments on this work. Furthermore, we are thankful for the feedback we received from various participants at the following events (in chronological order): 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.

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 with transferable utility. 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), medical school 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, Pathak and Roth, 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.[1] 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. In fact, it is conjectured to be the unique mechanism that satisfies all three properties (Lee and Sethuraman, 2011; Bade, 2016). 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 (Featherstone, 2015), but it is incompatible with strategyproofness, even without additional fairness requirements. Obviously, strategyproofness is in conflict with

[1]Specifically, Zhou (1990) showed that no (possibly random and possibly cardinal) assignment mechanism can satisfy strategyproofness, ex-ante efficiency, and symmetry.

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 desiderata.

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,[2] a rank-efficient mechanism for the assignment of teachers to positions through the Teach for America program (Featherstone, 2015), a mechanism for the assignment of on-campus housing at MIT that minimizes the number of unassigned units,[3] and the HBS Draft mechanism for the assignment of schedules at Harvard Business School (Budish and Cantillon, 2012). It is 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).[4] In this paper, we introduce *partial strategyproofness*, a new, relaxed notion of strategyproofness that is particularly suited to the analysis of assignment mechanisms. We first illustrate the definition of partial strategyproofness with a motivating example, and subsequently, 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

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

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

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

that agent 1 has utility 0 for its last choice c; and that the non-strategyproof Probabilistic

[2]Variants of the Boston mechanism are used in Minneapolis, Seattle, Lee County (Kojima and Ünver, 2014), San Diego, Freiburg (Germany) (Dur, Mennle and Seuken, 2018), and throughout the German state of Nordrhein Westfalen (Basteck, Huesmann and Nax, 2015).

[3]Source: MIT Division of Student Life, retrieved September 8, 2015: https://housing.mit.edu/, and personal communication

[4]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, 2018). We describe their relation to partial strategyproofness in Sections 4.3 and 4.4 respectively.

Serial mechanism (Bogomolnaia and Moulin, 2001) is used to assign the objects. If all agents report their preferences 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 $3/4 \cdot u_1(a) \geqslant u_1(b)$.

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 $r \cdot u_i(a) \geqslant u_i(b)$ holds whenever i prefers a to b (after appropriate normalization). A mechanism is *r-partially strategyproof* 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, agents 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 three main arguments that support this claim:

Partial strategyproofness has a compelling axiomatic motivation. We first prove that full strategyproofness can be decomposed into three simple axioms. Each of the 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 react to the swap at all. In other words, if the mechanism reacts to a swap at all, then it must directly affect the agent's probabilities for the objects that are swapped and the change must be monotonic in the agent's reported preferences.

2. A mechanism is *upper invariant* if no agent can improve its chances for objects that it likes more by misrepresenting its preferences for objects that it likes less. Precisely, the swap of a and b must not affect the agent's chances for any object that it strictly prefers to a.

3. A mechanism is *lower invariant* if no agent can affect its chances for less-preferred objects by swapping a and b. This axiom is the complement to upper invariance but for objects that the agent likes strictly less than b.

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 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 two most important of the three axioms.

Partial strategyproofness tightly parametrizes a broad spectrum of incentive concepts. 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:[5] It is implied by all of them, but at the same time it is the only one under which all pairs of lotteries over objects are comparable. In this sense, the corresponding *LD-strategyproofness* is a minimal notion of strategyproofness for random assignment mechanisms. We prove that LD-strategyproofness is the lower limit concept for r-partial strategyproofness as r approaches 0 (Proposition 2). The degree of strategyproofness thus parametrizes the entire spectrum of incentive concepts between the most demanding full strategyproofness and the minimal LD-strategyproofness.

Regarding tightness, observe that r-partial strategyproofness is formally equivalent to strategyproofness on the restricted domain of utility functions that satisfy URBI(r). We prove that this domain restriction is *maximal* for r-partially strategyproof mechanisms: Specifically, for any utility function that violates URBI(r), we can construct an r-partially

[5] *Trivial*, *deterministic*, *sure thing*, and *first-order stochastic* dominance (Aziz, Brandt and Brill, 2013).

strategyproof mechanism that is manipulable for an agent with this utility function (Proposition 1). Therefore, no other single-parameter measure conveys strictly more information about a mechanism's incentive properties than the parameter r (which we call the *degree of strategyproofness*).

Partial strategyproofness provides a unified perspective on incentives. Various relaxed notions of strategyproofness have been employed in prior work to describe the incentive properties of non-strategyproof assignment mechanisms. Partial strategyproofness provides a unified perspective on these because it implies all of them in meaningful ways (Theorem 3): r-partial strategyproofness for some $r > 0$ implies *weak SD-strategyproofness* (Bogomolnaia and Moulin, 2001) and *convex strategyproofness* (Balbuzanov, 2015); a positive degree of strategyproofness $r > 0$ implies *ε-approximate strategyproofness* (Carroll, 2013) (where ε is guaranteed to be close to 0 for r sufficiently close to 1); if the degree of strategyproofness of a mechanism converges to 1 in large markets, then this mechanism is *strategyproof in the large* (Azevedo and Budish, 2018); and the converse holds for neither of these implications. The relation to approximate strategyproofness is especially important because it allows us to further refine the strategic advice we can give: By definition of r-partial strategyproofness, agents are best off reporting their preferences truthfully if their preference intensities for any two objects differ sufficiently (by the factor r); otherwise, if they are close to being indifferent between some objects, then we now know that their potential gain from misreporting may be positive but it is bounded (by the corresponding value of ε).

We conclude that the axiomatic motivation, the ability to parametrize a broad spectrum of incentive properties, and the unified perspective on incentives make the partial strategyproofness concept an important addition to the mechanism designer's toolbox. To illustrate its usefulness in mechanism design, we provide two applications: First, we show that it yields the most demanding description of the incentive properties of the Probabilistic Serial mechanism in finite settings (refining (Bogomolnaia and Moulin, 2001; Balbuzanov, 2015)) as well as in the limit as markets get large (refining (Kojima and Manea, 2010; Azevedo and Budish, 2018)). Second, we show that the concept can be used to distinguish between two common variants of the Boston mechanism in terms of their incentive properties: Under the classical *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, 2017; Dur, 2015), 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, Mennle and Seuken, 2018). Using our new partial strategyproofness concept, we can now show formally that ABM indeed has better incentive properties than BM: When priorities are random, ABM is partially strategyproof while BM is not. These examples demonstrate that partial strategyproofness generates new insights in the analysis of non-strategyproof assignment mechanisms.

2. Model

A *setting* (N, M, q) consists of a set of *agents* N with $n = |N|$, a set of *objects* M with $m = |M|$, and *capacities* $q = (q_1, \ldots, q_m)$ (i.e., q_j units of object j are available). We assume $n \leqslant \sum_{j \in M} q_j$ (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 $P_{-i} \in \mathcal{P}^{N \backslash \{i\}}$ is a collection of preference orders of all agents except i. We extend agents' preferences over objects to preferences over lotteries via von Neumann–Morgenstern utility functions: Every agent i is endowed with a *utility function* $u_i : M \to \mathbb{R}^+$ that 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., $\sum_{i \in N} x_{i,j} \leqslant q_j$ 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 $x_{i,j} \geqslant 0$ 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 $f : \mathcal{P}^N \to \Delta(X)$ 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., $f : \mathcal{P}^N \to X$). 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 $\mathbb{E}_{f_i(P_i,P_{-i})}[u_i] = \sum_{j\in M} u_i(j) \cdot f_{i,j}(P_i, P_{-i})$.

3. A New Decomposition of Strategyproofness

In this section, we formally define strategyproofness, we introduce three axioms, and we show that strategyproofness decomposes into these axioms.

The most common and most demanding 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 $\mathbb{E}_{f_i(P_i,P_{-i})}[u_i] \geqslant \mathbb{E}_{f_i(P_i',P_{-i})}[u_i]$.

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

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

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*.

Next, we introduce three axioms that are related to incentives. Their formal definitions require 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 $P_i : a \succ b \succ c$ contains $P_i' : b \succ a \succ c$ but does not contain $P_i'' : c \succ a \succ b$). 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 $P_i : a \succ b \succ c \succ d$ is $U(b, P_i) = \{a\}$ and the lower contour set is $L(b, P_i) = \{c, d\}$.

Swapping two consecutively ranked objects in the true preference order (or equivalently, reporting a preference order from the neighborhood of the true preference order) is a basic misreport. The axioms we define limit the ways in which a mechanism can change an agent's assignment when this agent uses such a basic kind of 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 $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a \succ b$ but $P_i' : b \succ a$, 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 strictly when b is reportedly more preferred.

For deterministic mechanisms, swap monotonicity is equivalent to strategyproofness (Proposition 5 in Appendix E). 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 $x_i = (0, 1/2, 0, 1/2)$ for a, b, c, d respectively, and reporting a preference for c over b leads to $y_i = (1/2, 0, 1/2, 0)$. 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. In addition, Deferred Acceptance is lower invariant for agents on both sides (but it is only strategyproof for the proposing side).

Each of the three axioms affects incentives by preventing misreports from being beneficial in particular ways: Swap monotonicity forces mechanisms to change the probabilities 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 (⇒). Assume towards contradiction that a mechanism f is strategyproof but not upper invariant. Then there exist 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 (⇐). 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. □

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

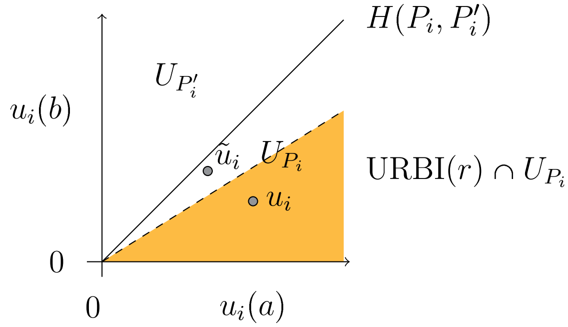

Figure 1: Illustration of uniformly relatively bounded indifference.

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 useful when verifying strategyproofness of mechanisms (e.g., see (Fack, Grenet and He, 2017)) or when encoding strategyproofness as constraints under the automated mechanism design paradigm (Sandholm, 2003).

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, 2017*a*).

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 degree of indifference between the agent's first and second choices, which determined whether the misreport was beneficial or not. The following domain restriction generalizes this intuition to agents who differentiate sufficiently between any two objects.

Definition 2. 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 $u_i(a) > u_i(b)$,

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

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

$$r \cdot u_i(a) \geqslant u_i(b). \tag{7}$$

In words, it requires that agent i values b at least a factor r less than a. Consider Figure 1 for a 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 3. 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 $\mathbb{E}_{f_i(P_i, P_{-i})}[u_i] \geqslant \mathbb{E}_{f_i(P_i', P_{-i})}[u_i]$.

In words, an r-partially strategyproof mechanisms makes truthful reporting a dominant strategy at least for those agents whose utility functions satisfy URBI(r). Since all utility functions satisfy URBI(1), 1-partial strategyproofness is equivalent to full strategyproofness, and since the set URBI(r) is smaller for smaller 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. We use the following lemma.

Lemma 1. *Given a setting (N, M, q) and a mechanism f, the following are equivalent:*

A. *For all agents $i \in N$, all preference profiles $(P_i, P_{-i}) \in \mathcal{P}^N$, and all misreports $P_i' \in \mathcal{P}$ with $f_i(P_i, P_{-i}) \neq f_i(P_i', P_{-i})$, there exists an object $a \in M$ such that $f_{i,a}(P_i, P_{-i}) > f_{i,a}(P_i', P_{-i})$ and $f_{i,j}(P_i, P_{-i}) = f_{i,j}(P_i', P_{-i})$ for all $j \in U(a, P_i)$,*

B. *f is swap monotonic and upper invariant.*

The proof of Lemma 1 is given in Appendix A. The lemma establishes a connection between the behavior of mechanisms under arbitrary misreports (Statement A) and local properties (swap monotonicity and upper invariance in Statement B). In this sense, it replaces the local sufficiency result of Carroll (2012) we used to prove Theorem 1.

To show Theorem 2, we prove equivalence between partial strategyproofness and Statement A of Lemma 1. For the fixed setting (N, M, q), let δ be the smallest non-zero variation in the assignment resulting from any change of report by any agent; formally,

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

Observe that δ is well-defined and strictly positive for any non-constant mechanism.

Necessity (partial strategyproofness $\Leftarrow$ A). 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 Statement A 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 δ. 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 δ 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

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

which is weakly negative if

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

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 δ-partially strategyproof.

Sufficiency (partial strategyproofness ⇒ A). Let f be r-partially strategyproof for some $r > 0$, and assume towards contradiction that f violates Statement A. This is equivalent to saying that there exist 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 δ. 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 δ and its last choice with probability $(1 - \delta)$ when misreporting. This makes

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

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

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

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. □

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 may violate lower invariance.

Remark 2. Two main arguments suggest dropping lower invariance and keeping swap monotonicity and upper invariance as the most 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 E). 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*.[6] In contrast, no mechanism can be upper invariant, lower invariant, ordinally efficient, and symmetric (Theorem 1 of Mennle and Seuken (2017*d*)); and no mechanism can be swap monotonic, lower invariant, ordinally efficient, anonymous, neutral, and non-bossy (Theorem 2 of Mennle and Seuken (2017*d*)). This means that these popular combinations of mechanism design axioms become unattainable when dropping swap monotonicity or upper invariance (instead of lower invariance).

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 subject for future research.

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 (Section 5.2) with priorities determined by a single uniform lottery 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

[6]See (Mennle and Seuken, 2017*d*) for formal definitions of these axioms.

URBI (1/3), since

$$\frac{u_i(c) - \min_{j \in M} u(j)}{u_i(b) - \min_{j \in M} u(j)} = 1/2 > 1/3. \tag{13}$$

Thus, the set of utility functions for which the adaptive Boston mechanism makes truthful reporting a dominant strategy is strictly larger than the set URBI(1/3).

Example 2 shows that, for some specific 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)**.** *For all settings (N, M, q) with $m \geqslant 3$ objects, all bounds $r \in (0, 1)$, all preference orders $\tilde{P}_i \in \mathcal{P}$, and all utility functions $\tilde{u}_i \in U_{\tilde{P}_i}$ that violate URBI(r), there exists a mechanism $\tilde{f}$ such that*

1. *$\tilde{f}$ is r-partially strategyproof and anonymous,*
2. *$\mathbb{E}_{\tilde{f}_i(\tilde{P}_i, P_{-i})}[\tilde{u}_i] < \mathbb{E}_{\tilde{f}_i(P'_i, P_{-i})}[\tilde{u}_i]$ for some misreport $P'_i \in \mathcal{P}$ and all $P_{-i} \in \mathcal{P}^{N \backslash \{i\}}$.*

Proof. We construct $\tilde{f}$ explicitly. If $\tilde{u}_i$ violates URBI(r), then

$$\frac{\tilde{u}_i(b) - \min_{j \in M} \tilde{u}_i(j)}{\tilde{u}_i(a) - \min_{j \in M} \tilde{u}_i(j)} = \tilde{r} \tag{14}$$

for some $a, b \in M$ and $\tilde{r} > r$. Observe that b cannot be i's last choice because $0 / (\tilde{u}_i(a) - \min_{j \in M} \tilde{u}_i(j)) \leqslant r$ is trivially satisfied. Define the mechanism $\tilde{f}$ by 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

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

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 $\tilde{f}$ is

r-partially strategyproof but manipulable for agent i with utility function $\tilde{u}_i$ (Lemma 2 in Appendix B). To construct an anonymous mechanism with the same properties, we randomly assign each agent to the role of the distinguished agent i. □

In words, 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 the following "measure" for incentive properties.

Definition 4. Given a setting (N, M, q) and a mechanism f, we define the *degree of strategyproofness of* f (*in the setting* (N, M, q)) as

$$\rho_{(N,M,q)}(f) = \max\{r \in [0,1] \mid f \text{ is } r\text{-partially strategyproof in } (N, M, q)\}.^{7} \qquad (15)$$

Observe that, for $0 \leqslant r < r' \leqslant 1$ we have URBI(r) $\subset$ URBI(r') by construction. Thus, a higher degree of strategyproofness corresponds to a stronger 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, then no single-parameter measure conveys strictly more information about the incentive properties of f.

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. In Section 5, we apply this comparison to differentiate three assignment mechanisms, namely the Probabilistic Serial mechanism and two variants of the Boston mechanism.[8] Furthermore,

[7]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.

[8]Pathak and Sönmez (2013) proposed a concept to compare mechanisms by their *vulnerability to manipulation*. As Proposition 6 in Appendix F shows, this comparison is consistent with (but not equivalent to) the comparison of mechanisms by their degrees of strategyproofness.

in (Mennle and Seuken, 2017*b*), we used it to quantify trade-offs between strategyproofness and efficiency that are achievable by hybrid mechanisms (i.e., convex combinations of a strategyproof mechanism and a manipulable but more efficient mechanism).

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 (Cho, 2018).

Definition 5. For a preference order $P_i \in \mathcal{P}$ and assignment vectors x_i, y_i, 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 $P_i' \in \mathcal{P}$, i's assignment $f_i(P_i, P_{-i})$ lexicographically dominates $f_i(P_i', P_{-i})$ at P_i.

In words, an agent with lexicographic preferences prefers any (arbitrarily small) increase in the probability for any more-preferred object to any (arbitrarily large) increase in the probability for any less-preferred object; and LD-strategyproofness requires that such agents do not wish to misreport their preferences. Aziz, Brandt and Brill (2013) formalized a hierarchy of dominance notions consisting of *trivial*, *deterministic*, *sure thing*, and *stochastic* dominance. It is easy to see that lexicographic dominance is implied by all of them. However, in contrast to these other notions, all assignment vectors are comparable under lexicographic dominance. In this sense, LD-strategyproofness represents a *minimal* notion of strategyproofness.

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* $LD\text{-}SP(N, M, q)$ *be the sets of mechanisms that are strategyproof,* r*-partially strategyproof, and LD-strategyproof in the setting* (N, M, q)*, respectively. Then*

$$SP(N, M, q) = \bigcap_{r<1} r\text{-}PSP(N, M, q), \tag{16}$$

$$LD\text{-}SP(N, M, q) = \bigcup_{r>0} r\text{-}PSP(N, M, q). \tag{17}$$

Proof. Towards equality (16), 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 (17) follows from Lemma 1: LD-SP(N, M, q) is the set of mechanisms that satisfy Statement A (by definition), and $\bigcup_{r>0} r$-PSP(N, M, q) is the set of swap monotonic and upper invariant mechanisms (by Theorem 2). □

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 of incentive concepts between the most demanding full strategyproofness and the minimal LD-strategyproofness.

An important consequence of Proposition 2 is that, for a fixed setting, any LD-strategyproof mechanism is already r-partially strategyproof for some $r > 0$. On a technical level, this means that it suffices to verify LD-strategyproofness in order to show partial strategyproofness. We exploit this in Proposition 3 to show that the Probabilistic Serial mechanism is partially strategyproof. On the other hand, Proposition 2 exposes LD-strategyproofness as being unnecessarily weak for the accurate description of the incentive properties of non-strategyproof assignment mechanisms. The concept ignores the parametric nature of the set of utility functions for which truthful reporting is a dominant strategy. In contrast, partial strategyproofness provides quantitative guarantees for the mechanism's incentive properties, and these guarantees are maximal in the sense of Proposition 1.

4.4. A Unified Perspective on Incentives

In this section, we demonstrate that partial strategyproofness provides a unified perspective on the incentive properties of non-strategyproof assignment mechanisms. Proposition 2 already shows that it implies LD-strategyproofness. Expanding on this insight, we consider other relaxed incentive requirements that have previously been discussed in the context of the assignment problem, namely weak SD-strategyproofness (Bogomolnaia and Moulin, 2001), convex strategyproofness (Balbuzanov, 2015), approximate strategyproofness (Carroll, 2013), and strategyproofness in the large (Azevedo and Budish, 2018). We formalize these requirements and we show in what sense they are implied by partial strategyproofness.

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 $\mathbb{E}_{f_i(P_i,P_{-i})}[u_i] \geqslant \mathbb{E}_{f_i(P_i',P_{-i})}[u_i]$. 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, 2015) 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 $P_{-i} \in \mathcal{P}^{N\backslash\{i\}}$ and all misreports $P_i' \in \mathcal{P}$, we have $\mathbb{E}_{f_i(P_i,P_{-i})}[u_i] \geqslant \mathbb{E}_{f_i(P_i',P_{-i})}[u_i]$.

Balbuzanov (2015) 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 *ε-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 $u_i \in U_{P_i}$, we have $\mathbb{E}_{f_i(P_i,P_{-i})}[u_i] \geqslant \mathbb{E}_{f_i(P_i',P_{-i})}[u_i] - \varepsilon$.

Finally, Azevedo and Budish (2018) recently proposed *strategyproofness in the large*, a concept which formalizes the intuition that agents who view themselves as price takers 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 $(N^n, M^n, q^n)_{n \geqslant 1}$ of settings with a constant set of objects ($M^n = M$), a growing agent population ($|N^n| = n$), growing capacities ($\min_{j \in M} q_j^n \to \infty$ as $n \to \infty$) that satisfy overall demand ($\sum_{j \in M} q_j^n \geqslant n$).

Definition 9. For a finite set of utility functions $\{u^1, \ldots, u^K\}$ and a sequence of settings $(N^n, M^n, q^n)_{n \geqslant 1}$ 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 $n \geqslant n_0$, no agent with a utility function from $\{u^1, \ldots, u^K\}$ can gain more than ε by misreporting.[9]

The following proposition summarizes in what sense partial strategyproofness implies all of the relaxed incentive concepts defined above.

Theorem 3 (Unification)**.** *Given a setting (N, M, q), the following hold:*

1. *r-partial strategyproofness for some $r > 0$ implies weak SD-strategyproofness and convex strategyproofness. The converse may not hold.*
2. *For all $\varepsilon > 0$ there exists $r < 1$ such that r-partial strategyproofness implies ε-approximate strategyproofness. The converse may not hold.*
3. *Fix a finite set of utility functions $\{u^1, \ldots, u^K\}$ and a sequence $(N^n, M^n, q^n)_{n \geqslant 1}$ with $|N^n| = n$, $M^n = M$, $\sum_{j \in M} q_j^n \geqslant n$, and $\min_{j \in M} q_j^n \to \infty$ as $n \to \infty$. If the degree of strategyproofness of a mechanism f converges to 1 as $n \to \infty$, then f is strategyproof in the large. The converse may not hold.*

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

[9]The original definition of strategyproofness in the larger (*SP-L*) 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, and Statement 3 in our Theorem 3 shows in what sense partial strategyproofness implies this more demanding version. Note that this strengthens our statement.

Statement 2. Let f be r-partially strategyproof for some $r > 0$. By Proposition 7 in Appendix G, this 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 $P_i : j_1 \succ \ldots \succ j_m$), $P_i' \in \mathcal{P}$, and $K \in \{1, \ldots, m\}$, we have

$$\sum_{k=1}^{K} r^k \cdot f_{i,j_k}(P_i, P_{-i}) \geqslant \sum_{k=1}^{K} r^k \cdot f_{i,j_k}(P_i', P_{-i}). \tag{18}$$

With $r > 0$, we obtain

$$\begin{aligned} \sum_{k=1}^{K} f_{i,j_k}(P_i', P_{-i}) - f_{i,j_k}(P_i, P_{-i}) &\leqslant \sum_{k=1}^{K} \left(1 - r^k\right) \cdot \left(f_{i,j_k}(P_i', P_{-i}) - f_{i,j_k}(P_i, P_{-i})\right) \\ &\leqslant \sum_{k=1}^{K} \left(1 - r^k\right) \leqslant \sum_{k=1}^{m} \left(1 - r^k\right). \end{aligned} \tag{19}$$

By choosing r close to 1, the last term in (19) can be made arbitrarily small. This implies a property that Liu and Pycia (2016) called ε-*strategy-proofness*, which is equivalent to ε-approximate strategyproofness (by Theorem 1 of Mennle and Seuken (2016)). To see why the converse does not hold, refer to our proof for 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 chose $\bar{r} < 1$ such that $u^k \in$ URBI($\bar{r}$) for all $u^k \in \{u^1, \ldots, u^K\}$. By convergence of the degree of strategyproofness, f is $\bar{r}$-partially strategyproof in all settings with sufficiently many agents, so that no agent can benefit from misreporting. 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 manipulable in an SD sense. □

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 already implied. Thus, proving either of the two weaker concepts does not yield additional insights about the mechanism's incentive properties.

Second, by definition, r-partially strategyproof mechanisms make truthful reporting a dominant strategy for all agents whose utility functions satisfy URBI(r). While it is possible that other agents benefit from misreporting, Statement 2 shows that their incentive to do so is limited. This allows the formulation of straightforward and honest strategic advice that can be given to agents who participate in partially strategyproof mechanisms: *They are best off reporting their preferences truthfully as long as their preference intensities for any two objects are sufficiently different; otherwise, if they are close to being indifferent between some objects, then their potential gain from misreporting may be positive but it is limited in the sense of approximate strategyproofness.* And with some additional analytical effort, numerical values for the parameters r and ε can be determined.

Third, convergence of the degree of strategyproofness to 1 is a straightforward formalization of the intuition that "incentives improve in large markets." It is reassuring to note that, by Statement 3, this formalization is consistent with strategyproofness in the large, a concept that Azevedo and Budish (2018) developed for this specific purpose.

5. Applications

In this section, we apply our new partial strategyproofness concept in two ways to derive new insights about important assignment mechanisms: 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 settings 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 largely elusive.

5.1. New Insights about the Probabilistic Serial Mechanism

The *Probabilistic Serial (PS)* (Bogomolnaia and Moulin, 2001) mechanism is one of the most well-studied mechanism 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 Proposition 2. □

Regarding the incentive properties of PS, Bogomolnaia and Moulin (2001) already showed that it is weakly SD-strategyproof (but not strategyproof), and Balbuzanov (2015) strengthened their result by showing that it is convex strategyproof. Since partial strategyproofness is strictly stronger than both properties, Proposition 3 establishes the most demanding description of its incentive properties in finite settings known to date.

PS has a special connection to the URBI(r) domain restriction: 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 (and only if) $3/4 \cdot u_1(a) < u_1(b)$. At the same time, the degree of strategyproofness of PS in this setting is also 3/4 (see Figure 2). Thus, the set of utility functions for which PS makes truthful reporting a dominant strategy is *exactly* URBI(3/4). A similar observation holds in a setting with 4 agents and 4 objects with unit capacity, where PS makes truthful reporting a dominant strategy for *exactly* those utility functions that satisfy URBI(1/2). In these settings, r-partial strategyproofness perfectly describes the incentive properties of PS.[10]

We have computed the degree of strategyproofness of PS in various settings, and Figure 2 shows the results.[11] Observe that these values increase as the settings get larger. Kojima and Manea (2010) already proved that, for an arbitrary, fixed utility function, PS makes truthful reporting a dominant strategy for agents with that utility function if there are sufficiently many copies of each object. Their result and our numerical

[10]In Appendix H, we discuss these observations in more detail as well as the curious fact that an analogous statement fails to holds for 5 agents and 5 objects with unit capacity.

[11]Note that the values in Figure 2 are exact calculations (rather than estimates or simulation results). These calculations are possible because of an equivalent formulation of partial strategyproofness in terms of a *discounted* notion of stochastic dominance (see Appendix G). The limited setting size for our calculations is owed to the fact that the algorithm must iterate over (a substantial subset of) all possible preference profiles (e.g., $(4!)^{16} > 10^{22}$ for $n = 16$ and $m = 4$).

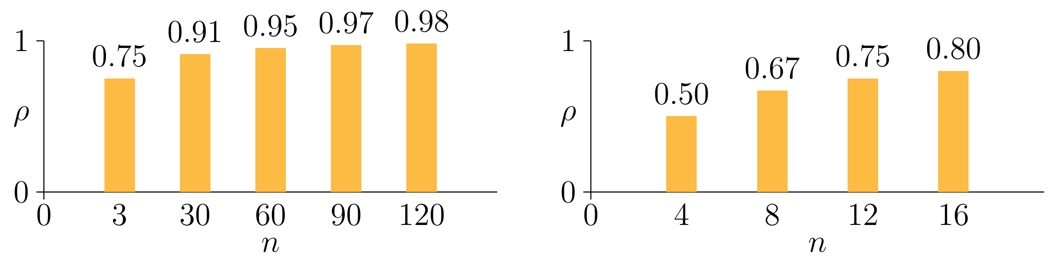

Figure 2: Plot of $\rho_{(N,M,q)}(\text{PS})$ for $m = 3$ (left) and $m = 4$ (right) objects, for varying numbers of agents n, and evenly distributed capacities $q_j = n/m$.

findings suggest the hypothesis that the degree of strategyproofness of PS converges to 1 as the settings get large, and Abächerli (2017) proved this hypothesis to be true.[12] In combination with Statement 3 of Theorem 3, this yields an elegant, parametric proof of the following corollary.

Corollary 1. *PS is strategyproof in the large.*

The partial strategyproofness concept thus also enables a refined understanding of the incentive properties of PS in large markets, providing additional structural information about the utility functions for which PS makes truthful reporting a dominant strategy.

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. In addition to the students' preferences, these mechanisms take into account priorities (Abdulkadiroğlu et al., 2006). While the priorities can depend on grades, siblings, or walk zones, they are almost always "coarse" in the sense that a substantial number of students fall into the same priority classes. Ties between such students must be broken randomly. In these situations, the school choice mechanisms can be modeled as random assignment mechanisms.

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. If a school receives more applications than it has seats, then it rejects the applications of students with the lowest priorities.

[12]Abächerli proved this as part of a Master's thesis supervised by the authors of the present paper.

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, Mennle and Seuken, 2018).[13] Under ABM, students apply to their *best available choice* in every round. Skipping exhausted schools therefore becomes ineffective.

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 of random priorities, the comparison is inconclusive (Dur, Mennle and Seuken, 2018). Interestingly, we can use partial strategyproofness to recover a meaningful distinction between BM and ABM by exploiting randomness in the priorities.

To state this result formally, we require additional notation: A *priority order* π_j is a strict order over agents, where $\pi_j : i \succ i'$ indicates that agent i has priority over agent i' at object j. We denote by Π 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 π 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 Π^M, and $\mathbb{P}$ *supports all single priority profiles* if $\mathbb{P}[\pi] > 0$ for all single priority profiles π. A *school choice mechanism* φ is a mapping $\varphi : \mathcal{P}^N \times \Pi^M \to X$ 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 $\varphi^{\mathbb{P}}(P) = \sum_{\pi \in \Pi} \mathbb{P}[\pi] \cdot \varphi(P, \pi)$. Then $f = \varphi^{\mathbb{P}}$ is a random

[13] See (Alcalde, 1996; Abdulkadiroğlu and Sönmez, 2003; Miralles, 2008; Harless, 2017; Dur, 2015) for prior work that studied ABM.

assignment mechanism, and it captures the strategic situation of agents who believe that priorities are drawn according to $\mathbb{P}$.

With this, we can formalize the distinction between BM and ABM in terms of their incentive properties by studying the mechanisms $\mathrm{BM}^{\mathbb{P}}$ and $\mathrm{ABM}^{\mathbb{P}}$.

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. However, $BM^{\mathbb{P}}$ violates swap monotonicity if $m \geqslant 4$, $n \geqslant 4$, and $\sum_{j \in M} q_j = n$.*

Proof. For all priority profiles $\pi \in \Pi^M$, upper invariance of $\mathrm{BM}(\cdot, \pi)$ and $\mathrm{ABM}(\cdot, \pi)$ is obvious. Since $\mathrm{BM}^{\mathbb{P}}$ and $\mathrm{ABM}^{\mathbb{P}}$ are simply convex combinations of these mechanisms for different priority profiles, they inherit this property. Next, observe that $\mathrm{ABM}(\cdot, \pi)$ is *monotonic* (i.e., for all $i \in N$, all $(P_i, P_{-i}) \in \mathcal{P}^N$, all $P_i' \in N_{P_i}$ with $P_i : a > b$ but $P_i : b > a$, we have $\mathrm{ABM}_{i,b}((P_i, P_{-i}), \pi) \leqslant \mathrm{ABM}_{i,b}((P_i', P_{-i}), \pi))$, and this property is again inherited by $\mathrm{ABM}^{\mathbb{P}}$. Lemma 3 in Appendix D shows that $\mathrm{ABM}_i^{\mathbb{P}}(P_i, P_{-i}) \neq \mathrm{ABM}_i^{\mathbb{P}}(P_i', P_{-i})$ implies existence of a single priority profile π with $\mathrm{ABM}_{i,b}((P_i, P_{-i}), \pi) < \mathrm{ABM}_{i,b}((P_i', P_{-i}), \pi)$. With this, monotonicity of $\mathrm{ABM}^{\mathbb{P}}$, and the fact that $\mathbb{P}$ supports all single priority profiles, we obtain swap monotonicity of $\mathrm{ABM}^{\mathbb{P}}$. Example 4 in Appendix D shows that $\mathrm{BM}^{\mathbb{P}}$ violates swap monotonicity in the respective settings. □

Proposition 4 provides a formal justification for the intuition that "ABM has better incentive properties than BM": If priorities are *minimally random* in the sense that all single priority profiles are possible, then $\mathrm{ABM}^{\mathbb{P}}$ is r-partially strategyproof for some positive $r > 0$ while $\mathrm{BM}^{\mathbb{P}}$ has a degree of strategyproofness of 0. This applies directly to various school choice markets where priorities are determined exclusively by a single uniform lottery (Lai, Sadoulet and de Janvry, 2009; Pathak and Sethuraman, 2011; Lauri, Põder and Veski, 2014), and it applies at least approximately in markets with coarse priorities and random tie-breaking (Erdil and Ergin, 2008; Dur et al., 2018).

For the design of school choice mechanisms, Proposition 4 yields an important insight: BM, ABM, and the Deferred Acceptance (*DA*) mechanism form a hierarchy in terms of their incentive properties. Interestingly, these three mechanisms also form a second hierarchy in terms of student welfare, which, however, points in the opposite direction.[14]

[14]Featherstone (2015) formalized *rank dominance* as an intuitive method to compare assignments in terms of societal welfare. In (Mennle and Seuken, 2017*c*), we used rank dominance to show that BM, ABM, and DA form a hierarchy in terms of student welfare with respect to the reported preferences.

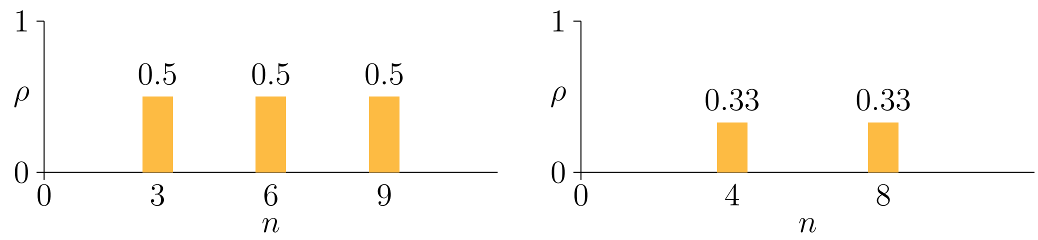

Figure 3: Plot of $\rho_{(N,M,q)}(\text{ABM}^{\mathbb{U}})$ for $m = 3$ (left) and $m = 4$ (right) objects, for varying numbers of agents n, and for $q_j = n/m$.

We thus learn that a choice between BM, ABM, and DA involves an implicit trade-off decision between incentives for truthtelling and student welfare.

For the *single uniform priority distribution* $\mathbb{U}$ (i.e., the distribution that selects all single priority profiles with equal probability), we have calculated the degrees of strategyproofness of $\text{ABM}^{\mathbb{U}}$ in various settings.[15] The results are shown in Figure 3. Observe that $\rho_{(N,M,q)}(\text{ABM}^{\mathbb{U}})$ is significantly lower than $\rho_{(N,M,q)}(\text{PS})$ (Figure 2). Furthermore, it does not increase as the settings grow, but it appears to remain constant (i.e., $\rho_{(N,M,q)}(\text{ABM}^{\mathbb{U}}) = 1/2$ for $m = 3$ objects and $n = 3, 6, 9$ agents; and $\rho_{(N,M,q)}(\text{ABM}^{\mathbb{U}}) = 1/3$ for $m = 4$ objects and $n = 4, 8$ agents). At least for the small settings for which we are able to compute the degree of strategyproofness of ABM, these numbers support our understanding that "PS has better incentive properties than ABM."

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

[15] 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 computed the degrees of strategyproofness of $\text{ABM}^{\mathbb{U}}$ only in settings with up to 9 agents and up to 4 objects.

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 URBI(r) domain restriction separates agents by this criterion. By requiring good incentives only on the restricted domain, partial strategyproofness focuses on those agents that are "conquerable" but leaves the more "difficult" agents aside. For any given mechanism, its degree of strategyproofness thus quantifies how many difficult agents must be left aside to establish truthtelling as a dominant strategy for all others.

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 for any two objects are sufficiently different; otherwise, if they are close to being indifferent between some objects, then their potential gain from misreporting is at least 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. In (Mennle and Seuken, 2017*b*), 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.

References

Abächerli, Daniel. 2017. "A Parametric Proof for Strategyproofness in the Large of the Probabilistic Serial Mechanism." Mimeo.

Abdulkadiroğlu, Atila, Alvin Roth, Parag Pathak, and Tayfun Sönmez. 2006. "Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access." Working Paper.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 1998. "Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems." *Econometrica*, 66(3): 689–702.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School Choice: A Mechanism Design Approach." *American Economic Review*, 93(3): 729–747.

Abdulkadiroğlu, Atila, Parag Pathak, and Alvin Roth. 2009. "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *American Economic Review*, 99(5): 1954–78.

Alcalde, Jose. 1996. "Implementation of Stable Solutions to Marriage Problems." *Journal of Economic Theory*, 69(1): 240–254.

Azevedo, Eduardo, and Eric Budish. 2018. "Strategy-proofness in the Large." Forthcoming in Review of Economic Studies.

Aziz, Haris, Felix Brandt, and Markus Brill. 2013. "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).*

Bade, Sophie. 2016. "Random Serial Dictatorship: The One and Only." Working Paper.

Balbuzanov, Ivan. 2015. "Convex Strategyproofness with an Application to the Probabilistic Serial Mechanism." *Social Choice and Welfare*, 1–10.

Basteck, Christian, Katharina Huesmann, and Heinrich Nax. 2015. "Matching Practices for Secondary Schools - Germany." matching-in-practice.eu, accessed May 16, 2015.

Birrell, Eleanor, and Rafael Pass. 2011. "Approximately Strategy-Proof Voting." In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI).*

Bogomolnaia, Anna, and Hervé Moulin. 2001. "A New Solution to the Random Assignment Problem." *Journal of Economic Theory*, 100(2): 295–328.

Budish, Eric. 2012. "Matching "versus" Mechanism Design." *ACM SIGecom Exchanges*, 11(2): 4–15.

Budish, Eric, and Estelle Cantillon. 2012. "The Multi-Unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard." *American Economic Review*, 102(5): 2237–2271.

Budish, Eric, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom. 2013. "Designing Random Allocation Mechanisms: Theory and Applications." *American Economic Review*, 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." *Social Choice and Welfare*, , (51).

Dur, Umut. 2015. "The Modified Boston Mechanism." Working Paper.

Dur, Umut, Scott Kominers, Parag Pathak, and Tayfun Sönmez. 2018. "Reserve Design: Unintended Consequences and The Demise of Boston's Walk Zones." Forthcoming in Journal of Political Economy.

Dur, Umut, Timo Mennle, and Sven Seuken. 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." *Journal of Economic Theory*, 151: 146–162.

Erdil, Aytek, and Haluk Ergin. 2008. "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice." *American Economic Review*, 98(3): 669–89.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2017. "Beyond Truth-Telling: Preference Estimation with Centralized School Choice." Working Paper.

Featherstone, Clayton. 2015. "Rank Efficiency : Investigating a Widespread Ordinal Welfare Criterion." Working Paper.

Harless, Patrick. 2017. "Immediate Acceptance with or without Skips? Comparing School Assignment Procedures." Working Paper.

Hashimoto, Tadashi, Daisuke Hirata, Onur Kesten, Morimitsu Kurino, and Utku Ünver. 2014. "Two Axiomatic Approaches to the Probabilistic Serial Mechanism." *Theoretical Economics*, 9(1): 253–277.

Kojima, Fuhito, and Mihai Manea. 2010. "Incentives in the Probabilistic Serial Mechanism." *Journal of Economic Theory*, 145(1): 106–123.

Kojima, Fuhito, and Utku Ünver. 2014. "The Boston School-Choice Mechanism: an Axiomatic Approach." *Economic Theory*, 55(3): 515–544.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry. 2009. "The Adverse Effects of Parents' School Selection Errors on Academic Achievement: Evidence from the Beijing Open Enrollment Program." *Economics of Education Review*, 28(4): 485–496.

Lauri, Triin, Kaire Põder, and Andre Veski. 2014. "Matching Practices for Elementary Schools - Estonia." accessed May 16, 2015.

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).*

Lee, Thiam, and Jay Sethuraman. 2011. "Equivalence Results in the Allocation of Indivisible Objects: A Unified View." Working Paper.

Liu, Qingmin, and Marek Pycia. 2016. "Ordinal Efficiency, Fairness, and Incentives in Large Markets." Working Paper.

Mennle, Timo, and Sven Seuken. 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, and Sven Seuken. 2016. "The Pareto Frontier for Random Mechanisms [Extended Abstract]." In *Proceedings of the 17th ACM Conference on Economics and Computation (EC).*

Mennle, Timo, and Sven Seuken. 2017*a*. "An Axiomatic Decomposition of Strategyproofness for Ordinal Mechanism with Indifferences." Research Note.

Mennle, Timo, and Sven Seuken. 2017*b*. "Hybrid Mechanisms: Trading Off Strategyproofness and Efficiency of Random Assignment Mechanisms." Working Paper.

Mennle, Timo, and Sven Seuken. 2017*c*. "Trade-offs in School Choice: Comparing Deferred Acceptance, the Naïve and the Adaptive Boston Mechanism." Working Paper.

Mennle, Timo, and Sven Seuken. 2017*d*. "Two New Impossibility Results for the Random Assignment Problem." Research Note.

Miralles, Antonio. 2008. "School Choice: the Case for the Boston Mechanism." Working Paper.

Pathak, Parag, and Jay Sethuraman. 2011. "Lotteries in Student Assignment: An Equivalence Result." *Theoretical Economics*, 6(1): 1–17.

Pathak, Parag, and Tayfun Sönmez. 2008. "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review*, 98(4): 1636–52.

Pathak, Parag, and Tayfun Sönmez. 2013. "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation." *American Economic Review*, 103(1): 80–106.

Roth, Alvin. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 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." *Journal of Economic Theory*, 52(1): 123–135.

Appendix

A. Proof of Lemma 1 for Theorem 2

Proof. Sufficiency ($A \Rightarrow B$). First, assume that f satisfies Statement A but violates upper invariance. Then there exist 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 $P'_i : b \succ a$, such that i's assignment for some $j \in U(a, P_i)$ changes. Let $\tilde{j}$ be the highest-ranking such object, then Statement A implies $f_{i,\tilde{j}}(P_i, P_{-i}) > f_{i,\tilde{j}}(P'_i, P_{-i})$. Moreover, Statement A imposes restrictions in the opposite direction when i changes its report from P'_i to P_i. Since $\tilde{j}$ remains the highest-ranking object for which the assignment changes, we get $f_{i,\tilde{j}}(P'_i, P_{-i}) > f_{i,\tilde{j}}(P_i, P_{-i})$, a contradiction.

Second, assume towards contradiction that f satisfies Statement A, but violates swap monotonicity. Again, there exist i, (P_i, P_{-i}), and P'_i as above such that the change in i's assignment violates swap monotonicity. This violation can occur in four ways:

1. i's assignment may not change for a and b, but for some other object. Then, by upper invariance, this object must be some $j \in L(b, P_i)$. We can choose $\tilde{j}$ to be the highest-ranking such object and derive a contradiction by a two-fold application of Statement A as in the proof of upper invariance.

2. i's assignment may change for a but not for b. Then, by upper invariance, a is the highest-ranking object for which i's assignment changes. The contradiction again arises by a two-fold application of Statement A.

3. i's assignment may change for b but not for a. This case is analogous to case 2, where the roles of P_i and P_i' are reversed.

4. i's assignment may change for both a and b, but in the wrong direction. If i's assignment for a increases, then a is the highest-ranking object for which i's assignment changes. This leads to a direct contradiction with Statement A. If i's assignment for b decreases, we obtain a contradiction by reversing the roles of P_i and P_i'.

Necessity ($B \Rightarrow A$). For this proof we introduce the notion of transitions between preference profiles. For two preference orders $P_i, P_i' \in \mathcal{P}$ a *transition from P_i' to P_i* is a finite sequence of preference orders $P^0, \ldots, P^K$ such that

- $P^0 = P_i'$ and $P^K = P_i$,
- for all $k \in \{0, \ldots, K-1\}$ we have $P^k \in N_{P^{k+1}}$.

Intuitively, such a transition resembles a series of consecutive swaps that transform the preference order P_i' into the preference order P_i. The *canonical transition (from P_i' to P_i)* is a particular transition that is inspired by the bubble-sort algorithm: Initially, we set $P^0 = P_i'$. The preference orders $P^1, \ldots, P^K$ are constructed in phases. In the first phase, we identify the highest ranking object under P_i that is not ranked in the same position under P_i', say j. Then we construct the preference orders $P^1, P^2, \ldots$ by swapping j with the respective next-more preferred objects. When j has reached the same position under P^k as under P_i, the first phase ends. Likewise, at the beginning of the second phase, we identify the object that is ranked highest under P_i of those objects that are not ranked in the same positions under P^k; and new preference orders are constructed by swapping this object up to its final position under P_i. Subsequent phases are analogous. The construction ends when the last created preference order P^K coincides with P_i.

Consider a fixed setting (N, M, q) and a mechanism f that is swap monotonic and upper invariant. Given an agent $i \in N$, a preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and a misreport $P_i' \in \mathcal{P}$, we must verify Statement A: If j is the highest-ranking object under P_i for which i's assignment changes, then i's assignment for j should decrease strictly if

i reports P_i' instead of P_i. Let

$$P_i \;:\; j_1 \succ \ldots \succ j_k \succ j_{k+1} \succ \ldots \succ j_m, \tag{20}$$

$$P_i' \;:\; j_1 \succ \ldots \succ j_k \succ j_{k+1}' \succ \ldots \succ j_m', \tag{21}$$

where P_i and P_i' coincide on the top-k choices but $j_{k+1} \neq j_{k+1}'$, and consider the canonical transition from P_i' to P_i. In the first phase, the object j_{k+1} is swapped upward in P_i' until it reaches the $(k+1)^{\text{st}}$ position. By upper invariance, i's assignment for the top-k choices $j_1, \ldots, j_k$ does not change during this phase. Moreover, by swap monotonicity, i's assignment for j_{k+1} cannot decrease. Initially, again by upper invariance, it does not change for the remainder of the canonical transition.

Suppose that i's assignment for j_{k+1} changes strictly during the first phase. Then a change of report from P_i to P_i' must lead to a strict decrease in i's assignment for j_{k+1}, and i's assignment for any of the more preferred objects $j_1, \ldots, j_k$ does not change. But this is precisely Statement A for the combination $i, (P_i, P_{-i}), P_i'$.

Alternatively, suppose that i's assignment for j_{k+1} does not change during the first phase. Then swap monotonicity implies that i's assignment cannot change at all during this phase. In this case, we proceed to the second phase, where the arguments are analogous. Finally, if i's assignment does not change in any phase, then it does not change at all, which would also be consistent with Statement A. □

B. 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 $-\delta_b \cdot u_i(a) + \delta_b \cdot u_i(b) \leqslant 0$. If $\hat{d} \neq a$ is i's reported last choice, then this change is $-\delta_a \cdot u_i(a) + \delta_b \cdot u_i(b) + (\delta_a - \delta_b) \cdot u_i(\hat{d})$. 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

$$\begin{aligned} & -\delta_a \cdot u_i(a) + \delta_b \cdot u_i(b) + (\delta_a - \delta_b) \cdot u_i(d) && (22) \\ = \; & -\delta_a \cdot \left(u_i(a) - \min_{j \in M} u_i(j) \right) + \delta_b \cdot \left(u_i(b) - \min_{j \in M} u_i(j) \right). && (23) \end{aligned}$$

This bound is guaranteed to be weakly negative if $\delta_a \geqslant r \cdot \delta_b$, 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

$$\begin{aligned} & -\delta_a \cdot \tilde{u}_i(a) + \delta_b \cdot \tilde{u}_i(b) + (\delta_a - \delta_b) \cdot \tilde{u}_i(d) && (24) \\ = \; & -\delta_a \cdot \left(\tilde{u}_i(a) - \min_{j \in M} \tilde{u}_i(j) \right) + \delta_b \cdot \left(\tilde{u}_i(b) - \min_{j \in M} \tilde{u}_i(j) \right), && (25) \end{aligned}$$

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

C. Example 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.

D. Example and Lemma Used in the Proof of Proposition 4

Example 4 (Violation of swap monotonicity by BM)**.** Consider a setting with $n \geqslant 4$ agents $N = \{1, 2, 3, 4, \ldots, n\}$, $m \geqslant 4$ objects $M = \{a, b, c, d, \ldots\}$, and capacities $q_j \geqslant 1$ such that $\sum_{j \in M} q_j = n$. Let agent 1 have preference order $P_1 : a \succ b \succ c \succ d \succ \ldots$, let agents 2 through $q_a + q_c + q_d$ have preference order $P_{\cdot} : a \succ b \succ \ldots$, and for all objects $j \in M \backslash \{a, b, d\}$, let there be q_j agents with preference order $P_{\cdot} : j \succ \ldots$. Then $\mathrm{BM}_1^{\mathbb{P}}(P_1, P_{-1}) = (x, y, 0, 1 - x - y, 0, \ldots)$ for some $x \in (0, 1), y \in (0, 1 - x)$ because $\mathbb{P}$ supports all single priority profiles. However, if i reports $P_i' : a \succ c \succ b \succ d \succ \ldots$, then $\mathrm{BM}_1^{\mathbb{P}}(P_1', P_{-1}) = (x, 0, 0, 1 - x, 0 \ldots)$. 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*

$$ABM_i^{\mathbb{P}}(P_i, P_{-i}) \neq ABM_i^{\mathbb{P}}(P_i', P_{-i}), \tag{26}$$

then there exists a priority profile π^* such that

$$ABM_{i,a}((P_i, P_{-i}), \pi^*) = ABM_{i,b}((P_i', P_{-i}), \pi^*) = 1. \tag{27}$$

Proof. From $\mathrm{ABM}_i^{\mathbb{P}}(P_i, P_{-i}) \neq \mathrm{ABM}_i^{\mathbb{P}}(P_i', P_{-i})$ it follows that there exists a priority profile π^0 with $\mathrm{ABM}_i((P_i, P_{-i}), \pi^0) \neq \mathrm{ABM}_i((P_i', P_{-i}), \pi^0)$. Let $\mathrm{ABM}_{i,j}((P_i, P_{-i}), \pi^0) = 1$ and $\mathrm{ABM}_{i,j'}((P_i', P_{-i}), \pi^0) = 1$, 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.

There are 6 cases:

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

$j = b$: Then monotonicity implies $j' = b$, 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 $k \in \{1, \ldots, K - 1\}$, let N^k be the agents who receive their object in round k, and let π_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 π^* where all objects have the same priority order π^*_j is a single priority profile with $\text{ABM}_{i,a}((P_i, P_{-i}), \pi^*) = \text{ABM}_{i,b}((P'_i, P_{-i}), \pi^*) = 1$.

$P_i : b \succ j$ and $P_i : b \succ j'$: If a and b are both exhausted at the end of round $K-1$, then i skips both objects, independent of its report. This implies $j = j'$, a contradiction. Next, if a is not exhausted at the end of round K, then i would receive b by reporting P'_i, again a contradiction. Analogously, we can rule out the case that a is not exhausted at the end of round K. Next, if a is exhausted at the end of round $K-1$ but b is not, then i skips a and applies to b, independent of its report. Symmetrically, the application process is independent of the report of i if b is exhausted at the end of round $K-1$ but b is not. Both imply contradictions.

$j = a$ 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 π^* be the same single priority profile as constructed in the third case. Then $\text{ABM}_{i,a}((P_i, P_{-i}), \pi^*) = \text{ABM}_{i,b}((P'_i, P_{-i}), \pi^*) = 1$.

$j' = b$ 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 π^* exists in all cases that do not imply contradictions. □

E. 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 $P_i' : b \succ a$. □

F. 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 are *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 $\rho_{(N,M,q)}(f) \geqslant \rho_{(N,M,q)}(g)$.*
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

$$\mathbb{E}_{f_i(P_i,P_{-i})}[u_i] < \mathbb{E}_{f_i(P_i',P_{-i})}[u_i], \tag{28}$$

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

$$\mathbb{E}_{g_i(P_i,P_{-i})}[u_i] < \mathbb{E}_{g_i(P_i'',P_{-i})}[u_i]. \tag{29}$$

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 $\rho_{(N,M,q)}(f) \geqslant \rho_{(N,M,q)}(g)$.

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

G. Discounted Dominance and Partial Strategyproofness

In this section, we prove an additional equivalence result between r-partial strategyproofness and an incentive concept induced by a new dominance notion we call *r-discounted dominance*. This dominance notion generalizes stochastic dominance but includes r as a discount factor, and it may be of independent interest to some readers.

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

$$\sum_{k=1}^{K} r^k \cdot x_{i,j_k} \geqslant \sum_{k=1}^{K} r^k \cdot y_{i,j_k}. \tag{30}$$

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 theorem 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 misreports $P_i' \in \mathcal{P}$. 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 $u_i \in U_{P_i} \cap$ URBI(r) we have $\sum_{j \in M} u_i(j) \cdot x_j \geqslant \sum_{j \in M} u_i(j) \cdot y_j$.

B. x_i r-discounted dominates y_i at P_i.

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

$$\sum_{l=1}^{m} u_i(j_l) \cdot (x_{j_l} - y_{j_l}) < 0. \tag{31}$$

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

$$S(K) = \sum_{k=1}^{K} u_i(j_k) \cdot (x_{j_k} - y_{j_k}) = \sum_{k=1}^{K} u_i(j_k) \cdot \delta_k. \tag{32}$$

By assumption, $S(m) = S(m-1) < 0$ (see (31)), so there exists a smallest value $K' \in \{1, \ldots, m-1\}$ such that $S(K') < 0$, but $S(k) \geqslant 0$ for all $k < K'$. Using Horner's method, we rewrite the partial sum and get

$$\begin{aligned} S(K') &= \sum_{k=1}^{K'} u_i(j_k) \cdot \delta_k = \left(\frac{S(K'-1)}{u_i(j_{K'})} + \delta_{K'} \right) \cdot u_i(j_{K'}) && (33) \\ &= \left(\left(\frac{S(K'-2)}{u_i(j_{K'-1})} + \delta_{K'-1} \right) \cdot \frac{u_i(j_{K'-1})}{u_i(j_{K'})} + \delta_{K'} \right) \cdot u_i(j_{K'}) && (34) \\ &= \left(\left(\ldots \left(\delta_1 \cdot \frac{u_i(j_1)}{u_i(j_2)} + \delta_2 \right) \cdot \frac{u_i(j_2)}{u_i(j_3)} + \ldots \right) \cdot \frac{u_i(j_{K'-1})}{u_i(j_{K'})} + \delta_{K'} \right) \cdot u_i(j_{K'}). && (35) \end{aligned}$$

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 $S(K'-1) \geqslant 0$ and $u_i(j_{K'-1}) > 0$, we must have that

$$\left(\frac{S(K'-2)}{u_i(j_{K'-1})} + \delta_{K'-1} \right) = \frac{S(K'-1)}{u_i(j_{K'-1})} \geqslant 0. \tag{36}$$

Therefore, when replacing $\frac{u_i(j_{K-1})}{u_i(j_K)}$ by $1/r$ in (35) 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

$$0 > S(K') \geqslant \left(\left(\dots\left(\frac{\delta_1}{r} + \delta_2\right) \cdot \frac{1}{r} + \dots\right) \cdot \frac{1}{r} + \delta_{K'}\right) \cdot u_i(j_{K'}) = \frac{u_i(j_{K'})}{r^{K'}} \cdot \sum_{k=1}^{K'} r^k \cdot \delta_k. \quad (37)$$

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

$$\sum_{k=1}^{K'} r^k \cdot \delta_k = \sum_{k=1}^{K'} r^k \cdot (x_{j_k} - y_{j_k}) \geqslant 0. \quad (38)$$

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

$$\sum_{k=1}^{K} r^k \cdot x_{j_k} < \sum_{k=1}^{K} r^k \cdot y_{j_k}, \quad (39)$$

and let K be the smallest rank for which inequality (39) is strict. Then the value

$$\Delta = \sum_{k=1}^{K} r^k \cdot (y_{j_k} - x_{j_k}), \quad (40)$$

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

$$u_i(j_k) = \begin{cases} Dr^k, & \text{if } k \leqslant K, \\ dr^k, & \text{if } K+1 \leqslant k \leqslant m-1, \\ 0, & k = m. \end{cases} \quad (41)$$

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

$$\begin{aligned} \sum_{j \in M} u_i(j) \cdot x_j - \sum_{j \in M} u_i(j) \cdot y_j &= \sum_{l=1}^{m} u(j_l) \cdot (x_{i,j_l} - y_{i,j_l}) && (42) \\ &= D \cdot \sum_{k=1}^{K} r^k \cdot (x_{j_k} - y_{j_k}) + d \cdot \sum_{k=K+1}^{m-1} r^k \cdot (x_{j_k} - y_{j_k}) \\ &\leqslant -D \cdot \Delta + d. && (43) \end{aligned}$$

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

This concludes the proof of Proposition 7. □

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-, and ST[16]-strategyproofness).

The dominance interpretation also unlocks the partial strategyproofness concept to algorithmic analysis: Recall that the definition of r-partial strategyproofness imposes inequalities that have to hold for all utility functions within the set URBI(r). This set is infinite, which makes 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 are satisfied (i.e., the inequalities (30) from Definition 11). These inequalities can also be used to encode r-partial strategyproofness as linear constraints to an optimization problem. This enables an automated search in the set of r-partially strategyproof mechanisms while optimizing for some other design objective under the automated mechanism design paradigm (Sandholm, 2003).

H. URBI(r) and the Probabilistic Serial Mechanism

3-by-3 Settings: Consider the same setting as in the motivating example in the introduction with three agents $N = \{1, 2, 3\}$ and three objects $M = \{a, b, c\}$ with unit capacity. Recall that

$$\mathrm{PS}_1(P_1, P_{-1}) = (3/4, 0, 1/4), \quad (44)$$

$$\mathrm{PS}_1(P_1', P_{-1}) = (1/2, 1/3, 1/6). \quad (45)$$

[16] *Sure thing* dominance; see (Aziz, Brandt and Brill, 2013)

The gain in expected utility for agent 1 from misreporting is therefore

$$-1/4 \cdot u_1(a) + 1/3 \cdot u_1(b) - 1/12 \cdot u_1(c). \tag{46}$$

This is weakly negative if and only if

$$\frac{u_1(b) - u_1(c)}{u_1(a) - u_1(c)} \leqslant 3/4, \tag{47}$$

which is the condition for u_1 to satisfy URBI(3/4). Next, recall that we have computed the degree of strategyproofness of PS in this setting to be $\rho_{(N,M,q)}(\text{PS}) = 3/4$ (Figure 2 in Section 5.1). Any agent therefore has a dominant strategy to report truthfully under PS in this setting *if and only if* that agent's utility function satisfies URBI(3/4). In other words, the set of utility functions for which PS induces good incentives is precisely URBI(3/4).

4-by-4 Settings (adapted from (Abächerli, 2017)): Next, we show a similar insight for a setting with four agents $N = \{1, 2, 3, 4\}$ and four objects $M = \{a, b, c, d\}$ with unit capacity. We have computed the degree of strategyproofness of PS in this setting to be $\rho_{(N,M,q)}(\text{PS}) = 1/2$ (Figure 2 in Section 5.1). Consider the preferences

$$P_1 \;:\; a \succ b \succ c \succ d, \tag{48}$$
$$P_2 \;:\; b \succ c \succ d \succ a, \tag{49}$$
$$P_3 \;:\; c \succ a \succ b \succ d, \tag{50}$$
$$P_4 \;:\; c \succ b \succ d \succ a, \tag{51}$$

and the misreport

$$P_1' \;:\; b \succ a \succ c \succ d \tag{52}$$

by agent 1. The resulting assignments are

$$\text{PS}_1(P_1, P_{-1}) = (3/4, 0, 0, 1/4), \tag{53}$$
$$\text{PS}_1(P_1', P_{-1}) = (1/2, 1/2, 0, 0). \tag{54}$$

Thus, agent 1's gain in expected utility from misreporting is

$$- 1/4 \cdot u_1(a) + 1/2 \cdot u_1(b) - 1/4 \cdot u_1(d). \tag{55}$$

This is weakly negative if and only if

$$\frac{u_1(b) - u_1(c)}{u_1(a) - u_1(c)} \leqslant 1/2. \tag{56}$$

Next, consider the preference profile

$$P_1, P_2 \ : \ a \succ b \succ c \succ d, \tag{57}$$
$$P_3, P_4 \ : \ a \succ c \succ d \succ b, \tag{58}$$

and the misreport

$$P_1' \ : \ a \succ c \succ b \succ d. \tag{59}$$

Then

$$\mathrm{PS}_1(P_1, P_{-1}) = (1/4, 1/2, 0, 1/4), \tag{60}$$
$$\mathrm{PS}_1(P_1', P_{-1}) = (1/4, 1/3, 1/3, 1/12), \tag{61}$$

so that the gain in expected utility for agent 1 from misreporting is

$$- 1/6 \cdot u_1(b) + 1/3 \cdot u_1(c) - 1/6 \cdot u_1(d). \tag{62}$$

This is weakly negative if and only if

$$\frac{u_1(c) - u_1(d)}{u_1(b) - u_1(d)} \leqslant 1/2. \tag{63}$$

Observe that conditions (56) and (63) are precisely equivalent to the requirement that u_1 satisfies URBI(1/2). Thus, since $\rho_{(N,M,q)}(\mathrm{PS}) = 1/2$, every agent with a utility function that satisfies URBI(1/2) has a dominant strategy to report truthfully. However, the two examples show that every agent with a utility function that violates URBI(1/2) can beneficially manipulate the mechanism in some situations. This implies that, in the setting with 4 agents and 4 objects in unit capacity, the set of utility functions for which PS makes truthful reporting a dominant strategy is precisely URBI(1/2).

5-by-5 Settings (adapted from (Abächerli, 2017)): In a setting with 5 agents $N = \{1, 2, 3, 4, 5\}$ and 5 objects $M = \{a, b, c, d, e\}$ with unit capacity, we again consider PS. Using the same algorithm as in Section 5.1, we can determine the degree of strategyproofness in this setting to be $\rho_{(N,M,q)}(\text{PS}) = 1/2$. The utility function u_i with

$$u_i(a) = 7.99, u_i(b) = 4, u_i(c) = 2, u_i(d) = 1, u_i(e) = 0 \tag{64}$$

violates URBI(1/2) because

$$\frac{u_i(a) - \min_{j \in M} u_i(j)}{u_i(b) - \min_{j \in M} u_i(j)} = \frac{4}{7.99} > 1/2. \tag{65}$$

However, exhaustive search over all possible preference profiles and misreports (using a computer) reveals that an agent with this utility function cannot benefit from misreporting. Thus, in the setting with 5 agents and 5 objects, the set of utility functions for which PS makes truthful reporting a dominant strategy is strictly larger than URBI(1/2). This contrasts tightness of this set of the 3-by-3 and 4-by-4 settings.