University of Zurich[UZH]

Department of Informatics

Trade-offs between Strategyproofness and Efficiency of Ordinal Mechanisms

Dissertation submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor / Doktorin der Wissenschaften, Dr. sc.
(corresponds to Doctor of Science, PhD)

presented by
Timo Mennle
from Germany

approved in February 2016

at the request of
Prof. Sven Seuken, Ph.D.
Prof. Tayfun Sönmez, Ph.D.
Prof. Eric Budish, Ph.D.

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, February 17, 2016

Chairwoman of the Doctoral Board: Prof. Elaine M. Huang, Ph.D.

For Judith

Abstract

There are some things that money cannot buy. For various reasons, moral or otherwise, society has set boundaries regarding the use of money for certain resources and transactions. Such restrictions often arise in situations that are of great importance to people's lives: subsidized housing must be assigned to tenants, seats at public schools must be assigned to students, or a new president must be elected. The design of mechanisms for these problems is plagued by severe impossibility results pertaining to strategyproofness. In this thesis we address the research question of how to trade off strategyproofness and other desiderata in the design of ordinal mechanisms. For the assignment domain we introduce the new relaxed incentive concept of *partial strategyproofness* which can be used to measure the incentive properties of non-strategyproof mechanisms. We employ this concept to show that a choice between three popular school choice mechanisms, the *Deferred Acceptance mechanism* and *two variants* of the *Boston mechanism*, involves an implicit trade-off between strategyproofness and efficiency. Next, we give conditions under which *hybrid mechanisms* facilitate meaningful trade-offs between strategyproofness and efficiency in the assignment domain. Finally, in the general ordinal domain we introduce a new framework to assess mechanisms by their *manipulability* and their *welfare deficit*. The welfare deficit is a measure for their ability to achieve another desideratum, such as efficiency, stability, or fairness. Within this framework the *Pareto frontier* consists of those mechanisms that trade off manipulability and deficit optimally. Our main result is a structural characterization of this Pareto frontier.

Zusammenfassung

Es gibt Dinge, die kann man nicht kaufen. Aufgrund moralischer oder ethischer Erwägungen ist in gewissen Märkten die Nutzung von Geld eingeschränkt oder vollständig ausgeschlossen. In solchen Märkten geht es häufig um Dinge, die für die betroffenen Personen von großer Bedeutung sind, wie der Zugang zu gefördertem Wohnraum, die Verteilung von Schülern auf öffentliche Schulen oder die Wahl eines neuen Präsidenten. Das Mechanismus-Design für diese Märkte unterliegt scharfen Unmöglichkeitstheoremen. In dieser Dissertation betrachten wir den Zielkonflikt zwischen Strategyproofness einerseits und dem Erreichen weiterer wünschenswerter Eigenschaften von Mechanismen andererseits. Für Märkte, in denen es um die Verteilung unteilbarer Güter geht, führen wir *partial strategyproofness* ein. Dies ist ein neues Konzept zur Bewertung der Anreizeigenschaften von Mechanismen, die nicht vollständig strategyproof sind. Wir wenden dieses Konzept auf drei Mechanismen an, die häufig für die Verteilung von Schülern auf öffentliche Schulen genutzt werden, den *Deferred-Acceptance-Mechanismus* und *zwei Varianten* des *Boston-Mechanismus*. Wir zeigen auf, dass eine Wahl zwischen diesen Mechanismen immer auch als Trade-off zwischen Anreizeigenschaften und ökonomischer Effizienz zu betrachten ist. Anschließend entwickeln wir Bedingungen, unter denen *hybride Mechanismen* einen solchen Trade-off ermöglichen. Für die Analyse allgemeiner ordinaler Mechanismen entwickeln wir zwei Maße. Das erste Maß quantifiziert die *Manipulierbarkeit* von Mechanismen; das zweite Maß (das *Defizit*) quantifiziert deren Fähigkeit, eine andere wünschenswerte Eigenschaft zu erreichen. Mit diesen Maßen definieren wir die *Pareto frontier*. Dies ist die Menge jener Mechanismen, die ein geringstmögliches Defizit erreichen, dabei aber eine beschränkte Manipulierbarkeit aufweisen. Unser Hauptresultat ist eine strukturelle Charakterisierung dieser Pareto frontier.

Citations to Previously Published Work

Parts of Chapter 2 have previously appeared in:

> **Mennle, Timo, and Sven Seuken.** 2014. "An Axiomatic Approach to Characterizing and Relaxing Strategyproofness of One-sided Matching Mechanisms." In *Proceedings of the 15th ACM Conference on Economics and Computation (EC).*

Acknowledgments

There are some things that money cannot buy; among these things are advisors like Sven Seuken. He has infected me with his scientific curiosity and his passion for market and mechanism design. His foresight and dedication have been the single most important factor in making my four years at graduate school a success. I can only hope that one day the guidance I give to my own students will be as priceless to them as Sven's guidance was to me.

I thank Tayfun Sönmez and Eric Budish for serving as external reviewers of this thesis and for their invaluable suggestions to improve it. I am deeply honored by having them on my dissertation committee.

For countless discussions that have shaped and refined my research I thank Atila Abdulkadiroğlu, Nick Arnosti, Haris Aziz, Ivan Balbuzanov, Christian Basteck, Craig Boutilier, Gabriel Carroll, John Dickerson, Umut Dur, Lars Ehlers, Aris Filos-Ratsikas, Bettina Klaus, Flip Klijn, Fuhito Kojima, Christian Kroer, Eric Maskin, Antonio Miralles, Bernardo Moreno, David Parkes, Ariel Procaccia, Baharak Rastegari, Alvin Roth, Arunava Sen, William Thomson, and Utku Ünver. My special thanks go to Albin Erlanson and Katharina Huesmann for their insightful yet actionable suggestions to improve the manuscript.

I thank the members of the Computation and Economics Research Group at the University of Zurich, Gianluca Brero, Ludwig Dierks, Dmitry Moor, Steffen Schuldenzucker, and Mike Shann who have created the most inspiring and motivating environment to work in. I thank Daniel Abächerli, Stefan Bublitz, Basil Philipp, Leonardo Stedile, and Michael Weiss for being wonderful students and for occasionally teaching me a lesson as well.

I gratefully acknowledge the financial support by the Swiss National Science Foundation and the Hasler Foundation.

I thank all members of my family for their love and encouragement, and I thank Kristin Michelitch and Thomas Müller for the deep friendships that we share. Most of all, I thank my parents Ulrike Vollmer-Rusche and Gerhard Vollmer who have given this wonderful life to me. You have done everything right!

Finally, I thank my beloved wife Judith Mennle, whose consistent support is doubtlessly responsible for much of the breadth and depth of this dissertation; and I thank my beloved daughter Tabea Mennle, who is probably responsible for some of its brevity.

Contents

1 Motivation and Overview of Results

1.1 Introduction

It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest.

– Adam Smith, *The Wealth of Nations*

The essence of *markets* is that self-interested individuals (called *agents*) engage in mutually beneficial transactions. The most prominent examples are financial markets where company shares or government bonds are bought and sold. In these markets, clear rules govern the interactions of the agents: the exchanges regulate how shares are issued, how bids are placed, and how trades are executed. The individual agents can have a variety of goals when participating in these markets, including raising capital, investing capital, or hedging risks. Often, the agents go to great lengths to achieve these goals as well as possible. In the language of *game theory*, they play a strategic game; what this game looks like is determined by the rules that govern the interactions in the market.

The research on *mechanism design* can be understood as the "inverse" of game theory: the objective of a mechanism designer is to set up these rules, called the *mechanism*, in such a way that a desirable outcome is achieved even if the agents act in their own self-interest. For example, governments can grant the rights to transmit signals on different frequency bands of the wireless spectrum. Mobile phone companies are interested in obtaining these rights. While the goal of the phone companies is to make a profit, the goal of the government may be to improve the country's communication infrastructure. Thus, it would prefer to grant the rights to those companies that will use them most effectively. However, information about their own effectiveness is typically private information of the companies. Governments can design auction mechanisms to overcome this problem. The aim of mechanism design is to identify those mechanisms that will yield the most desirable outcomes while taking into account that the agents have private information about their own preferences and will act in their own interest.

Money plays an important role in the markets discussed above. As Nobel Laureate Alvin Roth has pointed out in his Prize Lecture, in these markets "[t]he price does all the work. The New York Stock Exchange discovers a price at which supply equals demand. But, in lots of markets, prices don't do all the work." For various reasons, moral or otherwise, society has set boundaries regarding the use of money for certain resources and transactions. For example, buying and selling human organs for transplantation is illegal in most countries. Nonetheless, the *exchange of kidneys* from living human donors (without money) saves thousands of lives every year in the United States alone. Second, there is a general consensus that a child's access to the public education system should not depend on the income of its family. *School choice* markets (without money) arise because parents have diverse and private preferences over which public schools they would like their children to attend. Third, the purpose of subsidized housing is to provide adequate housing to those who are too poor to obtain such housing otherwise. While the potential tenants may have diverse preferences, simply offering this type of housing to the "highest bidder" would clearly defy the initial purpose. This gives rise to *housing markets* (without money). Finally, the outcomes of elections should not depend on the financial means of those who vote. This is reflected by the fact that in *democratic elections* (without money) all votes have the same weight and cannot be bought or sold.

Without money, it is often difficult for agents to identify and express their own preference intensities. Instead, the mechanisms in such markets typically elicit preference rankings from the agents (called *ordinal preferences*). In fact, mechanisms whose outcomes must be independent of the agents' levels of wealth are bound to be ordinal (Huesmann and Wambach, 2015; Ehlers et al., 2015). As this thesis concerns market design for markets without money, we study the design of ordinal mechanisms.

We begin with three instructive examples.

1.1.1 On-campus Housing at MIT

When starting university, it is often challenging for students to find suitable accommodation. Searching on the private market can be tedious and students' budgets are limited. For this reason, universities frequently offer accommodation on or near the campus that is tailored to the needs of students: it is affordable, conveniently located, and provides a student community. Rents for this kind of student housing are usually fixed and lower than rents for comparable housing on the private market. Therefore, the number of applicants typically exceeds the number of rooms. Since rents cannot be changed to balance supply and demand, an *assignment mechanism* must be used to assign the rooms

to the applicants.

The MIT Division of Student Life is responsible for this assignment mechanism at the Massachusetts Institute of Technology.[1] Until 2002 this mechanism worked as follows: first, students were asked to submit rank-ordered lists of acceptable (types of) rooms. Then a random number was drawn for each applicant. The applicant with the highest number received her first choice. Next, the applicant with the second-highest number received her best choice that was still available taking into account that the first applicant's room was no longer available. This process was repeated for each applicant in decreasing order of their random numbers. If for some applicant none of the remaining choices were acceptable, then this applicant remained unassigned. The process ended when all rooms had been filled or all applicants had been considered.

This mechanism is known as *Random Serial Dictatorship*. It has a number of appealing properties: first, truthful reporting is a dominant strategy for the applicants (i.e., the mechanism is *strategyproof*). Second, once each applicant is assigned to a room, there exist no Pareto improvements that make some applicant strictly better off without also making some other applicant strictly worse off at the same time (i.e., the mechanism is *ex-post efficient*). Third, it is "fair" in the sense that all applicants who submit the same preference orders also have the same chances of obtaining each of the rooms (i.e., the mechanism is *symmetric*).

While these three properties are intuitively appealing, Random Serial Dictatorship also has at least one significant drawback: consider two applicants, Alice and Bob, who compete for two rooms, `North` and `South`. Both prefer `North`, but Alice would also be happy to move into `South` while Bob finds `South` unacceptable. If Alice receives a higher random number, then she will be assigned to `North`, Bob will remain unassigned, and `South` will remain unoccupied. This assignment is *wasteful* because Bob could have been assigned to `North` if Alice had taken `South`. To reduce this kind of waste, the assignment mechanism at MIT was changed in 2003. Under the new mechanism, applicants still submit rank-ordered lists. Subsequently, however, an optimization algorithm determines an assignment of the applicants to the rooms, where the primary objective is to maximize the total number of applicants assigned to on-campus housing. Among all these "maximal" assignments, the algorithm then tries to identify one that makes students as happy as possible given the rank orders they submitted.

In the above example, the new mechanism would assign Bob to `North` and Alice to `South`, so that room `South` would not be wasted. However, the new mechanism is no

[1] Source: MIT Division of Student Life through https://housing.mit.edu/ and personal communication.

longer robust to strategic manipulation by applicants: to see this, suppose that Alice ranks only `North` and pretends that `South` is unacceptable for her. The algorithm is then restricted to assigning `North` to Alice or to Bob and leaving `South` unassigned. If Alice prefers a 50% chance for `North` to living in `South` for sure, then this strategic misreport will make her better off. If we expect applicants to misreport their preferences strategically, then we risk that the assignment decision is based on false preference information.

This simple example illustrates a fundamental conflict between different desiderata in the design of assignment mechanisms which can also be prove formally: *any* assignment mechanism that maximizes the number of students who receive on-campus housing is susceptible to strategic manipulation by the students. Conversely, any strategyproof mechanisms is wasteful. For student housing at MIT it appears that the administrators considered the reduction of waste a more important desideratum than strategyproofness. There can be good justifications for such a decision: for example, if outside options are much less attractive than acceptable on-campus housing options, then Alice's manipulation would not be beneficial for her. Alternatively, students may not possess the necessary information to determine a useful misreport, even if one exists.

The prevalence of non-strategyproof assignment mechanisms, like the new assignment mechanism for on-campus housing at MIT, gives rise to an important research question: *if assignment mechanisms are not strategyproof, how can we understand and compare their incentive properties?*

1.1.2 School Choice Markets

Our second example of a market where monetary transfers are restricted is the assignment of seats at public schools. Often students are simply assigned to the school that is located nearest to their home. The main motivation for this approach is to minimize the daily logistics of getting students to school and home again, which can be substantial in sparsely populated areas. However, in most cities, public schools are more densely distributed, and travel distances are no longer prohibitive for students to access different schools. Other factors, such as teaching quality, special programs, or differences in the teaching approach may also influence the parents' preferences about which schools they would like their children to attend. Under the default approach, parents can get their children into another school by moving to the catchment area of that school. Indeed, the presence of better public schools was found to have a significant, positive effect on housing prices (Fack and Grenet, 2010). Thus, access to good public schools can

effectively be "bought" by moving and paying the markup for the better school. This contradicts the express objective of most public education systems to offer equal access for all students, independent of their socioeconomic background.

School choice programs give parents the opportunity to make their preferences known so that they can be considered in the assignment process. This can (but does not always) eliminate the need for parents to move in order to access a better public school. A particular mechanism that finds wide-spread application in many cities around the world is the *Boston mechanism*.[2] Under this mechanism, parents first submit rank-ordered lists of public schools that they find acceptable. In the first round, all students "apply" to their reported first choice. If a school has enough seats to accommodate all applicants, these students are assigned to this school. If there are more applicants than open seats at some school, then the available seats are assigned to the applicants according to priorities and using random tie-breaking if necessary. The students who have not been assigned to their first choice enter the second round where their second choices are considered. Schools continue to accept applicants into open seats until their capacity is exhausted, and unassigned students enter the third round. This process continues until all seats are filled or until all remaining seats are unacceptable to all remaining students.

If parents report their preferences truthfully, then the Boston mechanism maximizes the number of students who get their first choice, and subsequently, it maximizes the number of second choices, third choices etc. From a welfare-perspective, this is intuitively appealing. However, it may not be in the parents' own best interest to report their preferences truthfully. Consider a situation with three students, Alice, Bob, and Charlie, who compete for three schools, `East`, `West`, and `Central`, with a single seat each. Suppose that their preferences are as follows:

- Alice and Bob prefer `East` to `West` to `Central`,
- Charlie prefers `West` to `East` to `Central`.

Assume that in case of ties, priority is given to students in alphabetic order of their names. If all students report truthfully, Alice and Charlie will get their respective first choices `East` and `West` in the first round. Bob will be rejected at `East` in the first round because Alice has higher priority. In the second round Bob will apply to `West`. But since the only seat is already taken, he will ultimately end up at `Central`, his last choice. Now suppose that Bob claimed that `West` was his first choice instead. He would receive

[2]The name stuck, despite the fact that this mechanism is no longer used in the city of Boston.

`West` in the first round, thereby obtaining a more preferred school. This illustrates the susceptibility of the Boston mechanism to strategic manipulation.

This manipulability of the Boston mechanism has a number of undesirable consequences: first, when Bob misreports in the above example, it *looks like* two first choices and one third choice are assigned. But in fact, one first, one second, and one third choice are assigned with respect to the *true* preferences. Thus, while Bob has benefited from his strategic behavior, the outcome is arguable less desirable for society as a whole. Second, simply determining their own true preferences may already be a challenging task for parents as this requires research about the quality and suitability of the different schools. On top of that, figuring out a good or even optimal manipulation strategy requires additional information about the preferences of other parents as well as a profound understanding of the mechanism. This makes participation in the Boston mechanism a complex strategic problem. Most parents would probably prefer to expend their cognitive resources on other matters. Third, not all parents have the time and ability to engage in this kind of strategic behavior. As a result, the children of non-strategic parents who report their preferences honestly may be at a disadvantage.

In their seminal paper, Abdulkadiroğlu and Sönmez (2003*b*) proposed the *Deferred Acceptance* mechanism as an alternative to overcome the issues that arise from the considerable manipulability of the Boston mechanism. Deferred Acceptance also collects rank-ordered lists and proceeds in rounds. However, the assignments in each round are "tentative" rather than "final." A student who applies to her second choice in the second round may be accepted at that school even if all the seats are already taken. To make space for the new student, the tentative acceptance of another student with lower priority is withdrawn. That student re-enters the mechanism and continues the application process in order of its reported preferences. Under the Deferred Acceptance mechanism, Bob would still have been rejected from `East` in the first round and applied to `West` in the second round. However, since Bob has a higher priority than Charlie, he would have displaced Charlie from `West`. Charlie (instead of Bob) would then have continued the application process, been rejected from `East`, and ultimately obtained `Central`. Under the Deferred Acceptance mechanism, it is optimal for parents to report their preferences truthfully, independent of the reports from the other parents. On the other hand, the mechanism no longer maximizes the number of first choices.

Juxtaposing the observations about the Boston mechanism and the Deferred Acceptance mechanism, it becomes evident that there exists a tension between incentives for truth-telling on the one hand and economic efficiency on the other hand. This raises questions

as to *how mechanism designers can make an informed decision about trade-offs between these two dimensions* and *whether there exist any mechanism design alternatives that have intermediate performance on both dimensions.*

1.1.3 Selecting a Host for the Next Olympic Games

Every four years, the International Olympic Committee (IOC) selects a host city for the next Olympic Games. To arrive at a decision, the IOC uses a voting mechanism called *Plurality with Run-offs*. This mechanism works as follows: initially, a set of alternatives is determined, based on objective criteria, most importantly the cities' general ability to host the event. Next, each member of the IOC casts one vote for his or her favorite alternative. If one alternative receives an absolute majority of the votes, this alternative is selected. Otherwise, the alternative with the least number of votes is removed from the choice set and the process is repeated.

Plurality with Run-offs is not strategyproof: for simplicity, let us suppose that the IOC has only five members[3] who are conveniently named Alice, Bob, Charlie, David, and Eva. They vote on three alternative, `Athens`, `Buenos Aires`, and `Cape Town`. Suppose further that the IOC members' true preferences are as follows:

- Alice and Bob prefer `Athens` to `Buenos Aires` to `Cape Town`,
- Charlie and David prefer `Cape Town` to `Athens` to `Buenos Aires`,
- Eva prefers `Buenos Aires` to `Cape Town` to `Athens`.

If they vote truthfully, `Buenos Aires` will be eliminated in the first round. Then, in the second round, Eva will vote for `Cape Town`, and so `Cape Town` will be selected. Alice could pretend that her first choice was `Buenos Aires` instead. Then `Athens` would be eliminated in the first round, and the vote from Bob would go to `Buenos Aires`, which would make `Buenos Aires` the winner. Thus, by ranking `Buenos Aires` in first position, Alice can change the winner from her last choice to her second choice.

This manipulability is not a problem as long as we have sufficient confidence in the honesty of the members of the IOC. However, if we worry that members might misreport their preferences strategically, then we face a similar dilemma as under the Boston school choice mechanism: while the decision may *look like* the right one with respect to the reported preferences, it can differ from the decision that we would have liked to make given the members' true preferences. One way to avoid the problem of strategic

[3] In September 2015, the IOC actually consisted of 134 members (IOC, 2015).

misreporting is to use a *Dictatorship* mechanism instead: one member of the IOC is chosen (e.g., randomly) and this member's most preferred alternative is selected. This mechanism is obviously strategyproof because the member who is chosen is happiest if her true first choice is selected, while the preferences of other members do not matter. However, this mechanism can obviously fail to select the socially preferred alternative: in the example this alternative would be `Cape Town`; but if Eva is chosen as the "dictator," then `Buenos Aires` will be selected. Moreover, it is unlikely that a dictatorial mechanism would be acceptable to the IOC.

Unfortunately, it is impossible to design a mechanism that always selects the correct alternative but is strategyproof at the same time. If Plurality with Run-offs is viewed as "too manipulable" and Dictatorship is "wrong too often," this raises the question *whether there exist mechanisms with intermediate performance on both dimensions; and if so, how we can identify the ones that make optimal trade-offs between the two dimensions.*

1.2 Problem Statement, Related Work, and Research Questions

The three examples about on-campus housing, school choice mechanisms, and collective decisions raise a number of questions.

First, we have seen that in the assignment problem, strategyproofness is in conflict with the desideratum to maximize the number of students who live on campus. Prior work has uncovered many of the restrictions that strategyproofness entails for the design of deterministic assignment mechanisms (Pápai, 2000; Ehlers and Klaus, 2006, 2007; Hatfield, 2009; Pycia and Ünver, 2014), as well as when the mechanisms are allowed to use randomization (Zhou, 1990; Abdulkadiroğlu and Sönmez, 1998; Lee and Sethuraman, 2011; Bade, 2014). Economists have also introduced notions of economic efficiency that are more demanding than the baseline ex-post efficiency of Random Serial Dictatorship. However, together with fairness these are incompatible with strategyproofness (Zhou, 1990; Bogomolnaia and Moulin, 2001; Featherstone, 2011).

To design mechanisms that achieve these efficiency notions, the requirement of strategyproofness must be relaxed. New incentive concepts have been proposed to study the incentive properties of these non-strategyproof mechanisms, such as *weak strategyproofness* (Bogomolnaia and Moulin, 2001), *lexicographic strategyproofness* (Cho, 2012), *convex strategyproofness* (Balbuzanov, 2015), and *strategyproofness in the large* (Azevedo and Budish, 2015). Like strategyproofness, these concepts are "binary" in the sense that a

mechanism either satisfies them or not. However, if two different mechanisms both violate a given incentive concept, then both mechanisms "look alike" in terms of their respective incentive properties. In domains with transferable utility, *approximate strategyproofness* (Lubin and Parkes, 2012) yields a *parametric* relaxation of strategyproofness. While approximate strategyproofness can be adapted for ordinal mechanisms (Birrell and Pass, 2011; Carroll, 2013), this concept does not reflect the specific structure of incentives under non-strategyproof assignment mechanisms. A relaxed notion of strategyproofness that is *parametric*, meaningful in *finite settings*, and exploits the particular structure of the *assignment problem* has remained elusive so far.

An important application of assignment mechanisms are school choice markets. Since the seminal paper by Abdulkadiroğlu and Sönmez (2003*b*), school choice mechanisms have attracted the attention of economists, and a growing body of research has had substantial impact on policy decisions (Abdulkadiroğlu, Pathak and Roth, 2005; Pathak and Sönmez, 2008; Abdulkadiroğlu, Pathak and Roth, 2009; Pathak and Sönmez, 2013). The debate on school choice mechanisms has largely been centered around the two mechanisms from our example, the *Boston mechanism* and the *Deferred Acceptance mechanism*. While the Deferred Acceptance mechanism is strategyproof, the Boston mechanism is not; and human agents were found to misreport under the Boston mechanism in the laboratory (Chen and Sönmez, 2006; Pais and Pinter, 2008) and in practice (Abdulkadiroğlu et al., 2006; Calsamiglia and Güell, 2014). At the same time, Abdulkadiroğlu, Che and Yasuda (2015) found that ex-ante welfare in equilibrium under the Boston mechanism may be higher. This raises the question whether a choice between the two mechanisms involves a trade-off between strategyproofness and efficiency.

In some school choice markets, the Boston mechanism is implemented in a subtly different fashion: in Amsterdam (until 2014), in Freiburg (Germany), and throughout the German state of Nordrhein-Westfalen a variant of the Boston mechanism is used where students automatically skip exhausted schools in the application process. This "adaptive" Boston mechanism has been largely overlooked by researchers so far (notable exceptions include (Dur, 2015; Harless, 2015)). In particular, there exists no satisfactory understanding of the *trade-offs between strategyproofness and efficiency* that are implicit in a choice between Deferred Acceptance and the two variants of the Boston mechanism. More generally, the *parametric trade-offs* between strategyproofness and efficiency of assignment mechanisms have remained largely unexplored.

Finally, our example involving the IOC has shown that conflicts between strategyproofness and other desiderata are not confined to the assignment domain, but they persist

for the design of ordinal mechanisms in general. The seminal Gibbard-Satterthwaite impossibility result showed that any decision mechanism that is strategyproof and *unanimous* (i.e., if all agents agree on a best alternative, then this alternative is selected) must be dictatorial (Gibbard, 1973; Satterthwaite, 1975). Gibbard (1977) extended this result to mechanisms that involve randomization. Many researchers have sought ways to circumvent these impossibility results. Important approaches include domain restrictions (Moulin, 1980; Chatterji, Sanver and Sen, 2013), limit results (Procaccia, 2010; Birrell and Pass, 2011), and computational hardness as a barrier to manipulation (Bartholdi, Tovey and Trick, 1989; Xia, 2011; Mossel and Rácz, 2014).

Neither of these three approaches is free of drawbacks: first, by definition, domain restrictions assume that agents' preferences are restricted. However, a mechanism that has "good" properties for the restricted preferences may lose these properties if reality is more general than the domain restriction. For example, the options for the location of a new hospital are typically not arranged on a straight line, as single-peakedness would require. Second, limit results are appealing in large settings, but the question remains when a setting is "large enough." Birrell and Pass (2011) go to great length to answer this question for a particular voting problem; and their is no simple answer to this question in general. Third, for many mechanisms, the agent's manipulation problem can be shown to be NP-hard. However, NP-hardness yields a worst-case notion of computational difficulty but says nothing about the average difficulty of finding a beneficial manipulation. In fact, Mossel and Rácz (2014) showed that almost all voting mechanisms are easy to manipulate *on average*.

A further concern with all three approaches is that insights are typically binary in the sense that a given combination of properties is either achievable by some mechanism or not; or that a particular mechanism either makes the agent's manipulation problem hard or not. What is missing so far is a *framework that quantifies the performance of mechanisms* with respect to their incentive properties on the one hand and their ability to achieve a certain desideratum (e.g., efficiency) on the other hand. Given such a framework, we would like to understand which mechanisms *trade off performance on the two dimensions optimally.*

Research Questions

In this thesis, we address the question of how we can trade off strategyproofness and efficiency or other desiderata in the design of assignment mechanisms and general ordinal mechanisms. Specifically, we ask:

Question 1. What is a good way to describe and compare the incentive properties of non-strategyproof assignment mechanisms?

Question 2. What are the implicit trade-offs between strategyproofness and efficiency when choosing between the Deferred Acceptance mechanism and the two variants of the Boston mechanism in school choice markets?

Question 3. How can we construct new assignment mechanisms that make smooth, parametric trade-offs between strategyproofness and efficiency?

Question 4. For general ordinal mechanisms, how can we design mechanisms that trade off strategyproofness and another desideratum optimally?

1.3 Contributions of this Thesis

This thesis consists of four papers, each of which addresses one of the research questions.

1.3.1 The Partial Strategyproofness Concept

In the first paper of this thesis, we introduce the *partial strategyproofness concept*. This is a new, relaxed notion of strategyproofness that captures the particular structure of incentives under non-strategyproof assignment mechanisms. If an assignment mechanism is strategyproof, then an agent who prefers an object a to another object b finds it beneficial to state this preference to the mechanism (as opposed to making the false claim that it "prefers b to a"). This is independent of whether it prefers a to b *just a little bit* or *a lot*. Under non-strategyproof mechanisms, however, it often depends on these *relative preference intensities* whether an agent finds a particular misreport beneficial: the closer it is to being indifferent between a and b, the more likely it is that this agent benefits from swapping the two objects. Following this intuition, we say that a mechanism is *partially strategyproof* if it makes truthful reporting a dominant strategy at least for those agents whose preference for a over b is sufficiently strong. However, if an agent is close to being indifferent between a and b, there is no requirement. In this sense, partial strategyproofness is a weaker condition than strategyproofness. The extent to which agents can be indifferent between objects is controlled by a numerical parameter r, the *degree of strategyproofness*. If $r = 1$, then the mechanism is strategyproof. If $r = 0$, then manipulation opportunities persist even for agents who distinguish very strongly

between different objects. Intermediate values of r correspond to intermediate incentive properties.

Partial strategyproofness can be derived axiomatically: we show that an assignment mechanism is strategyproof if and only if it satisfies three axioms, called *swap monotonicity*, *upper invariance*, and *lower invariance*. The larger class of partially strategyproof mechanisms arises by dropping the lower invariance axiom. Furthermore, we show that the degree of strategyproofness is a meaningful measure for the strength of the incentive properties of non-strategyproof assignment mechanisms: via a maximal domain result we show that partial strategyproofness is in a well-defined sense the strongest incentive guarantee that can be given for swap monotonic, upper invariant mechanisms.

Following this introduction of the partial strategyproofness concept, we demonstrate that it captures our intuitive understanding of what it means for a non-strategyproof assignment mechanism to have "good incentive properties:" first, while partial strategyproofness is by design weaker than strategyproofness, it implies many other incentive concepts that have been considered previously; among these are *weak strategyproofness* (Bogomolnaia and Moulin, 2001), *convex strategyproofness* (Balbuzanov, 2015), *approximate strategyproofness* (Carroll, 2013), and *lexicographic strategyproofness* (Cho, 2012), as well as *strategyproofness in the large* (if r converges to 1 in large markets) (Azevedo and Budish, 2015). At the same time, mechanisms that are perceived as having better incentive properties (e.g., because they satisfy one of these concepts) also satisfy partial strategyproofness, such as the Probabilistic Serial mechanism (Bogomolnaia and Moulin, 2001), the adaptive variant of the Boston mechanism (Mennle and Seuken, 2015*d*), and hybrid mechanisms (Mennle and Seuken, 2015*a*).

On the technical level, we show that partial strategyproofness has an alternative definition via a dominance concept. This makes the requirement algorithmically verifiable (for any finite setting) and allows the computation of the degree of strategyproofness measure r. Moreover, partial strategyproofness yields an interesting version of local sufficiency that unifies two prior local sufficiency results by Carroll (2012) and Cho (2012). Finally, we show how the partial strategyproofness concept can also be extended to deterministic mechanisms if agents are uncertain about the preference reports of the other agents.

By providing a useful relaxation of strategyproofness, this first paper of this thesis lays the foundation for the second and the third paper. There, we apply partial strategyproofness to study the incentive properties of school choice mechanisms and hybrid mechanisms, respectively.

1.3.2 Trade-offs in School Choice

The literature on the economics of school choice centers mainly around two mechanisms: the Boston mechanism and the Deferred Acceptance mechanism. One of the dilemmas that parents face under the Boston mechanism is that in the application process they may apply to schools that have already filled up in previous rounds. An application to such an "exhausted" school will definitely be rejected, but in the same round the seats at other schools will be further depleted by other applicants. Since any such application is futile, parents may as well remove any schools from their preference lists that are exhausted by the time they arrive at these schools in the application process. This straightforward manipulation will only improve their chances at other schools. A second variant of the Boston mechanism eliminates the need to manipulate in this fashion: under the *adaptive Boston mechanism*, parents automatically skip exhausted schools in the application process. This mechanism has been largely overlooked by researchers, but it is nonetheless in frequent use, e.g., in the German state of Nordrhein-Westfalen, in Amsterdam (until 2014, see (de Haan et al., 2015)), and in Freiburg (Germany).

In the second paper of this thesis, we study this adaptive Boston mechanism and compare it to the classical Boston mechanism and the Deferred Acceptance mechanism. When priorities are coarse, as they often are in school choice markets, and ties are broken randomly, we show that the adaptive Boston mechanism is partially strategyproof. In contrast, the classical Boston mechanism is merely upper invariant but not partially strategyproof, while the Deferred Acceptance mechanism is strategyproof. This hierarchical relationship is noteworthy because the superior incentive properties of the adaptive Boston mechanism do not surface when comparing the two Boston mechanisms by their vulnerability to manipulation (Pathak and Sönmez, 2013). Furthermore, we establish a second hierarchy in terms of economic efficiency that points in the opposite direction: whenever the outcomes of Deferred Acceptance and the classical Boston mechanism are comparable by the rank dominance relation, then the outcome of the Boston mechanism is (weakly) preferable. Surprisingly, the same comparison fails for the adaptive Boston mechanism with either Deferred Acceptance or the classical Boston mechanism. However, we recover the hierarchy by means of limit results and simulations.

The take-home message for mechanism designers is that choosing between the three mechanisms in school choice markets remains a question of trading off strategyproofness and efficiency. We do not advocate the use of either of the three mechanism. Instead, our insights allow mechanism designers to make a conscious and informed decision about the necessary and possible trade-offs.

1.3.3 Hybrid Assignment Mechanisms

The three mechanisms from the previous paper represent discrete points in the space of possible mechanisms. In contrast, in the third paper of this thesis, we present a constructive approach to construct new assignment mechanisms that *smoothly* trade off strategyproofness and efficiency. The basic idea of *hybrid mechanisms* is very simple: suppose that a mechanism designer has two different mechanisms to choose from. She can choose one or the other, or she can toss a coin to decide randomly which of the two mechanisms to use. If the agents do not know which mechanism will be used, any strategic misreporting on their part must take this uncertainty into account. Formally, hybrid mechanisms are simply convex combinations of two mechanisms, where the *mixing factor* β corresponds to the probability of using the second mechanism. If the first mechanisms is strategyproof and the second is more efficient, then intuition suggests that any hybrid of the two mechanisms inherits a share of the attractive properties from both mechanisms.

We formalize hybrid mechanisms and study the trade-off between strategyproofness and efficiency that can be achieved by this construction. To assess performance on the strategyproofness dimension, we employ the partial strategyproofness concept, and for the efficiency dimension, we use the standard notions of ordinal and rank dominance. We show that hybrid mechanisms enable a smooth trade-off between strategyproofness and efficiency as long as the pair of mechanisms satisfies a set of three conditions, which we call *hybrid-admissibility*. If the pair is hybrid-admissible, then any (arbitrarily small) relaxation of the strategyproofness requirement allows the inclusion of a non-trivial share of the more efficient mechanism in the hybrid. Conversely, we show that this guarantee is lost if we drop either of the three conditions that constitute hybrid-admissibility.

Applying this result, we show that Random Serial Dictatorship can be paired with the Probabilistic Serial mechanism or the adaptive Boston mechanism to obtain efficiency gains from relaxing strategyproofness. At the same time, pairs of Random Serial Dictatorship with either the classical Boston mechanism or rank efficient mechanisms violate hybrid-admissibility. In fact, their hybrids only yield a degenerate trade-off in the sense that including any non-trivial share of the more efficient mechanism (i.e., using any mixing factor $\beta > 0$) immediately causes the hybrid to have a degree of strategyproofness of 0.

These findings illustrate that the trade-offs obtainable from hybrid mechanisms, although intuitive, are by no means trivial. However, with hybrid-admissibility in place, we can construct interesting new assignment mechanisms with intermediate performance

that facilitate smooth trade-offs between strategyproofness and efficiency.

1.3.4 The Pareto Frontier

The fourth paper of this thesis is different in three ways: first, while the other papers specifically consider the assignment problem, the fourth paper studies *ordinal mechanisms* in general. Second, rather than focusing on economic efficiency, we consider a *wide range of desiderata* that may be in conflict with strategyproofness (including fairness and stability). Third, we consider mechanisms that make *optimal trade-offs* between strategyproofness and these desiderata.

To assess the performance of mechanisms with respect to strategyproofness and another desideratum, we introduce a new framework. This framework consists of two measures: the first is a measure for the *manipulability* of a mechanism, which is based on the approximate strategyproofness concept. Approximate strategyproofness captures the intuition that even though an agent may find a beneficial manipulation, the gain from this manipulation is "small." If agents incur some cost (e.g., moral or cognitive) when misreporting, then this cost can outweigh the benefit and thus promote truthful reporting. The second measure of our framework expresses how badly a mechanism fails at achieving a given desideratum: first, we define *welfare functions*, which quantify the value of selecting a given alternative when the agents have a given set of preferences. We use these to define the *welfare deficit* of a mechanism. A welfare deficit of 0 means that a mechanism satisfies the desideratum perfectly, while any higher deficits correspond to a stronger violation of the desideratum.

Equipped with these two measures, we can compare mechanisms by their manipulability and deficit: the lower the value of either measure, the more attractive the mechanism. In particular, we can study *optimal mechanisms* which are those mechanisms that have the lowest possible deficit, subject to a given bound on manipulability. These mechanisms form the *Pareto frontier* in the sense that their deficit cannot be reduced without accepting higher manipulability. Naturally, a mechanism designer would want to consider this Pareto frontier when contemplating trade-offs.

The main result of our fourth paper is a structural characterization of the Pareto frontier. We show that it can be described in terms of two building blocks: (1) we identify a finite set of *supporting manipulability bounds* B and the mechanisms that are optimal at each of them; (2) for all other bounds not in B, we show that mechanisms are optimal at these bounds *if and only if* they are hybrids of two mechanisms that are optimal at the two adjacent supporting manipulability bounds from B. We exploit this structure to

develop an algorithm that computes optimal mechanisms along the entire Pareto frontier. Our results unlock the Pareto frontier to further analytic, axiomatic, and algorithmic explorations.

1.4 Conclusion

Markets without money present an interesting and complex challenge for mechanism designers because severe impossibility results make strategyproofness an extremely restrictive requirement. In the presence of other desiderata, the question arises whether strategyproofness should be taken as an indispensable requirement, or whether we slightly can relax strategyproofness in return for better performance of our mechanisms on other dimensions. In this thesis, we have explored relaxations of strategyproofness for assignment mechanisms in particular and for ordinal mechanisms in general. Our results yield a better understanding of the incentive properties of non-strategyproof ordinal mechanisms, and they provide the conceptual and algorithmic means to trade off strategyproofness and other desiderata in the design of new mechanisms.

2 Partial Strategyproofness: An Axiomatic Approach to Relaxing Strategyproofness for Assignment Mechanisms

2.1 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 settings with transferable utility. Since the seminal paper of Hylland and Zeckhauser (1979), the assignment problem has attracted much attention from mechanism designers (e.g., Abdulkadiroğlu and Sönmez (1998); Bogomolnaia and Moulin (2001); Abdulkadiroğlu and Sönmez (2003*b*)). In practice, such problems often arise in situations that are of great importance to peoples' lives. For example, we must assign seats at public schools, positions in training programs, or accommodation in subsidized housing.

As mechanism designers, we care specifically about *efficiency*, *fairness*, and *strategyproofness*. Strategyproofness is the "gold standard" among incentive concepts. However, it is also often in conflict with other design objectives: Zhou (1990) showed that, unfortunately, it is impossible to achieve the optimum on all three dimensions simultaneously, which makes the assignment problem an interesting mechanism design challenge. 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, 2014). The more demanding *ordinal efficiency* is achieved by the Probabilistic Serial mechanism, but any mechanism that guarantees ordinal efficiency and symmetry will not be strategyproof (Bogomolnaia and Moulin, 2001). Finally, *rank efficiency*, an even stronger efficiency concept, can

be achieved via Rank Value mechanisms (Featherstone, 2011), but it is incompatible with strategyproofness even without additional fairness requirements. The fact that strategyproofness is in conflict with many desirable design objectives explains why market designers are interested in studying non-strategyproof mechanisms: to understand how to make useful trade-offs between different design objectives.

In practice, non-strategyproof mechanisms are ubiquitous. When assigning seats at public schools, it is frequently an explicit objective of administrators to assign as many students as possible to their top-1 or top-3 choices (Basteck, Huesmann and Nax, 2015). The *Boston mechanism* is often used for this assignment and when students report their preferences truthfully, it intuitively fares well with respect to this objective. However, this mechanism is known to be highly manipulable by strategic students. A second example is the Teach for America program which used a mechanism that aimed at rank efficiency when assigning new teachers to positions at different schools. While this mechanism was manipulable, the organizers were confident that the majority of preferences were reported truthfully because participants lacked the information that is necessary to determine beneficial misreports (Featherstone, 2011). For the allocation of courses at Harvard Business School, Budish and Cantillon (2012) demonstrated that the strategyproof Random Serial Dictatorship mechanism would lead to very unbalanced outcomes and that the non-strategyproof HBS Draft mechanism yields preferable results despite its manipulability.

The incompatibility of strategyproofness with other design objectives, such as ordinal or rank efficiency, and the fact that the mechanisms used in practice are frequently not strategyproof shows the need to study non-strategyproof mechanisms; and researchers have been calling for useful relaxed notions of strategyproofness for this purpose (e.g., Azevedo and Budish (2015); Budish (2012)). In this paper, we take an axiomatic approach to this research question and present *partial strategyproofness*, a relaxed notion of strategyproofness that exploits the structure of the assignment problem.

2.1.1 A Motivating Example

To obtain an intuition about the partial strategyproofness concept, consider a setting in which agents $1, 2, 3$ compete for objects a, b, c and their preferences are

$$\begin{aligned} P_1 &: a \succ b \succ c, \\ P_2 &: b \succ a \succ c, \\ P_3 &: b \succ c \succ a. \end{aligned}$$

Suppose that the non-strategyproof Probabilistic Serial mechanism (Bogomolnaia and Moulin, 2001) is used to assign the objects and that agents 2 and 3 report truthfully.[1] By reporting P_1 truthfully, agent 1 receives a, b, c with probabilities $(3/4, 0, 1/4)$, respectively. If instead agent 1 reports $P_1' : b \succ a \succ c$, these probabilities change to $(1/2, 1/3, 1/6)$. Now suppose that agent 1 has value 0 for its last choice c and higher values for the objects a and b. Whether or not the misreport P_1' increases agent 1's *expected utility* depends on its relative value for a over b: if $u_1(a)$ is close to $u_1(b)$, then agent 1 will find it beneficial to report P_1'. If $u_1(a)$ is significantly larger than $u_1(b)$, then agent 1 will want to report truthfully. Precisely, the manipulation is not beneficial if $\left(\frac{3}{4} - \frac{1}{2}\right) \cdot u_1(a) \geqslant \left(\frac{1}{3} - 0\right) \cdot u_1(b)$ or, equivalently, if $\frac{3}{4} \cdot u_1(a) \geqslant u_1(b)$. We observe that agent 1's incentive to manipulate hinges on its "degree of indifference" between objects a and b: the closer agent 1 is to being indifferent between a and b, the higher the incentive to misreport.

Partial strategyproofness captures this intuition by providing a parameter r. This value controls how close to indifferent agents may be between different objects but still have a dominant strategy to report their preferences truthfully. In the above example, $r = 3/4$ is the pivotal degree of indifference between a and b for agent 1. In fact, we will later see that the Probabilistic Serial mechanism is precisely 3/4-*partially strategyproof* in this situation. This means that it makes truthful reporting a dominant strategy for *any* agent whose degree of indifference between any two objects is at most 3/4.

2.1.2 One Familiar and Two New Axioms

In this paper, we first provide a decomposition of strategyproofness into three axioms to then arrive at our new partial strategyproofness concept by relaxing one of the axioms. To understand the three axioms, suppose an agent considers swapping two consecutive objects in its report, e.g., as agent 1 in the above example, from $P_1 : a \succ b \succ c$ to $P_1' : b \succ a \succ c$. Our axioms restrict the way in which the mechanism can react to this kind of change of report. The first axiom, *swap monotonicity*, requires that either the agent's assignment remains unchanged, or its probability for b must strictly increase and its probability for a must strictly decrease. This means that the mechanism is responsive to the agent's ranking of a and b and that the swap affects at least the objects a and b, if any. The second axiom, *upper invariance*, requires that the agent's probabilities for all objects that it strictly prefers to a do not change under the swap. This essentially means that the mechanism is *robust to manipulation by truncation*: falsely claiming

[1] For details on the Probabilistic Serial mechanism and the Simultaneous Eating algorithm see Appendix 2.A.

higher preference for an outside option does not improve the agent's assignment. Upper invariance was introduced by Hashimoto et al. (2014) (called *weak invariance* in their paper) as one of the axioms to characterize the Probabilistic Serial mechanism. Finally, we introduce *lower invariance* which requires that the agent's probabilities do not change for any object that it ranks below b. Our first main result is that assignment mechanisms are strategyproof if and only if they satisfy all three axioms.

2.1.3 Bounded Indifference and Partial Strategyproofness

Arguably, lower invariance is the least important one of the three axioms. Indeed, as we will show later, many non-strategyproof mechanisms that are viewed as having "good" incentive properties violate only lower invariance but satisfy swap monotonicity and upper invariance. To understand the incentives under mechanisms that are swap monotonic and upper invariant we define a new relaxed notion of strategyproofness: following the intuition from the motivating example, we require mechanisms to make truthful reporting a dominant strategy but on a restricted domain where agents have sufficiently different values for different objects. This domain restriction can be formalized as follows: a utility function satisfies *uniformly relatively bounded indifference* with respect to *bound* $r \in [0, 1]$ (URBI(r) for short) if, given $u(a) > u(b)$, the agent's (normalized) value for b is at least a factor r lower than its value for a (i.e., $r \cdot u(a) \geqslant u(b)$). We say that a mechanism is *r-partially strategyproof* if the mechanism makes truthful reporting a dominant strategy for any agent whose utility function satisfies URBI(r). Our second main result is the following equivalence: for any setting (i.e., number of agents, number of objects, and object capacities) a mechanism is swap monotonic and upper invariant if and only if it is r-partially strategyproof for some $r > 0$. Thus, partial strategyproofness is axiomatically motivated. Furthermore, it allows us to give honest and useful strategic advice to the agents: under any swap monotonic, upper invariant mechanism agents are best off by reporting truthfully as long as they are not too close to indifferent between different objects.

2.1.4 Maximality of URBI(r) and the Degree of Strategyproofness

Knowing that a given mechanism is r-partial strategyproofness yields a guarantee about the set of utility functions for which the mechanism will make truthful reporting a dominant strategy, namely all those utility functions that satisfy the URBI(r) constraint. However, no guarantee is given for utility functions that violate this constraint. For

our third main result, we show that the URBI(r) domain restriction is in fact *maximal* for r-partially strategyproof mechanisms. Specifically, there exists no larger set of utility functions for which guarantees of this form can be given without any additional information about the agents or the mechanisms. In this sense, r-partial strategyproofness as defined above is the strongest statement that we can make about the incentive properties of swap monotonic and upper invariant mechanisms.

By virtue of this maximality, partial strategyproofness induces a meaningful parametric measure for the incentive properties of non-strategyproof mechanisms: for a given mechanism we define its *degree of strategyproofness* ρ as the largest value of the bound r such that the mechanism remains r-partially strategyproof. We show that comparing mechanisms by their degree of strategyproofness is consistent with (but not equivalent to) the method for comparing mechanisms by their *vulnerability to manipulation* (Pathak and Sönmez, 2013). However, the degree of strategyproofness measure has two advantages: it is a *parametric* measure for incentive guarantees, while vulnerability to manipulation is a binary comparison; and it is algorithmically *computable*, while no algorithm is known to perform the comparison by vulnerability to manipulation. This makes the degree of strategyproofness a compelling new method to measure and compare the incentive properties of non-strategyproof mechanisms.

2.1.5 Properties of Partial Strategyproofness

With partial strategyproofness, we have defined a new, relaxed notion of strategyproofness for assignment mechanisms. The concept has a number of appealing properties which we formalize, prove, and discuss in this paper.

Dominance Interpretation of Partial Strategyproofness

When assignments are random, the agents' preferences must be extended to lotteries in some way. This is typically done via dominance notions, such as *(first order-)stochastic dominance (SD)* or *lexicographic dominance (DL)*. Any dominance notion in turn induces a strategyproofness notion that arises by requiring the outcomes from truthful reporting to dominate the outcomes from misreporting. For stochastic dominance, the resulting *SD-strategyproofness* is in fact equivalent to the definition of *strategyproofness* we use in this paper (which requires truthful reporting to maximize any agent's expected utility).

The question arises whether partial strategyproofness has an alternative definition in terms of a dominance relation. We present the notion of *r-partial dominance (r-PD)* which

is similar to stochastic dominance except that the influence of changes in the assignment of less preferred objects is discounted by the factor r. Our fourth main result is that r-partial strategyproofness is equivalent to r-PD-strategyproofness, the incentive concept induced by r-partial dominance. This equivalence has two important consequences: first, it allows an alternative definition of partial strategyproofness that is independent of the agents' utility functions. Thus, partial strategyproofness integrates nicely with other incentive concepts that are defined via dominance notions such as strong and weak SD-strategyproofness, and DL-strategyproofness. Second, the dominance interpretation provides an equivalent condition in terms of finitely many linear constraints. This makes partial strategyproofness algorithmically verifiable and enables the computation of the degree of strategyproofness measure.

Intermediateness of the Partial Strategyproofness Concept

We also study the relationship of partial strategyproofness and other incentive concepts, and we establish that it provides a unified view on the incentive properties of non-strategyproof assignment mechanisms: while partial strategyproofness is a weaker requirement than *strategyproofness*, it in turn implies many relaxed incentive concepts that have been proposed previously, namely *weak*, *convex*, and *approximate strategyproofness*, as well as *strategyproofness in the large* (if the degree of strategyproofness converges to 1). Moreover, we prove out fifth main result, the following equivalence: a mechanism is r-partially strategyproof for some $r > 0$ if and only if it is strategyproof for agents with lexicographic preferences (i.e., DL-strategyproof). Thus, the "upper" limit case of r-partial strategyproofness for $r = 1$ corresponds to strategyproofness, while the "lower" limit case for $r \to 0$ corresponds to DL-strategyproofness. In this sense, the degree of strategyproofness parametrizes the whole space of mechanisms between those that are strategyproof on the one side and those that are merely DL-strategyproof on the other side.

Local Sufficiency for Partial Strategyproofness

A *local* misreport is a swap of two consecutive objects in the reported preference order of an agent. For an incentive concept, such as strategyproofness or DL-strategyproofness, *local sufficiency* holds if it suffices to check only the local misreports in order to verify that a given mechanism satisfies the incentive concept. Local sufficiency is interesting from an axiomatic as well as from an algorithmic perspective: for instance, the axioms swap monotonicity, upper invariance, and lower invariance are based on swaps, which

makes them simple and accessible. Furthermore, the naïve way to verify algorithmically that a given mechanism satisfies a given incentive concept would be to iterate through all incentive constraints. This is typically a large number. If local sufficiency holds, it is enough to check only those constraints that arise from swaps.

For strategyproofness and DL-strategyproofness, local sufficiency was proven by Carroll (2012) and Cho (2012), respectively. Thus, local sufficiency holds for the two limit concepts of partial strategyproofness. This raises the question whether it also holds for partial strategyproofness. Our sixth main result is that r-local partial strategyproofness implies r^2-partial strategyproofness. Furthermore, the bound "2" is tight in the sense that there exists no $\epsilon > 0$ such that $r^{2-\epsilon}$-partial strategyproofness can also be guaranteed. This insight connects the prior local sufficiency results for strategyproofness and for DL-strategyproofness as it provides a unified proof for both results.

Applications to Deterministic Mechanisms

Partial strategyproofness crucially depends on the randomness of mechanisms. However, not all interesting assignment mechanisms involve randomization; other mechanisms may involve randomization but this may be insufficient to allow a straightforward application of partial strategyproofness. Nonetheless, partial strategyproofness can also be applied to study the incentive properties of non-strategyproof *deterministic* mechanisms and other mechanisms that are not "random enough." To this end, we consider a second source of randomness, namely the agents' uncertainty about the reports from other agents. Specifically, for our seventh main result, we give an axiomatic characterization of the mechanisms that are partially strategyproof for agents who are unsure about the preference reports from the other agents. In particular, these axioms are satsfied by the deterministic versions of the naïve and the adaptive Boston mechanisms (even if priorities are strict and fixed), as well as the HBS Draft mechanism and the Probabilistic Serial mechanism for multi-unit assignment.

Applications of Partial Strategyproofness

Finally, we demonstrate that our partial strategyproofness concept yields new insights about the incentive properties of many popular, non-strategyproof assignment mechanisms. First, the Probabilistic Serial mechanism is partially strategyproof which provides a more accurate understanding of its incentive properties than, e.g., weak SD-strategyproofness. Numerically, we show that the degree of strategyproofness of PS increases in larger settings, which is in line with (but not implied by) prior findings by Kojima and Manea (2010).

Second, the traditional "naïve" Boston mechanism[2] is known to be highly manipulable, and it is not partially strategyproof. However, a variant where agents automatically skip exhausted objects in the application process is in fact partially strategyproof. The degree of strategyproofness of this "adaptive" Boston mechanism is lower than that of PS and, therefore, it has *intermediate* incentive guarantees. Third, in (Mennle and Seuken, 2015*a*) we have introduced *hybrids* which are convex combinations of two mechanisms. Under certain technical conditions, we have shown that hybrids facilitate a trade-off between strategyproofness and efficiency where we use partial strategyproofness to quantify the incentive properties. Prior to the introduction of partial strategyproofness, no concept existed to study such trade-offs. These examples highlight that partial strategyproofness captures our intuitive understanding of what it means for a non-strategyproof mechanism to have "good" incentive properties.

In summary, our axiomatic treatment of the strategyproofness concept leads to a new way of thinking about how to relax strategyproofness for assignment mechanisms. The resulting partial strategyproofness concept is simple, tight, parametric, it integrates well with existing methods, and it differentiates nicely between many popular mechanisms.

Organization of this paper: In Section 2.2, we discuss related work. In Sections 2.3 and 2.4, we introduce our formal model and the three axioms. In Section 2.5, we present our axiomatic decompostion of strategyproofness, and in Section 2.6, we derive the new partial strategyproofness concept. In Section 2.7, we present our maximality result and the degree of strategyproofness measure. In Section 2.8, we give the dominance interpretation of partial strategyproofness. In Section 2.9, we compare partial strategyproofness to other incentive concepts, and in Section 2.10, we discuss local sufficiency. In Section 2.11, we extend partial strategyproofness to deterministic mechanisms. In Section 2.12, we apply our new concept to popular assignment mechanisms, and we conclude in Section 2.13.

2.2 Related Work

While the seminal paper on assignment mechanisms by Hylland and Zeckhauser (1979) proposed a mechanism that elicits agents' *cardinal* utilities, this approach has proven problematic because it is difficult if not impossible to elicit cardinal utilities in settings without money. For this reason, recent work has focused on *ordinal mechanisms* where

[2]We consider the Boston mechanism with no priorities and single uniform tie-breaking.

agents submit preference orders over objects. In fact, it has been shown that mechanisms whose outcomes have to be independent of the agents' levels of wealth are bound to be ordinal (Huesmann and Wambach, 2015; Ehlers et al., 2015). Throughout this paper, we only consider ordinal mechanisms.

For the case of deterministic assignment mechanisms, strategyproofness has been studied extensively. Pápai (2000) showed that the only group-strategyproof, ex-post efficient, reallocation-proof mechanisms are hierarchical exchanges. Characterizations of strategyproof, efficient, and reallocation-consistent (Ehlers and Klaus, 2006) or consistent (Ehlers and Klaus, 2007) mechanisms are also available. The only deterministic, strategyproof, ex-post efficient, non-bossy, and neutral mechanisms are known to be serial dictatorships (Hatfield, 2009). Furthermore, Pycia and Ünver (2014) showed that all group-strategyproof, ex-post efficient mechanisms are trading cycles mechanisms. Barbera, Berga and Moreno (2012) gave a decomposition of strategyproofness that is similar in spirit to ours but restricted to deterministic social choice domains.

For random social choice rules, Gibbard (1977) gave a decomposition of strategyproofness into the two properties *localized* and *non-perverse*. This (as well as any other) decomposition of strategyproofness is by definition equivalent to our decomposition. Here our contribution lies in the definition of simple and intuitive axioms that make the conditions accessible and straightforward to interpret. For random assignment mechanisms, Abdulkadiroğlu and Sönmez (1998) showed that Random Serial Dictatorship (RSD) is equivalent to the Core from Random Endowments mechanism for house allocation. Bade (2014) extended their result by showing that taking any ex-post efficient, strategyproof, non-bossy, deterministic mechanism and assigning agents to roles in the mechanism uniformly at random is equivalent to RSD. However, it is still an open conjecture whether RSD is the unique mechanism that is strategyproof, ex-post efficient, and anonymous (Lee and Sethuraman, 2011; Bade, 2014).

Besides the baseline requirement of ex-post efficiency, the research community has also introduced more demanding efficiency concepts, such as ordinal efficiency, which is achieved by the Probabilistic Serial (PS) mechanism (Bogomolnaia and Moulin, 2001). The PS mechanism has received considerable attention from researchers: Hashimoto et al. (2014) showed that PS with uniform eating speeds is in fact the unique mechanism that is ordinally fair and non-wasteful. Bogomolnaia and Moulin (2001) had already shown that PS is not strategyproof but Ekici and Kesten (2012) found that its Nash equilibria can lead to ordinally dominated outcomes. Incentive concerns for PS may be severe for small settings but they get less problematic for larger settings: Kojima and Manea

(2010) showed that for a fixed number of object types and a fixed agent PS makes it a dominant strategy for that agent to be truthful if the number of copies of each object is sufficiently large.

While ex-post efficiency and ordinal efficiency are the most well-studied efficiency concepts for assignment mechanisms, some mechanisms used in practice aim to achieve *rank efficiency* which is a further refinement of ordinal efficiency (Featherstone, 2011). However, no rank efficient mechanism can be strategyproof in general. Other popular mechanisms, like the Boston Mechanism (Ergin and Sönmez, 2006; Miralles, 2008), are highly manipulable but nevertheless in frequent use. Budish and Cantillon (2012) found practical evidence from combinatorial course allocation which suggests that using a non-strategyproof mechanism may lead to higher social welfare than using an ex-post efficient and strategyproof mechanism, such as RSD. The fact that strategyproofness is in conflict with many other design objectives challenges whether it should be taken as an indispensable requirement in mechanism design.

Given that strategyproofness is such a strong restriction, many researchers have tried to relax it. Bogomolnaia and Moulin (2001) used *weak SD-strategyproofness* to describe the incentive properties under PS and Balbuzanov (2015) showed that PS in fact satisfies the more demanding *convex strategyproofness*. Carroll (2013) adapted *approximate strategyproofness* for bounded utilities to quantify agents' incentives to manipulate in the voting domain. Azevedo and Budish (2015) proposed a desideratum called *strategyproof in the large* (SP-L) which formalizes the intuition that as the number of agents in the market gets large the incentives for an individual agent to misreport its preference order should vanish in the limit. Finally, Cho (2012) considered strategyproofness for agents with lexicographic preferences (DL-strategyproofness). We show that partial strategyproofness unifies these relaxations of strategyproofness: on the one hand, many non-strategyproof mechanisms that are generally viewed as "having better incentive properties," because they satisfy these various notions of strategyproofness, turn out to satisfy partial strategyproofness as well, such as Probabilistic Serial, a variant of the Boston Mechanism, and newly defined hybrid mechanisms. On the other hand, partial strategyproofness implies the other notions. Pathak and Sönmez (2013) introduced a general method to compare different mechanisms by their vulnerability to manipulation. The degree of strategyproofness measure we propose in this paper is consistent with (but not equivalent to) this method. However, our concept has two advantages: it is parametric and it is computable. We discuss the connection in more detail in Section 2.7.2.

Local sufficiency is a property of an incentive concept which holds when the absence of local manipulation opportunities implies the absence of manipulation opportunities "globally." Carroll (2012) and Cho (2012) showed that local sufficiency holds for strategyproofness and DL-strategyproofness, respectively. We prove a local sufficiency result for partial strategyproofness which bridges the gap between (and provides a unified proof for) both of these prior results.

2.3 Model

A *setting* $(N, M, \boldsymbol{q})$ consists of a set N of n *agents*, a set M of m *objects*, and a vector $\boldsymbol{q} = (q_1, \ldots, q_m)$ of *capacities* (i.e., there are q_j units of object j available). We assume that supply satisfies demand (i.e., $n \leqslant \sum_{j \in M} q_j$), since we can always add a dummy object with capacity n. Agents $i \in N$ have strict preference orders P_i over objects, where $P_i : a \succ b$ means that agent i prefers object a to object b. The set of all preference orders is denoted by $\mathcal{P}$. A *preference profile* $\boldsymbol{P} = (P_1, \ldots, P_n) \in \mathcal{P}^N$ is a collection of preference orders of all agents, and we denote by $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$ the collection of preference orders of all agents, except i. Agents' preferences are extended to lotteries via *von Neumann-Morgenstern (vNM) utilities* $u_i : M \to \mathbb{R}^+$. The utility function u_i is *consistent* with preference order P_i (denoted $u_i \sim P_i$) if $P_i : a \succ b$ whenever $u_i(a) > u_i(b)$. We denote by $U_{P_i} = \{u_i : u_i \sim P_i\}$ the set of all utility functions that are consistent with P_i.

A *(random) assignment* is a matching of objects to agents. It is represented by an $n \times m$-matrix $x = (x_{i,j})_{i \in N, j \in M}$ satisfying the *fulfillment constraint* $\sum_{j \in M} x_{i,j} = 1$, the *capacity constraint* $\sum_{i \in N} x_{i,j} \leqslant q_j$, and the *probability constraint* $x_{i,j} \in [0, 1]$ for all $i \in N, j \in M$. It is represented by an $n \times m$-matrix $x = (x_{i,j})_{i \in N, j \in M}$ satisfying the *fulfillment constraint* $\sum_{j \in M} x_{i,j} = 1$, the *capacity constraint* $\sum_{i \in N} x_{i,j} \leqslant q_j$, and the *probability constraint* $x_{i,j} \in [0, 1]$ for all $i \in N, j \in M$. The entry $x_{i,j}$ of the matrix x is the probability that agent i gets object j. An assignment is *deterministic* if all agents get exactly one full object (i.e., $x_{i,j} \in \{0, 1\}$ for all $i \in N, j \in M$). For any agent i, the ith row $x_i = (x_{i,j})_{j \in M}$ of the matrix x is called the *assignment vector of* i (or *i's assignment* for short). The Birkhoff-von Neumann Theorem and its extensions (Budish et al., 2013) ensure that, given any random assignment, we can always find a lottery over deterministic assignments that implements these marginal probabilities. Finally, let X and $\Delta(X)$ denote the spaces of all deterministic and random assignments, respectively.

A *(random) mechanism* is a mapping $\varphi : \mathcal{P}^N \to \Delta(X)$ that chooses an assignment based on a profile of reported preference orders. $\varphi_i(P_i, P_{-i})$ is the assignment vector

that agent i receives if it reports P_i and the other agents report P_{-i}. A mechanism is *deterministic* if if only selects deterministic assignments (i.e., $\varphi : \mathcal{P}^N \to X$). Note that mechanisms only receive preference profiles as input (i.e., only agents' preference orders) but no additional cardinal information. Thus, we consider *ordinal* mechanisms, where the assignment is independent of the actual vNM utilities. The expected utility for i is given by the dot product

$$\langle u_i, \varphi_i(P_i, P_{-i}) \rangle = \mathbb{E}_{\varphi_i(P_i, P_{-i})}[u_i] = \sum_{j \in M} u_i(j) \cdot \varphi_{i,j}(P_i, P_{-i}). \tag{1}$$

2.4 The Axioms

In this section, we introduce the axioms that we use to decompose and relax strategyproofness. To do so formally, we require the auxiliary concepts of *neighborhoods* and *contour sets*.

Definition 1 (Neighborhood)**.** For any two preference orders $P, P' \in \mathcal{P}$ we say that P and P' are *adjacent* if they differ by just a swap of two consecutive object; formally,

$$\begin{aligned} P &: a_1 \succ \ldots \succ a_k \succ a_{k+1} \succ \ldots \succ a_m, \\ P' &: a_1 \succ \ldots \succ a_{k+1} \succ a_k \succ \ldots \succ a_m. \end{aligned}$$

The set of all preference orders adjacent to P is the *neighborhood of* P, denoted N_P.

Definition 2 (Upper and Lower Contour Sets)**.** For a preference order $P \in \mathcal{P}$ with $P : a_1 \succ \ldots \succ a_k \succ \ldots \succ a_m$, the *upper contour set* $U(a_k, P)$ and the *lower contour set* $L(a_k, P)$ *of* a_k *at* P are the sets of objects that an agent with preference order P strictly prefers or likes strictly less than a_k, respectively; formally,

$$U(a_k, P) = \{a_1, \ldots, a_{k-1}\} = \{j \in M \mid P : j \succ a_k\}, \tag{2}$$

$$L(a_k, P) = \{a_{k+1}, \ldots, a_m\} = \{j \in M \mid P : a_k \succ j\}. \tag{3}$$

Swapping two consecutive objects in the true preference order (or equivalently, reporting a preference order from the neighborhood of the true preference order) is a basic manipulation that an agent could consider. Our axioms limit the way in which a mechanism can change the assignment of the reporting agent under this basic manipulation.

Axiom 1 (Swap Monotonic)**.** *A mechanism φ is* swap monotonic *if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$, one of the following holds:*

- *either $\varphi_i(P_i, P_{-i}) = \varphi_i(P_i', P_{-i})$,*
- *or $\varphi_{i,a_k}(P_i, P_{-i}) > \varphi_{i,a_k}(P_i', P_{-i})$*
  and $\varphi_{i,a_{k+1}}(P_i, P_{-i}) < \varphi_{i,a_{k+1}}(P_i', P_{-i})$.

Swap monotonicity is an intuitive axiom because it simply requires the mechanism to react to the swap in a *direct* and *responsive* way: the swap reveals information about the agent's relative ranking of a_k and a_{k+1}. Thus, if anything changes about the assignment for that agent, the probabilities for the objects a_k and a_{k+1} must be affected *directly*. In addition, the mechanism must *respond* to the agent's preferences by assigning more probability for the object the agent claims to like more and less probability for the object the agent claims to like less.

Swap monotonicity 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 strictly 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 first order-stochastic dominance sense, as Example 1 shows.

Example 1. Consider a mechanism where reporting $P : a \succ b \succ c \succ d$ leads to an assignment vector of $(0, 1/2, 0, 1/2)$ of a, b, c, d, respectively, and reporting $P' : a \succ c \succ b \succ d$ leads to $(1/2, 0, 1/2, 0)$. This is consistent with swap monotonicity.[3] However, the latter assignment vector first order-stochastically dominates the former at P.

While the swap monotonic mechanism in Example 1 is manipulable in a first order-stochastic dominance sense, the manipulations involves changes of the probabilities for other objects that the agent prefers to b. To prevent this we need an additional axiom.

Axiom 2 (Upper Invariant)**.** *A mechanism φ is* upper invariant *if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$, we have that i's assignment for objects from the upper contour set of a_k does not change (i.e., $\varphi_{i,j}(P_i, P_{-i}) = \varphi_{i,j}(P_i', P_{-i})$ for all $j \in U(a_k, P_i)$).*

[3] In Appendix 2.B, we give a swap monotonic continuation of this mechanism for all preference orders.

Intuitively, under upper invariance, an agent cannot influence its probabilities for obtaining one of its better choices by swapping two less preferred objects. Upper invariance was introduced by Hashimoto et al. (2014) as one of the central axioms to characterize the Probabilistic Serial mechanism. If a null object is present and the mechanism is individually rational, then upper invariance is equivalent to *truncation robustness*. Truncation robustness is a type of robustness to manipulation that is important in theory and application: it prevents that an agent can increase its chances of being assigned a more preferred object by bringing the null object up in its preference order. Many mechanisms from the literature satisfy upper invariance, including Random Serial Dictatorship, Probabilistic Serial, the Boston mechanism, and Deferred Acceptance (for the proposing agents), and the HBS Draft for multi-unit assignment.

Axiom 3 (Lower Invariant)**.** *A mechanism φ is* lower invariant *if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P'_i \in N_{P_i}$ from the neighborhood of P_i with $P_i : a_k \succ a_{k+1}$ and $P'_i : a_{k+1} \succ a_k$, we have that i's assignment for objects from the lower contour set of a_{k+1} does not change (i.e., $\varphi_{i,j}(P_i, P_{-i}) = \varphi_{i,j}(P'_i, P_{-i})$ for all $j \in L(a_{k+1}, P_i)$).*

Lower invariance complements upper invariance: it requires that an agent cannot influence its probabilities for obtaining any *less* preferred objects by swapping two *more* preferred objects. Lower invariance has a subtle effect on incentives: if agents had upward-lexicographic preferences (Cho, 2012), mechanisms that are not lower invariant will be manipulable for these agents, even if they are swap monotonic and upper invariant. Arguably, lower invariance is the least important axiom, but in Section 2.5, we show that it is exactly the missing link to guarantee strategyproofness. In Section 2.6, we drop lower invariance for our axiomatization of partially strategyproof mechanisms.

2.5 A New Decomposition of Strategyproofness

In this section, we review the strategyproofness concept. Then we present our first main result, an axiomatic decomposition of strategyproofness.

2.5.1 The Strategyproofness Concept

Ideally, a mechanism makes truthful reporting a dominant strategy for all agents. This requirement is formalized by strategyproofness.

Definition 3 (Strategyproof)**.** A mechanism φ is *strategyproof* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any utility function $u_i \in U_{P_i}$ that is consistent with P_i, we have

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_j(P_i', P_{-i}) \rangle \geqslant 0. \tag{4}$$

Alternatively, strategyproofness can be defined via the notion of stochastic dominance. For a preference order $P \in \mathcal{P}$ with $P : a_1 \succ \ldots \succ a_m$ and assignment vectors x and y, we say that x *(first order-)stochastically dominates* y *at* P if for all ranks $k \in \{1, \ldots, m\}$ we have

$$\sum_{l=1}^{k} x_{a_l} \geqslant \sum_{l=1}^{k} y_{a_l}. \tag{5}$$

This means that the probability of obtaining one's kth choice or better is weakly higher under x than under y. Intuitively, an agent with preference order P would unambiguously prefer x to y. The dominance is *strict* if in addition inequality (5) is strict for some k. A mechanism φ is *stochastic dominance-strategyproof (SD-strategyproof)* if truthful reporting always yields a stochastically dominant assignment vector for the reporting agent (i.e., $\varphi_i(P_i, P_{-i})$ stochastically dominates $\varphi_i(P_i', P_{-i})$ at P_i). Strategyproofness in the sense of expected utilities from Definition 3 and SD-strategyproofness are equivalent (Erdil, 2014), and we will simply refer to this requirement as *strategyproofness* for the rest of the paper.

For deterministic mechanisms, strategyproofness and swap monotonicity coincide.

Proposition 1. *A deterministic mechanism φ is strategyproof if and only if it is swap monotonic.*

The proof is straightforward and given in Appendix 2.F.1. Example 1 shows that this equivalence no longer holds for random mechanisms, since the mechanism in the example is swap monotonic but not strategyproof. Our following decomposition result shows which additional axioms are needed for strategyproofness of random mechanisms.

2.5.2 Decomposition Result

We are now ready to formulate our first main result, the decomposition of strategyproofness into the three axioms swap monotonicity, upper invariance, and lower invariance.

Theorem 1. *A mechanism φ is strategyproof if and only if it is swap monotonic, upper invariant, and lower invariant.*

Proof Outline (formal proof in Appendix 2.F.2). Assuming strategyproofness, we consider a swap of two consecutive objects in the report of some agent. Towards contradiction, we suppose that φ violates either upper or lower invariance. We show that this implies that the assignment for the manipulating agent from misreporting is not stochastically dominated by its assignment from truthful reporting, which contradicts strategyproofness. With upper and lower invariance established, swap monotonicity follows as well. For necessity, if φ satisfies the axioms, we show that any swap of two consecutive objects produces an assignment that is stochastically dominated by the assignment from reporting truthfully. Carroll (2012) showed that this *local* strategyproofness is sufficient. □

Theorem 1 illustrates why strategyproofness is so restrictive: if an agent swaps two consecutive objects in its preference order, the only thing that a strategyproof mechanism can do (if anything) is to increase that agent's probability for the object that is swapped up and decrease its probability for the object that is swapped down by the same amount.

In his seminal paper, Gibbard (1977) gave a decomposition of strategyproofness for random social choice rules that is similar in spirit to ours: he showed that any rule is strategyproof if and only if it is *localized* and *non-perverse*. We would like to point out that *any* characterization of the set of strategyproof mechanisms is by its very nature equivalent to any other characterization of this set. Our decomposition in Theorem 1 is appealing because of the choice of axioms, which are simple and straightforward. This makes the decomposition useful, e.g., when proving strategyproofness of new mechanisms, or when encoding strategyproofness as constraints to an optimization problem under the automated mechanism design paradigm (Sandholm, 2003).

Remark 1. Theorem 1 can be extended to the case where indifferences between different objects are possible. In the present paper we focus on a model that rules out indifferences, but we have given the extension in (Mennle and Seuken, 2014).

2.6 An Axiomatic Decomposition of Partial Strategyproofness

In the previous section, we have seen that swap monotonicity, upper invariance, and lower invariance are necessary and sufficient conditions for strategyproofness. Example 1 has shown that swap monotonicity and upper invariance are essential to guarantee at least truncation robustness and the absence of manipulations in a stochastic dominance-sense. Furthermore, we observed that lower invariance is the least intuitive and the

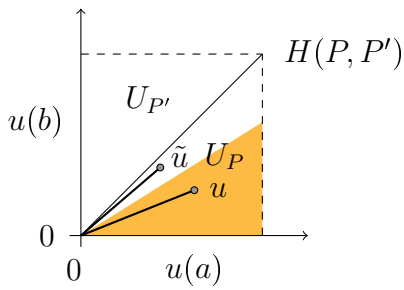

Figure 2.1: Geometric interpretation of uniformly relatively bounded indifference.

least important of the axioms. Obviously, mechanisms that violate lower invariance are not strategyproof. However, we show that such mechanisms still make truthful reporting a dominant strategy for a subset of the utility functions. This leads to a relaxed notion of strategyproofness, which we call *partial strategyproofness*: we show that swap monotonicity and upper invariance are equivalent to strategyproofness on utility functions that satisfy *uniformly relatively bounded indifference*.

2.6.1 Uniformly Relatively Bounded Indifference URBI(r)

Recall the motivating example from the introduction, where an agent was contemplating a misreport under the Probabilistic Serial mechanism. $r = 3/4$ was the pivotal degree of indifference, which determined whether the misreport was beneficial to the agent or not. Generalizing the idea from this example, we introduce the concept of *uniformly relatively bounded indifference*: loosely speaking, an agent must value any object at least a factor r less than the next better object (after appropriate normalization).

Definition 4 (URBI)**.** A utility function u satisfies *uniformly relatively bounded indifference with respect to bound* $r \in [0, 1]$ *(URBI(r))* if for any objects $a, b \in M$ with $u(a) > u(b)$, we have

$$r \cdot (u(a) - \min u) \geqslant u(b) - \min u, \tag{6}$$

where we write $\min u = \min_{j \in M} (u(j))$ for the utility of the last choice.

If $\min_{j \in M} (u(j)) = 0$, uniformly relatively bounded indifference has an intuitive interpretation because inequality (6) simplifies to $r \cdot u(a) \geqslant u(b)$. In words, this means that given a choice between two objects a and b the agents must value b at least a factor r less than a.

For a geometric interpretation of URBI(r), consider Figure 2.1: the condition means that the agent's utility function, represented by the vector u, cannot be arbitrarily close to the indifference hyperplane $H(P, P')$ between the sets of consistent utility functions U_P and $U_{P'}$, but it must lie within the shaded area in U_P. A different utility function in U_P, represented by the vector $\tilde{u}$, would violate URBI(r). For convenience we introduce the convention that URBI(r) denotes the set of all utility functions that satisfy uniformly relatively bounded indifference with respect to r, so that we can write "$u \in$ URBI(r)" to indicate that u satisfies the requirement with respect to bound r.

Remark 2. To gain some intuition about the "size" of the set URBI(r), consider a setting with $m = 3$ objects. Suppose that $\min_{j \in M}(u(j)) = 0$ and that the utilities for the first and second choice are determined by drawing a vector uniformly at random from $(0,1)^2 \backslash H(P, P')$ (i.e., from the open unit square excluding the indifference hyperplane). Then the share of utilities that satisfy URBI(r) is r. For example, if $r = 0.4$, the probability of drawing a utility function from URBI(0.4) is 0.4. In Figure 2.1, this corresponds to the area of the shaded triangle over the area of the larger triangle formed by the x-axis, the diagonal $H(P, P')$, and the vertical dashed line on the right.

2.6.2 Definition of r-Partial Strategyproofness

Using URBI(r), we now define our new relaxed notion of strategyproofness.

Definition 5 (Partially Strategyproof)**.** Given a setting $(N, M, \boldsymbol{q})$ and a bound $r \in [0, 1]$, a mechanism φ is *r-partially strategyproof (in the setting $(N, M, \boldsymbol{q})$)* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any utility function $u_i \in U_{P_i} \cap$ URBI(r) that is consistent with P_i and satisfies URBI(r), we have

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_j(P_i', P_{-i}) \rangle \geqslant 0. \tag{7}$$

When the setting is clear from the context, we simply write *r-partially strategyproof* without explicitly stating the setting. Furthermore, we say that φ is *partially strategyproof* if there exists some non-trivial bound $r > 0$ for which φ is r-partially strategyproof.

Definition 5 of partial strategyproofness is very similar to Definition 3 of strategyproofness; the only difference is that inequality (7) only has to hold for utility functions u_i that satisfy URBI(r). In this sense, r-partial strategyproofness for $r < 1$ is weaker than strategyproofness, but it is equivalent to strategyproofness for $r = 1$. In Section 2.9, we explore the connection of partial strategyproofness to other incentive concepts in detail.

2.6.3 Axiomatic Decomposition Result

In this section, we show our second main result that dropping lower invariance but requiring swap monotonicity and upper invariance leads to the larger class of partially strategyproof mechanisms.

Theorem 2. *Given a setting $(N, M, \mathbf{q})$, a mechanism φ is partially strategyproof (i.e., r-partially strategyproof for some $r > 0$) if and only if φ is swap monotonic and upper invariant.*

Proof outline (formal proof in Appendix 2.F.3). Suppose, an agent i has true preference order $P_i : a_1 \succ \ldots \succ a_K \succ a_{K+1} \succ \ldots \succ a_m$ and is considering a misreport P_i' that leaves the positions of the first K choices unchanged. We first show that under swap monotonicity and upper invariance, it suffices to consider misreports P_i' for which its assignment of a_{K+1} strictly decreases. The key insight comes from considering certain chains of swaps and their impact on the assignment (called *canonical transitions*, see Claim 1). Then we show that for sufficiently small $r \in (0, 1]$, the decrease in expected utility that is caused by the decrease in the assignment of a_{K+1} is sufficient to deter manipulation by any agent whose utility function satisfies URBI(r), even though its assignment for less preferred objects $a_{K+2}, \ldots, a_m$ may improve. Finally, we show that a strictly positive r can be chosen uniformly for all preference profiles and misreports. Thus, the bound r depends only on the mechanism and the setting.

To see necessity, we assume towards contradiction that the mechanism is not upper invariant. For any $r \in (0, 1]$ we construct a utility function that satisfies URBI(r) but for which the mechanism would be manipulable. The key idea is to make the agent want the object from the upper contour set strongly enough to remedy any other (negative) effects of the swap. Finally, using upper invariance, swap monotonicity follows as well. □

Theorem 2 yields an axiomatic motivation for the definition of partial strategyproofness: if we wish to retain truncation robustness and prevent manipulability in a stochastic dominance-sense, the set of partially strategyproof mechanisms arises naturally. It also shows what requiring strategyproofness on top of partial strategyproofness buys, namely lower invariance.

Finally, the equivalence also teaches us what straightforward and honest strategic advice we can give to the agents: if the mechanism is swap monotonic and upper invariant, we can tell agents that they are best off reporting their preferences truthfully as long as their cardinal valuations for different objects are sufficiently different.

Remark 3. In light of the interpretation of upper invariance as robustness to manipulation by truncation, dropping lower invariance suggests itself as the most sensible approach to relaxing strategyproofness. Alternatively, one could consider mechanisms that are swap monotonic and *lower* invariant but violate *upper* invariance. Naturally, this gives rise to a different class of non-strategyproof mechanisms, which is related to *upward* lexicographic-strategyproofness in a similar way in which partial strategyproofness is related to *downward* lexicographic-strategyproofness (see Theorem 5). We leave this approach to future research.

2.7 Maximality of URBI(r) and the Degree of Strategyproofness

r-partial strategyproofness of a mechanism φ requires that agents with a utility function in URBI(r) have a dominant strategy to report their preferences truthfully. However, this does not imply that the set of utility functions for which φ makes truthful reporting a dominant strategy is *exactly equal* to the set URBI(r). The following Example 2 shows that in general we cannot hope for an exact equality.

Example 2. Consider a setting with 4 agents and 4 objects a, b, c, d in unit capacity. In this setting, the adaptive Boston mechanism (ABM) (Mennle and Seuken, 2015*d*) is r-partially strategyproof for any $r \leqslant \frac{1}{3}$ but not r-partially strategyproof for any $r > \frac{1}{3}$. However, an agent i with preference order $P_i : a \succ b \succ c \succ d$ and consistent utility function $\tilde{u}_i = (6, 2, 1, 0)$ will not find a beneficial manipulation for any reports P_{-i} from the other agents. Thus, ABM makes truthful reporting a dominant strategy for an agent with utility function $\tilde{u}_i$ in this setting. But $\tilde{u}_i$ violates URBI (1/3), since

$$\frac{\tilde{u}_i(c) - \min_{j \in M}(\tilde{u}(j))}{\tilde{u}_i(b) - \min_{j \in M}(\tilde{u}(j))} = \frac{1-0}{2-0} = \frac{1}{2} > \frac{1}{3}. \tag{8}$$

This example can be verified using Algorithm 1 in Appendix 2.C.

2.7.1 Maximality of URBI(r) for Partially Strategyproof Mechanisms

Despite Example 2, the URBI(r) domain restriction is *maximal* in the following sense: consider a mechanism φ that is r-partially strategyproof for some bound $r \in (0, 1)$. we

show that, unless we are given additional structural information about φ, URBI(r) is in fact the largest set of utilities for which truthful reporting is guaranteed to be a dominant strategy. This maximality is our third main result.

Theorem 3. *For any setting $(N, M, \mathbf{q})$ with $m \geqslant 3$, any bound $r \in (0, 1)$, and any utility function $\tilde{u}_i$ (consistent with a preference order P_i) that violates URBI(r) there exists a mechanism $\tilde{\varphi}$ such that*

1. *$\tilde{\varphi}$ is r-partially strategyproof, but*
2. *there exist preferences of the other agents P_{-i} and a misreport P_i' such that*

$$\langle \tilde{u}_i, \tilde{\varphi}_i(P_i, P_{-i}) - \tilde{\varphi}_i(P_i', P_{-i}) \rangle < 0. \tag{9}$$

Furthermore, $\tilde{\varphi}$ can be chosen to satisfy anonymity.

Proof outline (formal proof in Appendix 2.F.4). If $\tilde{u}_i$ violates URBI(r), there must be a pair of objects $a, b \in M$ with $P_i : a \succ b$, such that

$$\frac{\tilde{u}_i(b) - \min_{j \in M} (\tilde{u}(j))}{\tilde{u}_i(a) - \min_{j \in M} (\tilde{u}(j))} = \tilde{r} > r. \tag{10}$$

We construct the mechanism $\tilde{\varphi}$ explicitly. $\tilde{\varphi}$ assigns a constant vector to agent i, except when i reports some preference order P_i' with $P_i' : b \succ a$. In that case $\tilde{\varphi}$ assigns less of a, more of b, and less of i's reported last choice (say, c) to i. Then $\tilde{\varphi}$ is swap monotonic and upper invariant. The re-assignment between a, b, and c must be constructed in such a way that i would want to manipulate if its utility is $\tilde{u}_i$ but would not want to manipulate if its utility satisfied URBI(r). We show that this is possible. Finally, by keeping all other agents' assignment vectors constant and randomizing over the roles of the agents in the mechanism, we obtain an anonymous mechanism with these properties. □

If some additional constraints are imposed on the space of possible mechanisms, the mechanism $\tilde{\varphi}$ constructed in the proof of Theorem 3 may no longer be feasible such that the counterexample fails. However, without any such constraints we cannot rule out the possibility that an agent with some utility function outside URBI(r) may want to manipulate an r-partially strategyproof mechanism. The following Corollary 1 makes this argument precise.

Corollary 1. *For any setting $(N, M, \boldsymbol{q})$ with $m \geqslant 3$ objects and any $r \in (0, 1)$, we have*

$$URBI(r) = \bigcap_{\substack{\varphi \text{ } r\text{-partially} \\ \text{strategyproof} \\ \text{in } (N,M,\boldsymbol{q})}} \left\{ u \;\middle|\; \begin{array}{c} \varphi \text{ makes truthful reporting a} \\ \text{dominant strategy for agents} \\ \text{with utility } u \text{ in } (N, M, \boldsymbol{q}) \end{array} \right\}. \tag{11}$$

This means that when considering the set of r-partially strategyproof mechanisms, the set of utilities for which *all* of them make truthful reporting a dominant strategy is *exactly equal* to URBI(r). Thus, there is no larger domain restriction for which all these mechanisms will also guarantee good incentives.

2.7.2 A Parametric Measure for Incentive Properties

The partial strategyproofness concept leads to a new, intuitive measure for the incentive properties of swap monotonic, upper invariant mechanisms: we consider the largest possible value r for which the mechanism is still r-partially strategyproof.

Definition 6 (Degree of Strategyproofness)**.** Given a setting $(N, M, \boldsymbol{q})$ and a swap monotonic, upper invariant mechanism φ, the *degree of strategyproofness of* φ is

$$\rho_{(N,M,\boldsymbol{q})}(\varphi) = \max \{r \in [0, 1] \mid \varphi \text{ is } r\text{-partially SP in } (N, M, \boldsymbol{q})\}. \tag{12}$$

Observe that for $0 \leqslant r' \leqslant r \leqslant 1$ we have URBI(r') $\subset$ URBI(r) by construction. A mechanism that is r-partially strategyproof will also be r'-partially strategyproof. Thus, a higher degree of strategyproofness corresponds to a stronger guarantee.

Remark 4. In (12), we use the *maximum* (rather than the *supremum*). This is possible because the constraint (7) in Definition 5 of partial strategyproofness is a weak inequality. Thus, the set of utilities for which a mechanism makes truthful reporting a dominant strategy is topologically closed. Consequently, there exists some maximal value $\rho > 0$, for which the mechanism is ρ-partially strategyproof, but it is not r-partially strategyproof for any $r > \rho$ (i.e., $\rho_{(N,M,\boldsymbol{q})}(\varphi)$ is well-defined).

Interpretation of the Degree of Strategyproofness

Maximality of the URBI(r) domain restriction (especially, Corollary 1) implies that when measuring the degree of strategyproofness of swap monotonic and upper invariant

mechanisms using $\rho_{(N,M,\boldsymbol{q})}(\varphi)$, no utility functions are omitted for which a guarantee could also be given. Thus, without further information on the mechanism, there cannot exist another single-parameter measure that conveys strictly more information about the incentive guarantees of φ. The degree of strategyproofness also allows for the comparison of two mechanisms: "$\rho_{(N,M,\boldsymbol{q})}(\varphi) > \rho_{(N,M,\boldsymbol{q})}(\psi)$" means that φ make truthful reporting a dominant strategy for a larger URBI(r) domain restriction than ψ.

From a quantitative perspective one might ask for "how many more" utility functions φ is guaranteed to have good incentives, compared to ψ. Recall Remark 2, where we considered URBI(0.4) in a setting with 3 objects, $\min_{j\in M}(u(j)) = 0$, and the remaining utilities for the first and second choice were chosen uniformly at random from the unit square. Suppose that $\rho_{(N,M,\boldsymbol{q})}(\varphi) = 0.8$ and $\rho_{(N,M,\boldsymbol{q})}(\psi) = 0.4$, then the set URBI(0.8) has "twice the size" of URBI(0.4). Thus, the guarantee for φ extends over twice as many utility functions as the guarantee for ψ. In this sense we can think of φ as being "twice as strategyproof" as ψ.

Relation of Degree of Strategyproofness and Vulnerability to Manipulation

Pathak and Sönmez (2013) proposed an interesting method for comparing mechanisms by their vulnerability to manipulation. An extension of this concept to the case of vNM utilities is straightforward: ψ is *strongly as manipulable as* φ if whenever an agent with utility u finds a beneficial misreport under φ, the same agent in the same situation also finds a beneficial misreport under ψ. The following Proposition 2 shows that vulnerability to manipulation and the degree of strategyproofness are *consistent* (but not equivalent).

Proposition 2. *For any setting $(N, M, \boldsymbol{q})$ and mechanisms φ and ψ, the following hold:*

1. *If ψ is strongly as manipulable as φ, then $\rho_{(N,M,\boldsymbol{q})}(\varphi) \geqslant \rho_{(N,M,\boldsymbol{q})}(\psi)$.*
2. *If $\rho_{(N,M,\boldsymbol{q})}(\varphi) > \rho_{(N,M,\boldsymbol{q})}(\psi)$ and φ and ψ are comparable by their vulnerability to manipulation, ψ is strongly more manipulable than φ.*

The proof as well as the formal definition of the *strongly as manipulable as*-relation for random assignment mechanisms are given in Appendix 2.F.5. Despite the consistency result, neither concept is always better at *strictly* differentiating mechanisms: a comparison by vulnerability to manipulation may be inconclusive when the degree of strategyproofness yields a strict winner; conversely, the degree of strategyproofness may indicate indifference (i.e., $\rho_{(N,M,\boldsymbol{q})}(\varphi) = \rho_{(N,M,\boldsymbol{q})}(\psi)$) when one of the mechanisms is in fact strongly *more manipulable* than the other.

An important difference between the two concepts is that the comparison by vulnerability to manipulation considers each preference profile separately, while the partial strategyproofness constraint must hold uniformly for *all* preference profiles. Thus, vulnerability to manipulation yields a *best response*-notion of incentives while the degree of strategyproofness yields a *dominant strategy*-notion of incentives. However, the degree of strategyproofness has two important advantages. First, Pathak and Sönmez (2013) have not presented a method to perform the comparison by vulnerability to manipulation algorithmically, and the definition of such a method is not straightforward. In contrast, $\rho_{(N,M,\boldsymbol{q})}$ is computable (see Remark 5 in Section 2.8 and Algorithm 2 in Appendix 2.C). Second, and more importantly, the degree of strategyproofness is a parametric measure while the *strongly as manipulable as*-relation is not. A mechanism designer could easily express a minimal acceptable degree of strategyproofness and then consider only mechanisms satisfying this constraint. A similar design approach using vulnerability to manipulation appears much more difficult as it would require the definition of a "benchmark mechanism" ψ with maximal acceptable manipulability and considering only mechanisms that are less manipulable than ψ.

2.8 A Dominance Interpretation of Partial Strategyproofness

Partial strategyproofness restricts the set of utility functions for which the mechanism must make truthful reporting a dominant strategy, but the definition is otherwise analogous to Definition 3 of strategyproofness. Furthermore, recall that strategyproofness is equivalent to SD-strategyproofness, the incentive concept induced by stochastic dominance. Our fourth main result shows that an analogous equivalence exists for r-partial strategyproofness. Specifically, it is equivalent to the incentive concept induced by a certain dominance notion. In this section, we formally define *r-partial dominance*, and we show that r-partial strategyproofness and r-partial dominance-strategyproofness are in fact the same.

Definition 7 (Partial Dominance)**.** For a preference order $P \in \mathcal{P}$ with $P : a_1 \succ \ldots \succ a_m$, a bound $r \in [0, 1]$, and assignment vectors x, y, we say that x *r-partially dominates y at P* if for all ranks $k \in \{1, \ldots, m\}$ we have

$$\sum_{l=1}^{k} r^l \cdot x_{a_l} \geqslant \sum_{l=1}^{k} r^l \cdot y_{a_l}. \tag{13}$$

Observe that for $r = 1$, this definition is precisely the same as stochastic dominance, since $1^l = 1$ for any l. However, for $r < 1$, the impact of less preferred objects is discounted by the factor r. Intuitively, inequality (13) can be interpreted as incentive constraint for certain "extreme" utility functions that satisfy URBI(r), put very high value on the first k objects, and next to no value on all other objects.

Analogous to stochastic dominance for SD-strategyproofness, we can use r-partial dominance to define r-partial dominance-strategyproofness.

Definition 8 (PD-Strategyproof)**.** Given a setting $(N, M, \boldsymbol{q})$ and a bound $r \in (0, 1]$, a mechanism φ is *r-partial dominance-strategyproof (r-PD-strategyproof)* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in \mathcal{P}$, $\varphi_i(P_i, P_{-i})$ r-partially dominates $\varphi_i(P_i', P_{-i})$ at P_i.

We are now ready to formally state our fourth main result, the equivalence of r-partial strategyproofness and r-PD-strategyproofness.

Theorem 4. *Given a setting $(N, M, \boldsymbol{q})$ and a mechanism φ, the following are equivalent:*

1. *φ is r-partially strategyproof,*
2. *φ is r-PD-strategyproof.*

Proof Outline (formal proof in Appendix 2.F.6). The challenge is that the sets URBI(r)$\cap$ U_P are unbounded, and therefore, they cannot be represented as convex polytopes with finitely many corner points. Nonetheless, the partial sums in the definition of partial dominance can be interpreted as the incentives to misreport that are induced by different "extreme utility functions." We prove that these extreme utilities are the *essential limit cases* that determine r-partial strategyproofness: we show that x is preferred to y by an agent with some utility u in URBI(r) if and only if this is also true for at least one of the extreme utilities. □

Theorem 4 means that the two requirements that (i) "a mechanism makes truthful reporting a dominant strategy for any agent with a utility function in URBI(r)," and (ii) "any assignment vector that an agent can obtain by misreporting is r-partially dominated by the assignment vector this agent can obtain by reporting truthfully," are in fact the same. This yields an alternative way of defining r-partial strategyproofness that does not rely on the agents' utility functions. This shows that the concept integrates nicely in the landscape of existing incentive concepts, most of which also rely on dominance notions (e.g., strong and weak SD-strategyproofness and DL-strategyproofness).

Moreover, the alternative definition of partial strategyproofness via partial dominance also unlocks the concept to algorithmic analysis. The original definition imposed inequalities that had to hold for all (infinitely many) utility functions within the set URBI(r). While this provides a good economic intuition, it makes algorithmic verification of r-partial strategyproofness infeasible via its original definition. However, by the equivalence from Theorem 4, it suffices to verify that all (finitely many) constraints for partial dominance are satisfied. The finite condition can also be used to encode r-partial strategyproofness as linear constraints to an optimization problem. This enables an automated search within the set of r-partially strategyproof mechanisms while optimizing for some other design objective under the automated mechanism design paradigm (Sandholm, 2003).

Remark 5. In Appendix 2.C we give algorithms that exploit the structure of r-PD-strategyproofness to verify whether a mechanism φ is r-partially strategyproof in a given setting (Algorithm 1), and to compute its degree of strategyproofness $\rho_{(N,M,\boldsymbol{q})}(\varphi)$ (Algorithm 2).

2.9 Intermediateness of Partial Strategyproofness

In this section, we study the relationship of partial strategyproofness to other incentive concepts that have been discussed previously in the context of assignment mechanisms or more broadly in domains with no monetary transfer. We demonstrate that our new concept takes an intermediate position *between* strategyproofness on one side and many other concepts on the other side: while partial strategyproofness is implied by strategyproofness, it in turn implies weak, convex, and approximate strategyproofness, as well as strategyproofness in the large (if the degree of strategyproofness converges to 1). Most importantly, while strategyproofness is the upper limit concept (for $r = 1$), we show that strategyproofness for lexicographic preferences is the lower limit concept for $r \to 0$. Figure 2.2 gives an overview of the relationships between the different incentive concepts.

2.9.1 Relation to Strategyproofness

We have already observed that r-partial strategyproofness for $r = 1$ is equivalent to strategyproofness. Thus, strategyproofness can be considered an upper limit case of r-partial strategyproofness.

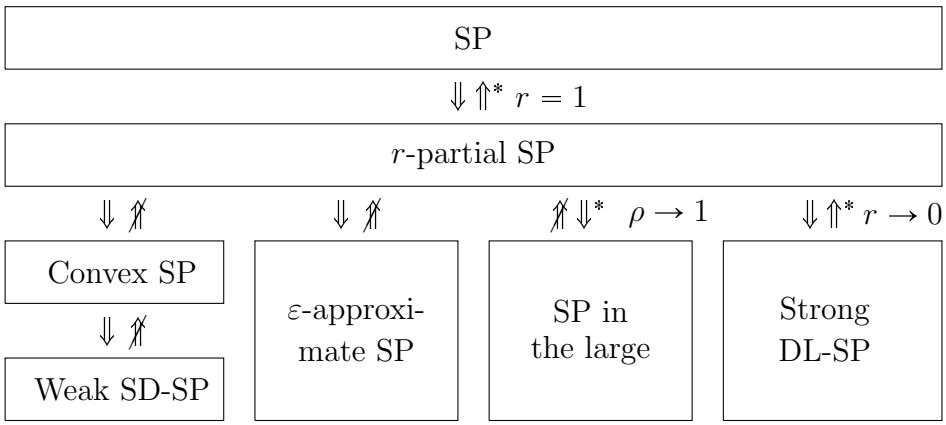

Figure 2.2: Relations between incentive concepts (*SP*: strategyproofness).

Remark 6. To obtain a more formal understanding of this limit case, let $\mathrm{SP}(N, M, \boldsymbol{q})$ and r-$\mathrm{PSP}(N, M, \boldsymbol{q})$ denote the sets of strategyproof and r-partially strategyproof mechanisms in the setting $(N, M, \boldsymbol{q})$, respectively. It is straightforward to see that

$$\mathrm{SP}(N, M, \boldsymbol{q}) = \bigcap_{r<1} r\text{-}\mathrm{PSP}(N, M, \boldsymbol{q}). \tag{14}$$

In words, any mechanisms that are r-partially strategyproof for *all* $r < 1$ must be strategyproof. In Section 2.9.5, we will prove a corresponding formal statement about the lower limit concept, DL-strategyproofness.

2.9.2 Relation to Weak SD-Strategyproofness

Weak SD-strategyproofness was employed by Bogomolnaia and Moulin (2001) to describe the incentive properties of the Probabilistic Serial mechanism. Recall that an assignment vector x stochastically dominates another assignment vector y at a preference order P if for any rank k an agent with preference order P is at least as likely to obtain its kth choice or better under x than under y. This dominance of x over y is strict if in addition for some rank the probability is strictly greater under x than under y. Under weakly SD-strategyproof mechanisms, agents cannot attain a strictly dominant assignment vector by misreporting; however, in contrast to strategyproof mechanisms, the assignment vectors do not need to be comparable by stochastic dominance.

Definition 9 (Weakly SD-Strategyproof)**.** A mechanism φ is *weakly SD-strategyproof* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in \mathcal{P}$,

agent i's assignment vector from truthful reporting is not strictly stochastically dominated by its assignment vector from reporting P_i'.

Weak SD-strategyproofness is equivalent to requiring that for a given preference profile (P_i, P_{-i}) and a potential misreport P_i', there exists a consistent utility $u_i \in U_{P_i}$ such that agent i would prefer reporting P_i to reporting P_i'. This is an extremely weak requirement because u_i may depend on P_i'. In other words, the mechanism might still offer an opportunity to manipulate to the agent with utility u_i. The only guarantee is that the particular report P_i' will not increase its expected utility. Thus, it is possible that for some preference order P_i, truthful reporting is not a dominant strategy, *independent* of agent i's utility functions.

We have shown that partial strategyproofness implies weak SD-strategyproofness.

Proposition 3. *Given a setting $(N, M, \boldsymbol{q})$, if a mechanism φ is partially strategyproof (i.e., r-partially strategyproof for some $r > 0$), then it is weakly SD-strategyproof. The converse may not hold.*

Proof Outline (formal proof in Appendix 2.F.7). We show that partial strategyproofness implies *convex strategyproofness* (Balbuzanov, 2015). This in turn implies weak SD-strategyproofness. Balbuzanov also gave an example of a mechanism that is weakly SD-strategyproof but violates convex strategyproofness. □

2.9.3 Relation to Approximate Strategyproofness

Approximate strategyproofness is a different relaxation of strategyproofness that has attracted interest in quasi-linear domains (Lubin and Parkes, 2012). Approximately strategyproof mechanisms may be manipulable, but there exists an upper bound on the gain that an agent can obtain by misreporting. The economic intuition behind this concept is that if the potential gain is small, agents might not be willing to collect the necessary information and deliberate about misreports but stick with truthful reporting instead. In this section, we formalize a notion of approximate strategyproofness that is meaningful for assignment mechanisms. Then we show that partial strategyproofness implies approximate strategyproofness but that the converse may not hold.

When payments are possible, money provides a canonical unit of measure for the gain from misreporting. However, the assignment domain does not permit payments, which makes the definition and interpretation of approximate strategyproofness more challenging. Here, we follow earlier work, where approximate strategyproofness for ordinal mechanisms was defined via bounded utilities (Birrell and Pass, 2011; Carroll, 2013).

Definition 10 (Approximately Strategyproof)**.** Given a setting $(N, M, \boldsymbol{q})$ and a bound $\varepsilon \in [0, 1]$, a mechanism φ is *ε-approximately strategyproof (in the setting $(N, M, \boldsymbol{q})$)* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any utility function $u_i \in U_{P_i}$ that is consistent with P_i and bounded between 0 and 1 (i.e., $u_i : M \to [0, 1]$), the gain in expected utility from reporting P_i' is upper-bounded by ε; formally,

$$\langle u_i, \varphi_i(P_i', P_{-i}) - \varphi_i(P_i, P_{-i}) \rangle \leqslant \varepsilon. \tag{15}$$

Note that if u_i was not bounded, the potential gain from manipulation under any non-strategyproof mechanism would instantly become arbitrarily large (Carroll, 2013). Since u_i is bounded between 0 and 1, a change of magnitude 1 in expected utility corresponds to getting one's first choice instead of one's last choice. Thus, "1" corresponds to the *maximal gain* from misreporting that *any* agent could obtain under an *arbitrary* mechanism. Relative to this value "1," the parameter ε is the *share of this maximal gain* by which any agent can at most improve its expected utility under an ε-approximately strategyproof mechanism. Furthermore, the gain will never exceed 1, which makes 1-approximate strategyproofness a void constraint that is trivially satisfied by any mechanism. Obviously, if φ is ε-approximately strategyproof, then it is also ε'-approximately strategyproof for any $\varepsilon' \geqslant \varepsilon$.

Proposition 4 shows that partial strategyproofness implies approximate strategyproofness, but the converse is not true in general.

Proposition 4. *Given a setting $(N, M, \boldsymbol{q})$, if a mechanism φ is r-partially strategyproof for some $r > 0$, then it is ε-approximately strategyproof for some $\varepsilon < 1$. The converse may not hold.*

Proof Outline (formal proof in Appendix 2.F.8). Using the axiomatic decomposition of partial strategyproofness from Theorem 2, we derive an upper bound for the gain from manipulation that any agent with bounded utilities can obtain by misreporting and we show that this bound can be chosen strictly below 1. To see that the converse does not hold, we construct a simple counter-example. □

Proposition 4 yields new insights for partially strategyproof mechanisms: initially, the definition of r-partial strategyproofness only required good incentives for agents whose utility functions satisfy the URBI(r) constraint. However, it imposes no restriction for agents with utilities outside this set. Proposition 4 shows that, even though these other agents may be able to benefit from misreporting, their incentive to do so is at least bounded by some $\varepsilon < 1$ in the sense of approximate strategyproofness.

2.9.4 Relation to Strategyproofness in the Large

Azevedo and Budish (2015) proposed *strategyproofness in the large* as an alternative when strategyproofness is incompatible with other essential design objectives. This incentive concept captures the intuition that the ability of any single agent to improve its own interim assignment (i.e., in expectation) by misreporting may vanish as more agents participate in a mechanism. For example, in school choice, where thousands of students compete for seats at a relatively small number of schools, this requirement may facilitate interesting design alternatives.

The model of Azevedo and Budish (2015) considered a finite set of vNM utility functions $\{u^1, \dots, u^K\}$. Strategyproofness in the large requires that for any $\varepsilon > 0$ there exists a number n_0 of agents such that in any setting with sufficiently many agents (i.e., $n \geqslant n_0$), no agent can gain more than ε by misreporting.[4] To apply this concept to the random assignment problem, we need to specify in what sense settings *get large*. To this end, we follow (Kojima and Manea, 2010) and (Azevedo and Budish, 2015) and keep the number of objects constant, but we let the number of agents grow and increase the objects' capacities such that supply satisfies demand. Thus, we consider a sequence of settings $(N^n, M^n, \boldsymbol{q}^n)_{n \geqslant 1}$ where the set of agents $\#N^n = n$ grows, the set of objects $M^n = M$ is held fixed, and capacities grow so that $n = \sum_{j \in M} q_j^n$, and $\min_{j \in M}(q_j^n) \to \infty$ for $n \to \infty$.

Proposition 5. *Fix any finite set of utility functions $\{u^1, \dots, u^K\} \subseteq \bigcup_{P \in \mathcal{P}} U_P$. If the degree of strategyproofness of φ converges to 1 as the settings grow (i.e., $\rho_{(N^n, M, \boldsymbol{q}^n)}(\varphi) \to 1$ for $n \to \infty$), then φ is strategyproof in the large with respect to $\{u^1, \dots, u^K\}$.*

Proof. Any consistent utility function u satisfies uniformly relatively bounded indifference for some (sufficiently large) $r < 1$. Let $\bar{r}$ be the largest of these values, such that $u^k \in \mathrm{URBI}(\bar{r})$ for all $k \in \{1, \dots, K\}$. Since by assumption, φ is $\bar{r}$-partially strategyproof in $(N^n, M, \boldsymbol{q}^n)$ for sufficiently large n, all agents will have a dominant strategy to report their preferences truthfully. □

Kojima and Manea (2010) showed that the incentives under the non-strategyproof Probabilistic Serial (PS) mechanism improve in larger settings: for any fixed utility function, PS eventually makes truthful reporting a dominant strategy for any agent with that utility function. Azevedo and Budish (2015) used this result to show that PS is in fact strategyproof in the large. In Section 2.12.2, we will show that PS is partially

[4] The original definition is more technical and involves probability measures over the other agents' preferences. However, this simplified version suffices to illustrate the connection with partial strategyproofness.

strategyproof (in finite settings). In combination, these insights suggest the following conjecture: *as settings grow in the way defined above, the degree of strategyproofness of PS converges to 1.* A proof of this conjecture would strengthen the result of Kojima and Manea (2010) because it would specify the precise way in which the set of utility functions with good incentives grows.[5] In combination with Proposition 5, it would also yield an elegant proof for the observation that PS is strategyproof in the large. In Section 2.12.2 we provide numerical evidence that supports this conjecture.

2.9.5 Relation to Lexicographic Strategyproofness

Finally, we compare our new partial strategyproofness concept to strategyproofness for agents with lexicographic preferences. In particular, we show that this is the lower limit concept of r-partial strategyproofness as $r \to 0$. The intuition of lexicographic preferences is that agents prefer any (arbitrarily small) increase in the probability for some object to any (arbitrarily large) increase in the probability for some less preferred object.

Definition 11 (DL-Dominance)**.** For preference order $P \in \mathcal{P}$ with $P : a_1 \succ \ldots \succ a_m$ and assignment vectors x, y, we say that x *(downward-)lexicographically dominates (DL-dominates)* y *at* P if either $x = y$, or for some rank $k \in \{1, \ldots, m\}$ we have $x_k > y_k$ and $x_l = y_l$ for all $l \leqslant k - 1$.

DL-dominance induces DL-strategyproofness in the same way in which stochastic dominance induces SD-strategyproofness.

Definition 12 (DL-Strategyproof)**.** A mechanism φ is *DL-strategyproof* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in \mathcal{P}$, the assignment from truthfully reporting P_i DL-dominates the assignment from misreporting P_i'.

Our fifth main result is Theorem 5, which yields an equivalence between partial strategyproofness and DL-strategyproofness.

Theorem 5. *Given a setting $(N, M, \boldsymbol{q})$, a mechanism φ is partially strategyproof (i.e., r-partially strategyproof for some $r > 0$) if and only if φ is DL-strategyproof.*

Proof Outline (formal proof in Appendix 2.F.9). The proof is analogous to the proof of Theorem 2, where we showed that partial strategyproofness is equivalent to the axioms

[5] Note that Theorem 1 from (Kojima and Manea, 2010) is precisely the statement that for any $\varepsilon > 0$, PS is ε-approximate strategyproofness in sufficiently large settings. However, by Proposition 4 this is strictly not sufficient for partial strategyproofness.

swap montonicity and upper invariance. The minimal change in the probability for the highest ranking object (for which there is any change) is now reflected by the strict change induced by DL-strategyproofness, if any. □

Theorem 5 demonstrates that for random assignment mechanisms, DL-strategyproofness is an unnecessarily weak concept. Imposing swap monotonicity and upper invariance already yields that the mechanism must be r-partially strategyproof for some positive bound r (Theorem 2). Partial strategyproofness specifies the precise structure of the incentive guarantees via the URBI(r) domain restriction. In contrast, DL-strategyproofness is a purely binary requirement that is either satisfied by a mechanism or not, but it ignores the parametric nature of the set of utility functions for which truthful reporting is guaranteed to be a dominant strategy.

A second interesting consequence of Theorem 5 is the fact that DL-strategyproofness is the lower limit concept for partial strategyproofness. In Remark 6, r-PSP$(N, M, \boldsymbol{q})$ was the set of mechanisms that are r-partially strategyproof in $(N, M, \boldsymbol{q})$. Similarly, let DL-SP$(N, M, \boldsymbol{q})$ be the set of DL-strategyproof mechanisms in that setting.

Corollary 2. *Given a setting $(N, M, \boldsymbol{q})$, we have*

$$\textit{DL-SP}(N, M, \boldsymbol{q}) = \bigcup_{r>0} r\textit{-PSP}(N, M, \boldsymbol{q}), \tag{16}$$

In words, any mechanisms that is DL-strategyproof must be r-partially strategyproof for some *strictly positive bound $r > 0$.*

Corollary 2 is the formal counterpart to Remark 6, where we showed that strategyproofness is the upper limit of partial strategyproofness in the sense that

$$\text{SP}(N, M, \boldsymbol{q}) = \bigcap_{r<1} r\text{-PSP}(N, M, \boldsymbol{q}). \tag{17}$$

2.10 Local Sufficiency of Partial Strategyproofness

For some incentives concepts, it suffices to check whether no local misreports are beneficial in order to establish that no misreports are beneficial at all. In this case, we say that the incentive concept satisfies *local sufficiency*. Local sufficiency simplifies the respective incentive concept from an axiomatic as well as from an algorithmic perspective. For assignment mechanisms, the *local* misreports are those in the neighborhood of the manipulating agent's true preference order, which arise by swapping two consecutive

objects. Carroll (2012) and Cho (2012) proved local sufficiency for strategyproofness and DL-strategyproofness, respectively.

Local sufficiency is an intriguing concept: it can be used to greatly reduce the complexity of the incentive concepts. As an example, recall that our axioms swap monotonicity, upper invariance, and lower invariance were simple and intuitive, in part because they restricted the behavior of the mechanisms only for local misreports. From a computational perspective, local sufficiency reduces algorithmic complexity because it reduces the number of constraints in the optimization problem that is used for searching optimal mechanisms under the automated mechanism design paradigm (Sandholm, 2003).

In this section, we prove an analogous local sufficiency result for partial strategyproofness. First, we formally define three notions of local strategyproofness.

Definition 13 (Locally Strategyproof & Locally DL-Strategyproof)**.** A mechanism φ is *locally strategyproof* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in N_{P_i}$ from the neighborhood of P_i, and any utility $u_i \in U_{P_i}$ that is consistent with P_i, we have

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_i(P_i', P_{-i}) \rangle \geqslant 0. \tag{18}$$

φ is *locally DL-strategyproof* if $\varphi_i(P_i, P_{-i})$ DL-dominates $\varphi_i(P_i', P_{-i})$.

Analogously, we can define a local variant of partial strategyproofness.

Definition 14 (Locally Partially Strategyproof)**.** Given a setting $(N, M, \boldsymbol{q})$ and a bound $r \in (0, 1]$, a mechanism φ is *r-locally partially strategyproof (in the setting $(N, M, \boldsymbol{q})$)* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in N_{P_i}$ from the neighborhood of P_i, and any utility $u_i \in U_{P_i} \cap \mathrm{URBI}(r)$ that is consistent with P_i and satisfies $\mathrm{URBI}(r)$, we have

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_i(P_i', P_{-i}) \rangle \geqslant 0. \tag{19}$$

and we say that φ is *locally partially strategyproof* if it is r-locally partially strategyproof for some non-trivial $r > 0$.

Facts 1 and 2 summarize the known local sufficiency results. Since local constraints are necessary, local sufficiency always implies equivalence.

Fact 1 (Carroll, 2012)**.** *Local strategyproofness is sufficient for strategyproofness.*

Fact 2 (Cho, 2012)**.** *Local DL-strategyproofness is sufficient for DL-strategyproofness.*

In combination with our equivalence result for partial strategyproofness and DL-strategyproofness (Theorem 5), Fact 2 immediately yields a weak notion of local sufficiency for partial strategyproofness.

Corollary 3. *Given a setting $(N, M, \boldsymbol{q})$, if a mechanism φ is r-locally partially strategyproof for some $r > 0$, then φ is r'-partially strategyproof for some $r' > 0$.*

Corollary 3 follows from the observation that local partial strategyproofness implies local DL-strategyproofness, which implies DL-strategyproofness, which in turn implies partial strategyproofness by Theorem 5. However, the (local) bound r and the (global) bound r' are not necessarily the same. Since r'-partial strategyproofness implies r'-local partial strategyproofness, we must have $r' \leqslant r$, but r' may still be (much) smaller than r. Our sixth main result establishes a precise connection between r and r'.

Theorem 6. *Given a setting $(N, M, \boldsymbol{q})$, if φ is r-locally partially strategyproof for some $r > 0$, then φ is r^2-partially strategyproof.*

Proof Outline (formal proof in Appendix 2.F.10). For any $u \in U_P$ that satisfies $\mathrm{URBI}(r^2)$ and any misreport P', we construct a line segment in the utility space that starts in u and ends in another utility function $v \in U_{P'}$. We then express the incentive to misreport P' (instead of P) for an agent with utility u as a telescoping sum over local incentive constraints along this line segment.[6] In this representation all but the first and the last term cancel out, such that it collapses to the required inequality. Since local incentive constraints are only available for utility functions inside $\mathrm{URBI}(r)$, we need to ensure that the line segment intersects the sets $U_{P_k} \cap \mathrm{URBI}(r)$ for every preference order P_k through which it passes. □

Theorem 6 means that r-local partial strategyproofness is sufficient to guarantee r'-partial strategyproofness, where $r' \leqslant r^2$. As a special case, we obtain that 1-local partial strategyproofness implies 1-partial strategyproofness, the local sufficiency result for strategyproofness of Carroll (2012). Furthermore, considering a sequence of bounds $(r_k)_{k \geqslant 1}$ that approaches 0, we obtain the local sufficiency result for DL-strategyproofness of Cho (2012) in the limit. Thus, Theorem 6 unifies both prior results.

The question remains whether Theorem 6 is tight or whether the bound $r' \leqslant r^2$ can be improved in any way. First, note that it is straightforward to construct a counter-example

[6] This step is inspired by Carroll (2012)'s proof of local sufficiency for strategyproofness.

showing that "$r' = r$" is out of the question, unless $r \in \{0, 1\}$. In fact, as we show in the next Theorem 7, the bound "$r' = r^2$" is tight in the sense that "2" is the smallest exponent for which a guarantee can be given.

Theorem 7. *Given a setting $(N, M, \boldsymbol{q})$ with $m \geqslant 4$ objects, for any $\varepsilon > 0$ there exists a bound $r \in (0, 1)$ and a mechanism φ such that*

1. *φ is r-locally partially strategyproof, but*
2. *φ is* not *$r^{2-\varepsilon}$-partially strategyproof.*

Proof Outline (formal proof in Appendix 2.F.11). The proof is constructive by giving φ explicitly. To find a suitable mechanism, we initially generated special instances of φ for fixed ε as solutions to a particular linear program. The main challenge was to subsequently infer the general structure of φ from the examples and to prove the required properties. □

Tightness by Theorem 7 means that "$r' = r^2$" is the *best polynomial bound* that allows a general statement about local sufficiency of the partial strategyproofness concept.[7]

2.11 An Extension of Partial Strategyproofness for Deterministic Mechanisms

Some assignment mechanisms do not involve randomization. For example, most results for school choice mechanisms were obtained under the assumption that priorities are fixed and strict. This makes the mechanisms *deterministic*. Since non-strategyproof, deterministic mechanisms play an important role, we would like to apply the partial strategyproofness concept to study the incentive properties of these mechanisms as well. To this end, we consider a second source of randomness, namely the agent's uncertainty about the reports from all other agents. This uncertainty makes the mechanisms random from the perspective of the agent, even if the mechanism itself is deterministic. We show in this section under what conditions (on the mechanisms) this injection of uncertainty induces partial strategyproofness from the agent's perspective.

Consider a setting $(N, M, \boldsymbol{q})$ and a (possibly deterministic) mechanism φ. Suppose that an agent $i \in N$ does not know exactly what the other agents are going to report, but

[7] Note that the value r in the counter-examples in the proof of Theorem 7 may depend on ε. We leave the exploration of the relationship between r and r' for *fixed* r to future research.

it has a probabilistic belief about these reports. Formally, i believes that $P_{-i} \in \mathcal{P}^{N\setminus\{i\}}$ is drawn from a distribution $\mathbb{P}$. Then from i's perspective, the mechanism that is relevant for its strategic considerations is random and given by $\varphi^{\mathbb{P}}$ with

$$\varphi^{\mathbb{P}}(P_i) = \sum_{P_{-i}\in\mathcal{P}^{N\setminus\{i\}}} \varphi_i(P_i, P_{-i}) \cdot \mathbb{P}[P_{-i}], \tag{20}$$

where $\varphi^{\mathbb{P}}(P_i)$ is simply i's *expected* assignment vector from reporting P_i. We say that $\mathbb{P}$ has *full support* if $\mathbb{P}[P_{-i}] > 0$ for all $P_{-i} \in \mathcal{P}^{N\setminus\{i\}}$.

Our main result in this section is a axiomatic characterization of the mechanisms φ (deterministic or otherwise) that admit the construction of partially strategyproof mechanisms $\varphi^{\mathbb{P}}$. This requires two new axioms.

Axiom 4 (Monotonic)**.** *A mechanism φ is* monotonic *if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any preference order $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$, the misreport P_i' weakly increases i's chances at a_{k+1} and weakly decreases i's chances at a_k.*

Monotonicity is a very natural requirement. It simply captures the intuition that bringing an object up in the preference order should not reduce the chances of obtaining this object. Swap monotonicity implies monotonicity, while the converse does not hold (e.g., for fixed, strict priorities the Boston mechanism is monotonic but not swap monotonic).

Axiom 5 (Sensitive)**.** *A mechanism φ is* sensitive *if for any agent $i \in N$, any preference orders $P_i \in \mathcal{P}$ and $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$, the following holds: if $\varphi_i(P_i, P_{-i}) \neq \varphi_i(P_i', P_{-i})$ for some $P_{-i} \in \mathcal{P}^{N\setminus\{i\}}$, then there exist $P_{-i}^k, P_{-i}^{k+1} \in \mathcal{P}^{N\setminus\{i\}}$ such that $\varphi_{i,a_k}(P_i, P_{-i}^k) \neq \varphi_{i,a_k}(P_i', P_{-i}^k)$ and $\varphi_{i,a_{k+1}}(P_i, P_{-i}^{k+1}) \neq \varphi_{i,a_{k+1}}(P_i', P_{-i}^{k+1})$.*

Intuitively, a mechanism is sensitive if the agent's relative preferences matter for the assignment of the respective objects: if the agent's assignment changes at all under a swap of a_k and a_{k+1}, then there must exist situations in which the two objects a_k and a_{k+1} are actually affected by this change. Again, swap monotonicity implies sensitivity, but the converse does not hold.

Our seventh main result characterizes the mechanisms, deterministic or not, for which uncertainty over the other agents' reports induces a partially strategyproof mechanism from the perspective of each individual agent.

Theorem 8. *A mechanism φ is upper invariant, monotonic, and sensitive if and only if $\varphi^{\mathbb{P}}$ is upper invariant and swap monotonic for all distributions $\mathbb{P}$ with full support.*

The formal proof is given in Appendix 2.F.12. The most useful consequence of Theorem 8 is an insight about the strategic situation of agents whose uncertainty about the other agents' reports is described by $\mathbb{P}$. Corollary 4 formalizes sufficient conditions under which these agents face a partially strategyproof mechanism.

Corollary 4. *Given a setting $(N, M, \boldsymbol{q})$, a distribution $\mathbb{P}$ with full support, and a mechanism φ that is upper invariant, monotonic, and sensitive, there exists $r > 0$ such that $\varphi^{\mathbb{P}}$ is r-partially strategyproof.*

This method of injecting randomness into the mechanism naturally extends the partial strategyproofness concept to mechanisms that are upper invariant, monotonic, and sensitive, even if they are deterministic. In Section 2.12, we use this extension to obtain partial strategyproofness of four mechanisms: first, for the school choice problem, these are the deterministic versions of the naïve and the adaptive Boston mechanism. Second, for the multi-unit assignment problem, the Probabilistic Serial mechanism and the HBS Draft mechanism are not swap monotonic. However, we show that all of them are upper invariant, monotonic, and sensitive.

2.12 Applications of Partial Strategyproofness

We now apply our new partial strategyproofness concept to a number of popular and new mechanisms. Table 2.1 provides an overview of our results.

2.12.1 Random Serial Dictatorship

Random Serial Dictatorship is known to be strategyproof. Thus, it satisfies all three axioms and is 1-partially strategyproof for any setting.

2.12.2 Probabilistic Serial

Upper invariance of Probabilistic Serial (PS) mechanism follows from Theorem 2 of Hashimoto et al. (2014). Our next Proposition 6 yields swap monotonicity.

Proposition 6. *PS is swap monotonic.*

Mechanism	Source of Randomness	UI	PSP	SP
Random Serial Dictatorship	Priorities (single, uniform)	✓	✓	✓
Probabilistic Serial	Mechanism	✓	✓	✗
Naïve Boston (random)	Priorities (single, uniform)	✓	✗	✗
Adaptive Boston (random)	Priorities (single, uniform)	✓	✓	✗
Rank Value	Mechanism (& Preferences)	✗	✗	✗
Hybrids of RSD & PS	Mixing (& other)	✓	✓	✗
Hybrids of RSD & ABM	Mixing (& other)	✓	✓	✗
Hybrids of RSD & NBM	Mixing (& other)	✓	✗	✗
Hybrids of RSD & RV	Mixing (& other)	✗	✗	✗
Naïve Boston (det.)	Preferences	✓	✓	✗
Adaptive Boston (det.)	Preferences	✓	✓	✗
Multi-unit PS	Preferences (& Mechanism)	✓	✓	✗
HBS Draft	Preferences (& Priorities)	✓	✓	✗

Table 2.1: Application of partial strategyproofness to popular and new mechanisms (*UI*: upper invariant, *PSP*: partially strategyproof, *SP*: strategyproof)

Proof Outline (formal proof in Appendix 2.F.13). We consider the times at which objects are exhausted under the Simultaneous Eating algorithm. Suppose an agent swaps two objects, e.g., from $P_i : a \succ b$ to $P_i' : b \succ a$. If anything changes about that agent's assignment, the agent will now spend *strictly more* time consuming b. We show that by the time b is exhausted, there will be *strictly less* of a available or there will be *strictly more competition* at a (relative to reporting $P_i : a \succ b$). □

Since PS is known to be manipulable (as evident from the motivating example in the introduction), it is not strategyproof, and hence, by Theorem 1, it cannot be lower invariant in general. However, since it is swap monotonic and upper invariant, it is partially strategyproof by Theorem 2. This is a stronger statement than weak SD-strategyproofness, and it is also stronger than recent findings by Balbuzanov (2015), who showed that PS is *convex strategyproof*.

Kojima and Manea (2010) have shown that for a fixed number of objects m and an agent i with a fixed utility function u_i, i will not want to misreport under PS if there are sufficiently many copies of each object. Since u_i is fixed, this does not mean that PS becomes strategyproof in some finite setting. However, we conjecture that the result of Kojima and Manea (2010) can be strengthened in the following sense: for m constant and $\min_{j \in M} q_j \to \infty$ we have $\rho_{(N,M,\boldsymbol{q})}(\text{PS}) \to 1$. We found numerical evidence that supports this conjecture. Figure 2.3 shows the development of the degree of strategyproofness $\rho_{(N,M,\boldsymbol{q})}(\text{PS})$ as the size of settings increases. We observe that in both cases, the degree

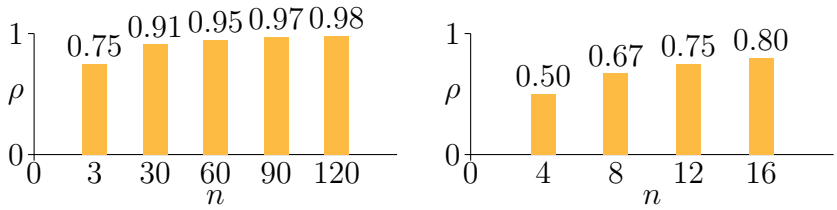

Figure 2.3: Plot of $\rho_{(N,M,\boldsymbol{q})}(\text{PS})$ for $m = 3$ (left) and $m = 4$ (right) objects, for varying numbers of agents n, and capacities $q_j = n/m$.

of strategyproofness increases and appears to converge to 1.

2.12.3 "Naïve" Boston Mechanism

We consider the Boston mechanism with priorities determined by a single, uniform lottery (Mennle and Seuken, 2015*d*). Intuitively, this mechanism is upper invariant because the object to which an agent applies in the kth round has no effect on the applications or assignments in previous rounds (see (Mennle and Seuken, 2015*d*) for a formal proof). The Boston mechanism is, however, neither swap monotonic nor lower invariant, as the following Example 3 shows. Thus, the it is not r-partially strategyproof for any $r > 0$.

Example 3. Consider the setting in which 4 agents, 4 objects with unit capacity, and preferences

$$\begin{aligned} P_1 &: a \succ b \succ c \succ d, \\ P_2 &: a \succ c \succ b \succ d, \\ P_3, P_4 &: b \succ c \succ a \succ d. \end{aligned}$$

Agent 1's assignment vector is $(1/2, 0, 0, 1/2)$ for the objects a, b, c, d, respectively. If agent 1 swaps b and c in its report, its assignment vector change to $(1/2, 0, 1/4, 1/4)$. First, note that the probability for b has not changed but the overall assignment has which violates swap monotonicity. Second, the assignment of d has changed, even though it is in the lower contour set of c, which violates lower invariance.

The Boston mechanism is "naïve," since it lets agents apply at their second, third, etc. choices, even if these have already been exhausted in previous rounds, such that agents "waste" rounds. Therefore, we refer to it as the *naïve Boston mechanism (NBM)*.

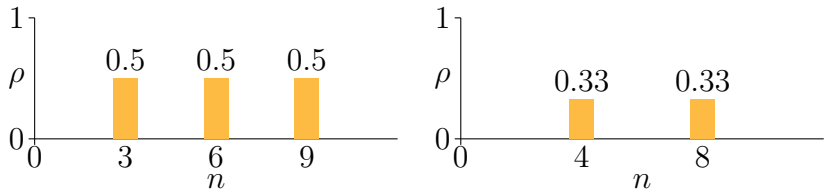

Figure 2.4: Plot of $\rho_{(N,M,\boldsymbol{q})}(\text{ABM})$ for $m = 3$ (left) and $m = 4$ (right) objects, for varying numbers of agents n, and for $q_j = n/m$.

2.12.4 Adaptive Boston Mechanism

Obvious manipulation strategies arise from this naïve approach of NBM: an agent who knows that its second choice will already be exhausted in the first round is better off ranking its third choice second because this will increase its chances at all remaining objects in a stochastic dominance sense without forgoing any chances at its second choice object. If instead, the agent automatically "skipped" exhausted objects in the application process, this manipulation strategy would no longer be effective.

In (Mennle and Seuken, 2015*d*) we have shown that such an *adaptive Boston mechanism (ABM)* is swap monotonic and upper invariant, and thus partially strategyproof. However, since ABM is not strategyproof, it cannot be lower invariant. Miralles (2008) used simulations to study how unsophisticated (truthful) agents are disadvantaged under the naïve Boston mechanism and found evidence that such an adaptive correction may be attractive. Indeed, in (Mennle and Seuken, 2015*d*) we have proven that ABM has *intermediate efficiency* between RSD and NBM. Since RSD is strategyproof while NBM is not even weakly strategyproof, ABM offers a trade-off between strategyproofness and efficiency. Partial strategyproofness enabled a formal understanding of this trade-off.

We have computed the degree of strategyproofness of ABM for various setting, and the results are shown in Figure 2.4. We observe that $\rho_{(N,M,\boldsymbol{q})}(\text{ABM})$ is significantly lower than $\rho_{(N,M,\boldsymbol{q})}(\text{PS})$ (Figure 2.3).[8] Furthermore, it does not grow for larger settings, but it appears to remain constant (i.e., $\rho_{(N,M,\boldsymbol{q})}(\text{ABM}) = 1/2$ for $m = 3$ objects and $n = 3, 6, 9$ agents, and $\rho_{(N,M,\boldsymbol{q})}(\text{ABM}) = 1/3$ for $m = 4$ objects and $n = 4, 8$ agents). Thus, ABM has *intermediate* incentive guarantees, which are stronger than those of NBM but weaker than those of PS or RSD.

[8] We computed the degree of strategyproofness for small settings with $m \in \{3, 4\}$ objects and $n \in \{3, 6, 9\}$ and $n \in \{4, 8\}$ agents, respectively. The computational cost of determining ABM is prohibitively high in larger settings.

2.12.5 Rank Efficient Mechanisms

Featherstone (2011) introduced rank efficiency, a strict refinement of ex-post and ordinal efficiency. Rank efficiency often closely reflects the welfare criteria used in practical applications, e.g., in school choice (Mennle and Seuken, 2015*d*), or in the assignment of teachers to schools (Featherstone, 2011). However, no rank efficient mechanism is even weakly strategyproof (Theorem 3 in (Featherstone, 2011)). Furthermore, any rank efficient mechanism will be neither swap monotonic, nor upper invariant, nor lower invariant (see Examples 4 and 5 in Appendix 2.D). Thus, they will not be partially strategyproof. Consequently, the attractive efficiency properties come at a price as such mechanisms will fail all of the axioms.

2.12.6 Hybrid Mechanisms

In (Mennle and Seuken, 2015*a*), we have shown how hybrid mechanisms can facilitate the trade-off between strategyproofness and efficiency for assignment mechanisms. The main idea of hybrid mechanisms is to consider convex combinations of two different mechanisms, one of which has good incentives while the other brings good efficiency properties. Under certain technical conditions, the resulting hybrid mechanisms are partially strategyproof but can also improve efficiency beyond the ex-post efficiency of Random Serial Dictatorship. Furthermore, the trade-off is scalable in the sense that the mechanism designer can accept a lower degree of strategyproofness in exchange for more efficiency. Note that prior to the introduction of partial strategyproofness, no measure existed to evaluate the incentive properties of such hybrid mechanisms.

2.12.7 Deterministic Boston Mechanisms

Even though random tie-breaking plays an important role in school choice mechanisms, many insights arise from the study of deterministic variants of these mechanisms. This is particularly true for the two variants of the Boston mechanism. We have already observed that when priorities are determined by a single, uniform lottery, then the adaptive Boston mechanism is partially strategyproof while the naïve Boston mechanism is not. However, if priorities are fixed and strict, both mechanisms are deterministic and manipulable. Consequently, both mechanisms have a degree of strategyproofness of 0 in this case.

In order to compare the two deterministic Boston mechanisms by their incentive properties, we must resort to a second source of randomness. In (Mennle and Seuken, 2015*d*), we have shown that both mechanisms are in fact upper invariant, monotonic, and

sensitive. Thus, we can apply Theorem 8 and get that $\text{NBM}^{\mathbb{P}}$ and $\text{ABM}^{\mathbb{P}}$ are partially strategyproof for any distribution $\mathbb{P}$ on the preference reports $\mathcal{P}^{N\setminus\{i\}}$ with full support. This allows a comparison of their incentive properties, even if priorities are not random (or not "random enough"). Fixing a setting $(N, M, \boldsymbol{q})$, a distribution $\mathbb{P}$, and a priority structure, a mechanism designer can compute the degree of strategyproofness of each mechanism to identify which has the better incentive guarantees.

2.12.8 Multi-unit Assignment Mechanisms

The multi-unit assignment problem is an important extension of the assignment problem, where each agent receives a bundle of K objects. For example, university students typically take several different courses each semester. If the number of participants in each course is limited, it may not be possible to accommodate every student's favorite schedule. Instead, the university may use a mechanism to elicit preference over courses from the students and assign course schedules.

Under the *HBS Draft* mechanism, agents take turns to draw objects that they like. After all agents have taken one object, the order in which they draw their second object is reversed (Budish and Cantillon, 2012). A second mechanism for the multi-unit assignment problem is a straightforward extension of the *Probabilistic Serial* mechanism (Heo, 2014): as in the case of single-unit demand, agents collect probability shares of objects. However, in the multi-unit version, they collect a total of K units (instead of 1). Moreover, two events trigger the move of an agent to another object: first, this still happens when an object is exhausted. Second, this also happens when the agent has collected probability of 1 of the object it is currently consuming.

Neither the HBS Draft nor the Probabilistic Serial mechanism are strategyproof for the multi-unit assignment problem. Furthermore, while both mechanisms are upper invariant and may involve randomization, they are not swap monotonic. However, they satisfy monotonicity and sensitivity.

Proposition 7. *Given a setting $(N, M, \boldsymbol{q})$ with $\sum_{j\in M} q_j = n \cdot K$, the Probabilistic Serial mechanism and the HBS Draft mechanism for the K-unit assignment problem are upper invariant, monotonic, and sensitive. However, neither of them is swap monotonic.*

The formal proof is given in Appendix 2.E.

By Theorem 8, we can apply the partial strategyproofness concept to both mechanisms when the preference reports from the other agents are drawn from a distribution $\mathbb{P}$ with

full support. In future work, it will be interesting to study the relation of their respective degrees of strategyproofness numerically and analytically.

2.13 Conclusion

In this paper, we have presented a new, axiomatic approach to studying the incentive properties of non-strategyproof assignment mechanisms.

First, we have shown that a mechanism is strategyproof if and only if it satisfies the three axioms swap monotonicity, upper invariance, and lower invariance. This illustrates why strategyproofness is such a strong requirement: if an agent swaps two consecutive objects, e.g., from "$a \succ b$" to "$b \succ a$," in its reported preference order, the only thing that a strategyproof mechanism can do (if anything) is to increase that agent's assignment of b and decrease its assignment of a by the same amount.

Towards relaxing strategyproofness, we have shown that by dropping the least important axiom, lower invariance, the class of partially strategyproof mechanisms emerges: a mechanism is *r-partially strategyproof* if it makes truthful reporting a dominant strategy for agents who have sufficiently different values for different objects. These are precisely the agents whose vNM utility functions satisfy the URBI(r) constraint. The set of partially strategyproof mechanisms is characterized by the axioms swap monotonicity and upper invariance. Consequently, under a swap monotonic, upper invariant mechanism agents are best off when reporting truthfully if they are not too indifferent between different object. This provides an economic intuition for the partial strategyproofness concept and allows us to give honest and useful strategic advice to agents.

We have proven that the URBI(r) domain restriction is maximal: for r-partially strategyproof mechanisms, URBI(r) is the largest set of utility functions for which one can guarantee that truthful reporting is a dominant strategy without knowledge of further properties of the mechanism. This maximality result has allowed us to define the degree of strategyproofness, a meaningful measure for incentive properties when mechanisms are not strategyproof. This measure is parametric, computable, and it is consistent with the method of comparing mechanisms by their vulnerability to manipulation (Pathak and Sönmez, 2013).

Next, we have established three appealing properties for partial strategyproofness: first, many important incentive concepts are defined via dominance notions, including strategyproofness, which is equivalent to stochastic dominance-strategyproofness. We have defined the notion of *r-partial dominance*, which is similar to stochastic domi-

nance, but the influence of less preferred choices is discounted by the factor r. We have shown that r-partial strategyproofness can alternatively be defined via this dominance notion. With an alternative definition that is independent of utility functions, partial strategyproofness integrates nicely with other dominance-based incentive concepts. Furthermore, this equivalence unlocks partial strategyproofness to algorithmic analysis. We have provided algorithms (see Appendix 2.C) that exploit this equivalence to verify partial strategyproofness and to compute the degree of strategyproofness of a mechanism (in any fixed setting).

Second, comparing partial strategyproofness to existing incentive concepts, we have established it as an *intermediate* concept: it is implied by *strategyproofness*, but in turn it implies many other incentive concepts, such as *weak*, *convex*, and *approximate strategyproofness*, as well as *strategyproofness in the large* if the degree of strategyproofness approaches 1 as the settings grow. Moreover, we have shown that strategyproofness and DL-strategyproofness are the upper and lower limit concepts, respectively. Thus, partial strategyproofness parametrizes the whole space between strategyproof and SL-strategyproof mechanisms.

Third, we have considered local sufficiency, a property of incentive concepts that is appealing from an axiomatic as well as from an algorithmic perspective. We have shown that r-local partial strategyproofness implies r^2-partial strategyproofness, and that the bound "2" in this implication is tight. Put differently, there exists no $\varepsilon > 0$ such that $r^{2-\varepsilon}$-partial strategyproofness can also be guaranteed. Prior work has shown that local sufficiency holds for both limit concepts, strategyproofness and DL-strategyproofness; our local sufficiency result for partial strategyproofness yields a unified proof for both statements.

Finally, we have applied the partial strategyproofness concept to gain a better understanding of the incentive properties of popular, non-strategyproof mechanisms. We have shown that the Probabilistic Serial mechanism is partially strategyproof, which is a significantly better description of its incentive properties than weak SD-strategyproofness. While the Boston mechanism in its naïve form (NBM) is not even weakly strategyproof, an adaptive variant (ABM) is in fact partially strategyproof, and we have presented numerical evidence that ABM has intermediate incentive guarantees, which are stronger than those of NBM but weaker than those of PS. Rank Value mechanisms violate all three axioms and are therefore not partially strategyproof. These examples demonstrate that partial strategyproofness reflects our intuitive understanding of what it means for a non-strategyproof mechanism to have "good" incentive properties. We have argued

that partial strategyproofness can also be used to measure the incentive properties of new hybrid mechanisms, which enable a parametric trade-off between strategyproofness and efficiency of random assignment mechanisms. Last, we have demonstrated that the partial strategyproofness concept can be extended to deterministic (and "less random") mechanisms, such as the deterministic variants of the two Boston mechanisms, as well as the HBS Draft mechanism and the Probabilistic Serial mechanism for the multi-unit assignment problem.

Our new partial strategyproofness concept has an axiomatic motivation and, as we have shown in this paper, it is in multiple ways a powerful addition to the toolbox of the mechanism designer. We believe this will lead to new insights in the analysis of existing non-strategyproof assignment mechanisms and facilitate the design of new ones.

Appendix for Chapter 2

2.A Probabilistic Serial Mechanism

The *Probabilistic Serial* mechanism (Bogomolnaia and Moulin, 2001) first collects the preference reports from all agents. Then it treats the objects *as if* they were divisible and uses the following *simultaneous eating algorithm* to determine a random assignment.

- At time 0 all agents begin consuming probability shares of their respective first choice objects at equal speeds.
- At some time, t_1 say, some object is exhausted. At this time all the agents who were consuming this object before t_1 move on to their respective second choice.
- Every time an object is exhausted the agents consuming this object move on to their respective next choices.
- If some agent's next choice is also exhausted, it immediately continues to its next best *available* choice.
- At time 1 all agents will have collected a total of 1 in probability shares from different objects.
- The entry $x_{i,j}$ in the assignment matrix is equal to the amount of j that i was able to consume during the process.

2.B Continuation of Example 1

We construct a mechanism (from the perspective of a single agent) that is swap monotonic but manipulable in a first order-stochastic dominance sense. For this consider a setting with 4 objects a, b, c, d. If the agent reports a preference of b over c then its assignment is $(0, 1/2, 0, 1/2)$ for a, b, c, d, respectively; if the agents reports a preference of c over b, then its assignment changes to $(1/2, 0, 1/2, 0)$.

First, note that this is a continuation of the mechanism from Example 1. Therefore, it is manipulable in a first order-stochastic dominance sense. Second, it is swap monotonic. To see this observe that the only swap that will cause a change of assignment is a swap of b and c. In this case the probabilities for getting b and c have to change strictly and in the right direction, which is obviously the case.

2.C Algorithms for Verifying Partial Strategyproofness and Computing the Degree of Strategyproofness

ALGORITHM 1: Verify r-partial strategyproofness

```
Input: setting (N, M, q), mechanism φ, bound r
Variables: agent i, preference profile (P_i, P_{-i}), misreport P'_i, vector Δ, counter k,
  choice function ch : {1, ..., m} → M
begin
    for i ∈ N, (P_i, P_{-i}) ∈ 𝒫^N, P'_i ∈ 𝒫 do
        ∀j ∈ M : Δ_j ← φ_{i,j}(P_i, P_{-i}) − φ_{i,j}(P'_i, P_{-i})
        for k ∈ {1, ..., m} do
            if Σ_{l=1}^{k} r^l · Δ_{ch(l)} < 0 then
                return false
            end
        end
    end
    return true
end
```

Remark 7. Algorithm 1 is straightforward: it verifies all (finitely many) r-partial dominance constraints.

Algorithm 2 optimistically sets ρ to 1, then iterates through all partial dominance constraints. Each constraint is understood as a polynomial $f(r)$ in r with y-intersect $f(0) = \Delta_{k_{\min}} > 0$. If the current guess of ρ is too high, one of the polynomials will have a negative value $f(\rho) < 0$ at ρ. In this case, the guess of ρ is updated to the smallest positive real root of f. Note that since $f(0)$ is always strictly positive, this root exists, is positive, and can be found in polynomial time, e.g., using the LLL-algorithm (Lenstra, Lenstra and Lovász, 1982)

ALGORITHM 2: Compute $\rho_{(N,M,\boldsymbol{q})}(\varphi)$

Input: setting $(N, M, \boldsymbol{q})$, mechanism φ
Variables: agent i, preference profile (P_i, P_{-i}), misreport P_i', vector Δ, counters $k, k_{\min}, l$, bound ρ, polynomial f, choice function $\mathrm{ch} : \{1, \ldots, m\} \to M$
begin
  $\rho \leftarrow 1$
  for $i \in N, (P_i, P_{-i}) \in \mathcal{P}^N, P_i' \in \mathcal{P}$ **do**
    $\forall j \in M : \Delta_j \leftarrow \varphi_{i,j}(P_i, P_{-i}) - \varphi_{i,j}(P_i', P_{-i})$
    $k_{\min} \leftarrow \min\{k : \Delta_{\mathrm{ch}(k)\neq 0}\}$
    for $k \in \{k_{\min}, \ldots, m\}$ **do**
      $f(r) \leftarrow \sum_{l=1}^{k} r^{l-k_{\min}} \cdot \Delta_{\mathrm{ch}(l)}$
      $\rho \leftarrow \min(\rho, \min\{r | r > 0 \text{ with } f(r) = 0\})$
    end
  end
  return ρ
end

2.D Examples for Rank Efficient Mechanism

Example 4. Consider the setting $N = \{1, 2, 3, 4\}$, $M = \{a, b, c, d\}$, $q_j = 1$, and the preference profile

$$\begin{aligned} P_1 &: a \succ d \succ c \succ b, \\ P_2 &: a \succ b \succ d \succ c, \\ P_3 &: b \succ c \succ d \succ a, \\ P_4 &: c \succ a \succ b \succ b. \end{aligned}$$

The unique rank efficient assignment is $d \to 1, a \to 2, b \to 3, c \to 4$. Suppose agent 1 changes its report to

$$P_1'' \;:\; a \succ c \succ b \succ d.$$

Now the only rank efficient assignment is $a \to 1, d \to 2, b \to 3, c \to 4$. The reports P_1 and P_1'' differ by two swaps: $d \leftrightarrow c$ and $d \leftrightarrow b$. Thus, at least one of these swaps must have increased the likelihood of getting object a for agent 1. This contradicts upper invariance. Also, under no report out of $P_1, P_1' : a \succ c \succ d \succ b, P_1''$ did agent 1 have any probability of getting objects b or c. Hence, the swap that changes the assignment involved a change of position of either object b or c, but the probability for each remained

zero, a contradiction to swap monotonicity.

Example 5. Consider the setting $N = \{1, 2, 3, 4, 5\}$, $M = \{a, b, c, d, e\}$, $q_j = 1$, and the preference profile

$$\begin{aligned} P_1 &: a \succ c \succ b \succ d \succ e, \\ P_2 &: c \succ b \succ a \succ d \succ e, \\ P_3 &: c \succ a \succ b \succ e \succ d, \\ P_4 &: a \succ c \succ b \succ e \succ d, \\ P_5 &: e \succ a \succ b \succ c \succ d. \end{aligned}$$

The unique rank efficient assignment is $d \to 1, b \to 2, c \to 3, a \to 4, e \to 5$.

Agent 1 could change its report to

$$P_1'' \;:\; b \succ a \succ c \succ d \succ e,$$

in which case $b \to 1, d \to 2, c \to 3, a \to 4, e \to 5$ is the unique rank efficient assignment. Hence, either the swap $c \leftrightarrow b$ or the swap $a \leftrightarrow b$ changed the assignment for d, a contradiction to lower invariance.

2.E Multi-unit Assignment Mechanisms

To introduce *K-unit assignment mechanisms* formally, we must extend our model slightly. Instead of one object, each agent should receive a bundle of K objects. We assume that the agents have additive valuations over bundles. Then it is meaningful to consider ordinal mechanism, where each agent submits a preference order over objects. An assignment is represented by an $n \times m$-matrix x, where $\sum_{i \in N} x_{i,j} = q_j$ for all $j \in M$, $\sum_{j \in M} x_{i,j} = K$ for all $i \in N$, and with the additional constraint that each agent should receive at most one copy of each object (i.e., $x_{i,j} \in [0, 1]$). By virtue of the Birkhoff-von Neumann Theorem and its extensions (Budish et al., 2013), these assignments are implementable via lotteries. We restrict attention to the assignment of "scarce" objects, i.e., objects for which the capacity q_j is lower than the total number of agents n. This is justified because for any objects with $q_j \geqslant n$ we can simply distribute one copy to each agent, independent of their preferences.

2.E.1 Probabilistic Serial Mechanism for K-unit Assignment

For the K-unit assignment problem, the *Probabilistic Serial* mechanism takes as input a preference profile $\boldsymbol{P}$ and determines an assignment as follows:

- The objects are treated *as if* they were divisible. All agent begin consuming *probability shares* of their respective most preferred object at equal speeds.
- When an agent has consumed a total of 1 in probability shares from an object, this agent leaves the object and continues consuming probability shares of its next most preferred object that still has remaining capacity.
- When an object is completely consumed, the agents from this object move on to their respective next most preferred objects that still have remaining capacity.
- This continues until all agents have collected probability shares that sum to K.
- The entry $x_{i,j}$ of the resulting assignment is given by the amount of shares of j that i managed to consume in this process.

2.E.2 HBS Draft Mechanism for K-unit Assignment

For the K-unit assignment problem, the *HBS Draft* mechanism is defined with respect to fixed priority order π over agents. Given π, the mechanism takes a preference profile $\boldsymbol{P}$ as input and determines an assignment as follows:

- At the beginning of the first *pass*, the agent with the highest priority (according to π) draws one copy of the object that it prefers most.
- The agent with the second highest priority draws one copy of the object that it prefers most out of all available objects.
- In the order given by π, all agents draw copies of their respective most preferred available objects.
- The first pass ends, when the last agent has drawn its first object.
- The second pass works like the first pass, but the order of π is reversed: the agent with the *lowest* priority draws a second object (but no second copy of object that it has drawn in the first pass).
- Then the agent with the second-lowest priority draws a second object, and so on.
- The third, fourth, etc. pass are analogous. In odd passes, the drawing order is π, while in even passes, the drawing order is the reverse of π.

- The process continues until all agents have received a total of K objects, which is the case after exactly K passes.

The HBS Draft mechanism can also be understood as a random mechanism if the priority order π is determined randomly.

Proof of Proposition 7. Given a setting $(N, M, \boldsymbol{q})$ with $\sum_{j \in M} q_j = n \cdot K$, the Probabilistic Serial mechanism and the HBS Draft mechanism for the K-unit assignment problem are upper invariant, monotonic, and sensitive. However, neither of them is swap monotonic.

Upper invariance and monotonicity of these mechanisms are straightforward. The proofs of sensitivity are more challenging. We first show sensitivity of Probabilistic Serial for multi-unit assignment.

Consider any preference orders $P_i \in \mathcal{P}, P_i' \in N_{P_i}$ with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$. Construct the preference profile (P_i, P_{-i}) by setting $P_{i'} = P_i$ for all $i' \neq i$. The assignment $\mathrm{PS}(P_i, P_{-i})$ simply gives all agents the same assignment vector. Precisely, $\mathrm{PS}_{i,j}(P_i, P_{-i}) = q_j/n$. Suppose that all agents start consuming shares of a_k at τ_{k-1} and start consuming shares of a_{k+1} at τ_{k+1}. If i reports P_i' instead, it will start consuming shares of a_{k+1} already at τ_{k-1}, but all other agents will arrive strictly later. Thus, i's assignment of a_{k+1} increases strictly. Furthermore, a_k will be exhausted by the other agents. Therefore, by the time i finishes consuming a_{k+1}, it will not receive any more shares of a_k. Thus, i's assignment of a_k decreases strictly. Since this construction holds for any preference order P_i, sensitivity of PS follows.

Next, we show sensitivity of the HBS Draft mechanism for multi-unit assignment. Consider any preference orders $P_i \in \mathcal{P}, P_i' \in N_{P_i}$ with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$. Suppose that some preference reports $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$, the assignment of i changes between P_i and P_i'.

First, observe that this change must involve either a_k or a_{k+1}: towards contradiction, assume that $\mathrm{HBSD}_i(P_i, P_{-i}) \neq \mathrm{HBSD}_i(P_i', P_{-i})$, but $\mathrm{HBSD}_{i,a_k}(P_i, P_{-i}) = \mathrm{HBSD}_{i,a_k}(P_i', P_{-i})$ and $\mathrm{HBSD}_{i,a_{k+1}}(P_i, P_{-i}) = \mathrm{HBSD}_{i,a_{k+1}}(P_i', P_{-i})$. Since all agents get exactly one object in each pass, the passes in which i received the different objects, including a_k and a_{k+1}, must be the same under P_i and P_i'. But then the order in which all agents received their objects must also remain the same, which implies that the assignment of i has not changes; a contradiction.

There are different ways in which the change can involve a_k and a_{k+1}. The following table gives an overview:

$\text{HBSD}_i(P_i, P_{-i})$		$\text{HBSD}_i(P'_i, P_{-i})$		
a_k	a_{k+1}	a_k	a_{k+1}	Comment
0	0	0	1	Case 1 (impossible)
1	0	0	0	Case 1 (impossible)
1	0	0	1	Case 2
1	0	1	1	Case 3
1	1	0	1	Symmetric to Case 3

The omitted cases can be ruled out, because they violate monotonicity or because there is no change. The remaining cases are discussed next.

Case 1: When i gets neither a_k nor a_{k+1}, i tried to draw both objects in the same pass. Since i did not get a_k, it immediately attempted to draw a_{k+1}, which it did not get either. Thus, if i tries to draw a_{k+1} first (in the same pass), it will still not receive it there, move on to a_k immediately, and also not get it. This shows that Case 1 is impossible.

Case 2: If it holds that $\text{HBSD}_{i,a_k}(P_i, P_{-i}) = \text{HBSD}_{i,a_{k+1}}(P'_i, P_{-i}) = 1$ and $\text{HBSD}_{i,a_{k+1}}(P_i, P_{-i}) = \text{HBSD}_{i,a_k}(P'_i, P_{-i}) = 0$, then we have already found preference reports P_{-i} where the change in the assignment of a_k and a_{k+1} is strict for both. This is what we need to show for sensitivity.

Case 3: This is the most complex case. By reporting P_i, i gets a_k, but by reporting P'_i, it gets both a_k and a_{k+1}. First, we "reduce" the preference reports P_{-i} to P^1_{-i} in such a way that agents first rank all the objects that they actually received. All other objects, which they do not receive are ranked below. The application process and resulting assignment of $\text{HBSD}(P_i, P_{-i})$ and $\text{HBSD}(P_i, P^1_{-i})$ are exactly the same, except that no agent (except i) tries to draw an object that is already exhausted.

Now, consider the pass in which i draws a_k (when reporting P_i). Note that i is not the last agent to draw an object in this pass (otherwise, i would draw a_k and then immediately draw again; thus, i would receive a_k and a_{k+1}, independent of the order in which it ranks them). At the beginning of the pass, a_k is either the object that i prefers most of all the objects with remaining capacity, or there are other objects with remaining capacity that i would prefer to a_k.

Case (I): Suppose that a_k is the object that i prefers most of all objects with remaining capacity at the beginning of the pass. Let q'_{a_k} be the remaining

capacity of a_k at the beginning of the pass. There are exactly $q_{a_k} - q'_{a_k}$ agents who have already received a_k in prior rounds. All other $(n-1) - (q_{a_k} - q'_{a_k})$ agents (except i) would attempt draw a_k if they ranked it in this round; and $(n-1) - (q_{a_k} - q'_{a_k}) \geqslant 1$ since the initial capacity q_{a_k} of a_k was at most $n-1$. We obtain the preference reports P^2_{-i} by changing the draw of q'_{a_k} agents in this round to a_k (if they do not already draw a_k). Finding such agents is possible, because $(n-1) - (q_{a_k} - q'_{a_k}) \geqslant q'_{a_k}$. In particular, we ensure that the last agent to draw in this round (as argued above, this agent cannot be i) draws the last copy of a_k.

With P^2_{-i} constructed in this way, i can draw either a_k or a_{k+1} in this round. But in the subsequent round, a_k will be exhausted. Thus, $\text{HBSD}_{i,a_k}(P_i, P^2_{-i}) = 1$, but $\text{HBSD}_{i,a_k}(P'_i, P^2_{-i}) = 0$. This shows that the assignment of a_k may be affected, which is what we needed for directness.

Case (II): Now suppose that at the beginning of the pass where i draws a_k there are other objects with remaining capacity that i would prefer to a_k. The reason for i to draw a_k instead must be that some other agents draw these objects between the beginning of the pass and i's draw. We can select an agent who draws one of the objects that i prefers to a_k (x, say) and change this agent's draw: denote by c the object that i draws after drawing a_k under $\text{HBSD}(P_i, P_{-i})$. Instead of x, the agent draws the copy of c that must still be available. Furthermore, there exists an agent who draws the last copy of a_{k+1} between the draw of i in this and the next pass. We change this agent's draw to a_k. This yields the preference reports P^3_{-i}.

Observe that under the preference profile (P_i, P^3_{-i}), i will draw x instead of a_k, one agent will draw c instead of x, and one agent will draw a_k instead of a_{k+1}. Thus, at the beginning of the next pass, there are no copies of objects that i prefers to a_k. Moreover, at the time when i draws next, there is one copy of a_{k+1} and one copy of a_k. This means that the next pass is a Case (I) pass, so that we can complete the proof by the construction as in Case (I).

This concludes the proofs of sensitivity for both mechanisms.

Finally, we show that neither of the mechanisms is swap monotonic. For strict, fixed priorities, NBM and ABM are deterministic mechanisms. On these mechanisms, swap monotonicity is equivalent to strategyproofness. Conversely, since the mechanisms are not strategyproof, they cannot be swap monotonic. To see that Probabilistic Serial and

the HBS Draft mechanism are not swap monotonic, consider the following: let there be four objects in unit capacity, a, b, c, d, two agents with preferences

$$\begin{aligned} P_1 &: a \succ b \succ c \succ d, \\ P_2 &: c \succ b \succ d \succ a, \end{aligned}$$

and each agent should get two objects. Under both mechanisms, the assignments are

$$\mathrm{PS}(P_1, P_2) = \mathrm{HBSD}(P_1, P_2) = \begin{pmatrix} 1 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1 & 1/2 \end{pmatrix}. \tag{21}$$

If agent 1 reports $P_1' : b \succ a \succ c \succ d$ instead, the assignments change to

$$\mathrm{PS}(P_1', P_2) = \mathrm{HBSD}(P_1', P_2) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}. \tag{22}$$

Thus, agent 1's assignments of b increase strictly under the swap, but its assignments for a do not decrease strictly. □

2.F Omitted Proofs

2.F.1 Proof of Proposition 1

Proof of Proposition 1. A deterministic mechanism φ is strategyproof if and only if it is swap monotonic.

Let φ be strategyproof, and consider an agent $i \in N$, a preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and a misreport $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a_k \succ a_{k+1}$, but $P_i' : a_{k+1} \succ a_k$, such that $\varphi_i(P_i, P_{-i}) \neq \varphi_i(P_i', P_{-i})$. Let $j \in M$ be the object that i obtains with truthful reporting, and let $j' \neq j$ be the object i obtains by reporting P_i'. There are four cases:

Case $j = a_k, j' = a_{k+1}$: This is consistent with swap monotonicity.

Case $j = a_k, j' \neq a_{k+1}$: If $P_i : j' \succ a_{k+1}$, then $P_i : j \succ a_k$. Thus, i can obtain an object that it strictly prefers to a_k by reporting P_i', a contradiction to strategyproofness. If $P_i : a_{k+1} \succ j'$, then an agent with preference order P_i' could obtain an object (namely a_k) which it strictly prefers to j' by reporting P_i instead of reporting P_i' truthfully, again a contradiction.

Case $j \neq a_k, j' = a_{k+1}$**:** This is symmetric to the previous case.

Case $j \neq a_k, j' \neq a_{k+1}$**:** If $P_i : j \succ j'$, then $P_i' : j \succ j'$ as well. Thus, an agent with preference order P_i' could manipulate by reporting P_i. Conversely, if $P_i : j' \succ j$, then i with preference order P_i could manipulate.

Thus, any strategyproof mechanism will be swap monotonic.

To see necessity, observe that swap monotonicity requires the assignment of a_k and a_{k+1} to change strictly (if there is any change at all under the swap). But if their assignment changes strictly, this must be a change from 0 to 1 or from 1 to 0. Thus, upon any swap, a swap monotonic deterministic mechanism can only change the assignment by assigning the object that has been brought up in the ranking instead of the object that has been brought down. Since any misreport can be decomposed into a sequence of swaps, the mechanism must be strategyproof. □

2.F.2 Proof of Theorem 1

Proof of Theorem 1. A mechanism φ is strategyproof if and only if it is swap monotonic, upper invariant, and lower invariant.

φ SP $\Rightarrow$ φ SM, UI, LI: First, assume towards contradiction that φ is not upper invariant. Then there exists some agent $i \in N$, some preference profile $\boldsymbol{P} = (P_i, P_{-i}) \in \mathcal{P}^N$, and some misreport $P_i' \in N_{P_i}$ in the neighborhood of agent i's true preference order such that

$$\begin{aligned} P_i &: a_1 \succ \ldots \succ a_{k-1} \succ a_k \succ a_{k+1} \succ a_{k+1} \succ \ldots \succ a_m, \\ P_i' &: a_1 \succ \ldots \succ a_{k-1} \succ a_{k+1} \succ a_k \succ a_{k+1} \succ \ldots \succ a_m, \end{aligned}$$

and for some $l < k$ we have $\varphi_{i,a_l}(P_i, P_{-i}) \neq \varphi_{i,a_l}(P_i', P_{-i})$. Without loss of generality, we can assume $\varphi_{i,a_l}(P_i, P_{-i}) < \varphi_{i,a_l}(P_i', P_{-i})$ (otherwise, we invert the roles of P_i and P_i'), and we let l be the minimal rank for which this inequality is strict. This implies that $\varphi_i(P_i, P_{-i})$ does not even weakly stochastically dominate $\varphi_i(P_i', P_{-i})$ at P_i, since

$$\sum_{P_i : j \succ a_l} \varphi_{i,j}(P_i', P_{-i}) > \sum_{P_i : j \succ a_l} \varphi_{i,j}(P_i, P_{-i}), \tag{23}$$

a contradiction to SD-strategyproofness of φ.

Second, a similar argument yields lower invariance of φ: again we find i, $\boldsymbol{P} = (P_i, P_{-i})$, $P_i' \in N_{P_i}$, and $l > k+1$ such that without loss of generality $\varphi_{i,a_l}(P_i, P_{-i}) > \varphi_{i,a_l}(P_i', P_{-i})$, and l is the largest rank for which this inequality is strict. Then $\varphi_i(P_i, P_{-i})$ does not even weakly stochastically dominate $\varphi_i(P_i', P_{-i})$ at P_i, since

$$\sum_{P_i : j > a_l} \varphi_{i,j}(P_i', P_{-i}) > \sum_{P_i : j > a_l} \varphi_{i,j}(P_i, P_{-i}). \tag{24}$$

Third, we observe that upper & lower invariance of φ imply that for any swap the mechanism may only change agent i's assignment for a_k and a_{k+1}, and therefore

$$\varphi_{i,a_k}(P_i, P_{-i}) - \varphi_{i,a_k}(P_i', P_{-i}) = \varphi_{i,a_{k+1}}(P_i', P_{-i}) - \varphi_{i,a_{k+1}}(P_i, P_{-i}). \tag{25}$$

If the change in probability $\varphi_{i,a_k}(P_i, P_{-i}) - \varphi_{i,a_k}(P_i', P_{-i})$ was negative, then a swap of a_k and a_{k+1} (from P_i to P_i') would simply give agent i more probability of a_k and less probability for a_{k+1}. But in this case, $\varphi_i(P_i', P_{-i})$ would strictly stochastically dominate $\varphi_i(P_i, P_{-i})$ at P_i, again a contradiction.

φ SM, UI, LI $\Rightarrow$ φ SP: First, consider any local misreport, i.e., an agent $i \in N$, a preference profile $\boldsymbol{P} = (P_i, P_{-i}) \in \mathcal{P}^N$, and a misreport $P_i' \in N_{P_i}$ in the neighborhood of agent i's true preference, such that $P_i : a \succ b$, but $P_i' : b \succ a$. Since φ satisfies all three axioms, we get that

- $\varphi_{i,a}(P_i, P_{-i}) \geqslant \varphi_{i,a}(P_i', P_{-i})$,
- $\varphi_{i,b}(P_i, P_{-i}) \leqslant \varphi_{i,b}(P_i', P_{-i})$,
- $\varphi_{i,a}(P_i, P_{-i}) - \varphi_{i,a}(P_i', P_{-i}) = \varphi_{i,b}(P_i', P_{-i}) - \varphi_{i,b}(P_i, P_{-i})$, and
- $\varphi_{i,j}(P_i, P_{-i}) = \varphi_{i,j}(P_i', P_{-i})$ for all $j \neq a, b$.

Consequently, $\varphi_i(P_i, P_{-i})$ stochastically dominates $\varphi_i(P_i', P_{-i})$ at P_i. This implies local strategyproofness of φ, which in turn implies strategyproofness of φ (Carroll, 2012).

□

2.F.3 Proof of Theorem 2

Proof of Theorem 2. Given a setting $(N, M, \mathbf{q})$, a mechanism φ is partially strategyproof (i.e., r-partially strategyproof for some $r > 0$) if and only if φ is swap monotonic and upper invariant.

Throughout the proof, we fix a setting $(N, M, \mathbf{q})$. First, we define

$$\delta = \min \left\{ |\varphi_j(P_i, P_{-i}) - \varphi_j(P_i', P_{-i})| \;\middle|\; \begin{array}{c} \forall i \in N, (P_i, P_{-i}) \in \mathcal{P}^N, \\ P_i' \in \mathcal{P}, j \in M : \\ |\varphi_j(P_i, P_{-i}) - \varphi_j(P_i', P_{-i})| > 0 \end{array} \right\}. \tag{26}$$

This is the smallest non-vanishing value by which the allocation of any object changes between two different preference orders that any agent could report. Since N, M, and $\mathcal{P}$ are finite, δ must be strictly positive (otherwise φ is constant).

φ SM, UI $\Rightarrow$ φ r-PSP for some $r > 0$: We must show that there exists $r \in (0, 1]$ such that no agent with utility in URBI(r) can benefit from misreporting. Suppose, agent i with preference order

$$P_i \;:\; a_1 \succ \ldots \succ a_K \succ b \succ c_1 \succ \ldots c_L$$

is considering the misreport P_i', and without loss or generality let b be the most preferred object for which the allocation changes, i.e., for all $k = 1, \ldots, K$

$$\varphi_{i,a_k}(P_i, P_{-i}) = \varphi_{i,a_k}(P_i', P_{-i}), \tag{27}$$

$$\varphi_{i,b}(P_i, P_{-i}) \neq \varphi_{i,b}(P_i', P_{-i}), \tag{28}$$

Such an object must exist, because otherwise the allocations would be equal under both reports and P_i' would not be a beneficial misreport.

Claim 1. *The allocation for b weakly decreases, i.e., $\varphi_{i,b}(P_i, P_{-i}) \geqslant \varphi_{i,b}(P_i', P_{-i})$.*

Since the allocation for b must change by assumption, a weak decrease implies a strict decrease. Thus, reporting P_i' instead of P_i will necessarily decrease the probability that agent i gets object b by at least δ. Non of the probabilities for the objects $a_1, \ldots, a_K$ are affected. Hence, in the best case (for agent i), all remaining probability is concentrated on c_1, i.e., the maximum utility gain for agent i is upper-bounded by

$$-\delta u_i(b) + u_i(c_1) - (1 - \delta) \min u_i. \tag{29}$$

The misreport P_i' is guaranteed *not* to be beneficial if the value in (29) is less than or equal to 0, or equivalently,

$$u_i(c_1) - \min u_i < \delta(u_i(b) - \min(u_i)). \tag{30}$$

This sufficient condition is satisfied by all utilities in URBI(r) with the choice of $r \leqslant \delta$. Consequently, the mechanism φ is r-partially strategyproof for any $r \leqslant \delta$.

It remains to be proven that Claim 1 holds.

Proof of Claim 1: Consider a preference order

$$P \;:\; a_1 \succ \ldots \succ a_m.$$

A transition from P to another preference order P' is a finite sequence of preference orders that starts with P and terminates with P', and in each step the relative ranking of exactly two objects is inverted. Formally,

$$P_0, P_1, \ldots, P_{K-1}, P_K, \tag{31}$$

is a *transition* from P to P' if

- $P_0 = P$ and $P_K = P'$,
- for all $k \in \{0, \ldots, K-1\}$ we have $P_k \in N_{P_{k+1}}$ and $P_{k+1} \in N_{P_k}$.

The *canonical transition* is a particular transition between two preference orders that is inspired the *bubble-sort* algorithm:

P_0**:** Set $P_0 = P$

P_k**:** Determine P_k based on P_{k-1} as follows:

- Let r be the rank where P_{k-1} and P' differ for the first time, i.e.,

$$P_{k-1} \;:\; j_1 \succ \ldots \succ j_{r-1} \succ j_r \succ \ldots, \tag{32}$$
$$P' \;:\; j_1 \succ \ldots \succ j_{r-1} \succ c \succ \ldots, \tag{33}$$

  such that $j_r \neq c$, and let c be the rth choice object under P'.
- Find c in the ranking P_{k-1}
- Construct P_k by swapping c up one rank, i.e., if

$$P_{k-1} \;:\; j_1 \succ \ldots \succ a \succ b \succ c \succ \ldots, \tag{34}$$
$$\text{then let } P_k \;:\; j_1 \succ \ldots \succ a \succ c \succ b \succ \ldots. \tag{35}$$

P_K**:** Stop if $P_k = P'$ for some k, then set $K = k$

To prove the Claim, consider the first part of canonical transition from P_i' to P_i: a_1 is swapped with its predecessors (under P_i') until it reaches its final position at the front of the ranking (under P_i). With each swap the share of a_1 that the agent receives from the respective misreport can only increase or stay constant, because the mechanism is swap monotonic. On the other hand, once a_1 is at the front of the ranking, the allocation of a_1 will remain unchanged during the rest of the transition. This is because φ is upper invariant, i.e., no change of order below the first position can affect the allocation of the first ranking object. Since by assumption the allocation for a_1 did not change between P_i and P_i', none of the swaps involving a_1 will have any effect on the allocation of a_1. But by swap monotonicity this means that none of these swaps will have any effect at all.

Next consider the second part of transition, where a_2 is brought into second position by swapping it upwards. The same argument can be applied to show that the overall allocation must remain unchanged under any of the swaps involving a_2. The same is true for $a_3, \ldots, a_K$. Thus, we arrive at a preference order

$$P_i'' \quad : \quad a_1 \succ \ldots \succ a_K \succ c_1' \succ \ldots c_{L'}' \succ b \succ c_{L'+1}' \succ \ldots \succ c_L'.$$

Under P_i'' all of the objects a_k are in the same positions as under P_i, b holds some position below its rank in P_i, and some of the c_l are ranking above b (possibly in a different order). From the previous argument we know that the overall allocation is the same between P_i' and P_i''. The next steps of the canonical transition will swap b with its predecessors until it reaches its final position just below a_K (as under P_i). During any of these swaps, the allocation for b has to increase weakly by swap monotonicity of φ.

Finally, any subsequent swaps in the transition occur strictly below b, and therefore, the allocation for cannot change any more until P_i is reached. Therefore, the allocation for b weakly increases between P_i' and P_i. □

φ r-PSP for some $r > 0 \Rightarrow \varphi$ SM, UI:

Upper invariance: Suppose φ is r-partially strategyproof for some fixed $r > 0$, i.e., no agent whose utility function satisfies URBI(r) can benefit from misreporting. We need to show that φ is upper invariant. Suppose, the agent has preference

order

$$P \;:\; \ldots \succ a \succ b \succ \ldots \succ c \succ d \succ \ldots.$$

Assume towards contradiction that a swap of c and d changes the allocation of some object ranked before c, and let a be the most preferred such object. Define δ as in (80), then without loss of generality the allocation of a increases by at least δ due to this swap (if it decreases, consider the reverse swap). This means that by swapping c and d, an agent with preference order P can gain at least probability δ for object a. Because a was the highest ranking object for which the allocation changed, the worst thing that can happen from the agent's perspective is that it looses all of its chances to get b and gets its last choice instead. Hence,

$$\delta u(a) - u(b) + (1-\delta)\min u \tag{36}$$

is a lower bound for the benefit that the agent can have from swapping c and d in its report. This misreport is guaranteed to be strictly beneficial if the value in (29) is strictly positive, or equivalently, if

$$u(b) - \min u < \delta(u(a) - \min u). \tag{37}$$

But for any $r \in (0,1]$, the set URBI(r) will contain a utility function satisfying this condition. This is a contradiction to the assumption that no agent with a utility function in URBI(r) will have a strictly beneficial manipulation. Consequently, φ must be upper invariant.

Swap monotonicity Suppose φ is r-partially strategyproof for some fixed $r \in (0,1]$. We know already that φ must be upper invariant. Towards contradiction, assume that upon a swap of two consecutive objects by some agent, the mechanism violates swap monotonicity, i.e., consider the preference order

$$P \;:\; \ldots \succ a \succ b \succ c \succ \ldots \succ d \succ d' \succ \ldots,$$

and let a and b be the objects that change position under the swap. For φ to violate swap monotonicity, one of the following must hold:

1. the allocation of a increases,

2. the allocation of a remains constant, and the allocation of b increases,
3. the allocation of a remains constant, and the allocation of b decreases,
4. the allocations of a and b remain constant, but the allocation changes for some object $d \neq a, b$,
5. the allocations of both a and b decrease.

We now consider each case separately and show that they all lead to contradictions.

Because of upper invariance, we know that the allocation of objects ranking above a cannot be affected. Therefore, in case 1, the agent can gain at least δ probability of getting a, with δ defined as in (80). Then the worst thing (for the agent) that could happen is that it looses all its chances of getting anything but its least preferred object. Hence,

$$\delta u(a) - u(b) + (1 - \delta) \min u \tag{38}$$

is a lower bound for the benefit the agent can have from swapping a and b. But as in the proof of upper invariance, this leads to a contradiction.

In case 2, the agent gains at least δ probability for b, but may loose shares in the next lower ranking object c. Again, the lower bound for the benefit is

$$\delta u(b) - u(c) + (1 - \delta) \min u \tag{39}$$

which leads to a contradiction. Note that if b is the lowest ranking object, this case is impossible.

Case 3 is symmetric to case 2, and we can consider the reverse swap instead.

In case 4, let d be the highest ranking object for which the allocation changes, which must lie after b because of upper invariance. Then without loss of generality, the agent can increase its chances of getting d by at least δ, but potentially looses all chances for the next lower ranking object d'. This again leads to a contradiction.

For case 5, we consider the reverse swap, which is covered by case 1.

In conclusion, we have shown that none of the cases 1 through 5 can occur under a mechanism that is r-partially strategyproof. Therefore, the mechanism must satisfy strict swap monotonicity. This concludes the proof of Theorem 2. □

2.F.4 Proof of Theorem 3

Proof of Theorem 3. For any setting $(N, M, \mathbf{q})$ with $m \geqslant 3$, any bound $r \in (0,1)$, and any utility function $\tilde{u}_i$ (consistent with a preference order P_i) that violates URBI(r) there exists a mechanism $\tilde{\varphi}$ such that

1. *$\tilde{\varphi}$ is r-partially strategyproof, but*

2. *there exist preferences of the other agents P_{-i} and a misreport P_i' such that*

$$\langle \tilde{u}_i, \tilde{\varphi}_i(P_i, P_{-i}) - \tilde{\varphi}_i(P_i', P_{-i}) \rangle < 0. \tag{40}$$

Furthermore, $\tilde{\varphi}$ can be chosen to satisfy anonymity.

By assumption, $\tilde{u}$ violates URBI(r). Thus, for some pair a, b of consecutive objects in the preference order P_i corresponding to $\tilde{u}$ we have

$$\frac{\tilde{u}(b) - \min \tilde{u}}{\tilde{u}(a) - \min \tilde{u}} = \tilde{r} > r. \tag{41}$$

Additionally, b is not the last choice of i, since the constraint $\frac{0}{\tilde{u}(a) - \min \tilde{u}} \leqslant r$ is trivially satisfied. We now need to define the mechanism $\tilde{\varphi}$ that offers a manipulation to agent i if its utility function is $\tilde{u} \in U_{P_i}$, but would not offer any manipulation to agent i if it has any utility satisfies URBI(r) (and possibly a different preference order). For partial strategyproofness, an agent should not have a beneficial manipulation for any set of reports from the other agents. Thus, it suffices to specify $\tilde{\varphi}$ for a single set of reports P_{-i}, where only agent i can vary its report. The allocation for i must then be specified for any possible report $\hat{P}_i$ from i.

We define $\tilde{\varphi}_i(\cdot, P_{-i})$ as follows:

- For a report $\hat{P}_i$ with $a \succ b$,

$$\tilde{\varphi}_i(\hat{P}_i, P_{-i}) = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right). \tag{42}$$

- For a report $\hat{P}_i$ with $b \succ a$, we adjust the original allocation by

$$\tilde{\varphi}_i(\hat{P}_i, P_{-i})(a) = \frac{1}{m} + \delta_a, \tag{43}$$

$$\tilde{\varphi}_i(\hat{P}_i, P_{-i})(b) = \frac{1}{m} + \delta_b, \tag{44}$$

$$\tilde{\varphi}_i(\hat{P}_i, P_{-i})(d) = \frac{1}{m} + \delta_d, \tag{45}$$

  where $\delta_a < 0$, $\delta_b > -\delta_a$, $\delta_d = -\delta_a - \delta_b < 0$. Here d denotes the last choice. In case $a = d$, both δ_a and δ_d are added. Note that if the last object changes, the allocation for the new last object is decreased (by adding δ_d), and the allocation of the previous last object is increased (by adding δ_d).

This mechanism is upper invariant: swapping the order of a and b induces a change in the allocation of a,b, and the last object d. Therefore no higher ranking object is affected. Swapping the last and the second to last object also only changes the allocation for these two object.

This mechanism is also swap monotonic: swapping a and b changes the allocation for both objects in the correct way, since $\delta_a < 0, \delta_b > 0$. Swapping the last to objects also changes the allocation appropriately, since $\delta_d < 0$. No other change of report changes the allocation.

Now we analyze the incentives for the different possible utility functions i could have:

Case $u_i = \tilde{u} \in U_{P_i}$: In this case, the true preference order is $P_i : a \succ b$. Swapping a and b in its order is beneficial for i if

$$\delta_a \tilde{u}(a) + \delta_b \tilde{u}(b) + \delta_d \tilde{u}(d) = \delta_a \left(\tilde{u}(a) - \min \tilde{u}\right) + \delta_b \left(\tilde{u}(b) - \min \tilde{u}\right) > 0 \tag{46}$$

$$\Leftrightarrow \delta_a > -\delta_b \frac{\tilde{u}(b) - \min \tilde{u}}{\tilde{u}(a) - \min \tilde{u}}. \tag{47}$$

(47) is satisfied if

$$\delta_a > -\delta_b \cdot \tilde{r}, \tag{48}$$

since $\frac{\tilde{u}(b) - \min \tilde{u}}{\tilde{u}(a) - \min \tilde{u}} = \tilde{r}$ by construction.

Case $u_i \in$ URBI(r), $P_i : a \succ b$: Swapping a and b should no longer be beneficial for i. This is the case if

$$\delta_a u_i(a) + \delta_b u_i(b) + \delta_d u_i(d) = \delta_a \left(u_i(a) - \min u_i\right) + \delta_b \left(u_i(b) - \min u_i\right) \leqslant 0 \tag{49}$$

$$\Leftrightarrow \delta_a \leqslant -\delta_b \frac{u_i(b) - \min u_i}{u_i(a) - \min u_i}. \tag{50}$$

(50) is satisfied if

$$\delta_a \leqslant -r, \tag{51}$$

since $\frac{u_i(b)-\min u_i}{u_i(a)-\min u_i} \leqslant r$ by construction.

Case $u_i \in$ URBI(r), $P_i : b \succ a$: Swapping b and a to $a \succ b$ should not be beneficial for i. This is the case if

$$\delta_b u_i(b) + \delta_a u_i(a) + \delta_d u_i(d) = \delta_a \left(u_i(a) - \min u_i\right) + \delta_b \left(u_i(b) - \min u_i\right) \geqslant 0 \tag{52}$$

$$\Leftrightarrow \delta_a \geqslant -\delta_b \frac{u_i(b) - \min u_i}{u_i(a) - \min u_i}. \tag{53}$$

(53) is satisfied if

$$\delta_a \geqslant -\frac{\delta_b}{\tilde{r}}, \tag{54}$$

since $\frac{u_i(a)-\min u_i}{u_i(b)-\min u_i} \leqslant r$ for agents with $b \succ a$ by construction.

This means that if δ_a and δ_b satisfy (48), (51), and (54), the mechanism $\tilde{\varphi}$ is in fact what we are looking for. Given some $\delta_b > 0$, we can choose δ_a appropriately, since $r < 1$, $r < \tilde{r}$, and

$$-\delta_b \cdot \tilde{r} < -\delta_b \cdot r \Leftrightarrow r < \tilde{r}, \qquad -\delta_b r > \frac{-\delta_b}{r} \Leftrightarrow r^2 < r. \tag{55}$$

Fixing the allocation for agent i in this manner, we can distribute all remaining probability for the objects evenly across all other agents, independent of their reports P_{-i}. Then no other agent but i has any influence on their own allocation, i.e., the mechanism is constant from the perceptive from all other agents but i. Finally, we can select one of the agents uniformly at random to take the role of i. The resulting mechanism will be anonymous, r-partially strategyproof for all agents, but manipulable for any agent with utility function $\tilde{u}$. □

2.F.5 Proof of Proposition 2

Definition 15 (Vulnerability to Manipulation of Random Assignment Mechanisms)**.** For a given setting $(N, M, \boldsymbol{q})$ and two mechanisms φ, ψ, we say that ψ is *strongly as manipulable as* φ if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any

utility function $u_i \in U_{P_i}$, if there exists a misreport $P_i' \in \mathcal{P}$ such that

$$\langle u_i, \varphi_i(P_i, P_{-i}) \rangle < \langle u_i, \varphi_i(P_i', P_{-i}) \rangle, \tag{56}$$

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

$$\langle u_i, \psi_i(P_i, P_{-i}) \rangle < \langle u_i, \psi_i(P_i'', P_{-i}) \rangle. \tag{57}$$

In words, any agent that has an incentive to manipulate φ would in the same situation also want to manipulate ψ.

Proof of Proposition 2. For any setting $(N, M, \boldsymbol{q})$ and mechanisms φ and ψ, the following hold:

1. *If ψ is strongly as manipulable as φ, then $\rho_{(N,M,\boldsymbol{q})}(\varphi) \geqslant \rho_{(N,M,\boldsymbol{q})}(\psi)$.*
2. *If $\rho_{(N,M,\boldsymbol{q})}(\varphi) > \rho_{(N,M,\boldsymbol{q})}(\psi)$ and φ and ψ are comparable by their vulnerability to manipulation, ψ is strongly more manipulable than φ.*

To see 1., note that if φ is strongly as manipulable as ψ (in the sense of Definition 15., then any agent who can manipulate ψ also finds a manipulation to φ. Thus, the set of utilities on which ψ makes truthful reporting a dominant strategy cannot be larger than the set of utilities on which φ does the same. This in turn implies $\rho_{(N,M,\boldsymbol{q})}(\varphi) \geqslant \rho_{(N,M,\boldsymbol{q})}(\psi)$.

For 2., observe that if $\rho_{(N,M,\boldsymbol{q})}(\varphi) > \rho_{(N,M,\boldsymbol{q})}(\psi)$, then there exists a utility function $\tilde{u}$ in URBI $\big(\rho_{(N,M,\boldsymbol{q})}(\varphi)\big)$, which is not in URBI $\big(\rho_{(N,M,\boldsymbol{q})}(\psi)\big)$, and for which ψ is manipulable, but φ is not. Thus, φ cannot be strongly as manipulable as ψ, but the reverse is possible. □

2.F.6 Proof of Theorem 4

Theorem 4 is a direct consequence of the following Lemma.

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

1. *For all utility functions $u \in U_{P_i}$ that satisfy URBI(r) we have $\langle u, x - y \rangle \geqslant 0$,*
2. *x r-partially dominates y at P.*

Proof. Let $P : a_1 \succ \ldots \succ a_m$.

2. ⇒ 1.: Assume towards contradiction that 2. holds, but for some utility $u \in U_{P_i}$ satisfying URBI(r), we have

$$\langle u, x - y \rangle = \sum_{l=1}^{m} u(a_l) \cdot (x_{a_l} - y_{a_l}) < 0. \tag{58}$$

Without loss of generality, we can assume $\min u = 0$. Let $\delta_l = x_{a_l} - y_{a_l}$ and let

$$S(k) = \sum_{l=1}^{k} u(a_l) \cdot (x_{a_l} - y_{a_l}) = \sum_{l=1}^{k} u(a_l) \cdot \delta_l. \tag{59}$$

By assumption, $S(m) < 0$, so there exists a smallest value $K \in \{1, \dots, m\}$ 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

$$\begin{aligned}
& S(K) && (60)\\
= & \sum_{l=1}^{K} u(a_l) \cdot \delta_l && (61)\\
= & \left(\frac{S(K-1)}{u(a_K)} + \delta_K \right) \cdot u(a_K) && (62)\\
= & \left(\left(\frac{S(K-2)}{u(a_{K-1})} + \delta_{K-1} \right) \cdot \frac{u(a_{K-1})}{u(a_K)} + \delta_K \right) \cdot u(a_K) && (63)\\
= & \left(\left(\dots \left(\delta_1 \cdot \frac{u(a_1)}{u(a_2)} + \delta_2 \right) \cdot \frac{u(a_2)}{u(a_3)} + \dots \right) \right. && (64)\\
& \left. \cdot \frac{u(a_{K-1})}{u(a_K)} + \delta_K \right) \cdot u(a_K). && (65)
\end{aligned}$$

Since u satisfies URBI(r), the fraction $\frac{u(a_{K-1})}{u(a_K)}$ is lower-bounded by r^{-1}. But since $u(a_{K-1}) > 0$ and $S(K-1) \geqslant 0$, we must have that

$$\left(\frac{S(K-2)}{u(a_{K-1})} + \delta_{K-1} \right) \geqslant 0, \tag{66}$$

and therefore, when replacing $\frac{u(a_{K-1})}{u(a_K)}$ by r^{-1} in (63) we only make the term smaller.

By the same argument, we can replace all the terms $\frac{u(a_{k-1})}{u(a_k)}$ and obtain

$$\begin{aligned} 0 &> S(K) && (67)\\ &\geqslant \left(\left(\dots\left(\frac{\delta_1}{r}+\delta_2\right)\cdot\frac{1}{r}+\dots\right)\cdot\frac{1}{r}+\delta_K\right)\cdot u(a_K) && (68)\\ &= \frac{u(a_K)}{r^K}\cdot\sum_{l=1}^{K} r^l\cdot\delta_l. && (69)\end{aligned}$$

This is a contradiction to r-partial dominance, since by 2.,

$$\sum_{l=1}^{K} r^l\cdot(x_{a_l}-y_{a_l}) = \sum_{l=1}^{K} r^l\cdot\delta_l \geqslant 0. \tag{70}$$

1. ⇛ 2.: Assume towards contradiction that 1. holds, but x does not r-partially dominate y at P, i.e., for some $k\in\{1,\dots,m\}$, we have

$$\sum_{l=1}^{k} r^l\cdot\varphi_{i,a_l}(P_i,P_{-i}) < \sum_{l=1}^{k} r^l\cdot\varphi_{i,a_l}(P_i',P_{-i}), \tag{71}$$

and let k is the smallest rank for which inequality (71) is strict. Then the value

$$\delta = \sum_{l=1}^{k} r^l\cdot(x_{i,a_l}-y_{i,a_l}), \tag{72}$$

is strictly positive. Let u be a utility function that is consistent with P and has values

$$u(a_l) = \begin{cases} Dr^l, & \text{if } l\leqslant k,\\ dr^l, & \text{if } k+1\leqslant l\leqslant m-1,\\ 0, & l=m.\end{cases} \tag{73}$$

This utility function satisfies URBI(r) as long as $D\geqslant d$. Furthermore, the difference

in utility between x and y is

$$\begin{aligned}
& \langle u, x_i - y_i \rangle && (74)\\
= & \sum_{l=1}^{m} u(a_l) \cdot (x_{i,a_l} - y_{i,a_l}) && (75)\\
= & D \sum_{l=1}^{k} r^l \cdot (x_{i,a_l} - y_{i,a_l}) + d \sum_{l=k+1}^{m-1} r^l \cdot (x_{i,a_l} - y_{i,a_l}) && (76)\\
\geqslant & D\delta - d. && (77)
\end{aligned}$$

Since $\delta > 0$, we can choose $D > \frac{d}{\delta}$ such that this change is strictly positive, a contradiction. □

2.F.7 Proof of Proposition 3

Proof of Proposition 3. Given a setting $(N, M, \boldsymbol{q})$, if a mechanism φ is partially strategyproof (i.e., r-partially strategyproof for some $r > 0$), then it is weakly SD-strategyproof. The converse may not hold.

Convex strategyproofness (Balbuzanov, 2015) requires that for any agent i and any preference order $P_i \in \mathcal{P}$, the set of utility functions $u_i \in U_{P_i}$ which make truthfully reporting P_i a dominant strategy for I (independent of the other agents' reports P_{-i}) is a non-empty (convex) subset of U_{P_i}. For any r-partially strategyproof mechanism this is precisely the set of utilities $\text{URBI}(r) \cap U_{P_i} \neq \varnothing$. Thus, partial strategyproofness implies convex strategyproofness in the sense that the set of utilities for which truthful reporting must be a dominant strategy is specified.

Balbuzanov (2015) gives an example of a mechanism that is weakly SD-strategyproof, but not convex strategyproof therefore not partially strategyproof either. □

2.F.8 Proof of Proposition 4

Proof of Proposition 4. Given a setting $(N, M, \boldsymbol{q})$, if a mechanism φ is r-partially strategyproof for some $r > 0$, then it is ε-approximately strategyproof for some $\varepsilon < 1$. The converse may not hold.

Consider a fixed setting $(N, M, \boldsymbol{q})$ and fixed $r > 0$, and let φ be a mechanism that is r-partially strategyproof in this setting. Let

$$\delta = \min \left\{ |\varphi_j(P_i, P_{-i}) - \varphi_j(P_i', P_{-i})| \;\middle|\; \begin{array}{c} \forall i \in N, (P_i, P_{-i}) \in \mathcal{P}^N, \\ P_i' \in \mathcal{P}, j \in M : \\ |\varphi_j(P_i, P_{-i}) - \varphi_j(P_i', P_{-i})| > 0 \end{array} \right\} \quad (78)$$

be the smallest amount by which the probability for any object changes for any agent under any misreport. As in the proof of Theorem 2, this value must be strictly positive. From Theorem 2 we also know that φ must be upper invariant and swap monotonic. Thus, under any manipulation, there is some highest ranking object a for which the manipulating agent's probability decreases strictly. The magnitude of this decrease is at least δ, and for all more preferred objects than a the probabilities do not change.

In the worst case (from the mechanism designer's perspective), the manipulating agent can loose δ probability for a, but at the same time, it will convert probability for its last choice object (d, say) to probability for its next choice below a (b, say). Setting $u(a) = 1$, $u(b)$ close to 1, and $u(c) = 0$, the gain from any manipulation is bounded by

$$u(b) - (1 - \delta)u(d) - \delta u(a) \leqslant 1 - \delta. \tag{79}$$

Thus, the agent can improve its utility by at most $1 - \delta < 1$, i.e., φ is ε-approximately strategyproof for $\varepsilon = 1 - \delta$.

To see that the converse may not hold we construct a mechanism that is ε-approximately strategyproof, but not partially strategyproof. Suppose, there are only 2 objects, a and b. If the agent reports $P : a \succ b$, then the mechanism assigns $(1/2, 1/2)$ for a and b, respectively. If instead the agent reports $P' : b \succ a$, the mechanism assigns $(1/2 + \varepsilon, 1/2 - \varepsilon)$. For $\varepsilon > 0$, this mechanism is manipulable in a stochastic dominance sense, and therefore not partially strategyproof. However, the maximal gain from manipulation is ε if $u(a) = 1$ and $u(b) = 0$. Therefore, the mechanism ε-approximately strategyproof. □

2.F.9 Proof of Theorem 5

Proof of Theorem 5. Given a setting $(N, M, \boldsymbol{q})$, a mechanism φ is partially strategyproof (i.e., r-partially strategyproof for some $r > 0$) if and only if φ is DL-strategyproof.

Consider a fixed setting $(N, M, \boldsymbol{q})$ and fixed $r > 0$, and let φ be a mechanism that is r-partially strategyproof in this setting. Let

$$\delta = \min \left\{ |\varphi_j(P_i, P_{-i}) - \varphi_j(P_i', P_{-i})| \;\middle|\; \begin{array}{c} \forall i \in N, (P_i, P_{-i}) \in \mathcal{P}^N, \\ P_i' \in \mathcal{P}, j \in M : \\ |\varphi_j(P_i, P_{-i}) - \varphi_j(P_i', P_{-i})| > 0 \end{array} \right\} \tag{80}$$

be the smallest amount by which the probability for any object changes for any agent

under any misreport. As in the proof of Theorem 2, this value must be strictly positive. From the Theorem 2 we also know that φ must be upper invariant and swap monotonic. Thus, under any manipulation, there is some highest ranking object a for which the manipulating agent's probability decreases strictly. The magnitude of this decrease is at least δ, and for all objects that the agent prefers strictly to a the probabilities do not change. This immediately implies DL-strategyproofness of φ.

To see that the other direction also holds, let δ be defined as above. By DL-strategyproofness, the highest-ranking object for which there is any change in probability under any misreport must be assigned with lower probability under the misreport. Thus, we can proceed analogously to the proof of necessity in Theorem 2. □

2.F.10 Proof of Theorem 6

Proof of Theorem 6. Given a setting $(N, M, \boldsymbol{q})$, if φ is r-locally partially strategyproof for some $r > 0$, then φ is r^2-partially strategyproof.

To prove this Theorem, we must verify that an r-locally partially strategyproof mechanism φ satisfies the conditions for r^2-partial strategyproofness, i.e., for any agent $i \in N$, any preference profile $\boldsymbol{P} = (P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any utility function $u_i \in U_{P_i}$ with $u_i \in \mathrm{URBI}(r^2)$ the inequality

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_i(P_i', P_{-i}) \rangle \geqslant 0 \tag{81}$$

holds. Without loss of generality, we can assume that $\min u = 0$, since the manipulation incentives are exactly the same for an agent with utility function $\tilde{u} = u - \min u$.

To simplify notation, we fix an arbitrary combination of agent, preference profile, misreport, and utility to satisfy these preconditions. We drop the index i on the preference orders, utility functions, and mechanism, and we omit the preferences of the other agents. With this simplification, inequality (81) becomes

$$\langle u, \varphi(P) - \varphi(P') \rangle \geqslant 0. \tag{82}$$

Recall that U_P denotes the set of utility functions that are consistent with P, i.e.,

$$U_P = \left\{ w : M \to \mathbb{R}^+ \mid w \sim P \right\}, \tag{83}$$

and U denotes the utility space, i.e., the union of all consistent utility functions

$$U = \bigcup_{P \in \mathcal{P}} U_P. \tag{84}$$

We say that a (utility) function $w : M \to \mathbb{R}^+$ *implies indifference* between two different objects $a, b \in M$ if $w(a) = w(b)$, and we denote by $W = \{w : M \to \mathbb{R}^+\}$ the *extended utility space*, i.e., the set of all possible (utility) functions, including those that imply indifferences.

For the proof, we have fixed a preference order P and a consistent utility function $u \in U_P$. Let v be a utility function that is consistent with the misreport P', and let

$$\mathrm{co}(u, v) = \{u_\alpha = (1 - \alpha)u + \alpha v \mid \alpha \in [0, 1]\} \tag{85}$$

be the convex line segment in W that connects u and v. This line segment "starts" in U_P, then (for increasing α) traverses the extended utility space W and eventually "ends" at v in $U_{P'}$. $\mathrm{co}(u, v)$ is said to *pass* a preference order P'' if for some value $\alpha \in [0, 1]$ we have that u_α is consistent with P'', or equivalently, if $u_\alpha \in U_{P''}$. By construction, $\mathrm{co}(u, v)$ passes a sequence of preference orders $P = P_0, P_1, \ldots, P_{K-1}, P_K = P'$ in this order, i.e., as α increases, u_α is first consistent with P_0, then with P_1, etc. until it is consistent with $P_K = P'$. Note that intermittently, it is possible that u_α is not consistent with any preference order as it may imply indifferences. By linearity we have that for any two objects $a, b \in M$ with $u(a) > u(b)$ but $v(a) < v(b)$, there exists a *unique* $\alpha \in (0, 1)$ for which u_α implies indifference between a and b, for any smaller $\alpha^- < \alpha$ we have $u_{\alpha^-}(a) > u_{\alpha^-}(b)$, and for any larger $\alpha^+ > \alpha$ we have $u_{\alpha^+}(a) < u_{\alpha^+}(b)$.

We are now ready to formally define two important requirements:

- We say that $\mathrm{co}(u, v)$ makes *no simultaneous transitions* if for any three different objects $a, b, c \in M$ we have

$$\mathrm{co}(u, v) \cap \{w \in W \mid w(a) = w(b) = w(c)\} = \varnothing, \tag{86}$$

  i.e., for no value of the parameter α does u_α imply indifference between all three objects a, b, c. Intuitively, this means that two consecutive preference orders P_k, P_{k+1} in the sequence $(P_0, \ldots, P_K)$ differ by exactly one swap of two consecutive objects.

- We say that $\mathrm{co}(u, v)$ *passes P_k in URBI(r)* if it passes P_k and there exists some $\alpha_k \in [0, 1]$ such that $u_{\alpha_k} \in U_{P_k} \cap \mathrm{URBI}(r)$. This means that the line segment

contains at least one utility function that is consistent with P_k and in addition satisfies the URBI(r)-constraint.

Recall that P is a preference order, u a utility consistent with P that satisfies URBI(r^2), and the mechanism φ is r-locally partially strategyproof.

Claim 2. *There exists $v \in U_{P'} \cap URBI(r)$ such that*

i. $co(u, v)$ makes no simultaneous transitions,

ii. if $co(u, v)$ passes a preference order P'', then it passes P'' in $URBI(r)$.

Using Claim 2, we can now show the inequality

$$\langle u, \varphi(P) - \varphi(P') \rangle \geqslant 0. \tag{87}$$

We will show this by writing the left side as a telescoping sum over local incentive constraints, where all but the first and the last terms cancel out, such that it collapses to yield the inequality. This idea is inspired by the proof of local sufficiency for strategyproofness in (Carroll, 2012).

Consider the utility function v constructed in Claim 2 and the convex line segment co(u, v). Let $\alpha_0 = 0$, $\alpha_K = 1$, and for each $k \in \{0, \ldots, K\}$ let α_k be the parameters for which $u_{\alpha_k} \in U_{P_k} \cap$ URBI(r), which exist by Claim 2ii. For any $k \in \{0, \ldots, K-1\}$ we know that the preference order P_k and P_{k+1} are neighbors of each other, i.e., $P_k \in N_{P_{k+1}}$ and $P_{k+1} \in N_{P_k}$ (by Claim 2i). Thus, by r-local partial strategyproofness of φ we obtain

$$\langle u_{\alpha_k}, \varphi(P_k) - \varphi(P_{k+1}) \rangle \geqslant 0 \tag{88}$$

and

$$\left\langle u_{\alpha_{k+1}}, \varphi(P_{k+1}) - \varphi(P_k) \right\rangle \geqslant 0. \tag{89}$$

Multiplication by α_k and $-\alpha_{k+1}$, respectively, and then adding both inequalities yields

$$\left\langle \alpha_k u_{\alpha_{k+1}} - \alpha_{k+1} u_{\alpha_k}, \varphi(P_{k+1}) - \varphi(P_k) \right\rangle \geqslant 0. \tag{90}$$

Now, observe that $\alpha_k u_{\alpha_{k+1}} - \alpha_{k+1} u_{\alpha_k} = (\alpha_k - \alpha_{k+1}) u$, and therefore

$$\langle u, \varphi(P_{k+1}) - \varphi(P_k) \rangle \geqslant 0 \tag{91}$$

for all $k \in \{0, \ldots, K-1\}$. Summing over all k, we get

$$\langle u, \varphi(P) - \varphi(P') \rangle = \sum_{k=0}^{K-1} \langle u, \varphi(P_{k+1}) - \varphi(P_k) \rangle \geqslant 0. \tag{92}$$

We now proceed to prove Claim 2.

Proof of Claim 2. The proof for the existence of v is constructive. Recall that for a preference order P the *rank* of an object j under P'' is the position that a holds in the ranking, i.e.,

$$\mathrm{rank}_{P''}(j) = \#\{j \in M \mid P'' : j \succ a\} + 1. \tag{93}$$

Define $v : M \to \mathbb{R}^+$ by setting

$$v(j) = C^{m - \mathrm{rank}_{P'}(j)} \tag{94}$$

for any $j \in M$. If $C > 1$, then $v \in U_{P'}$. Furthermore, for sufficiently large C, $v \in \mathrm{URBI}(r)$, since for any $a, b \in M$ with $P' : a \succ b$

$$\frac{v(b) - \min v}{v(a) - \min v} = \frac{C^{\mathrm{rank}_{P'}(b)} - 1}{C^{\mathrm{rank}_{P'}(a)} - 1} = o\left(1/C\right). \tag{95}$$

It remains to be shown that for sufficiently large C, statements i and ii from the Claim hold.

To prove both statements, we require the concept of the *canonical transitions* (same as in Claim 1 in the proof of Theorem 2). A transition is a finite sequence of preference orders that starts and terminates with given preference orders and in each step the relative ranking of exactly two consecutive objects is inverted. Formally,

$$P_0, P_1, \ldots, P_{K-1}, P_K, \tag{96}$$

is a *transition* from P to P' if

- $P_0 = P$ and $P_K = P'$,
- for all $k \in \{0, \ldots, K-1\}$ we have $P_k \in N_{P_{k+1}}$ and $P_{k+1} \in N_{P_k}$.

The *canonical transition* is a particular transition between two preference orders that is inspired the *bubble-sort* algorithm:

P_0**:** Set $P_0 = P$

P_k**:** Determine P_k based on P_{k-1} as follows:

- Let r be the rank where P_{k-1} and P' differ for the first time, i.e.,

$$P_{k-1} \;:\; j_1 \succ \ldots \succ j_{r-1} \succ j_r \succ \ldots, \tag{97}$$
$$P' \;:\; j_1 \succ \ldots \succ j_{r-1} \succ c \succ \ldots, \tag{98}$$

  such that $j_r \neq c$, and let c be the rth choice object under P'.
- Find c in the ranking P_{k-1}
- Construct P_k by swapping c up one rank, i.e., if

$$P_{k-1} \;:\; j_1 \succ \ldots \succ a \succ b \succ c \succ \ldots, \tag{99}$$
$$\text{then let } P_k \;:\; j_1 \succ \ldots \succ a \succ c \succ b \succ \ldots. \tag{100}$$

P_K**:** Stop if $P_k = P'$ for some k, then set $K = k$

Besides the canonical transition, we formalize *transition times*. Suppose that for two objects $a, b \in M$ we have $P : a \succ b$, but $P' : b \succ a$, such that $u(a) > u(b)$, but $v(a) < v(b)$. Recall that in this case, there exists a unique parameter α for which $u_\alpha(a) = u_\alpha(b)$, for any smaller $\alpha^- < \alpha$ we have $u_{\alpha^-}(a) > u_{\alpha^-}(b)$, and for any larger $\alpha^+ > \alpha$ we have $u_{\alpha^+}(a) < u_{\alpha^+}(b)$. The line segment $\mathrm{co}(u, v)$ "pierces" the hyperplane of indifference between a and b at the point u_α, i.e., it transitions from preference orders that prefer a to b to preference orders that prefer b to a. Formally, the *transition time* $\alpha(a, b, 1)$ is the parameter for which $u_{\alpha(a,b,1)}(a) = u_{\alpha(a,b,1)}(b)$. Extending this notation, we define $\underline{\alpha}(a, b, r)$ as the *first* time when u_α violates the URBI(r) constraint for $a \succ b$, i.e.,

$$\underline{\alpha}(a, b, r) = \inf \left\{ \alpha \in [0, 1] \;\middle|\; \frac{u_\alpha(b) - \min u_\alpha}{u_\alpha(a) - \min u_\alpha} > r \right\}, \tag{101}$$

and $\overline{\alpha}(b, a, r)$ as the *last* time when u_α violates the URBI(r) constraint for $b \succ a$, i.e.,

$$\overline{\alpha}(a, b, r) = \sup \left\{ \alpha \in [0, 1] \;\middle|\; \frac{u_\alpha(a) - \min u_\alpha}{u_\alpha(b) - \min u_\alpha} > r \right\}. \tag{102}$$

Obviously,

$$\underline{\alpha}(a, b, r) < \alpha(a, b, 1) < \overline{\alpha}(b, a, r), \tag{103}$$

i.e., as α increases, u_α violates URBI(r) for $a \succ b$ at some time, then subsequently it transitions from $a \succ b$ to $b \succ a$, and finally it no longer violates the URBI(r) constraint

for $b \succ a$.

We are now ready to formulate Claims 3, 4, and 5, which are needed to establish statement i (no simultaneous transitions) and statement ii (passing all preference orders in URBI(r)), respectively, and the fact that only pairs of objects are relevant that rank differently under P and P'.

Claim 3. *For sufficiently large C, co(u, v) induces the canonical transition*

$$P_0 = P, P_1, \ldots, P_{K-1}, P_K = P'. \tag{104}$$

Claim 4. *For sufficiently large C, if $\alpha(a, b, 1) < \alpha(c, d, 1)$, then*

$$\overline{\alpha}(a, b, r) \leqslant \underline{\alpha}(c, d, r). \tag{105}$$

Claim 5. *If $P : a \succ b$ and $P' : a \succ b$ and $u, v \in$ URBI(r), then for all $\alpha \in [0, 1]$*

$$\frac{u_\alpha(b) - \min u_\alpha}{u_\alpha(a) - \min u_\alpha} \leqslant r. \tag{106}$$

Since co(u, v) induces a *transition* by Claim 3, we already know that for all pairs $(a, b) \neq (c, d)$ we have $\alpha(a, b, 1) \neq \alpha(c, d, 1)$. Thus, co$(u, v)$ makes no simultaneous transitions.

If a is preferred to b under both P and P', then by Claim 5 the URBI(r) constraint for a over b is satisfied for any α. Suppose now that $P : a \succ b$, $P : c \succ d$, $P' : b \succ a$, $P' : d \succ c$, and $\alpha(a, b, 1) < \alpha(c, d, 1)$. Then co$(u, v)$ "enters" a new set of consistent utility functions U_{P_k} at time $\alpha(a, b, 1)$, where P_k differs from P_{k-1} by a swap of a and b, and it "leaves" U_{P_k} at time $\alpha(c, d, 1)$, where P_k differs from P_{k+1} by a swap of c and d. In this case the URBI(r) constraint for b over a is satisfied after time $\overline{\alpha}(a, b, r) > \alpha(a, b, 1)$, and the URBI($r$) constraint for c over d is satisfied before time $\underline{\alpha}(c, d, r) < \alpha(c, d, 1)$. Claim 4 yields the constraint for c over d holds "long enough" for the constraint for b over a to be restored. Thus, at any time $\alpha_k \in [\overline{\alpha}(a, b, r), \underline{\alpha}(c, d, r)] \neq \varnothing$, both constraints are satisfied. Iterated application of this argument yields that for any $k \in \{0, \ldots, K\}$, there exists some α_k for which u_{α_k} satisfies URBI(r) with respect to preference order P_k.

This concludes the proof of Claim 2. □

We now provide the proofs of Claims 3 and 4. Claim 5 is obvious.

Proof of Claim 3. First, we formulate an equivalent condition for co(u, v) to induce the

canonical condition in terms of transition times $\alpha(a, b, 1)$.

Claim 6. *The following are equivalent:*

1. *$co(u, v)$ induces the canonical transition*

$$P_0 = P, P_1, \ldots, P_{K-1}, P_K = P'. \tag{107}$$

2. *For any $a, b, c, d \in M$ with $P : a \succ b$, $P : c \succ d$, $P' : b \succ a$, $P' : d \succ c$,*
   i. *if $P' : b \succ d$, then $\alpha(a, b, 1) < \alpha(c, d, 1)$,*
   ii. *if $b = d$ and $P : c \succ a$, then $\alpha(a, b, 1) < \alpha(c, d, 1)$.*

Proof of Claim 6. First, we show sufficiency ("$\Rightarrow$"). To see that 2i holds, observe that since $P' : b \succ d$, b will be "brought up" by bubble sort before d is ever swapped up against another object. Since $P : c \succ d$, the swap of $c \leftrightarrow d$ is such a swap, and therefore, it has to occur after the swap $a \leftrightarrow b$. 2ii follows by observing that from $b = d$ and $P : c \succ a$ we get that $P : c \succ a \succ b$, but ultimately $P' : b \succ (a, c)$. The bubble sort algorithm will bring b up by swapping it with a before it swaps b and c.

To see necessity ("$\Leftarrow$"), let $(a \leftrightarrow b)$ and $(c \leftrightarrow d)$ be two swaps that occur at $\alpha(a, b, 1)$ and $\alpha(c, d, 1)$, respectively. If $P' : b \succ d$, then 2i implies that $(a \leftrightarrow b)$ occurs before $(c \leftrightarrow d)$, which is consistent with the canonical transition. By symmetry, the case $P' : d \succ b$ also follows. Next, observe that any case not covered by this argument involves identity of b and d, i.e., $b = d$. If $a = c$ as well, then there is nothing to show, so assume $P : a \succ c$, where 2ii implies the correct behavior. The last remaining case where $b = d$ and $P : c \succ a$ follows again by symmetry. □

We now verify that the sequence of types through which $\mathrm{co}(u, v)$ passes is indeed a canonical transition. Let $a, b, c, d \in M$ be such that $P : a \succ b$, $P : c \succ d$, $P' : b \succ a$, $P' : d \succ c$, and either $P' : b \succ d$ (as in 2i of Claim 6) or $b = d$ and $P : c \succ a$ (as in 2ii of Claim 6). We can write

$$\alpha(a, b, 1) = \frac{u(a) - u(b)}{u(a) - u(b) + v(b) - v(a)} \tag{108}$$

and

$$\alpha(c, d, 1) = \frac{u(c) - u(d)}{u(c) - u(d) + v(d) - v(c)}, \tag{109}$$

and we need to show that

$$\begin{aligned}
& \alpha(a,b,1) < \alpha(c,d,1) && (110)\\
\Leftrightarrow\ & (u(a)-u(b))\,(u(c)-u(d)+v(d)-v(c)) && (111)\\
& \quad < (u(c)-u(d))\,(u(a)-u(b)+v(b)-v(a)) && (112)\\
\Leftrightarrow\ & (u(a)-u(b))\,(v(d)-v(c)) < (u(c)-u(d))\,(v(b)-v(a)) && (113)\\
\Leftrightarrow\ & \frac{u(a)-u(b)}{u(c)-u(d)} < \frac{v(d)-v(c)}{v(b)-v(a)}. && (114)
\end{aligned}$$

If $P : b \succ d$, the left side of (114) grows faster than C, i.e.,

$$\frac{v(d)-v(c)}{v(b)-v(a)} = \frac{C^{m-\mathrm{rank}_{P'}(d)} - C^{m-\mathrm{rank}_{P'}(c)}}{C^{m-\mathrm{rank}_{P'}(b)} - C^{m-\mathrm{rank}_{P'}(a)}} = \omega(C), \tag{115}$$

since $\mathrm{rank}_{P'}(b) < \mathrm{rank}_{P'}(d)$, $\mathrm{rank}_{P'}(b) < \mathrm{rank}_{P'}(a)$, and $\mathrm{rank}_{P'}(d) < \mathrm{rank}_{P'}(c)$. Similarly, if $b = d$ and $P : c \succ a$, we obtain that

$$\frac{v(d)-v(c)}{v(b)-v(a)} = \frac{C^{m-\mathrm{rank}_{P'}(b)} - C^{m-\mathrm{rank}_{P'}(c)}}{C^{m-\mathrm{rank}_{P'}(b)} - C^{m-\mathrm{rank}_{P'}(a)}} = \omega(C). \tag{116}$$

Since for sufficiently large C, the right side in (114) is not small, we can ensure that $\alpha(a,b,1) < \alpha(c,d,1)$ whenever the conditions of 2i or 2ii of Claim 6 are satisfied. □

Proof of Claim 4. First we define a "conservative estimate" for the violation times $\overline{\alpha}(a,b,r)$ and $\underline{\alpha}(c,d,r)$. Let

$$s(a,b,\alpha) = \frac{u_\alpha(b)}{u_\alpha(a)} \tag{117}$$

and observe that $s(a,b,\alpha)$ is continuous and strictly monotone in α and $s(a,b,\alpha(a,b,1)) = 1$. Thus, we can define the inverse $\alpha(a,b,s)$ for which $s(a,b,\alpha(a,b,s)) = s$ for any value of s that is attained by $s(a,b,\alpha)$. In particular for $\alpha = 0$, $s(a,b,0) = \frac{u(b)}{u(a)} \leqslant r$ and for $\alpha = 1$, $s(a,b,1) = \frac{v(b)}{v(a)} > \frac{1}{r}$, so $\alpha(a,b,s)$ is well-defined for all values $s \in \left[r, \frac{1}{r}\right]$. In fact, we can solve

$$\frac{u_{\alpha(a,b,s)}(b)}{u_{\alpha(a,b,s)}(a)} = s \tag{118}$$

for $\alpha(a,b,s)$ and obtain the expression

$$\alpha(a,b,s) = \frac{su(a)-u(b)}{su(a)-u(b)+v(b)-sv(a)}. \tag{119}$$

Using $\min u_\alpha \geqslant 0$,

$$s(a,b,\alpha) = \frac{u_\alpha(b)}{u_\alpha(b)} \leqslant r \tag{120}$$

implies

$$\frac{u_\alpha(b) - \min u_\alpha}{u_\alpha(b) - \min u_\alpha} \leqslant r, \tag{121}$$

and therefore,

$$\overline{\alpha}(a,b,r) \leqslant \alpha(b,a,r) \text{ and } \alpha(c,d,r) \leqslant \underline{\alpha}(c,d,r). \tag{122}$$

We now show that for sufficiently large C, $\alpha(b,a,r) \leqslant \alpha(c,d,r)$ holds. Recall that we are considering objects $a,b,c,d \in M$, where $P : a \succ b$, $P : c \succ d$, $P' : b \succ a$, and $P' : d \succ c$, so that the required inequality can be rewritten equivalently as

$$\alpha(b,a,r) \leqslant \alpha(c,d,r) \Leftrightarrow \frac{u(a) - ru(b)}{ru(c) - u(d)} \leqslant \frac{rv(b) - v(a)}{v(d) - rv(c)}. \tag{123}$$

By Claim 3, $\mathrm{co}(u,v)$ induces the canonical transition for sufficiently large C. Thus, by Claim 6, $\alpha(a,b,1) < \alpha(c,d,1)$ holds if

i. either $P' : b \succ d$,

ii. or $b = d$ and $P : c \succ a$.

In case i we observe that the left side of (123) is constant, but the right side grows for growing C, i.e., it is in $\omega(C)$. Therefore, (123) is ultimately satisfied for sufficiently large C.

In case ii the right side converges to r (from below) as C becomes large. Thus, it suffices to verify

$$\frac{u(a) - ru(b)}{ru(c) - u(b)} \leqslant r \tag{124}$$

$$\Leftrightarrow \quad u(a) - ru(b) \leqslant r^2 u(c) - ru(b) \tag{125}$$

$$\Leftrightarrow \quad 0 \leqslant r^2 u(c) - u(a). \tag{126}$$

Using the assumption that u satisfies URBI(r^2), $\min u = 0$, and $P : c \succ a$ we get that

$$\frac{u(c)}{u(a)} \leqslant r^2 \;\Leftrightarrow\; r^2 u(c) - u(a) \geqslant 0. \tag{127}$$

□

This concludes the proof of Theorem 6. □

2.F.11 Proof of Theorem 7

Proof of Theorem 7. Given a setting $(N, M, \boldsymbol{q})$ with $m \geqslant 4$ objects, for any $\varepsilon > 0$ there exists a bound $r \in (0,1)$ and a mechanism φ such that

1. *φ is r-locally partially strategyproof, but*
2. *φ is* not *$r^{2-\varepsilon}$-partially strategyproof.*

Consider a mechanism φ that selects the following assignments:

$$\begin{aligned}
\varphi(a \succ \ldots) &= (\alpha, 0, 0, 1-\alpha)\,, && (128)\\
\varphi(b \succ \ldots) &= (0, \beta, 0, 1-\beta)\,, && (129)\\
\varphi(d \succ \ldots) &= (0, 0, 0, 1)\,, && (130)\\
\varphi(c \succ d \succ \ldots) &= (0, 0, \gamma_c, 1-\gamma_c)\,, && (131)\\
\varphi(c \succ a \succ d \succ b) &= (1-\gamma_c-\gamma_d, 0, \gamma_c, \gamma_d)\,, && (132)\\
\varphi(c \succ b \succ \ldots) = \varphi(c \succ a \succ b \succ d) &= (1-\gamma_c-\gamma_d, \gamma_d, \gamma_c, 0) && (133)
\end{aligned}$$

for the objects a, b, c, d, respectively, where

$$\begin{aligned}
\alpha, \beta, \gamma_c, \gamma_d &\in [0,1], && (134)\\
s &= \tfrac{1}{r}, && (135)\\
\beta &= s\alpha, && (136)\\
\gamma_c &= \frac{(1-\alpha)}{(s-1)\,(s\,(s+1)-1)}, && (137)\\
\gamma_d &= \frac{s\,(s+1)\,(1-\alpha)}{s\,(s+1)-1}. && (138)
\end{aligned}$$

Observe that φ is entirely specified by the values of r and α. We will now show that for sufficiently small $r > 0$ we can chose α such that

1. φ is feasible,
2. φ is r-locally partially strategyproof,
3. but not $r^{2-\varepsilon}$-partially strategyproof.

First, we verify 1. that φ is feasible.

Claim 7. *For $s > 1$, φ is feasible if and only if $\alpha \in \left[\frac{s}{s^3-s+1}, \frac{1}{s}\right]$.*

Preferences	I	II	III	IV	V	VI	VII	VIII	IX
$a \succ \ldots$	*	*	*	*					
$b \succ \ldots$	*				*	*			
$d \succ \ldots$		*			*		*		
$c \succ d \succ \ldots$							*	*	
$c \succ a \succ d \succ b$			*						*
$c \succ b \succ \ldots$ or $c \succ a \succ b \succ d$				*		*		*	*

Table 2.2: Cases for local manipulations

Proof of Claim 7. Note that for $s > 1$ and $\alpha < 1$, γ_c and γ_d are positive. We must ensure that $\beta = s\alpha \leqslant 1$, which is the case if and only if $\alpha \leqslant \frac{1}{s}$. Next, we give a condition for $\gamma_c + \gamma_d \leqslant 1$, which in turn implies feasibility of the mechanism. This inequality holds if and only if $\alpha \geqslant \frac{s}{s^3-s+1}$. Observing that $\frac{1}{s} > \frac{s}{s^3-s+1}$ for $s > 1$, we have that the mechanism φ is feasible if and only if $\alpha \in \left[\frac{s}{s^3-s+1}, \frac{1}{s}\right] \neq \varnothing$. □

Second, we give equivalent conditions for r-local partial strategyproofness of φ, i.e., 2.

Claim 8. *For sufficiently small r, the following are equivalent:*

- *φ is feasible and r-locally partially strategyproof,*
- $\alpha \in I_s = \left[\frac{s^4-s^3}{s^5+2s^4-s^2-s-1}, \frac{s^3-s+\frac{s^2}{s-1}+1}{s^4+s^3-s^2+s+\frac{s^2}{s-1}}\right]$.

Furthermore, for sufficiently small $r > 0$, $I_s \neq \varnothing$.

Proof of Claim 8. We use Lemma 1 to establish r-partial dominance for any manipulation by just a swap, which in turn yields r-local partial strategyproofness. We only need to consider swaps that lead to a change of the assignment, otherwise there is nothing to show. In the following, δ_k denotes the adjusted kth partial sum, i.e., for $P : j_1 \succ \ldots \succ j_m$,

$$\delta_k = \sum_{l=1}^{k} s^{k-l} \left(\varphi_{j_l}(P) - \varphi_{j_l}(P')\right) = r^{-k} \left(\sum_{l=1}^{k} r^l \left(\varphi_{j_l}(P) - \varphi_{j_l}(P')\right)\right). \tag{139}$$

Observe that positivity of $\delta_1, \delta_2, \delta_3$ is equivalent to r-partial dominance of $\varphi_{j_l}(P)$ over $\varphi_{j_l}(P')$ at P by Lemma 1. Table 2.2 lists all the cases we need to consider.

I. • $a \succ b \succ \ldots \rightsquigarrow b \succ a \succ \ldots$:

$$\delta_1 = \alpha \geqslant 0, \tag{140}$$
$$\delta_2 = s\alpha - \beta = 0 \geqslant 0, \tag{141}$$
$$\delta_3 = (1-\alpha) - (1-\beta) = \beta - \alpha \geqslant 0. \tag{142}$$

For δ_3, we assumed that the third choice was d, otherwise there is nothing to show.

• $b \succ a \succ \ldots \rightsquigarrow a \succ b \succ \ldots$:

$$\delta_1 = \beta \geqslant 0, \tag{143}$$
$$\delta_2 = s\beta - \alpha = \alpha(s^2 - 1) \geqslant 0, \tag{144}$$
$$\delta_3 = s^2\beta - s\alpha + \alpha - \beta = \alpha(s^3 - 2s + 1) \geqslant 0. \tag{145}$$

For δ_3, we assumed that the third choice was d, otherwise there is nothing to show.

II. $a \succ \ldots \leftrightsquigarrow d \succ \ldots$: $\varphi(a \succ \ldots)$ first-order stochastically dominates $\varphi(d \succ \ldots)$ for all preference orders where a is preferred to d, and vice versa.

III. • $a \succ c \succ d \succ b \rightsquigarrow c \succ a \succ d \succ b$:

$$\delta_1 = \alpha - (1 - \gamma_c - \gamma_d) \tag{146}$$
$$= \alpha - 1 + (1-\alpha)\left(\frac{(s-1)^{-1} + s(s+1)}{s(s+1) - 1}\right) \geqslant 0, \tag{147}$$

since

$$(s-1)^{-1} + s(s+1) \geqslant s(s+1) - 1 \;\Leftrightarrow\; (s-1)^{-1} \geqslant -1. \tag{148}$$

$$\delta_2 = s(\alpha - 1 + \gamma_c + \gamma_d) - \gamma_c \tag{149}$$
$$= (1-\alpha)s\left(\frac{(s-1)^{-1} + s(s+1) - (s-1)^{-1}s^{-1}}{s(s+1) - 1} - 1\right)$$
$$= (1-\alpha)s\left(\frac{s(s+1) + s^{-1}}{s(s+1) - 1} - 1\right) \geqslant 0, \tag{150}$$
$$\delta_3 = s\delta_2 \geqslant 0. \tag{151}$$

- $c \succ a \succ d \succ b \rightsquigarrow a \succ c \succ d \succ b$:

$$\begin{aligned}
\delta_1 &= \gamma_c \geqslant 0, && (152)\\
\delta_2 &= s\gamma_c + 1 - \gamma_c - \gamma_d - \alpha && (153)\\
&= (a-\alpha)\left(1 + \frac{1 - s(s+1)}{s(s+1)-1}\right) = 0 && (154)\\
\delta_3 &= \gamma_d - 1 + \alpha, && (155)\\
&= (1-\alpha)\left(\frac{s(s+1)}{s(s+1)-1} - 1\right) \geqslant 0. && (156)
\end{aligned}$$

IV. - $a \succ c \succ b \succ d \rightsquigarrow c \succ a \succ b \succ d$:

$$\begin{aligned}
\delta_1 &= \alpha - (1 - \gamma_c - \gamma_d) \geqslant 0, && (157)\\
\delta_2 &= s(\alpha - 1 + \gamma_c + \gamma_d) - \gamma_c \geqslant 0, && (158)
\end{aligned}$$

as in case III, and

$$\begin{aligned}
\delta_3 &= s^2(\alpha - 1 + \gamma_c + \gamma_d) - s\gamma_c + (1-\alpha) - \gamma_d && (159)\\
&= (1-\alpha)\left(1 - s^2 + \frac{s + (s^2-1)s(s+1)}{s(s+1)-1}\right) && (160)\\
&= (1-\alpha)\left(\frac{s^2+s-1}{s(s+1)-1}\right) = 1 - \alpha \geqslant 0. && (161)
\end{aligned}$$

- $c \succ a \succ b \succ d \rightsquigarrow a \succ c \succ b \succ d$:

$$\begin{aligned}
\delta_1 &= \gamma_c \geqslant 0 && (162)\\
\delta_2 &= s\gamma_c + (1 - \gamma_c - \gamma_d) - \alpha && (163)\\
&= (1-\alpha)\left(1 + \frac{\frac{s}{s-1} - \frac{1}{s-1} - s(s+1)}{s(s+1)-1}\right) && (164)\\
&= (1-\alpha)\left(1 + \frac{1 - s(s+1)}{s(s+1)-1}\right) && (165)\\
&= (1-\alpha)(1-1) = 0 \geqslant 0, && (166)\\
\delta_3 &= 0 + \gamma_d \geqslant 0. && (167)
\end{aligned}$$

V. $b \succ \ldots \leftrightsquigarrow d \succ \ldots$: $\varphi(b \succ \ldots)$ first-order stochastically dominates $\varphi(d \succ \ldots)$ for all preference orders where b is preferred to d, and vice versa.

VI. • $b \succ c \succ \ldots \rightsquigarrow c \succ b \succ \ldots$: We begin with δ_3 as its positivity will also imply positivity of δ_1 and δ_2. Furthermore, the strictest condition arises from the preference order $b \succ c \succ a \succ d$.

$$\begin{aligned} \delta_3 &= s^2(\beta - \gamma_d) + s(-\gamma_c) + (-1 + \gamma_c + \gamma_d) \\ &= \alpha \left(\frac{s^5 + 2s^4 - s^2 - s - 1}{s(s+1) - 1} \right) - \frac{s^4 - s^3}{s(s+1) - 1} \geqslant 0 \end{aligned} \tag{168}$$

holds if and only if

$$\alpha \geqslant \frac{s^4 - s^3}{s^5 + 2s^4 - s^2 - s - 1}. \tag{169}$$

• $c \succ b \succ \ldots \rightsquigarrow b \succ c \succ \ldots$:

$$\delta_1 = \gamma_c \geqslant 0. \tag{170}$$

We can consider the case where d is the third choice as this condition is stirctly stronger than if a is the third choice. It suffices to consider δ_3 as its positivity implies positivity of δ_2.

$$\delta_3 = s^2\gamma_c + s\gamma_d - s\beta - (1 - \beta) \tag{171}$$

$$= \alpha \left(\frac{-s^4 - s^3 + s^2 - s - \frac{s^2}{s-1}}{s(s+1) - 1} \right) \tag{172}$$

$$+ \left(\frac{s^3 - s + \frac{s^2}{s-1} + 1}{s(s+1) - 1} \right) \geqslant 0 \tag{173}$$

holds if and only if

$$\alpha \leqslant \frac{s^3 - s + \frac{s^2}{s-1} + 1}{s^4 + s^3 - s^2 + s + \frac{s^2}{s-1}}. \tag{174}$$

VII. $d \succ c \succ \ldots \leftrightsquigarrow c \succ d \succ \ldots$: $\varphi(d \succ \ldots)$ first-order stochastically dominates $\varphi(c \succ d \succ \ldots)$ for all preference orders where d is preferred to c, and vice versa.

VIII. • $c \succ d \succ b \succ a \rightsquigarrow c \succ b \succ d \succ a$:

$$\delta_1 = \gamma_c - \gamma_c \geqslant 0, \tag{175}$$

$$\delta_2 = 1 - \gamma_c - 0 \geqslant 0, \tag{176}$$

$$\delta_3 = s(1 - \gamma_c) + \gamma_d \geqslant 0. \tag{177}$$

- $c \succ b \succ d \succ a \rightsquigarrow c \succ d \succ b \succ a$:

$$\begin{aligned}
\delta_1 &= \gamma_c - \gamma_c \geqslant 0, && (178)\\
\delta_2 &= \gamma_d - 0 \geqslant 0, && (179)\\
\delta_3 &= s\gamma_d - (1-\gamma_c) && (180)\\
&= \alpha\left(\frac{-s^2(s+1) - \frac{1}{s-1}}{s(s+1)-1}\right) && (181)\\
&\quad + \left(\frac{s^2(s+1) - \frac{1}{s-1} - s(s+1) + 1}{s(s+1)-1}\right), && (182)
\end{aligned}$$

which is positive if and only if

$$\alpha \leqslant \frac{s^3 - s + 1 - \frac{1}{s-1}}{s^3 + s^2 - \frac{1}{s-1}}. \tag{183}$$

IX. - $c \succ a \succ d \succ b \rightsquigarrow c \succ a \succ b \succ d$:

$$\begin{aligned}
\delta_1 &= \gamma_c - \gamma_c \geqslant 0, && (184)\\
\delta_2 &= 1 - \gamma_c - \gamma_d - 1 + \gamma_c + \gamma_d \geqslant 0, && (185)\\
\delta_3 &= \gamma_d \geqslant 0. && (186)
\end{aligned}$$

- $c \succ a \succ b \succ d \rightsquigarrow c \succ a \succ d \succ d$:

$$\begin{aligned}
\delta_1 &= \gamma_c - \gamma_c \geqslant 0, && (187)\\
\delta_2 &= 1 - \gamma_c - \gamma_d - 1 + \gamma_c + \gamma_d \geqslant 0, && (188)\\
\delta_3 &= \gamma_d \geqslant 0. && (189)
\end{aligned}$$

In summary, all local incentive constraints are satisfied if and only if

$$\begin{aligned}
&\frac{s^4 - s^3}{s^5 + 2s^4 - s^2 - s - 1} \leqslant \alpha && (190)\\
\leqslant\ &\min\left\{\frac{s^3 - s + 1 - \frac{1}{s-1}}{s^3 + s^2 - \frac{1}{s-1}}, \frac{s^3 - s + \frac{s^2}{s-1} + 1}{s^4 + s^3 - s^2 + s + \frac{s^2}{s-1}}\right\}. && (191)
\end{aligned}$$

The stronger upper bound is the second: asymptotically, as s grows, it behaves like $\frac{1}{s+1}$, which converges to 0, while the first bound converges to 1. The stronger upper bound

is also stronger than the upper bound for feasibility, since $\frac{1}{s+1}$ is smaller than $\frac{1}{s}$ The lower bound behaves like $\frac{1}{s+2}$, which is greater than $\frac{1}{s^2-1}$, the asymptotic of the lower bound for feasibility. Finally, observe that the lower bound behaves like $\frac{1}{s+2}$, which is strictly less than the asymptotic of the upper bound $\frac{1}{s+1}$. Thus, for sufficiently large s, α can be chosen such that φ is r-locally partially strategyproof, which in turn implies feasibility. □

It remains to show that for given $\varepsilon > 0$, there exist r and α such that φ is r-locally partially strategyproof (and therefore feasible), but not $r^{2-\varepsilon}$-partially strategyproof, i.e., 3. To see this, we let $\tilde{s} = s^{2-\varepsilon}$ and consider the preference order $a \succ b \succ c \succ d$ and the non-local misreport $c \succ a \succ b \succ d$. If φ is $\tilde{r}$-partially strategyproof, then in particular we must have $\delta_3 \geqslant 0$ for this manipulation. However, extensive algebraic transformations yield

$$\delta_3 = \tilde{s}^2 (\alpha - 1 + \gamma_c + \gamma_d) + \tilde{s} (-\gamma_d) + (-\gamma_c) \tag{192}$$

$$= (1-\alpha) \left(\frac{-s^{5-\varepsilon} + s^{5-2\varepsilon} + s^{3-\varepsilon} - 1}{s^3 - 2s + 1} \right). \tag{193}$$

Since the leading term with exponent $5 - \varepsilon$ has negative sign, this value is negative for sufficiently large s, and this negativity of δ_3 is independent of α.

In conclusion, given a value of $\varepsilon > 0$, we can find $r > 0$ and $\alpha \in (0, 1)$ such that the resulting mechanism φ is feasible and r-locally partially strategyproof, but it is not $r^{2-\varepsilon}$-partially strategyproof. □

2.F.12 Proof of Theorem 8

Proof of Theorem 8. A mechanism φ is upper invariant, monotonic, and sensitive if and only if $\varphi^{\mathbb{P}}$ is upper invariant and swap monotonic for all distributions $\mathbb{P}$ with full support.

First, we show sufficiency ("⇒"): for some fixed preference reports P_{-i}, the mechanism $f_i(\cdot, P_{-i})$ is upper invariant. Thus, $f^{\mathbb{P}}$ is the convex combination of a finite number of upper invariant mechanisms.

Consider preference orders $P_i \in \mathcal{P}$ and $P_i' \in N_{P_i}$ with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$. If $f^{\mathbb{P}}(P_i) = f^{\mathbb{P}}(P_i')$, then there is nothing to show for swap monotonicity. Else, if $f^{\mathbb{P}}(P_i) \neq f^{\mathbb{P}}(P_i')$, then there must exist some $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$ with $f_i(P_i, P_{-i}) \neq f_i(P_i', P_{-i})$. We get from monotonicity of f that $f_{i,a_k}(P_i, P_{-i}) \geqslant f_{i,a_k}(P_i', P_{-i})$ and $f_{i,a_{k+1}}(P_i, P_{-i}) \leqslant f_{i,a_{k+1}}(P_i', P_{-i})$ for all $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$. Moreover, by sensitivity, there exist preference reports $P_{-i}^k, P_{-i}^{k+1} \in \mathcal{P}^{N \setminus \{i\}}$ for which the differences are actually strict, i.e., $f_{i,a_k}(P_i, P_{-i}^k) >$

$f_{i,a_k}(P_i', P_{-i}^k)$ and $f_{i,a_{k+1}}(P_i, P_{-i}^{k+1}) < f_{i,a_{k+1}}(P_i', P_{-i}^{k+1})$. Since $\mathbb{P}$ has full support, it follows that $f_{a_k}^{\mathbb{P}}(P_i) > f_{a_k}^{\mathbb{P}}(P_i')$ and $f_{a_{k+1}}^{\mathbb{P}}(P_i) < f_{a_{k+1}}^{\mathbb{P}}(P_i')$. This is precisely swap monotonicity.

Next, we show necessity ("$\Leftarrow$"): towards contradiction, assume that f is not upper invariant, then there exist a preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, a preference order $P_i' \in N_{P_i}$ with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$, and an object $a_l \in U(a_k, P_i)$, such that $f_{i,a_l}(P_i, P_{-i}) > f_{i,a_l}(P_i', P_{-i})$. Choose $\mathbb{P}$ such that $\mathbb{P}[P_{-i}] = 1-\varepsilon$ and $\mathbb{P}[P_{-i}'] = \varepsilon/(m!-1)$ for all other $P_{-i}' \in \mathcal{P}^{N\setminus\{i\}}$. Then $f_{a_l}^{\mathbb{P}}(P_i) > f_{a_l}^{\mathbb{P}}(P_i')$ for sufficiently small $\varepsilon > 0$, i.e., $f^{\mathbb{P}}$ is not upper invariant for $\mathbb{P}$ with full support, a contradiction.

Monotonicity of f follows by an analogous argument. Finally, assume towards contradiction that f is not sensitive. Thus, there exists a pair of preference orders $P_i \in \mathcal{P}, P_i \in N_{P_i}$ with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$, such that for some preference reports, P_{-i}, $f_i(P_i, P_{-i}) \neq f_i(P_i', P_{-i})$, but $f_{i,a_k}(P_i, P_{-i}^k) = f_{i,a_k}(P_i', P_{-i}^k)$ for all $P_{-i}^k \in \mathcal{P}^N$ (without loss of generality, otherwise, we reverse the roles of P_i and P_i'). Again, choose $\mathbb{P}$ such that $\mathbb{P}[P_{-i}] = 1-\varepsilon$ and $\mathbb{P}[P_{-i}'] = \varepsilon/(m!-1)$ for all other $P_{-i}' \in \mathcal{P}^{N\setminus\{i\}}$. Then $f_{a_k}^{\mathbb{P}}(P_i) \neq f_{a_k}^{\mathbb{P}}(P_i')$ for sufficiently small $\varepsilon > 0$, i.e., $f^{\mathbb{P}}$ is not swap monotonic for the distribution $\mathbb{P}$ with full support, a contradiction. This concludes the proof. □

2.F.13 Proof of Proposition 6

Proof of Proposition 6. PS is swap monotonic.

Suppose agent i is considering the following two reports that only differ by the ordering of two objects x and y:

$$\begin{aligned} P_i &: a_1 \succ \ldots \succ a_K \succ x \succ y \succ b_1 \succ \ldots \succ b_L, \\ P_i' &: a_1 \succ \ldots \succ a_K \succ y \succ x \succ b_1 \succ \ldots \succ b_L. \end{aligned}$$

The Probabilistic Serial mechanism is implemented via the Simultaneous Eating algorithm; objects are continuously consumed as time progresses. Let τ_j be the time when object j is exhausted under report P_i, and τ_j' the time when j is exhausted under report P_i'.

If $\tau_A = \max(\tau_{a_k}, k \leqslant K) \geqslant \min(\tau_x, \tau_y)$, the last of the objects a_k is exhausted only after the first of x and y is exhausted. By upper invariance, $\tau_A = \tau_A'$. This means that by the time i arrives at x (under report P_i) or at y under report P_i', one of them is already exhausted. Thus, i will proceed directly to the respective other object. The consumption pattern does not differ between the two reports, so the assignment does not change.

Now suppose that $\tau_A < \tau_y \leqslant \tau_x$. Then i received no shares of y under P_i. But under P_i', it consumes shares of y from τ_A until $\tau_y' > \tau_A$. Thus, i's share in y strictly increases.

Furthermore, i consumed shares of x from τ_A until τ_x under report P_i. Under report P_i', i arrives at x only later at $\tau_y' > \tau_A$. The same agents that consumed x under report P_i will also be consuming x under report P_i' and at the same times. In addition, there may be some agents who arrive together with i from y. Thus, under report P_i' agent i faces strictly more competition for weakly less capacity of x, implying that its share of x will strictly decrease. Note that if i faced no competition at y, it was the only agent at y, and thus consumes it until time 1. In this case the assignment will also decrease, because i arrived later under report P_i'.

Finally, suppose that $\tau_A < \tau_x < \tau_y$. Under report P_i', agent i will arrive strictly earlier at y, i.e., the competing agents will be the same and arrive at the same times or later (if they arrived from x). Thus, the assignment for y will strictly increase under report P_i'. Furthermore, i might not receive any shares of x under report P_i', a strict decrease. Otherwise, the argument why i receives strictly less shares of x under P_i' is the same as for the case "$\tau_A < \tau_y \leqslant \tau_x$". $\square$

3 Trade-offs in School Choice: Comparing Deferred Acceptance, the Naïve and the Adaptive Boston Mechanism

Abstract

We compare three school choice mechanisms in terms of their strategyproofness and efficiency properties: *Deferred Acceptance (DA)*, the *naïve Boston mechanism (NBM)*, and the *adaptive Boston mechanism (ABM)*. Here NBM is the classical Boston mechanism, while ABM is different in that students automatically skip exhausted schools in the application process. When priorities are determined by a single uniform lottery, the three mechanisms form two hierarchies: regarding strategyproofness, we show that ABM satisfies the intermediate incentive requirement of *partial strategyproofness*. Regarding efficiency, we show that NBM rank dominates DA whenever this comparison is possible. Furthermore, using new limit arguments and simulations, we establish that ABM has *intermediate efficiency* between NBM and DA. Many of our results continue to hold for general priority structures and other tie-breakers. Our results reveal the inherent trade-off between strategyproofness and efficiency that market designers face when choosing between these three school choice mechanisms.

3.1 Introduction

Each year, millions of children enter a new public school. However, the capacities of public schools are limited and therefore, the students' individual wishes can almost never be accommodated perfectly. When students are allowed to express preferences over schools, administrators face the challenge of designing a *market*. In this market

scarce resources (the seats at public schools) must be assigned to self-interested agents (the students), who have heterogeneous, private preferences over these resources. In particular, administrators must devise a *school choice mechanism* which is a procedure that determines an assignment of students to schools taking into account the students' preferences but usually without any monetary transfers. Since the seminal paper by Abdulkadiroğlu and Sönmez (2003*b*), school choice mechanisms have attracted the attention of economists, and a growing body of research has had substantial impact on policy decisions.

3.1.1 Boston 'versus' Deferred Acceptance Mechanism

Two particular mechanisms have received the lion's share of the attention: the *Boston mechanism* and the *Deferred Acceptance mechanism* (*DA*).[1] Both mechanisms collect preference reports from the students and then assign seats to students in rounds.

Under the Boston mechanism, students apply to their favorite school in the first round. If a school has sufficient capacity to accommodate all applications in the first round, all applications are accepted. Otherwise, the school accepts applications following some priority order until its capacity is exhausted, and then it rejects all remaining applications. Students who were rejected in the first round apply to their second choice school in the second round. The process then repeats until all students have received a school or all schools have reached capacity. Variants of the Boston mechanism for school choice are ubiquitous (e.g., in Spain (Calsamiglia and Güell, 2014), in Germany (Basteck, Huesmann and Nax, 2015), and in many school districts in the United States (Ergin and Sönmez, 2006)).

The main motivation for letting parents choose the schools for their children through a school choice mechanism is *student welfare*. Popular measures for student welfare are the number of students who received their top choice or one of their top-k choices. Intuitively, the Boston mechanism fares well on this criterion as long as parents submit their preferences truthfully. It assigns as many applicants as possible to their first choices, then does the same with second choices in the second round, and so on. The mechanism owes much of its popularity to the intuitive way in which it attempts to increase this measure of student welfare.

On the other hand, it is susceptible to strategic manipulation by students. In particular, it was found to disadvantage honest participants, and the equilibria of the induced type

[1] The name *Boston mechanism* has stuck with researchers, despite the fact that as of 2005 the Boston mechanism is no longer used in Boston.

revelation game can have undesirable welfare properties (Ergin and Sönmez, 2006). Concerns about the manipulability of the Boston mechanism led to its abandonment in some cities in the US and around the world. In England, "first-preference-first" mechanisms (essentially the Boston mechanism) were even declared illegal in 2007 because they were believed to give unfair advantage to more sophisticated parents (Pathak and Sönmez, 2013).

The *(Student Proposing) Deferred Acceptance* mechanism (*DA*) has been proposed as an alternative to the Boston mechanism. Under DA, students also apply to schools in rounds. However, the acceptance at any school is *tentative* rather than final. If in any subsequent round a student applies to a school with no free capacity, she is not automatically rejected. Instead, she will be accepted at that school if another student who has been tentatively accepted at the same school has lower priority. In this case, the tentative acceptance of a student with lowest priority is revoked, and this student enters the next round of the application process. In each round previously rejected students continue applying to the school they prefer most out of all the schools that have not rejected them yet. When no more new applications are received by any school, all tentative acceptances are finalized.

DA makes truthful reporting a dominant strategy for students, which alleviates concerns about strategic manipulation. However, this *strategyproofness* comes at a cost: unlike the Boston mechanism, DA does not maximize the assigned number of first choices, then subsequently the number of second choices etc. In this paper, we capture this efficiency difference formally: we prove that the Boston mechanism *rank dominates* the Deferred Acceptance mechanism whenever the assignments are comparable at a given preference profile.

3.1.2 The Adaptive Boston Mechanism

So far, research on the Boston mechanism has largely focused on the classic *"naïve" Boston mechanism* (*NBM*) described above, where students apply to their kth choice school in the kth round. However, the Boston mechanism is sometimes used in a subtly different fashion: instead of applying to their kth choice in the kth round, in each round students apply to their *most preferred school that still has available capacity.*

For example, in the city of Freiburg, Germany, approximately 1'000 students transition from primary schools to one of ten secondary schools each year. Initially, they are asked to apply to their first choice school. If this application is successful, their assignment is finalized. Students whose applications were rejected receive a list of schools that still

have seats available. They are then asked to apply to one of the schools from this list in the second round. This process repeats in subsequent rounds. The procedure resembles the Boston mechanism, except that students are barred from applying to schools that have no more open seats. This alteration leads to an *adaptive Boston mechanism* (*ABM*). ABM eliminates the risk of "wasting one round" by applying to an already exhausted school. Most school districts in the state of Nordrhein-Westfalen, Germany, also use a school choice process that exhibits such an adaptive characteristic: parents receive a single slip of paper that they submit to their preferred school. If their application at that school is rejected, the application slip is returned and they can use it to apply to another school. However, before they apply to another school, parents are encouraged to call ahead and verify whether that school still has seats available. The adaptive variant of the Boston mechanism was also used for admission to secondary schools in Amsterdam but was replaced by DA in 2015 (de Haan et al., 2015).

On the one hand, ABM removes some obvious opportunities for manipulation that arise under NBM from the possibility of applying to exhausted schools. On the other hand, a student can obtain her third choice in the second round which may prevent another student from getting her second choice in that round. Consequently, one would expect the adaptive Boston mechanism to take an intermediate position *between* the Deferred Acceptance and the naïve Boston mechanism in terms of strategyproofness and efficiency. In this paper, we formalize and prove this intuition. Thereby, we establish ABM as an intermediate alternative that trades off strategyproofness and efficiency against each other in the design of school choice mechanisms.

3.1.3 Stability, Coarse Priorities, and Tie-breaking in School Choice

In two-sided markets, *stability* of the matching is often essential to prevent unraveling. In an unstable matching, some participants can benefit by breaking away and matching outside the mechanism. This is especially problematic if both sides have strategic interests in the match (e.g., doctors and hospitals). However, in contrast to two-sided markets, schools in school choice markets don't express preferences over students. Instead, *priorities* take the place of the schools' preferences. Usually, these priorities are exogenously determined and outside the control of the schools.[2] In particular, administrators have sufficient control over schools to ensure that they do not circumvent the procedure and

[2]One notable exception is the school choice market in New York City, where school principals develop preferences over students, e.g., based on reading scores or attendance (Abdulkadiroğlu, Pathak and Roth, 2005).

match with students outside of the mechanism. In this paper, we put stability aside and focus on the properties *strategyproofness* and *efficiency*.

In prior work about school choice mechanisms, most results have been obtained under the assumption that priorities are fixed and strict at all schools. However, as Kojima and Ünver (2014) have pointed out, this assumption is almost always violated: priorities are typically *coarse* because they are based only on neighborhoods or siblings. Recently, the role of priorities has been further de-emphasized; for example, walk-zone priorities in Boston were abandoned in 2013 (Dur et al., 2014). Coarse priorities put many students in the same priority class. This necessitates tie-breaking when two students with equal priority compete for a seat at some school. This introduces uncertainty into the mechanism. In this paper, we explicitly model the uncertainty that arises from coarse priorities and random tie-breaking by considering the *probabilistic* assignments before the tie-breaking procedure has been implemented.

In some school choice markets there are no initial priorities but the priorities are randomly generated by a single uniform lottery. For example, in 1999, the 15 neighborhoods of the Beijing Eastern City District used the naïve Boston mechanism to assign students to middle schools and priorities were determined by a single uniform lottery (Lai, Sadoulet and de Janvry, 2009). Similarly, the second phase of the school choice procedure in New York City used Deferred Acceptance with priorities derived in this way (Pathak and Sethuraman, 2011). Finally, most cities in Estonia employ the same procedure for the assignment of children to elementary schools (Lauri, Pöder and Veski, 2014). We say that a school choice market satisfies *assumption* $\mathbb{U}$ if priorities are determined by a single uniform lottery. While most markets do not satisfy the assumption $\mathbb{U}$ completely, the coarse nature of priorities in school choice is arguably closer to $\mathbb{U}$ than to the assumption of some strict, fixed priorities. All our findings in the present paper hold *at least* under the assumption $\mathbb{U}$, but most of them generalize to arbitrary priority structures with or without randomization.

3.1.4 A Motivating Example

In this paper we study a trio of popular school choice mechanisms: Deferred Acceptance, the naïve, and the adaptive Boston mechanism. We uncover their relationship in terms of strategyproofness and efficiency. To obtain an intuition about this relationship, consider a market with 4 students, conveniently named 1, 2, 3, 4, and 4 schools, named a, b, c, d,

with a single seat each. Suppose that the students' preferences are

$$P_1 \;:\; a \succ b \succ c \succ d, \tag{194}$$

$$P_2, P_3 \;:\; a \succ c \succ b \succ d, \tag{195}$$

$$P_4 \;:\; b \succ a \succ c \succ d, \tag{196}$$

where $P_i : x \succ y$ indicates that student i prefers school x to school y. Furthermore, suppose that priorities are determined by a single uniform lottery. If the students report truthfully, the probabilities of each student obtaining each of the seats are the following:

	NBM				**ABM**				**DA**			
i	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***
1	1/3	0	0	2/3	1/3	0	1/3	1/3	1/3	1/4	1/6	1/4
2	1/3	0	1/2	1/6	1/3	0	1/3	1/3	1/3	1/24	3/8	1/4
3	1/3	0	1/2	1/6	1/3	0	1/3	1/3	1/3	1/24	3/8	1/4
4	0	1	0	0	0	1	0	0	0	2/3	1/12	1/4

First, we evaluate the incentives for truth-telling: DA is known to be strategyproof for students, and therefore, no student can improve her assignment by misreporting. Under NBM, when reporting truthfully, student 1 has no chance of obtaining her second choice b or third choice c. By swapping b and c in her report, she would receive a, c, or d with probability 1/3 each. Thus, by misreporting she can obtain an assignment that she prefers in a first order-stochastic dominance sense. Finally, observe that under ABM, swapping b and c is no longer a beneficial manipulation for student 1: the misreport would have no effect on her assignment because ABM automatically skips the exhausted school b for student 1 anyways. However, by ranking school c in *first* position, she can receive c with certainty (holding the other students' reports fixed). If she had utilities of $15, 10, 9, 0$ for a, b, c, d, respectively, her *expected utility* would improve from 8 to 9. However, if her utility for c was 6 instead, her expected utility would decrease from 7 to 6. Thus, whether or not ranking c first is a useful manipulation for student 1 depends on "how strongly" she prefers a over c.

Next, we compare student welfare under the three mechanisms. Consider the *rank distributions* (Featherstone, 2011), that is the expected numbers of students who receive their kth choice for $k = 1, 2, 3, 4$. These are:

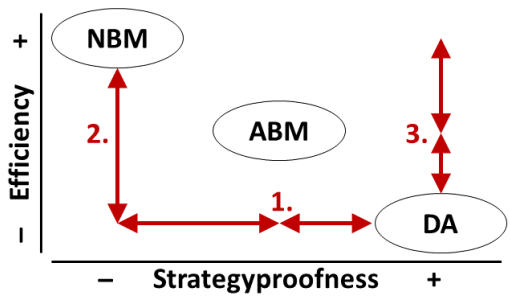

Figure 3.1: Overview of contributions (informal): 1. partial strategyproofness, 2. comparable rank dominance, 3. limit results and simulations.

Mechanism	$k = 1$	$k = 2$	$k = 3$	$k = 4$
NBM	2	1	0	1
ABM	2	2/3	1/3	1
DA	5/3	1	1/3	1

Observe that ABM and NBM assign the same number of first choices, but ABM assigns strictly fewer second choices and strictly more third choices. Thus, the assignment under NBM is "more efficient" in the sense that its rank distribution first order-stochastically dominates the rank distribution of ABM. Similarly, DA assigns a lower number of first choices and strictly more second choices than ABM, but for ranks 3 and 4 the rank distributions coincide. Thus, the rank distribution of ABM first order-stochastically dominates the rank distribution of DA. Consequently, with the rank distribution as a criterion for student welfare, NBM is *more efficient* than ABM which in turn is *more efficient* than DA.

The above example provides the intuition for our main result: NBM, ABM, and DA form a hierarchy with respect to strategyproofness with DA being fully strategyproof, NBM being manipulable in a stochastic dominance sense, and ABM taking an intermediate position; but at the same time the three mechanisms also form a hierarchy with respect to efficiency with NBM being the most efficient mechanism, DA being least efficient, and ABM again taking an intermediate position.

In this paper we show that the intuition from the motivating example generalizes in a non-trivial way. Figure 3.1 depicts our contributions. From a broader perspective, our results yield the important take-home message that choosing between NBM, ABM, and DA for school choice markets remains a question of trading off strategyproofness and efficiency: if strategyproofness is a hard requirement, DA is the mechanism of choice.

When the weaker partial strategyproofness is also acceptable, ABM can be employed to harness improvements in the rank distribution. Finally, if manipulability is not a concern, NBM offers further efficiency gains over both DA and ABM. We do not advocate superiority of either mechanism, but instead our insights allow mechanism designers to make a conscious and informed decision about this trade-off.

Remark 8. We would like to point out that the present paper differs substantially from two related papers by Dur (2015) and Harless (2015). We highlight the distinctions in our discussion of related work.

Organization of this paper: In Section 3.2, we discuss related work. In Section 3.3, we introduce our formal model and basic concepts, and in Section 3.4, we formally define the mechanisms DA, NBM, and ABM. In Sections 3.5, and 3.6 we compare them by their incentive and efficiency properties, respectively, and Section 3.7 concludes.

3.2 Related Work

The naïve Boston mechanism has received significant attention because it is frequently used for the assignment of students to public schools in many school districts around the world. The mechanism has been heavily criticized for its manipulability: for the case of strict priorities, Abdulkadiroğlu and Sönmez (2003*b*) showed that NBM is neither strategyproof nor stable. They suggested the Deferred Acceptance mechanism (Gale and Shapley, 1962) as an alternative that is stable and strategyproof for students. Ergin and Sönmez (2006) showed that with full information, NBM has undesirable equilibrium outcomes. Experimental studies, such as those conducted by Chen and Sönmez (2006) and Pais and Pinter (2008), revealed that it is indeed manipulated more frequently by human subjects than strategyproof alternatives.

Kojima and Ünver (2014) provided an axiomatic characterization of the naïve Boston mechanism for the case of fixed, strict priorities. However, they also pointed out that the assumption of fixed, strict priorities is usually violated in school choice problems. Some recent work has considered coarse priorities, revealing a number of surprising properties: Abdulkadiroğlu, Che and Yasuda (2015) demonstrated that in a setting with no priorities and perfectly correlated preferences, NBM can lead to higher *ex-ante* welfare than Deferred Acceptance in equilibrium. Similarly, simulations conducted by Miralles (2008) illustrated that with single uniform tie-breaking and no priorities, equilibria of the naïve Boston mechanism can yield higher welfare ex-ante. It has remained an open

research question if and how the Boston mechanism can be understood to have preferable efficiency for general priority structures. The present paper addresses this question: we show that NBM *rank dominates* the Deferred Acceptance mechanism whenever the two mechanisms are comparable at a given preference profile.

While the majority of prior work was focused on the naïve Boston mechanism, the idea of an adaptive adjustment has previously been discussed as well. Alcalde (1996) studied a "now-or-never" mechanism for two-sided marriage markets, where men propose to their most preferred *available* partner in each round. Miralles (2008) informally argued that an adaptive order of applications may improve the position of unsophisticated (i.e., truthful) students.

For the case when priorities are strict and fixed, Dur (2015) provided an axiomatic characterization of the adaptive Boston mechanism. Furthermore, he showed that NBM is at least *as manipulable as* ABM in the sense of (Pathak and Sönmez, 2013). If some students are unacceptable to some schools, then he also presented an example showing that NBM is in fact *more manipulable* than ABM. Harless (2015) provided insights about efficiency, robustness, and solidarity properties of ABM. The present paper differs from both papers (Dur, 2015) and (Harless, 2015) in three ways: first, we consider the *more general problem* where priorities can be coarse and ties may be broken randomly. Second, we show that the incentive properties of ABM are *strictly better* than those of NBM, even if all students are acceptable at all schools (Dur's example relies critically on the fact that some students are unacceptable for some schools). Furthermore, a comparison by vulnerability to manipulation remains inconclusive in the general domain (as we show in Section 3.5.1). Third, we give *more* results, such as the intermediate efficiency of ABM through limit results and simulations, and the proofs that neither ABM nor NBM lie on the efficient frontier; and our counter-examples to illustrate in-comparabilities are in parts *stronger* than those in (Harless, 2015). Our findings yield a comprehensive understanding of the hierarchical relationships of all three mechanisms on both dimensions of strategyproof *and* efficiency. To the best of our knowledge, ours is the first paper to consider the adaptive Boston mechanism in the general domain. We are also the first to formally establish its intermediate position in terms of strategyproofness and efficiency between DA and NBM.

Severe impossibility results restrict the design of school choice mechanisms that are strategyproof, efficient, and fair at the same time. Under the assumption $\mathbb{U}$ (i.e., single uniform tie-breaking and no priorities) the school choice problem becomes formally equivalent to the random assignment problem (Hylland and Zeckhauser, 1979). Zhou

(1990) showed that one cannot hope to design a random assignment mechanism that is strategyproof, ex-ante efficient, and anonymous. While Random Serial Dictatorship is at least strategyproof, anonymous, and *ex-post* efficient, it is conjectured to be the unique mechanism with these properties (Lee and Sethuraman, 2011; Bade, 2014). Bogomolnaia and Moulin (2001) introduced the Probabilistic Serial mechanism, which is ordinally efficient but only *weakly* strategyproof. Moreover, they showed that no strategyproof, symmetric mechanism can also be ordinally efficiency. Finally, Featherstone (2011) formalized rank efficiency, which is a strict refinement of ordinal efficient. He presented Rank Value mechanisms, which are rank efficient, but he also showed that strategyproofness and rank efficiency are incompatible even without additional fairness requirements. Since the random assignment problem is a special case of the school choice problem, these restrictive impossibility results also apply to school choice mechanisms. Thus, one cannot hope to design school choice mechanisms that achieve the optimum on all dimensions, but instead trade-offs are called for. Our finding that NBM, ABM, and DA form hierarchies with respect to strategyproofness and efficiency reveal the trade-offs that are implicit in any decision between these mechanisms.

3.3 Preliminaries

3.3.1 Basic Notation

Let N be a set of n students and let M be a set of m schools. We will usually use i to refer to particular students and j or $a, b, c, \ldots$ to refer to particular schools. Each school j has a *capacity* of q_j seats, and we assume that there are enough seats to accommodate all students (i.e., $n \leqslant q_1 + \ldots + q_m$). Otherwise we can add a dummy school with capacity n. Students have strict preferences P_i over schools, where $P_i : a \succ b$ means that i *prefers* school a over school b. The set of all possible preference orders is denoted by $\mathcal{P}$. A *preference profile* $\boldsymbol{P} = (P_1, \ldots, P_n) \in \mathcal{P}^N$ is a collection of preferences of all students, and we denote by P_{-i} the collection of preferences of all students except i, so that we can write $\boldsymbol{P} = (P_i, P_{-i})$.

3.3.2 Deterministic and Probabilistic Assignments

A *deterministic assignment* of students to schools is represented by an $n \times m$-matrix $x = (x_{i,j})_{i \in N, j \in M}$, where the entry $x_{i,j}$ is equal to 1 if student i holds a seat at school j, and 0 otherwise. An assignment x is *feasible* if all students receive a seat at some

school (i.e., $\sum_{j \in M} x_{i,j} = 1$ for all students $i \in N$) and no school's seats are assigned beyond its capacity (i.e., $\sum_{i \in N} x_{i,j} \leqslant q_j$ for all schools $j \in M$). We denote by X the set of all feasible, deterministic assignments. From the perspective of a student, school choice mechanisms usually involve some randomness, most commonly because of random tie-breaking. We capture this by considering *probabilistic assignments*, which are lotteries over deterministic assignments. Extending the notation for deterministic assignments, we represent any probabilistic assignment as a matrix x, where the entry $x_{i,j}$ has a value between 0 and 1. $x_{i,j}$ is interpreted as the *probability* that student i is assigned a seat at school j.[3] Let $\Delta(X)$ denote the set of all feasible probabilistic assignments. For some student i and some probabilistic assignment $x \in \Delta(X)$, we denote by x_i the *assignment vector* of student i. This is the ith row $x_i = (x_{i,j})_{j \in M}$ of the matrix x containing the probabilities with which i is assigned to each of the schools.

3.3.3 Priorities and Random Tie-breaking

In school choice markets, schools usually do not report preferences over students, but instead priorities take the role of preferences on the school-side. Typically, some coarse priority requirements are imposed exogenously (e.g., based on neighborhoods or siblings), and the remaining ties are broken randomly. We model this structure explicitly by introducing *admissible priority distributions*.

A *priority order* π is a strict ordering of the students, where $i \ \pi \ i'$ means that student i has priority over student i', and we denote by Π the set of all possible priority orders. A *priority profile* is a collection of priority orders $\boldsymbol{\pi} = (\pi_j)_{j \in M} \in \Pi^M$, where each priority order π_j is associated with the school j. These priority profiles take the place of the preferences of the schools in the mechanisms that we consider.

We model exogenous priority requirements (such as neighborhood or sibling priorities) by restricting the set of *admissible priority profiles* $\boldsymbol{\Pi} \subseteq \Pi^M$. As an example, suppose that all students from the neighborhood of school j (denoted N_j) should have priority over any other students at school j. In this case, the set of admissible priority profiles $\boldsymbol{\Pi}$ would consist only of those $\boldsymbol{\pi} = (\pi_j, \pi_{-j})$ for which π_j gives preference to students from N_j; formally, for all $i \in N_j$ and $i' \in N \backslash N_j$, we have $i \ \pi_j \ i'$.

While a particular set $\boldsymbol{\Pi}$ can reflect coarse exogenous priority requirements, random

[3] By virtue of the Birkhoff-von Neumann Theorem (Birkhoff, 1946) and its extensions (Budish et al., 2013), it suffices to consider the matrix representation of probabilistic assignments, as they are always implementable. In fact, the way in which we construct DA, NBM, and ABM already prescribes a canonical way of implementing the resulting probabilistic assignments.

tie-breaking introduces uncertainty about which priority profile is ultimately used by the mechanism. A *priority distribution* $\mathbb{P}$ is a probability distribution over the set of priority profiles Π^N. $\mathbb{P}$ is said to be $\boldsymbol{\Pi}$-*admissible* if the support of $\mathbb{P}$ only contains admissible priority profiles; formally, $\text{supp}(\mathbb{P}) = \left\{\boldsymbol{\pi} \in \Pi^M \mid \mathbb{P}[\boldsymbol{\pi}] > 0\right\} \subseteq \boldsymbol{\Pi}$. Throughout the paper, we will use admissible priority distributions to incorporate coarse priority requirements as well as random tie-breaking into the mechanisms we study.

As mentioned in the introduction, some school choice markets operate with no priority requirements and ties are broken by a single uniform lottery, and in many, this reflects the situation at least *approximately*. A *single priority profile* is a priority profile $\boldsymbol{\pi}$ that selects the same priority order at all schools; formally, $\boldsymbol{\pi} = (\pi, \ldots, \pi)$ for some $\pi \in \Pi$. Otherwise, $\boldsymbol{\pi}$ is called a *multiple priority profile*. $\mathbb{P}$ is a *single priority distribution* if it only randomizes over single priority profiles, or equivalently, $\text{supp}(\mathbb{P})$ contains only single priority profiles. If in addition, ties are broken uniformly at random, we obtain the *single uniform priority distribution* $\mathbb{U}$, which is the uniform distribution over all single priority profiles; formally,

$$\mathbb{U}[\boldsymbol{\pi}] = \begin{cases} \frac{1}{n!}, & \text{if } \boldsymbol{\pi} = (\pi, \ldots, \pi) \text{ with } \pi \in \Pi, \\ 0, & \text{else.} \end{cases} \tag{197}$$

Under $\mathbb{U}$, all students are part of a single, large priority class, each student draws a unique random number, and conflicting preferences at any school are resolved using these numbers. All our results in this paper hold at least for the single uniform priority distribution, but most of them continue to hold for general priority structures and tie-breakers.

3.3.4 Construction of School Choice Mechanisms

We study mechanisms that arise by specifying the way in which they handle the students' preferences for each priority profile. A *deterministic school choice mechanism* is a mapping

$$\varphi : \Pi^M \times \mathcal{P}^N \to X. \tag{198}$$

that receives as input a priority profile $\boldsymbol{\pi} \in \Pi^M$ and student preferences $\boldsymbol{P} \in \mathcal{P}^N$ and selects a deterministic assignment based on this input. For a given (random) priority distribution $\mathbb{P}$ the respective *probabilistic school choice mechanism* (or just *mechanism* for short) is the mapping

$$\varphi^{\mathbb{P}} : \mathcal{P}^N \to \Delta(X) \tag{199}$$

that receives the students' preferences $\boldsymbol{P}$ and selects the probabilistic assignment

$$\varphi^{\mathbb{P}}(\boldsymbol{P}) = \sum_{\boldsymbol{\pi} \in \Pi^m} \varphi(\boldsymbol{\pi}, \boldsymbol{P}) \cdot \mathbb{P}[\boldsymbol{\pi}]. \tag{200}$$

In practice, this mechanism can easily be implemented by the following procedure: first, collect the preference reports $\boldsymbol{P}$ from the students. Second, choose a priority profile $\boldsymbol{\pi}$ randomly according to $\mathbb{P}$ but independent of $\boldsymbol{P}$. Third, assign students to schools according to $\varphi(\boldsymbol{\pi}, \boldsymbol{P})$. All mechanisms defined in the next section are constructed in this way. With slight abuse of notation we will also write $\varphi^{\boldsymbol{\pi}}$ for the deterministic mechanism $\varphi(\boldsymbol{\pi}, \cdot)$. For the sake of readability, we sometimes omit the superscript $\mathbb{P}$ and simply write φ when $\mathbb{P}$ is arbitrary or clear from the context.

3.3.5 Incentive Constraints

We briefly review the two common incentive requirements *strategyproofness* and *weak strategyproofness*. Let student $i \in N$ have preference order P_i, and let x_i and y_i be two assignment vectors for i. We say that x_i *(first order)-stochastically dominates* y_i *at* P_i if for all schools $j \in M$, i is at least as likely to receive a school she prefers to j under x_i than under y_i. Formally, for all schools $j \in M$

$$\sum_{j' \in M:\ P_i : j' \succ j} x_i \geqslant \sum_{j' \in M:\ P_i : j' \succ j} y_i. \tag{201}$$

x_i *strictly* stochastically dominates y_i at P_i if inequality (201) is strict for some j.

Definition 16 (Strategyproof)**.** A mechanism φ is *strategyproof* if misreporting ones preferences leads to an assignment vector that is stochastically dominated by the assignment vector obtained from truthful reporting.

Bogomolnaia and Moulin (2001) introduced weak strategyproofness to describe the incentives under the non-strategyproof Probabilistic Serial mechanism. While strategyproofness requires that students are unambiguously worse-off when misreporting, weak strategyproofness requires that students are not unambiguously better off when misreporting. This captures the idea that they may not want to manipulate as long as the benefit from the misreport is not obvious.

Definition 17 (Weakly Strategyproof)**.** A mechanism φ is *weakly strategyproof* if no student can obtain a strictly stochastically dominant assignment vector by misreporting her preferences.

3.4 School Choice Mechanisms

We now introduce the three mechanisms that we study in this paper, the Deferred Acceptance mechanism and two variants of the Boston mechanism.

3.4.1 Student Proposing Deferred Acceptance Mechanism

For any preference profile $\boldsymbol{P}$ and priority profile $\boldsymbol{\pi}$, the *(Student Proposing) Deferred Acceptance* mechanism selects the assignment $\mathrm{DA}(\boldsymbol{\pi}, \boldsymbol{P})$ as follows:

Round 1: All students apply to their first choice according to P_i. Each school j then processes all applications it has received:

- If j has sufficient capacity, all applications to j are *tentatively* accepted.
- If j has less capacity than applications, it *tentatively* accepts applications from students with highest priority according to π_j until all seats are filled. All other applications are *permanently* rejected.

Round k: Students who were not tentatively accepted at some school at the end of round $k-1$ apply to the best school (according to P_i) that has not permanently rejected them so far. The set of *candidates* at school j is comprised of the new applicants at j as well as all students who were tentatively accepted at j at the end of round $k-1$.

- If j has sufficient capacity, all candidates are tentatively accepted.
- If j has less capacity than candidates, it tentatively accepts candidates with highest priority according to π_j until all seats are filled. All other candidates are permanently rejected.

Termination: When no school receives new applications, the current assignment given by the tentative acceptances is finalized.

An important aspect of DA is that students' tentative acceptances can be revoked in subsequent rounds if this is necessary to accommodate applications from students with higher priority.

Remark 9. For a *single priority distribution* $\mathbb{P}$, the mechanism $\mathrm{DA}^{\mathbb{P}}$ is equivalent to a *Serial Dictatorship (SD)* mechanism where the order in which the students get to pick their schools is the (single) priority order drawn from $\mathbb{P}$ (Erdil, 2014). This makes the school choice problem formally equivalent to the *random assignment problem*.

For the special case of the *single uniform priority distribution* $\mathbb{U}$, this yields the well-known *Random Serial Dictatorship (RSD)* mechanism (i.e., $\mathrm{DA}^{\mathbb{U}} = \mathrm{RSD}$) (Morrill, 2013). Thus, our findings also shed light on the trade-offs between strategyproofness and efficiency in the random assignment domain as well.

3.4.2 Naïve Boston Mechanism

For any preference profile $\boldsymbol{P}$ and priority profile $\boldsymbol{\pi}$, the *naïve Boston mechanism* selects the assignment $\mathrm{NBM}(\boldsymbol{\pi}, \boldsymbol{P})$ as follows:

Round 1: All students apply to their first choice according to P_i. Each school j then processes all applications it has received:

- If j has sufficient capacity, all applications to j are *permanently* accepted.
- If j has less capacity than applications, it *permanently* accepts applications from students with highest priority according to π_j until all seats are filled. All other applications are *permanently* rejected.

Round k: All students who have not been permanently accepted at some school in rounds $1, \ldots, k-1$ apply to their kth choice school according to P_i; analogous to round 1, applicants are *permanently* accepted into the unoccupied seats at each school j according to π_j. When all seats at j are filled, the remaining applicants are *permanently* rejected.

Termination: When no school receives new applications, the assignment is finalized.

This mechanism works similarly to DA, but in contrast to DA the acceptance of a student in any round is *final* and cannot be revoked in subsequent rounds.

3.4.3 Adaptive Boston Mechanism

Under the "traditional" naïve Boston mechanism, students may apply to schools that have no more unfilled seats. When applying to such an exhausted school, a student has no chance of obtaining a seat at that school, *independent* of her priority at that school. While the student is making her futile application, the capacities of the other schools are further reduced. Thus, the student effectively wastes one round in which she could have competed for other schools instead. Under the *adaptive Boston mechanism (ABM)*, students automatically skip exhausted schools and instead apply to their most preferred

available schools in each round. Thus, a student's application may still be rejected, but the rejection always depends on her priority.

For a given preference profile $\boldsymbol{P}$ and priority profile $\boldsymbol{\pi}$, the *adaptive Boston mechanism* selects the assignment $\mathrm{ABM}(\boldsymbol{\pi}, \boldsymbol{P})$ as follows:

Round 1: All students apply to their first choice according to P_i. Each school j then processes all applications it has received:

- If j has sufficient capacity, all its applications are *permanently* accepted.
- If j has less capacity than applications, it *permanently* accepts applications from students with highest priority according to π_j until all seats are filled. All other applications are *permanently* rejected.

Round k: All students who have not been permanently accepted at some school in rounds $1, \ldots, k-1$ apply to their best choice according to P_i *out of those schools that have positive remaining capacity*; analogous to round 1, applicants are *permanently* accepted into the unoccupied seats at each school j according to π_j. When all seats at j are filled, the remaining applicants are *permanently* rejected.

Termination: When no school receives new applications, the assignment is finalized.

3.5 Incentives for Truth-telling

In this section, we present our first main result that ABM has intermediate incentive properties: while it is not strategyproof, it satisfies the intermediate incentive requirement of *partial strategyproofness* (Mennle and Seuken, 2015*b*). This establishes a hierarchy of manipulability between DA, ABM, and NBM. Our finding is in contrast with the rather surprising fact that a comparison by *vulnerability to manipulation* (Pathak and Sönmez, 2013) fails to differentiate between NBM and ABM, except in very special cases.

3.5.1 Failure of Comparison by Vulnerability to Manipulation

It is well-known that DA is strategyproof (Roth, 1982) while NBM is not even weakly strategyproof (Proposition 11 in Appendix 3.B.1). Even though ABM is not fully strategyproof, intuitively, it should have better incentive properties than NBM: under ABM, students automatically skip exhausted schools, which removes some obvious opportunities for manipulation. The motivating example in the introduction further

supports this intuition. However, a formal justification for this intuition has remained elusive so far.

One may hope to obtain a distinction via the *vulnerability to manipulation* concept (Pathak and Sönmez, 2013). There are three ways to conduct this comparison. Unfortunately, neither of them delivers satisfactory results: first, for fixed priority profiles $\boldsymbol{\pi} \in \Pi^M$, NBM$^{\boldsymbol{\pi}}$ is indeed *as manipulable as* ABM$^{\boldsymbol{\pi}}$ (Dur, 2015). However, the strict comparison that NBM$^{\boldsymbol{\pi}}$ is *more manipulable* than ABM$^{\boldsymbol{\pi}}$ only holds if some students are unacceptable at some schools. Second, the comparison cannot be strengthened to the statement that NBM$^{\boldsymbol{\pi}}$ is *strongly* as manipulable as ABM$^{\boldsymbol{\pi}}$ (see Examples 8 & 9 in Appendix 3.A.3). Third, when ties are broken randomly, as is common in school choice, then the mechanisms are not even comparable by the weaker *as manipulable as*-relation (see Examples 6 & 7 in Appendix 3.A.2).

This highlights that a different approach must be taken to obtain a conclusive comparison of ABM and NBM by their incentive properties. To this end, we employ the partial strategyproofness concept, which we review in the next section.

3.5.2 Review of Partial Strategyproofness

In (Mennle and Seuken, 2015*b*), we have shown that strategyproofness can be decomposed into three simple axioms. These axioms restrict the way in which a mechanism may change the assignment of some student when that student changes her report by swapping two consecutive schools in her reported preference order, e.g., from $P : a \succ b$ to $P' : b \succ a$. φ is called *upper invariant* if this swap leaves the student's assignment unchanged for any school that she strictly prefers to a, and φ is called *lower invariant* if it leaves her assignment unchanged for any school that she likes strictly less than b. Finally, φ is called *swap monotonic* if the swap either does not lead to a change of the student's assignment at all, or if it induces any change, then her probability for a must decrease strictly, and her probability for b must increase strictly.

Fact 3 (Mennle and Seuken, 2015*b*)**.** *A mechanism is strategyproof* if and only if *it is upper invariant, swap monotonic, and lower invariant.*

Now suppose that a student i has a vNM utility function u_i that is consistent with her preference order P_i. We say that u_i satisfies *uniformly relatively bounded indifference* with respect to *indifference bound* $r \in [0, 1]$ (URBI(r)) if for any schools $a, b \in M$ with

$P_i : a \succ b$ we have that

$$r\left(u_i(a) - \min_{j \in M} u_i(j)\right) \geqslant u_i(b) - \min_{j \in M} u_i(j), \tag{202}$$

This implies that the factor difference between i's (normalized) preference intensity for a over b is a least $1/r$. Lower r means that the student differentiates more strongly, while higher r allows her to be closer to indifferent between the two schools a and b.

Definition 18 (Partially Strategyproof)**.** For a given *setting* (i.e., set of students, set of schools, and school capacities), a mechanism φ is *r-partially strategyproof* if truthful reporting is a dominant strategy for any student whose vNM utility u_i satisfies URBI(r). φ is *partially strategyproof* if it is r-partially strategyproof for some $r > 0$.

Fact 4 (Mennle and Seuken, 2015*b*)**.** *For a given setting, a mechanism is partially strategyproof* if and only if *it is swap monotonic and upper invariant.*

Partial strategyproofness is implied by strategyproofness, and it implies weak strategyproofness (Bogomolnaia and Moulin, 2001), convex strategyproofness (Balbuzanov, 2015), approximate strategyproofness (Carroll, 2013), strategyproofness in the large (for $r \to 1$) (Azevedo and Budish, 2015), and lexicographic strategyproofness (Cho, 2012). Thus, partial strategyproofness can be understood as an intermediate incentive requirement. We further discuss its implications in the context of our partial strategyproofness result for ABM in Section 3.5.3.

3.5.3 Partial Strategyproofness of ABM

Our first main result formally establishes that the incentive properties of ABM are in fact *intermediate* between those of DA and NBM.

Theorem 9. *ABM$^{\mathbb{U}}$ is partially strategyproof but not strategyproof.*

Proof Outline (formal proof in Appendix 3.B.2). We prove partial strategyproofness of ABM$^{\mathbb{U}}$ by showing upper invariance and swap monotonicity and using Fact 4. For upper invariance, we first show that ABM$^{\boldsymbol{\pi}}$ is upper invariant for any priority profile $\boldsymbol{\pi}$, and then we show that this property is inherited by any mechanism that randomly selects the priority profile $\boldsymbol{\pi}$ according to some priority distribution $\mathbb{P}$. The more challenging proof is swap monotonicity: we first observe that ABM$^{\boldsymbol{\pi}}$ is always *monotonic* (i.e., bringing a school up in one's ranking never decreases the chances of obtaining that

school). Next, given any priority profile $\boldsymbol{\pi}$ such that $\text{ABM}^{\boldsymbol{\pi}}$ changes the manipulating student's assignment under a swap of some schools (a and b, say), we construct a single priority profile $\boldsymbol{\pi}^*$ such that under $\text{ABM}^{\boldsymbol{\pi}^*}$, the manipulating student receives either a or b, depending on the relative ranking of a and b in her report. Thus, the change in probability for a and b is *strict* because $\boldsymbol{\pi}^*$ is chosen with positive probability. □

Theorem 9 has a number of interesting consequences. First, partial strategyproofness is the strongest incentive requirement (for finite markets) that has been shown to hold for the celebrated Probabilistic Serial (PS) mechanism (Mennle and Seuken, 2015*b*). Thus, from an axiomatic perspective, Theorem 9 means that the incentive properties of $\text{ABM}^{\mathbb{U}}$ are in the same class as those of the PS mechanism.

Second, partial strategyproofness implies weak strategyproofness: a student cannot obtain a stochastically dominant assignment by misreporting her preferences. Put differently, any manipulation will necessarily involve a trade-off on the part of the student between probabilities for different schools. This is illustrated by the example in the introduction: recall that student 1 could obtain schools a, b, d with probability 1/3 each under truthful reporting, or she could obtain her third choice school c with certainty by ranking c in first position. By misreporting, the student had to "sacrifices" all probability for her first choice a in order to convert chances to obtain her last choice d into chances to obtain c. Whether or not she would prefer this manipulation to reporting truthfully depends on her relative preference intensities for the different schools. Theorem 9 teaches us that any manipulation will take such a form, and no student can gain *unambiguously* (in a first order-stochastic dominance sense) from misreporting.

Third, partial strategyproofness by Theorem 9 implies that $\text{ABM}^{\mathbb{U}}$ makes truthful reporting a dominant strategy for all students who differentiate sufficiently between different schools. Formally, for any setting, there exists $r > 0$ such that any student whose vNM utility satisfies $\text{URBI}(r)$ will have a dominant strategy to be truthful. Thus, even though $\text{ABM}^{\mathbb{U}}$ is not strategyproof, we can give honest and useful strategic advice to the students: they are best off reporting their preferences truthfully as long as they are not too close to indifferent between any two schools.

Remark 10. It is worth noting that NBM satisfies the upper invariance axiom, which is essentially equivalent to *truncation robustness*: students cannot improve their chances of obtaining a better school by "truncating" their preference reports and falsely claiming that some lower ranking schools are unacceptable (Hashimoto et al., 2014). However, NBM violates swap monotonicity (Proposition 11 in Appendix 3.B.1), and therefore it

Property	$DA^{\mathbb{U}}$	$ABM^{\mathbb{U}}$	$NBM^{\mathbb{U}}$
Upper invariant	✓	✓	✓
Swap monotonic	✓	✓	✗
Lower Invariant	✓	✗	✗
Partially strategyproof	✓	✓	✗
Strategyproof	✓	✗	✗

Table 3.1: Incentive properties of mechanisms

cannot be partially strategyproof. Table 3.1 provides and overview of the properties that each of the mechanisms violate or satisfy.

Generality of Theorem 9: We have proven that Theorem 9 continues to hold for a larger class of priority distributions. We say that a priority distribution $\mathbb{P}$ *supports all single priority profiles* if any single priority profile is selected with positive probability. This means that $\mathbb{P}[(\pi, \ldots, \pi)] > 0$ for all $\pi \in \Pi$, but multiple priority profiles may also be selected. Our proof of Theorem 9 covers the more general statement:

Theorem 10. *For any priority distribution $\mathbb{P}$ that supports all single priority profiles, $ABM^{\mathbb{P}}$ is partially strategyproof.*

3.5.4 Partial Strategyproofness for Arbitrary Priority Distribution

Partial strategyproofness of $ABM^{\mathbb{P}}$ by Theorem 10 hinges on the randomness induced by the random choice of the priority profile. However, if the priority distribution is not sufficiently random (or if the priority profile is fixed), then $ABM^{\mathbb{P}}$ may no longer be partially strategyproof. To obtain a meaningful comparison of $ABM^{\mathbb{P}}$ an $NBM^{\mathbb{P}}$ even in this case, we can consider a second source of randomness: if students are uncertain about the preference reports from the other students, then they face a random mechanism because the outcome depends on the unknown reports P_{-i}. This random mechanism is guaranteed to be partially strategyproof if the original mechanism is upper invariant, monotonic, and sensitive[4] (see Theorem 8 of (Mennle and Seuken, 2015*b*)). The following Proposition 8 shows that NBM and ABM both satisfy these conditions.

[4] A mechanism is *sensitive* if for any swap that changes student i's assignment for some P_{-i}, there also exist P'_i, P''_{-i} such that the swap changes that i's assignment of each of the two objects that she swaps, respectively.

Proposition 8. *For any priority distribution* $\mathbb{P}$*, the mechanisms* $NBM^{\mathbb{P}}$ *and* $ABM^{\mathbb{P}}$ *are upper invariant, monotonic, and sensitive.*

A formal proof is given in Appendix 3.B.3.

It follows that the partial strategyproofness concept is applicable to $\text{ABM}^{\mathbb{P}}$ and $\text{NBM}^{\mathbb{P}}$ even for priority distributions that do not support all single priority profiles. While this insight does not in itself establish a distinction between $\text{ABM}^{\mathbb{P}}$ and $\text{NBM}^{\mathbb{P}}$ in terms of their incentive properties, it implies that we can compare the two mechanisms using the *degree of strategyproofness* measure. This comparison will be an interesting subject for future research.

3.6 Efficiency Comparison

As is common in the study of school choice mechanisms, we assess the welfare properties of school choice mechanisms via *dominance* and *efficiency* notions. Specifically, we compare the resulting assignments $\text{DA}^{\mathbb{P}}(\boldsymbol{P})$, $\text{NBM}^{\mathbb{P}}(\boldsymbol{P})$, and $\text{ABM}^{\mathbb{P}}(\boldsymbol{P})$ at the preference profile $\boldsymbol{P}$.

3.6.1 Decoupling Strategyproofness and Efficiency

Since DA is strategyproof (for students), one would expect students to report their preferences truthfully. Therefore, it is meaningful to assess the efficiency of DA by analyzing the resulting assignments at the true preference profiles. However, NBM and ABM are not strategyproof, and thus strategic students would not necessarily report truthfully. Assuming fully rational student behavior, the most accurate assessment of efficiency would be to consider the assignments that arise *in equilibrium*; a mechanism would be considered to be "more efficient" if the resulting assignments dominate those of another mechanism *in equilibrium* with respect to the *true preferences*. However, for the general case with coarse priorities and random tie-breaking, the shape of the equilibria under NBM and ABM is an open research question and beyond the scope of this paper.

For the case of fixed, strict priority profiles, it is known that the Nash equilibrium outcomes of $\text{NBM}^{\boldsymbol{\pi}}$ are weakly dominated by those of $\text{DA}^{\boldsymbol{\pi}}$ (Ergin and Sönmez, 2006). Conversely, for the single uniform priority distribution $\mathbb{U}$, the equilibrium outcomes under $\text{NBM}^{\mathbb{U}}$ at some vNM utility profiles can be preferred by all students to those under $\text{DA}^{\mathbb{U}}$ *ex-ante* (Miralles, 2008; Abdulkadiroğlu, Che and Yasuda, 2015). These "contradictory" findings illustrate that even if equilibria were known for the general case,

an efficiency comparison "in equilibrium" will not indicate a clear preference for either of the mechanisms.

The fact that an unambiguous equilibrium analysis is presently unavailable (and beyond the scope of this paper) motivates the alternative approach we take in this paper: we *decouple* strategyproofness and efficiency by considering each dimension separately. Concretely, we compare the efficiency of DA, NBM, and ABM when each mechanism is applied to the true preference profile. This comparison is straight-forward and analogous to the assessment of non-strategyproof mechanisms in prior work, e.g., the celebrated Probabilistic Serial mechanism (Bogomolnaia and Moulin, 2001) is ordinally efficient only when applied to the true preference profile, and the men-proposing Deferred Acceptance mechanism (Gale and Shapley, 1962) is men-optimal only if applied to the true preferences profile (in particular, when women do not misreport). Similarly, characterization results for non-strategyproof mechanisms often rely on axioms that involve the true preferences, such as *ordinal fairness* (Hashimoto et al., 2014) and *respect of preference rankings* (Kojima and Ünver, 2014). The assignments resulting from the respective mechanisms satisfy the characterizing properties only when the mechanisms are applied to the true preferences. Until a more comprehensive equilibrium analysis becomes available, these (and our) insights serve as a useful second-best to inform decisions about school choice mechanisms.

3.6.2 Dominance and Efficiency Concepts

We now review the dominance and efficiency notions for our analysis. In the following, $\boldsymbol{P}$ is a preferences profile, x and y are assignments, and φ is a probabilistic mechanism.

Definition 19 (Ex-post Efficient)**.**

1. For deterministic x, y, x *ex-post dominates* y *at* $\boldsymbol{P}$ if all students weakly prefer their school under x to their school under y. This dominance is *strict* if the preference is strict for at least one student.
2. A deterministic x is *ex-post efficient at* $\boldsymbol{P}$ if it is not strictly ex-post dominated by any other deterministic assignment at $\boldsymbol{P}$.
3. A probabilistic x is *ex-post efficient at* $\boldsymbol{P}$ if it can be written as a convex combination of ex-post efficient, deterministic assignments.
4. φ is *ex-post efficient* if $\varphi(\boldsymbol{P})$ is ex-post efficient at $\boldsymbol{P}$ for all $\boldsymbol{P} \in \mathcal{P}^N$.

Ex-post efficiency can be viewed as a baseline requirement in the school choice problem. Conceivably, an administrator will be hard pressed to explain why two students received their particular placements if they would actually prefer the seats at the respective other schools.[5]

Definition 20 (Ordinally Efficient)**.**

1. *x ordinally dominates y at $\boldsymbol{P}$* if for all students $i \in N$ the respective assignment vector x_i weakly stochastically dominates y_i at P_i. This dominance is *strict* if x_i strictly stochastically dominates y_i for at least one student i.
2. x is *ordinally efficient at $\boldsymbol{P}$* if it is not strictly ordinally dominated at $\boldsymbol{P}$ by any other assignment.
3. φ is *ordinally efficient* if $\varphi(\boldsymbol{P})$ is ordinally efficient at $\boldsymbol{P}$ for all $\boldsymbol{P} \in \mathcal{P}^N$.

Ordinal efficiency formalizes the idea of Pareto optimality for probabilistic assignments. It is equivalent to the absence of trade cycles that can be shown to benefit all students (strict for some) when only the students' ordinal preferences are known. This efficiency notion has been used by Bogomolnaia and Moulin (2001) to describe the efficiency advantages of the non-strategyproof Probabilistic Serial mechanism over the strategyproof Random Serial Dictatorship mechanism. Ordinal efficiency implies ex-post efficiency, but the converse does not hold.

Definition 21 (Rank Efficient)**.**

1. Let $\text{rank}_{\boldsymbol{P}}(i,j)$ denote the *rank* of school j in the preference order P_i, which is the number of schools that i weakly prefers to j. Let

$$d_k^x = \sum_{i \in N} \sum_{j \in M:\ \text{rank}_{\boldsymbol{P}}(i,j) \leqslant k} x_{i,j} \tag{203}$$

   be the expected number of students who receive their kth choice under x (given $\boldsymbol{P}$). Then the vector $d^x = (d_1^x, \ldots, d_m^x)$ is called the *rank distribution of x at $\boldsymbol{P}$*.

2. *x rank dominates y at $\boldsymbol{P}$* if for all ranks $k \in \{1, \ldots, m\}$ we have

$$\sum_{r=1}^{k} d_k^x \geqslant \sum_{r=1}^{k} d_k^y. \tag{204}$$

[5] Parents of secondary school students in Amsterdam have gone to court in 2015 over the fact that multiple tie-breaking led to an ex-post *in*efficient assignment of students (de Haan et al., 2015).

The rank dominance of x over y is *strict* if inequality (204) holds for all k and is strict for some.

3. x is *rank efficient at* $\boldsymbol{P}$ if it is not rank dominated by any other assignment at $\boldsymbol{P}$.
4. φ is *rank efficient* if $\varphi(\boldsymbol{P})$ is rank efficient at $\boldsymbol{P}$ for all $\boldsymbol{P} \in \mathcal{P}^N$.

Rank efficiency represents a strict refinement of ordinal efficiency. To illustrate the difference, consider the trade-cycle interpretations of both efficiency concepts: an assignment is ordinally efficient if it does not admit trade cycles of probability shares that are unambiguously preferred by all students, given their ordinal preferences. However, it may still be possible to identify trade cycles that hurt one student but create a "much higher" benefit for another student, e.g., if assigning student i to her 2nd rather than 1st choice allows us to assign another student to her 1st rather than 3rd choice, this would improve overall rank distribution (from $(1, 0, 1)$ to $(1, 1, 0)$), but it hurts student i. Rank efficient assignments do not even admit improvements by these kinds of "tough decisions." Focusing on the rank distribution closely resembles welfare criteria that are frequently used in practice: many school choice procedures in Germany have the express objective of assigning as many students to one of their top-3 or top-5 choices (Basteck, Huesmann and Nax, 2015). Similarly, rank efficiency resembles the informal objective of the matching procedure of the Teach-for-America program (Featherstone, 2011).

Finally, we define what it means for a mechanism to be on the efficient frontier.

Definition 22 (Efficient Frontier)**.**

1. φ *ordinally (or rank) dominates* another mechanism φ' if the assignment $\varphi(\boldsymbol{P})$ ordinally (or rank) dominates $\varphi'(\boldsymbol{P})$ at $\boldsymbol{P}$ for all $\boldsymbol{P} \in \mathcal{P}^N$. This is *strict* if in addition, $\varphi(\boldsymbol{P})$ strictly ordinally (or rank) dominates $\varphi'(\boldsymbol{P})$ at $\boldsymbol{P}$ for some preference profile $\boldsymbol{P} \in \mathcal{P}^N$.
2. For some set Φ of mechanisms, we say that $\varphi \in \Phi$ is *on the efficient frontier with respect to ordinal (or rank) dominance, subject to* Φ, if it is not strictly ordinally (or rank) dominated by any other mechanism φ' from Φ.

Intuitively, mechanisms on the efficient frontier are as efficient as possible while simultaneously satisfying other design requirements described by Φ, e.g., strategyproofness or fairness. Put differently, it may be possible to design a mechanism φ' that outperforms φ in terms of efficiency, but any such φ' would necessarily violate other criteria (i.e., equivalently lie outside Φ).

Some assignments are not comparable by any of the dominance concepts defined above. Consequently, some mechanisms may not be comparable at *every* preference profile. The strict extension of the dominance notions to mechanisms from Definition 22 will yield inconclusive results. This in-comparability also occurs for NBM, ABM, and DA. To overcome this difficulty, we formalize the idea that a mechanism should be considered *more efficient* if it dominates another mechanism at least at those preference profile *at which the two mechanisms are comparable.*

Definition 23 (Comparable Dominance)**.** *φ comparably ordinally (or rank) dominates* another mechanism φ' if $\varphi(\boldsymbol{P})$ ordinally (or rank) dominates $\varphi'(\boldsymbol{P})$ at $\boldsymbol{P}$ for all $\boldsymbol{P} \in \mathcal{P}^N$ whenever $\varphi(\boldsymbol{P})$ and $\varphi'(\boldsymbol{P})$ are comparable by ordinal (or rank) dominance.

Comparable dominance is less demanding than "perfect" dominance, as the decision in favor of one mechanism is based solely on those profiles where a comparison is possible. The advantage is that more mechanisms become comparable but the distinction remains unambiguous.

3.6.3 Efficiency of NBM

Regarding the efficiency of NBM, we first establish that the traditional dominance and efficiency notions may not differentiate between NBM and DA. We then present our second main result (Theorem 11) that NBM comparably rank dominates DA. This formalizes our intuition from the motivating example in the introduction that "NBM is more efficient than DA."

The following Fact 5 summarizes the known efficiency properties of DA.

Fact 5.

1. *$DA^{\mathbb{P}}$ is not ex-post efficient in general (Roth and Sotomayor, 1990). For any single priority distribution $\mathbb{P}$, $DA^{\mathbb{P}}$ is ex-post efficient but may not be ordinally or rank efficient (Bogomolnaia and Moulin, 2001; Featherstone, 2011).*
2. *Among all strategyproof, symmetric mechanisms, $DA^{\mathbb{U}} = RSD$ is not on the efficient frontier with respect to ordinal dominance when the total number of seats at schools exceeds the number of students (Erdil, 2014).*

In Proposition 9 we provide an analogous assessment for NBM.

Proposition 9.

1. *For any priority distribution* $\mathbb{P}$*, NBM*$^{\mathbb{P}}$ *is ex-post efficient but may not be ordinally or rank efficient.*
2. *Among all upper invariant mechanisms, NBM*$^{\mathbb{U}}$ *is not on the efficient frontier with respect to ordinal (or rank) dominance. Equivalently, there exists an upper invariant mechanism that strictly ordinally (and rank) dominates NBM*$^{\mathbb{U}}$*.*

Proof Outline (formal proof in Appendix 3.B.4). We prove ex-post efficiency of NBM by showing that for each preference profile $\boldsymbol{P}$ and each priority profile $\boldsymbol{\pi}$, we can construct a *single* priority profile $\boldsymbol{\sigma} = (\sigma, \ldots, \sigma)$ such that $\text{NBM}^{\boldsymbol{\pi}}(\boldsymbol{P}) = \text{SD}^{\boldsymbol{\sigma}}(\boldsymbol{P})$. Since SD is ex-post efficient, $\text{NBM}^{\boldsymbol{\pi}}(\boldsymbol{P})$ must be ex-post efficient at $\boldsymbol{P}$.[6]

To see the second claim, we construct a mechanism NBM^{+} that is essentially equal to $\text{NBM}^{\mathbb{U}}$, except for a certain set of preference profiles. For these preference profiles, NBM^{+} chooses the assignment selected by the Probabilistic Serial mechanism instead. It is easy to show that NBM^{+} ordinally dominates $\text{NBM}^{\mathbb{U}}$, and that this dominance may be strict. Since ordinal dominance implies rank dominance, $\text{ABM}^{\mathbb{U}}$ is also rank dominated by ABM^{+}. To show that NBM^{+} satisfies upper invariance, we show that for any swap, the mechanism's changes to the assignment are consistent with upper invariance. In particular, this is true for transitions between the special preference profiles where NBM^{+} and $\text{NBM}^{\mathbb{U}}$ choose different assignments and those preference profiles where the assignments of both mechanisms are equal. □

Juxtaposing Proposition 9 about the efficiency of NBM to Fact 5 about the efficiency of DA reveals that it may be difficult to differentiate between the two mechanisms: for single priority distributions, both mechanisms are ex-post efficient but fail more demanding efficiency requirements, and neither of them is on the efficient frontier subject to their respective incentive properties. Our second main result Theorem 11 resolves this ambiguity as it uncovers the efficiency advantage of NBM over DA in terms of *comparable rank dominance.*

Theorem 11. *NBM*$^{\mathbb{U}}$ *strictly comparably rank dominates DA*$^{\mathbb{U}}$*:*

1. *NBM*$^{\mathbb{U}}(\boldsymbol{P})$ *rank dominates DA*$^{\mathbb{U}}(\boldsymbol{P})$ *at* $\boldsymbol{P}$ *for any* $\boldsymbol{P} \in \mathcal{P}^N$ *where NBM*$^{\mathbb{U}}(\boldsymbol{P})$ *and DA*$^{\mathbb{U}}(\boldsymbol{P})$ *are comparable by rank dominance,*
2. *there exists a preference profile* $\boldsymbol{P} \in \mathcal{P}^N$ *such that NBM*$^{\mathbb{U}}(\boldsymbol{P})$ *strictly rank dominates DA*$^{\mathbb{U}}(\boldsymbol{P})$ *at* $\boldsymbol{P}$*.*

[6] Even though ex-post efficiency of NBM appears straight-forward, we are not aware of any formal proof, and we therefore give a proof for completeness.

Proof Outline (formal proof in Appendix 3.B.4). We need to show that if $\mathrm{DA}^{\mathbb{U}}(\boldsymbol{P})$ and $\mathrm{NBM}^{\mathbb{U}}(\boldsymbol{P})$ are comparable by rank dominance, then this comparison favors NBM. For any single priority profile $\boldsymbol{\pi}$ we show that if $\mathrm{DA}^{\boldsymbol{\pi}}$ and $\mathrm{NBM}^{\boldsymbol{\pi}}$ assign the same number of first choices, they will in fact assign these first choices to the same students. We can remove these students and the corresponding schools and proceed by induction, carefully handling the case when some school has capacity zero. Averaging over priority profiles, we find that if DA weakly rank dominates NBM, then the assignment from both mechanism must be the same. Therefore, DA never strictly rank dominates NBM. The comparison is strict for the preference profile considered in the motivating example in the introduction. □

The rank distribution is frequently used as a welfare criterion by administrators and researchers (e.g., see (Featherstone, 2011; Budish and Cantillon, 2012)). In line with this approach, Theorem 11 shows that $\mathrm{NBM}^{\mathbb{U}}$ yields the more appealing assignments whenever the results of $\mathrm{NBM}^{\mathbb{U}}$ and $\mathrm{DA}^{\mathbb{U}}$ are comparable. This is the case, for instance, in the motivating example from the introduction, where $\mathrm{DA}^{\mathbb{U}}$ assigns fewer first choices than $\mathrm{NBM}^{\mathbb{U}}$ but also assigns third choices where $\mathrm{NBM}^{\mathbb{U}}$ does not.

The significance of Theorem 11 is further emphasized by the fact that traditional efficiency notions are unable to differentiate between the two mechanisms (Fact 5 and Proposition 9). The clear preference for $\mathrm{NBM}^{\mathbb{U}}$ in terms of efficiency from Theorem 11 contrasts with the clear preference for $\mathrm{DA}^{\mathbb{U}}$ in terms of strategyproofness. In this sense, we have identified the "cost of strategyproofness" that a mechanism designer incurs when choosing Deferred Acceptance over the naïve Boston mechanism.

Simulation Results: In addition to the theoretical insights from Theorem 11, one may want to understand the *strength* of the dominance of $\mathrm{NBM}^{\mathbb{U}}$ over $\mathrm{DA}^{\mathbb{U}}$: if the assignments from both mechanisms are comparable, how often does the outcome of $\mathrm{NBM}^{\mathbb{U}}$ have a *strictly* better rank distribution?

For settings with $n = m \in \{3, \ldots, 10\}$ and $q_j = 1$ for all $j \in M$, we sampled 100'000 preference profiles $\boldsymbol{P}$ uniformly at random for each value of n. Whenever $\mathrm{NBM}^{\mathbb{U}}(\boldsymbol{P})$ and $\mathrm{DA}^{\mathbb{U}}(\boldsymbol{P})$ were comparable by rank dominance at $\boldsymbol{P}$, we determined whether the dominance was weak or strict. Figure 3.2 shows the results. For small n, the rank distributions of both assignments are frequently equivalent because both mechanisms often produce the same assignments. As n increases, however, the efficiency advantage of $\mathrm{NBM}^{\mathbb{U}}$ over $\mathrm{DA}^{\mathbb{U}}$ quickly becomes apparent: for $n = 8$ or more students, the share of profiles where $\mathrm{NBM}^{\mathbb{U}}$ produces a *strictly* better rank distribution is above 95% (subject

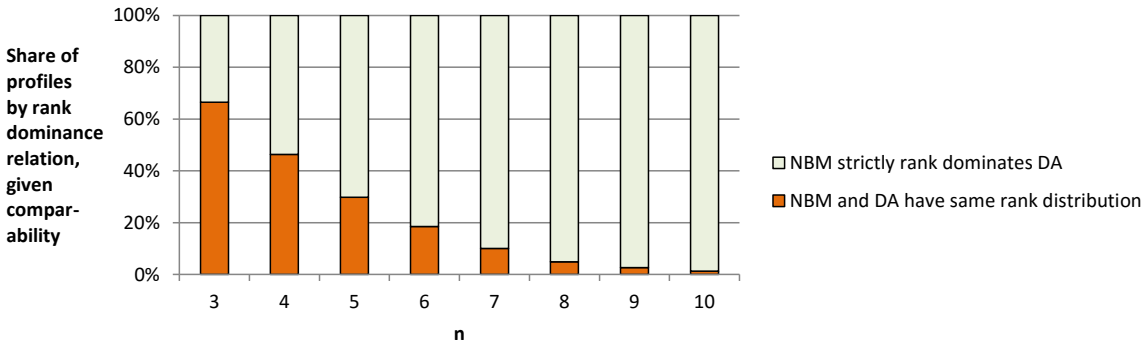

Figure 3.2: Share of preference profiles by rank dominance relation between $\text{NBM}^{\mathbb{U}}$ and $\text{DA}^{\mathbb{U}}$ (conditional on comparability), for settings with unit capacity and $n = m$ students.

to comparability), and it keeps increasing as n grows.

Generality of Theorem 11:

Theorem 12. *For* any *priority distribution* $\mathbb{P}$*, NBM*$^{\mathbb{P}}$ *comparably rank dominates DA*$^{\mathbb{P}}$*.*

Proof. For any fixed single priority profile, we first prove the following statement: if $\text{DA}^{\boldsymbol{\pi}}$ rank dominates $\text{NBM}^{\boldsymbol{\pi}}$ at some preference profile, then both mechanisms must select the same assignment at that preference profile. This insight allows us to formulate an "extension"-argument and prove comparable rank dominance for any *single priority distribution*. For the case of fixed *multiple* priority profiles, Harless (2015) proved the statement about coincidence of the assignments. Observing that his proof is essentially analogous to ours, we can re-use our "extension"-argument to obtain comparable rank dominance of $\text{NBM}^{\mathbb{P}}$ over $\text{DA}^{\mathbb{P}}$ for *arbitrary priority distributions* $\mathbb{P}$. □

To see the significance of Theorem 12, recall that DA may be ex-post *in*efficient for multiple priority profiles (Fact 5), while NBM is always ex-post efficient. This difference can be interpreted as an efficiency advantage of NBM, but the distinction only holds for multiple priority profiles. The generalized comparison in Theorem 12 identifies the efficiency advantage of NBM, *independent* of the priority distribution.

In summary, we have formally established the efficiency advantage of NBM over DA. When administrators face a decision on whether to implement NBM or DA in a school choice setting, this difference can be interpreted as the "cost of strategyproofness" that one incurs when choosing the strategyproof DA over the non-strategyproof NBM.

3.6.4 Efficiency of ABM

In this section we study the efficiency of the *adaptive* Boston mechanism, ABM. We first observe that (analogous to NBM) traditional efficiency notions do not differentiate between ABM and DA, nor between ABM and NBM.[7]

Proposition 10.

1. *For any priority distribution* $\mathbb{P}$*, ABM*$^{\mathbb{P}}$ *is ex-post efficient but may not be ordinally or rank efficient.*
2. *Among all partially strategyproof mechanisms, ABM*$^{\mathbb{U}}$ *is not on the efficient frontier with respect to ordinal (or rank) dominance. Equivalently, there exists a partially strategyproof mechanism that strictly ordinally (and rank) dominates ABM*$^{\mathbb{U}}$*.*

Proof Outline (formal proof in Appendix 3.B.5). The proof is analogous to the proof of Proposition 9 for NBM. When constructing the mechanism ABM^+, we must also ensure that ABM^+ is swap monotonic. This requires careful analysis of the transition cases where the preference profile changes from one with $\text{ABM}^+ = \text{ABM}^{\mathbb{U}}$ to another with $\text{ABM}^+ = \text{PS}$. □

Proposition 10 about ABM is analogous to Proposition 9 about NBM and Fact 5 about DA. These results leave all three mechanisms looking similar in terms of traditional measures of efficiency. The natural next step is to attempt another comparison by comparable rank dominance. However, surprisingly, this turns out to be inconclusive between both pairs ABM and DA as well as ABM and NBM (Section 3.6.4). Despite this in-comparability, limit arguments and simulations allow us to tease out the intuitive but well-hidden efficiency differences between ABM and the other mechanisms (Section 3.6.4). This establishes our third main result, the intermediate efficiency properties of ABM *between* DA and NBM.

In-comparability of ABM by Comparable Rank Dominance

One would expect that ABM comparably rank dominates DA and is comparably rank dominated by NBM. Indeed, in the motivating example discussed in the introduction, NBM rank dominates ABM, which in turn rank dominates DA. Consequently, it is at least possible that such a comparison of NBM, ABM, and DA is strict and points in the expected direction.

[7] Dur (2015) and Harless (2015) independently proved ex-post efficiency of $\text{ABM}^{\boldsymbol{\pi}}$ for fixed priority profiles $\boldsymbol{\pi}$.

Surprisingly, however, it can also point in the opposite direction: in Examples 11 and 12 (in Appendix 3.B.5) we present preference profiles $\boldsymbol{P}$ and $\boldsymbol{P}'$ such that

1. $\mathrm{DA}^{\mathbb{U}}(\boldsymbol{P})$ strictly rank dominates $\mathrm{ABM}^{\mathbb{U}}(\boldsymbol{P})$ at $\boldsymbol{P}$, and
2. $\mathrm{ABM}^{\mathbb{U}}(\boldsymbol{P}')$ strictly rank dominates $\mathrm{NBM}^{\mathbb{U}}(\boldsymbol{P}')$ at $\boldsymbol{P}'$.

Remark 11. Note that by Theorem 11, there cannot exist a single preference profile $\boldsymbol{P}$ at which 1. and 2. hold simultaneously. In the light of this restriction, our counter-examples are very general: in both examples, we give a single priority profile under which the respective dominance relation holds. Independently, Harless (2015) presented similar examples, but his examples rely critically on multiplicity of the priority profiles. Since our examples do not make use of this multiplicity, they show that comparability cannot be recovered, even when restricting attention to single priority profiles or the single uniform priority distribution.

These examples teach us that the identification of the efficiency differences between ABM and the other mechanisms requires a more subtle approach, since even the rather flexible requirement of comparable rank dominance does not enable a comparison.

Comparing ABM and DA: Comparable Rank Dominance in Large Markets

The surprising fact that $\mathrm{DA}^{\mathbb{U}}$ may strictly rank dominate $\mathrm{ABM}^{\mathbb{U}}$ at some preference profile (Example 11) raises the question how frequently this "unexpected" dominance relation occurs. Interestingly, complete enumeration (using a computer) has revealed that $\mathrm{DA}^{\mathbb{U}}$ does not dominate $\mathrm{ABM}^{\mathbb{U}}$ for any setting with less than 6 schools (assuming unit capacities). Furthermore, while such cases are possible for 6 or more schools, they turn out to be extremely rare. We now present our third main result that the share of preference profiles where $\mathrm{DA}^{\mathbb{U}}$ dominates $\mathrm{ABM}^{\mathbb{U}}$ vanishes in the limit as markets get large.

For our limit results, Theorems 13 and 14, we consider two independent notions of how market size increases: the first notion is adopted from (Kojima and Manea, 2010), where the number of schools is fixed, but the number of seats at each school grows as well as the number of students who demand them. This approximates school choice settings, where the capacity of public schools frequently exceeds 100 seats per school. For the second notion, all schools have unit capacity, but the numbers of schools and students increase. This resembles *house allocation problems*, where every "house" is different and can only be assigned once.

Theorem 13 (School Choice)**.** *Let $(N^k, M^k, \mathbf{q}^k)_{k\geqslant 1}$ be a sequence of settings such that*

- *the set of schools does not change (i.e., $M^k = M$ for all k),*
- *the capacity of each school increases (i.e., $\min_{j\in M} q_j^k \to \infty$ for $k \to \infty$),*
- *the number of students equals the number of seats (i.e., $|N^k| = \sum_{j\in M} q_j^k$).*

Then the share of preference profiles where $DA^{\mathbb{U}}$ rank dominates $ABM^{\mathbb{U}}$ (even weakly) vanishes in the limit:

$$\lim_{k\to\infty} \frac{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k} : DA^{\mathbb{U}}(\boldsymbol{P}) \text{ rank dominates } ABM^{\mathbb{U}}(\boldsymbol{P}) \text{ at } \boldsymbol{P}\}}{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k}\}} = 0. \tag{205}$$

Theorem 14 (House Allocation)**.** *Let $(N^k, M^k, \mathbf{q}^k)_{k\geqslant 1}$ be a sequence of settings such that*

- *the number of schools equals the number of students (i.e., $|M^k| = |N^k| = k$),*
- *all schools have unit capacity (i.e., $q_j^k = 1$ for $j \in M^k$).*

Then the share of preference profiles where $DA^{\mathbb{U}}$ rank dominates $ABM^{\mathbb{U}}$ (even weakly) vanishes in the limit:

$$\lim_{k\to\infty} \frac{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k} : DA^{\mathbb{U}}(\boldsymbol{P}) \text{ rank dominates } ABM^{\mathbb{U}}(\boldsymbol{P}) \text{ at } \boldsymbol{P}\}}{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k}\}} = 0. \tag{206}$$

Proof Outline (formal proofs in Appendix 3.B.5 and 3.B.5). For Theorem 13, we prove the stronger statement that the share of profiles where $\mathrm{DA}^{\mathbb{U}}$ assigns the same number of *first choices* as $\mathrm{ABM}^{\mathbb{U}}$ converges to zero. We give a bound for this share in terms of multinomial coefficients. Here, we must separately treat the conditional probabilities of the different cases that schools are un-demanded, under-demanded, over-demanded, or exactly exhausted as first choices. Theorem 14 follows in a similar fashion, but its proof is more involved as it requires the new notion of *overlap* for preference profiles. An upper bound for the share of profiles can then be established using 2-associated Stirling numbers of the second kind and variants of the Stirling approximation. □

The surprising finding that $\mathrm{DA}^{\mathbb{U}}$ may strictly rank dominate $\mathrm{ABM}^{\mathbb{U}}$ (Example 11) raised doubts about the idea that that $\mathrm{ABM}^{\mathbb{U}}$ is intuitively the more efficient mechanism. However, our limit results, Theorems 13 and 14, show that Example 11 is pathological. As markets grow, preference profiles at which $\mathrm{DA}^{\mathbb{U}}$ rank dominates $\mathrm{ABM}^{\mathbb{P}}$ even weakly become rare. Consequently, for larger markets, we can be confident that $\mathrm{ABM}^{\mathbb{U}}$ will

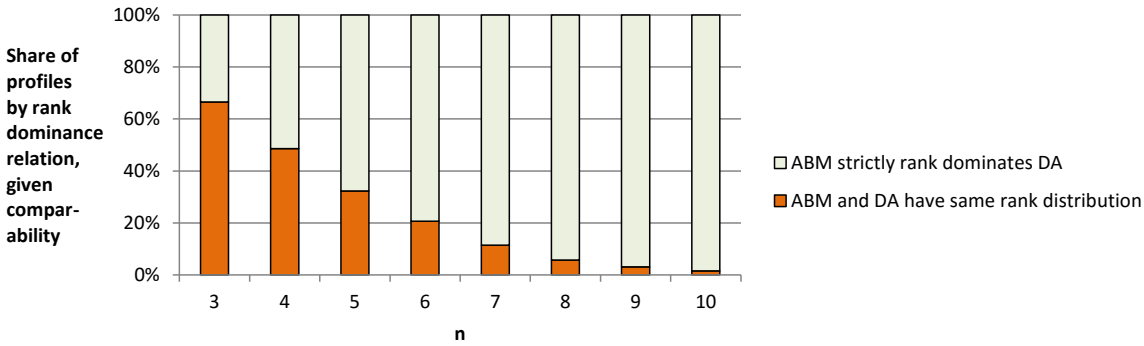

Figure 3.3: Share of preference profiles by rank dominance relation between $\text{ABM}^{\mathbb{U}}$ and $\text{DA}^{\mathbb{U}}$ (conditional on comparability), for settings with unit capacity and $n = m$ students.

produce the more efficient assignments whenever the results of both mechanisms are comparable. While our theoretical results yield this confidence for *sufficiently large markets*, the following simulation results provide reassurance that convergence occurs quickly. Thus, the theoretical possibility of rank dominance of $\text{DA}^{\mathbb{U}}$ over $\text{ABM}^{\mathbb{U}}$ is not a relevant concern in markets of *any size*.

Simulation Results: As in Section 3.6.3, we conducted simulations to complement our theoretical results: for settings with $n = m \in \{3, \ldots, 10\}$ and $q_j = 1$ for all $j \in M$, we sampled 100'000 preference profiles uniformly at random for each value of n. We then determined the rank dominance relation between the resulting assignments under $\text{ABM}^{\mathbb{U}}$ and $\text{DA}^{\mathbb{U}}$ at the sampled preference profiles whenever they were comparable. Figure 3.3 shows the results.

The insights are very similar to those for $\text{NBM}^{\mathbb{U}}$ and $\text{DA}^{\mathbb{U}}$ (Figure 3.2). First, we observe that the share of profiles where $\text{ABM}^{\mathbb{U}}$ does not yield a strictly preferable rank distribution decreases rapidly and is under 5% for 8 or more students (conditional on comparability). Second, despite the existence of Example 11, where $\text{DA}^{\mathbb{U}}$ strictly rank dominates $\text{ABM}^{\mathbb{U}}$, this situation did not occur even once in the entire sample. This suggests rapid convergence in both limit results (Theorems 13 and 14), and there is no need to worry about exceptions in practice.

To summarize, we have found that concerns about cases where $\text{DA}^{\mathbb{U}}$ dominates $\text{ABM}^{\mathbb{U}}$ can be dismissed as pathological; $\text{ABM}^{\mathbb{U}}$ is *essentially* the more efficient mechanism.

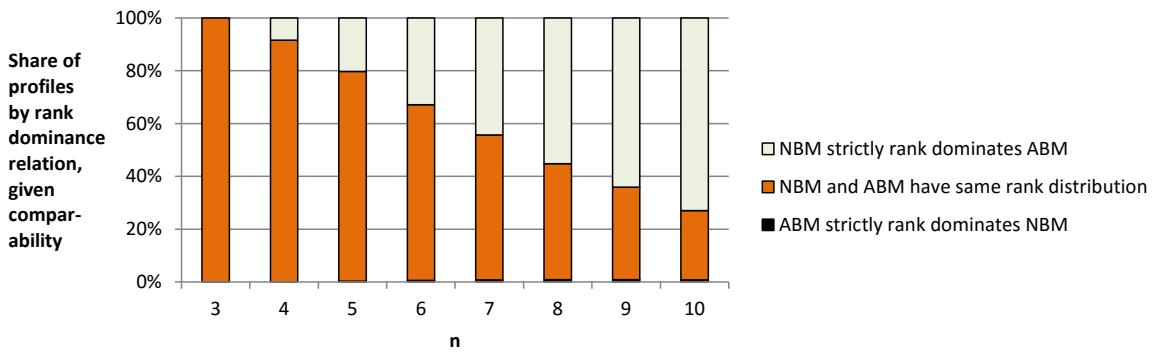

Figure 3.4: Share of preference profiles by rank dominance relation between $\text{NBM}^{\mathbb{U}}$ and $\text{ABM}^{\mathbb{U}}$ (conditional on comparability), for settings with unit capacity and $n = m$ students.

Comparing ABM and NBM: the "Cost of Partial Strategyproofness"

In Section 3.6.3, we have identified a "cost of strategyproofness:" DA is comparably rank dominated by NBM, a price one must pay when full strategyproofness is achieved using DA. The motivating example suggests a similar relationship between ABM and NBM. This intuition is further supported by the observation that NBM considers only kth choices in the kth round, while ABM may also consider less preferred choices. However, surprisingly, Example 12 has shown that $\text{ABM}^{\mathbb{U}}$ may strictly rank dominate $\text{NBM}^{\mathbb{U}}$. Consequently, the "cost of partial strategyproofness" between ABM and NBM cannot be identified analogously in terms of comparable rank dominance.

Restricting attention to the single uniform priority distribution $\mathbb{U}$, we now present evidence from simulations which shows that *$\text{NBM}^{\mathbb{U}}$ is usually the dominant mechanism*: conditional on comparability, dominance of $\text{NBM}^{\mathbb{U}}$ over $\text{ABM}^{\mathbb{U}}$ occurs much more frequently than the opposite case. Therefore, when choosing $\text{ABM}^{\mathbb{U}}$ over $\text{NBM}^{\mathbb{U}}$, the fact that $\text{NBM}^{\mathbb{U}}$ rank dominates $\text{ABM}^{\mathbb{U}}$ much more frequently can be considered a price that we pay for *partial strategyproofness*.

The setup for our simulation is analogous to those in Sections 3.6.3 and 3.6.4: for each $n = m \in \{3, \ldots, 10\}$ we sampled 100'000 preference profiles uniformly at random in settings with unit capacities ($q_j = 1$ for all $j \in M$). Conditional on comparability, we determined which of the mechanisms had a better rank distribution at each profile. The results are given in Figure 3.4. First, we observe that share of preference profiles where $\text{NBM}^{\mathbb{U}}$ has a strict advantage over $\text{ABM}^{\mathbb{U}}$ grows for growing numbers of students (light gray bars); this value increases continuously and reaches $\sim 75\%$ for $n = 10$. Second,

the share of profiles where $\text{ABM}^{\mathbb{U}}$ strictly rank dominates $\text{NBM}^{\mathbb{U}}$ is small (i.e., always below 1%). However, recall that for DA and ABM the same simulation did not yield a single instance of a preference profile where $\text{DA}^{\mathbb{U}}$ strictly rank dominates $\text{ABM}^{\mathbb{U}}$. Thus, instances where $\text{ABM}^{\mathbb{U}}$ strictly rank dominates $\text{NBM}^{\mathbb{U}}$ are rare but not as obviously negligible as was the case for the comparison with $\text{DA}^{\mathbb{U}}$. Thus, the efficiency advantage of $\text{NBM}^{\mathbb{U}}$ over $\text{ABM}^{\mathbb{U}}$ is less pronounced. Nonetheless, from the perspective of the market designer, the efficiency differences identified by our simulation constitute a "cost of partial strategyproofness" when choosing $\text{NBM}^{\mathbb{U}}$ over $\text{ABM}^{\mathbb{U}}$.

3.6.5 Extensions of the Efficiency Comparison

So far, we have given insights about the efficiency comparison of DA, ABM, and NBM whenever the three mechanisms are comparable by rank dominance. The results justify our intuition that NBM is the most efficient mechanism, while ABM has intermediate efficiency below NBM but above DA. However, they leave open what happens when a comparison by rank dominance is not possible. In future research, it will be interesting to also compare the mechanisms by a different measure for student welfare, namely the *expected share of students who get one of their top-k choices*. We conjecture that for a fixed k and sufficiently large settings (with sufficiently many schools with sufficiently high capacities), this value is highest under NBM, lowest under DA, and intermediate under ABM.

3.7 Conclusion

In this paper, we have studied the Deferred Acceptance mechanism, as well as the traditional (naïve) and a new (adaptive) variant of the Boston mechanism for school choice. These mechanisms are of particular interest as all three of them are frequently used in practice. In particular, ABM has been largely ignored in the literature. We have established that these mechanisms form two hierarchies, one with respect to strategyproofness, and one with respect to efficiency.

First, DA is strategyproof for students, while NBM is not even weakly strategyproof. We have proven that ABM satisfies the *intermediate* requirement of partial strategyproofness (subject to the technical condition that the priority distribution *supports all single priority profiles*). This result is contrasted by the fact that a comparison by *vulnerability to manipulation* does not differentiate between these mechanisms in terms of their incentive

properties.

Second, we have shown that NBM offers unambiguous efficiency gains over DA: the resulting assignments rank dominate those from DA whenever they are comparable. This advantage persists independent of the priority distribution. Via simulations (for the single uniform priority distribution) we found that when the dominance relation holds, it is almost certainly strict, e.g., in 95% of the comparable cases for $n = m \geqslant 8$ or more schools and students. In the light of the manipulability of NBM, this efficiency difference can be interpreted as the "cost of strategyproofness" that a market designer incurs when choosing DA over NBM.

Third, we have found ABM to have *intermediate* efficiency between NBM and DA. This result required the use of limit arguments and simulation because (surprisingly) a comparison by the rather flexible comparable rank dominance concept with either DA or NBM was inconclusive.

Throughout the paper, it has become apparent that traditional methods frequently fail to differentiate between the three popular school choice mechanisms: while a comparison by vulnerability to manipulation is inconclusive for NBM and ABM, partial strategyproofness provides a clear formal argument for the intuition that "ABM has better incentive properties than NBM." Similarly, all mechanisms we considered are ex-post efficient (for any single priority distributions) but neither ordinally nor rank efficient, and neither of them is on the efficient frontier subject to the respective incentive properties. Nonetheless, a comparison using comparable rank dominance, limit arguments, and simulation revealed an efficiency hierarchy: NBM has the most appeal in terms of rank dominance, but ABM is still more appealing than DA. The failure of traditional methods to differentiate between the three mechanisms illustrates that the toolbox of concepts is not yet complete; by considering *comparable dominance*, *limit results*, and *partial strategyproofness*, we have applied new instruments to help market designers tackle the analysis of three common mechanisms and facilitate decisions in practice.

The general lesson to be learned from our results is that a decision between DA, NBM, and ABM requires a non-trivial trade-off between strategyproofness and efficiency: when partial (as opposed to full) strategyproofness is acceptable, ABM can be used to obtain a preferable rank distribution; and if manipulability is not a concern, then NBM offers the most appealing efficiency advantages. Rather than indicating preference for any one of these mechanisms, our insights endow market designers with the means to make a conscious decision about this trade-off.

Appendix for Chapter 3

3.A Comparison by Vulnerability to Manipulation

In this section we compare NBM and ABM by their vulnerability to manipulation. We show that this comparison is inconclusive, except in the simplest case. Our findings do not diminish the value of the vulnerability to manipulation concept, but they highlight the fact that there is no "one size fits all" solution and that a wider array of concepts is required to perform a meaningful comparison in this case: for NBM and ABM, the partial strategyproofness concept is able to clearly differentiate between the two mechanisms.

3.A.1 Formalization of Vulnerability to Manipulation

We first review notions for the *comparison of mechanisms by their vulnerability to manipulation*, introduced by Pathak and Sönmez (2013). For a deterministic mechanisms φ and ψ, these are as follows.

Definition 24 (Pathak and Sönmez, 2013)**.** A deterministic mechanism φ is *manipulable by i at $\boldsymbol{P} \in \mathcal{P}^N$* if there exists a report $P_i' \in \mathcal{P}$ such that i strictly prefers her assigned school under $\varphi_i(P_i', P_{-i})$ to her school under $\varphi_i(P_i, P_{-i})$. φ is *manipulable at $\boldsymbol{P} \in \mathcal{P}^N$* if it is manipulable by some student i at $\boldsymbol{P}$.

Definition 25 (Pathak and Sönmez, 2013)**.** A deterministic mechanism φ is *as manipulable as ψ* if for any preference profile $\boldsymbol{P}$, φ is manipulable at $\boldsymbol{P}$ whenever ψ is manipulable at $\boldsymbol{P}$. φ is *strongly as manipulable as ψ* if for any preference profile $\boldsymbol{P}$, φ is manipulable by i at $\boldsymbol{P}$ whenever ψ is manipulable by i at $\boldsymbol{P}$.

φ is *more manipulable than ψ* if φ is as manipulable as ψ and there exists a preference profile $\boldsymbol{P}$ such that φ is manipulable at $\boldsymbol{P}$ but ψ is not.

Analogously, φ is *strongly more manipulable than ψ* if φ is strongly as manipulable as ψ and there exist a preference profile $\boldsymbol{P}$ and a student i such that φ is manipulable by i at $\boldsymbol{P}$ but ψ is not.

The original definitions consider deterministic mechanisms. For these, ordinal preference reports reflect all information that explains the students' strategic behavior. However, for probabilistic school choice mechanisms (again, denoted by φ and ψ), we need to extend the students' preferences to *lotteries*. This is achieved by assuming that students have *vNM utility functions* underlying their preferences. Each student i with preference order P_i has a utility function $u_i : M \to \mathbb{R}^+$ such that $u_i(j) > u_i(j')$ whenever $P_i : j \succ j'$ (denoted $u_i \in U_{P_i}$). Given reports P_{-i} from the other students, student i wants to report $P_i' \in \mathcal{P}^N$ such that her expected utility

$$\langle u_i, \varphi_i(P_i', P_{-i})\rangle = \sum_{j \in M} u_i(j) \cdot \varphi_{i,j}(P_i', P_{-i}) \tag{207}$$

is as high as possible. To compare probabilistic mechanisms, we extend the definitions in such a way that they coincide with the Definitions 24 and 25 from (Pathak and Sönmez, 2013) in the deterministic case.

Definition 26. A mechanism φ is *manipulable by i at $\boldsymbol{u}$* $= (u_i, u_{-i})$ (where $u_{i'} \in U_{P_{i'}}$ for all $i' \in N$) if there exists a report $P_i' \in \mathcal{P}$ such that

$$\langle u_i, \varphi_i(P_i', P_{-i}) - \varphi_i(P_i, P_{-i})\rangle > 0. \tag{208}$$

φ is *manipulable at* $\boldsymbol{u} = (u_1, \ldots, u_n)$ if it is manipulable by some student i at $\boldsymbol{u}$.

Definition 27. A mechanism φ is *as manipulable as* ψ if for any utility profile $\boldsymbol{u}$, φ is manipulable at $\boldsymbol{u}$ whenever ψ is manipulable at $\boldsymbol{u}$. φ is *strongly as manipulable as* ψ if for any utility profile $\boldsymbol{u}$, φ is manipulable by i at $\boldsymbol{u}$ whenever ψ is manipulable by i at $\boldsymbol{u}$.

φ is *more manipulable than* ψ if φ is as manipulable as ψ and there exists a utility profile $\boldsymbol{u}$ such that φ is manipulable at $\boldsymbol{u}$ but ψ is not.

Analogously, φ is *strongly more manipulable than* ψ if φ is strongly as manipulable as ψ, and there exist a utility profile $\boldsymbol{u}$ and a student i such that φ is manipulable by i at $\boldsymbol{u}$ but ψ is not.

In the case of a fixed strict priority profile $\boldsymbol{\pi}$, Dur (2015) showed that NBM$^{\boldsymbol{\pi}}$ is as manipulable as ABM$^{\boldsymbol{\pi}}$. One could hope to obtain a better understanding of the incentive properties of the two mechanisms in two ways: first, we can ask whether one of the mechanisms is *as manipulable as* the other under random priority distributions. Second, we can ask whether a stronger relation holds, namely whether NBM$^{\boldsymbol{\pi}}$ is *strongly as manipulable as* ABM$^{\boldsymbol{\pi}}$. The following Sections show that the answer to both questions is

negative: comparability of NBM and ABM "ends" with the weaker notion for comparison and fixed priority profiles.

3.A.2 Failure of Weak Comparison for Random Priority Distributions

First, we investigate whether comparability can be recovered through randomization, e.g., when $\mathbb{U}$ is the uniform distribution over all single priority profiles. Recall that a probabilistic mechanism φ is *as manipulable as* φ' if it is manipulable at least at the same utility profiles as φ. Even though this is arguably the weakest way in which we can extend the least demanding concept from (Pathak and Sönmez, 2013) to probabilistic mechanisms, we find that $\text{NBM}^{\mathbb{U}}$ and $\text{ABM}^{\mathbb{U}}$ are in-comparable in this sense. Specifically,

- for some utility profile $\boldsymbol{u}$, $\text{NBM}^{\mathbb{U}}$ is manipulable but $\text{ABM}^{\mathbb{U}}$ is not (Example 6),
- for some utility profile $\boldsymbol{u}'$, $\text{ABM}^{\mathbb{U}}$ is manipulable but $\text{NBM}^{\mathbb{U}}$ is not (Example 7).

Example 6. Consider the setting with 4 students $N = \{1, \dots, 4\}$, 4 schools $M = \{a, b, c, d\}$, each with unit capacities, and the preference profile

$$\begin{aligned} P_1 &: a \succ b \succ c \succ d, \\ P_2, P_3 &: a \succ c \succ b \succ d, \\ P_4 &: b \succ \dots . \end{aligned}$$

Student 1's assignment is $(1/3, 0, 0, 2/3)$ for the schools a through d, respectively. If student 1 swaps b and c in her report, the assignment will be $(1/3, 0, 1/3, 1/3)$. The outcome from this manipulation stochastically dominates the outcome from truthful reporting. Thus, student 1 will want to misreport under $\text{NBM}^{\mathbb{U}}$ (independent of her underlying utility).

Under $\text{ABM}^{\mathbb{U}}$, the outcome for student 1 is $(1/3, 0, 1/3, 1/3)$, independent of whether or not she swaps b and c. Now suppose that all students have utility $u = (9, 3, 1, 0)$ for their first, second, third, and fourth choice, respectively. Then no student has an incentive to deviate from truthful reporting under $\text{ABM}^{\mathbb{U}}$. Therefore, at this utility profile, $\text{NBM}^{\mathbb{U}}$ is vulnerable to manipulation but $\text{ABM}^{\mathbb{U}}$ is not.

The last example showed that $\text{ABM}^{\mathbb{U}}$ is *not* as manipulable as $\text{NBM}^{\mathbb{U}}$. The next example shows that $\text{NBM}^{\mathbb{U}}$ is *not* as manipulable as $\text{ABM}^{\mathbb{U}}$ either. This is surprising in the light of the Dur's comparability result for any fixed priority profile.

	Gain from misreport P_1^*	
Misreport P_1^*	under NBM	under ABM
$a \succ e \succ d \succ c \succ f \succ b$	-2.1	+0.0
$a \succ c \succ e \succ d \succ f \succ b$	-0.4	**+1.1**
$a \succ c \succ d \succ e \succ f \succ b$	-0.3	**+1.1**
$a \succ d \succ e \succ c \succ f \succ b$	-9.5	-7.7
$a \succ d \succ c \succ e \succ f \succ b$	-8.1	-7.7

Table 3.2: Change in expected utility from misreports for student 1 in Example 7.

Example 7. Consider the setting with 6 students $N = \{1, \dots, 6\}$, 6 schools $M = \{a, \dots, f\}$, each with unit capacities, and the preference profile

$$\begin{aligned} P_1', P_2' &: a \succ e \succ c \succ d \succ f \succ b, \\ P_3', P_4' &: a \succ e \succ d \succ c \succ f \succ b, \\ P_5' &: b \succ c \succ \dots, \\ P_6' &: b \succ d \succ \dots. \end{aligned}$$

Suppose that all students have utility $u' = (120, 30, 19, 2, 1, 0)$ for their first through sixth choice, respectively.

Consider the incentives of the students under NBM$^{\mathbb{U}}$: keeping the reports of all other students constant, student 5 has no incentive to deviate, and the same holds for student 6. If student 1 does not rank a in first position, she looses all chances at a and may at best get her second choice e for sure, which is not an improvement under the particular utility chosen. Also, it is easy to check that changing the position of f or b will never be beneficial for student 1. Thus, any beneficial manipulation for student 1 will involve only changes in the order of the schools e, c, d. However, none of these misreports are beneficial, which can be seen in Table 3.2 (middle column). Due to symmetry, none of the other student 2, 3, 4 have an incentive to misreport either.

Under ABM$^{\mathbb{U}}$, however, student 1 does have an incentive to misreport, which can also be seen in Table 3.2 (bold values in right column).

3.A.3 Failure of Strong Comparison for Fixed Priority Profiles

Now we address the second question regarding the strong comparison. We show that NBM$^{\boldsymbol{\pi}}$ is not *strongly as manipulable as* ABM$^{\boldsymbol{\pi}}$, i.e., there exist preference profiles $\boldsymbol{P}, \boldsymbol{P}'$,

priority profiles $\boldsymbol{\pi}, \boldsymbol{\pi}'$, and students i, i' such that

- $\text{NBM}^{\boldsymbol{\pi}}$ is manipulable by i at $\boldsymbol{P}$ but $\text{ABM}^{\boldsymbol{\pi}}$ is not (Example 8),
- $\text{ABM}^{\boldsymbol{\pi}'}$ is manipulable by i' at $\boldsymbol{P}'$ by i' but $\text{NBM}^{\boldsymbol{\pi}'}$ is not (Example 9).

Example 8. Consider a setting with students $N = \{1, \ldots, 4\}$, schools $M = \{a, \ldots, d\}$, each with unit capacity, the preference profile

$$\begin{aligned} P_1 &: a \succ \ldots, \\ P_2 &: b \succ \ldots, \\ P_3 &: a \succ b \succ c \succ d, \\ P_4 &: a \succ c \succ \ldots, \end{aligned}$$

and the single priority profile $1 \; \pi \ldots \pi \; 4$. Then student 3 will get d under $\text{NBM}^{\boldsymbol{\pi}}$, but if student 3 reports

$$P_3' \;:\; a \succ c \succ \ldots,$$

she will get c instead, a strict improvement. Under $\text{ABM}^{\boldsymbol{\pi}}$ and truthful reporting, b is exhausted by student 2 in the first round, and therefore, both students 3 and 4 apply for c in the second round, where 3 gets c. It is clear that due to her low priority, student 3 can not get a better school than c under $\text{ABM}^{\boldsymbol{\pi}}$ with any misreport. Thus, truthful reporting is a best response for student 3 under $\text{ABM}^{\boldsymbol{\pi}}$ but not under $\text{NBM}^{\boldsymbol{\pi}}$ (when all other students report truthfully).

Example 9. Consider a setting with 5 students $N = \{1, \ldots, 5\}$, 5 schools $M = \{a, \ldots, e\}$, each with unit capacities, the preference profile

$$\begin{aligned} P_1' &: a \succ \ldots, \\ P_2' &: b \succ \ldots, \\ P_3' &: d \succ \ldots, \\ P_4' &: a \succ b \succ d \succ c \succ e, \\ P_5' &: a \succ b \succ c \succ d \succ e, \end{aligned}$$

and the single priority profile $1 \; \pi' \ldots \pi' \; 5$. Under $\text{NBM}^{\boldsymbol{\pi}'}$ and truthful reporting, student 5 will get c, and there is no false report that will provide a better school, since a and b are exhausted in the first round (if all other students report truthfully). However, under

$\text{ABM}^{\boldsymbol{\pi}'}$, student 5 will get e. By ranking c first instead, she can get c, which is better than e.

Note that in Examples 8 and 9, we considered a *single* priority profile, which is a special case of a general multiple priority profile. Therefore, the counter-examples also show the in-comparability of $\text{NBM}^{\boldsymbol{\pi}}$ and $\text{ABM}^{\boldsymbol{\pi}}$ by the *strongly as manipulable as*-relation in general.

Remark 12. A natural next step to further understand the vulnerability of both mechanisms to manipulation is a quantitative analysis. This analysis should ask *how often* a mechanism is manipulable, i.e., given a prior over the students' utility profiles, how likely is each mechanism manipulable. In (Mennle et al., 2015) we have studied $\text{NBM}^{\mathbb{U}}$, $\text{ABM}^{\mathbb{U}}$, and the Probabilistic Serial mechanism in this way. We have found that under $\text{ABM}^{\mathbb{U}}$ truth-telling is a best response for all students significantly more often than under $\text{NBM}^{\mathbb{U}}$. This result is robust to changes in the size of the setting, the correlation of the preferences, and the underlying distributions in the utility model.

3.B Examples and Proofs

3.B.1 Incentives under the Naïve Boston Mechanism

Example 10. Consider a setting with 4 students $N = \{1, \dots, 4\}$, 4 schools $M = \{a, \dots, d\}$, each with unit capacity, i.e., $q_j = 1$ for all $j \in M$, and preference profile

$$\begin{aligned} P_1, P_2 &: \; a \succ b \succ c \succ d, \\ P_3, P_4 &: \; b \succ c \succ a \succ d. \end{aligned}$$

Student 2's assignment vector is $\text{NBM}_2^{\mathbb{U}}(\boldsymbol{P}) = \left(\frac{1}{2}, 0, 0, \frac{1}{2}\right)$ for a, b, c, d, respectively. However, if she reports

$$P_2' \;:\; a \succ c \succ b \succ d$$

instead, her assignment vector will be $\text{NBM}_2^{\mathbb{U}}(P_2', P_{-2}) = \left(\frac{1}{2}, 0, \frac{1}{4}, \frac{1}{4}\right)$, which strictly stochastically dominates $\text{NBM}_2^{\mathbb{U}}(\boldsymbol{P})$. Thus, $\text{NBM}^{\mathbb{U}}$ is not even weakly strategyproof. When we consider the single priority profile $\boldsymbol{\pi} = (\pi, \dots, \pi)$ with $1 \; \pi \dots \pi \; 4$, we obtain in the same way that $\text{NBM}^{\boldsymbol{\pi}}$ is manipulable and neither swap monotonic nor lower invariant.

Proposition 11. *For any priority distribution $\mathbb{P}$, NBM$^{\mathbb{P}}$ is upper invariant, but it may violate swap monotonicity and lower invariance.*

Proof. Example 10 shows that NBM is neither swap monotonic, nor lower invariant: swapping b and c changed student 2's assignment vector, but her probability for b remained unchanged, which violates swap monotonicity. By swapping b and c, student 1 managed to change her probability for d, which violates lower invariance.

To see upper invariance, first consider a fixed priority profile $\boldsymbol{\pi} \in \Pi^M$ and let a and b be two adjacent schools in student i's preference order. Before applying to a or b, i applies to all the schools that she strictly prefers to a and b in previous rounds. The order of applications is the same, independent of the order in which a and b appear in her preference report. Since her chances of being accepted at a better school only depend on these previous rounds, these chances do not change if she swaps a and b. Thus, NBM$^{\boldsymbol{\pi}}$ is upper invariant for any fixed priority profile.

Recall that NBM$^{\mathbb{P}}$ randomly determines a priority profile $\boldsymbol{\pi}$ according to $\mathbb{P}$. Since NBM$^{\boldsymbol{\pi}}$ is upper invariant for any fixed $\boldsymbol{\pi}$ that may be drawn from $\mathbb{P}$, the probabilities for the preferred schools are unaffected by the swap of a and b. Thus, NBM$^{\mathbb{P}}$ is upper invariant for *any* priority distribution $\mathbb{P}$. □

3.B.2 Incentives under the Adaptive Boston Mechanism

Theorem 9 is a direct consequence of the following Lemma 2.

Lemma 2. *ABM$^{\mathbb{U}}$ is upper invariant and swap monotonic but not lower invariant.*

Proof. We prove the stronger statement that ABM$^{\mathbb{P}}$ is upper invariant and swap monotonic for any priority distribution $\mathbb{P}$ that supports all single priority profiles. The proof of upper invariance (for arbitrary priority profiles) is exactly the same as for NBM$^{\mathbb{P}}$ (Proposition 11). Next, we show that for arbitrary priority profiles $\boldsymbol{\pi}$, ABM$^{\boldsymbol{\pi}}$ is *monotonic*: if a student swaps the order of two adjacent schools a and b, then her probability for acceptance at b cannot strictly decrease, and her probability for acceptance at a cannot strictly increase.

Claim 9. *For any priority distribution $\mathbb{P}$, ABM$^{\mathbb{P}}$ and NBM$^{\mathbb{P}}$ are monotonic.*

Proof of Claim 9. Recall that for some fixed $\boldsymbol{\pi}$, ABM$^{\boldsymbol{\pi}}$ is deterministic. If the student ranks a below b, she will apply to a in a later round. Thus, if she did not receive a under the original report, she will not receive a under the new report either. Similarly,

she cannot decrease her chances at b by applying to b in an earlier round. For the probabilistic mechanism $\text{ABM}^{\mathbb{P}}$, this means monotonicity.

For $\text{NBM}^{\mathbb{P}}$, the arguments are exactly the same. □

Now suppose that $\mathbb{P}$ supports all single priority profiles (i.e., $\mathbb{P}[(\pi, \ldots, \pi)] > 0$ for all priority orders $\pi \in \Pi$). We now have to show that if the swap of a and b changes the student's probability vector, then the change is in fact strict for a and b. We show this by constructing a single priority profile $\boldsymbol{\pi}^* = (\pi^*, \ldots, \pi^*)$, such that under $\text{ABM}^{\boldsymbol{\pi}^*}$ she receives either school a or school b, whichever she ranked first. Since $\boldsymbol{\pi}^*$ is selected with positive probability, this implies swap monotonicity.

Claim 10. *Let P_i be student i's preference order, where schools $a, b \in M$ are ranked consecutively and $P_i : a \succ b$, let P_i' be the same preference order, except that $P_i' : b \succ a$, and let P_{-i} be any collection of preference orders from the other students.*

If there exists a priority profile $\boldsymbol{\pi}$ such that

$$ABM_i^{\boldsymbol{\pi}}(P_i, P_{-i}) \neq ABM_i^{\boldsymbol{\pi}}(P_i', P_{-i}), \tag{209}$$

then there exists a single priority profile $\boldsymbol{\pi}^* = (\pi^*, \ldots, \pi^*)$ such that

$$ABM_{i,a}^{\boldsymbol{\pi}^*}(P_i, P_{-i}) = ABM_{i,b}^{\boldsymbol{\pi}^*}(P_i', P_{-i}) = 1, \tag{210}$$

i.e., i receives a by reporting P_i truthfully and b by falsely reporting P_i'.

Proof of Claim 10. Let j denote the school that i receives when reporting P_i, and let j' be the school that she receives when reporting P_i', i.e, $\text{ABM}_{i,j}^{\boldsymbol{\pi}}(P_i, P_{-i}) = 1$ and $\text{ABM}_{i,j'}^{\boldsymbol{\pi}}(P_i', P_{-i}) = 1$. Consider the following cases:

- If $j = j'$, this violates the assumption that the assignment changes.
- If $P_i : j \succ a$, then $j = j'$ by upper invariance, a contradiction.
- If $j = b$, then $j' = b$ by monotonicity from Claim 9.
- If $j = c$ for some $c \in M$ with $P_i : b \succ c$, then $j' = c$. To see this, consider the order in which i applies to the schools. If she applies to a first (in round k, say), she is rejected, which implies that a is exhausted by the end of round k. The same holds for b. Since the mechanism lets i skip exhausted schools, she will not apply to b or a, respectively, in round $k + 1$ after being rejected from a or b, respectively. Therefore, the order of her applications remains unchanged after she was rejected

in round k. This means that none of the other students are affected by this change of report, and therefore, the assignment does not change for i.

Thus, it must be the case that $j = a$ and $P_i : b \succeq j'$. Let N_k be the students who received their kth choice school under $\text{ABM}^{\boldsymbol{\pi}}$ and let π' be some (i.e., any) priority order such that $i' \; \pi' \; i''$ whenever $i' \in N_k$ and $i'' \in N_{k'}$ for some $k < k'$. Observe that under $\text{ABM}^{\boldsymbol{\pi}'}$ with the single priority profile $\boldsymbol{\pi}' = (\pi', \ldots, \pi')$ constructed from π', all students from N_k will receive their kth choice, i.e., $\text{ABM}^{\boldsymbol{\pi}'}(\boldsymbol{P}) = \text{ABM}^{\boldsymbol{\pi}}(\boldsymbol{P})$. If $j' = b$, set $\boldsymbol{\pi}^* = \boldsymbol{\pi}'$. Otherwise, let π^* be the priority order that arises by taking π' and inserting student i just before the last student whose application to b was accepted. Then, under $\text{ABM}^{\boldsymbol{\pi}^*}(\boldsymbol{P}')$, i will get b. □

This concludes the proof of Lemma 2. □

3.B.3 Incentives for Arbitrary Priority Distributions

Proof of Proposition 8

Proof of Proposition 8. For any priority distribution $\mathbb{P}$, the mechanisms $NBM^{\mathbb{P}}$ and $ABM^{\mathbb{P}}$ are upper invariant, monotonic, and sensitive.

Upper invariance follows from Lemma 2 and Remark 10. Monotonicity is straightforward.

The proofs of sensitivity are more challenging. We first show sensitivity of $\text{NBM}^{\boldsymbol{\pi}}$ for arbitrary priority profiles $\boldsymbol{\pi}$. Sensitivity of $\text{NBM}^{\mathbb{P}}$ then follows with monotonicity and upper invariance.

Suppose that for some preferences $P_i \in \mathcal{P}, P_i' \in N_{P_i}$ with $P_i : a_k \succ a_{k+1}$ and $P_i' : a_{k+1} \succ a_k$ and some preference reports $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$ we have that $\text{NBM}_i^{\boldsymbol{\pi}}(P_i, P_{-i}) \neq \text{NBM}_i^{\boldsymbol{\pi}}(P_i', P_{-i})$.

First, we show that either a_k or a_{k+1} must be affected by this change. Suppose towards contradiction that neither of them is affected. Since i was not accepted at either school under either preference order, no other student who applied to a_k or a_{k+1} in the same rounds as i received a different assignment under the two reports P_i and P_i'. In round $k + 2$, the situation is exactly the same under both reports. Thus, the subsequent application process remains unchanged, so that i receives the same assignment under both reports, a contradiction.

If i receives a_k under P_i and a_{k+1} under P_i', then there is nothing to show. If i receives a_{k+1} under P_i, then it must also receive a_{k+1} under P_i' by monotonicity, and there is no

change. Thus, we can assume that i receives a_k under P_i and $c \neq a_{k+1}$ under P_i'. If a_{k+1} is i's last choice under P_i, then sensitivity follows from monotonicity.

Next, suppose that a_k is the first choice of i. Consider preference reports P_{-i}^1 where each student (except i) ranks the school first that it received under $\text{NBM}^{\boldsymbol{\pi}}(P_i, P_{-i})$. We define P_{-i}^2 by changing the first choice of all students who rank a_{k+1} first under P_{-i}^1. Instead, let them rank c first. If the other students' reports are P_{-i}^2, then i would obtain a_k by reporting P_i and it would obtain a_{k+1} by reporting P_i'.

Finally, assume that a_k is not the first choice of i, and a_{k+1} is not its last choice. Observe that under the original preference profile (P_i, P_{-i}), i obtains a_k in the kth round, whereas under the preference profile (P_i', P_{-i}), i obtains c in some later round. Since a_{k+1} was exhausted before or in round k under (P_i', P_{-i}), there exists a student i' who received a_{k+1}. Thus, i' must rank a_{k+1} first under P_{-i}^1. There are 2 cases:

Case 1: Suppose that c was not exhausted under (P_i, P_{-i}). Then we change the first choice of i' to c to obtain P_{-i}^3. Under the preference profile (P_i, P_{-i}^3), i will be rejected from all schools $a_1, \ldots, a_{k-1}$. In the kth round, i will receive either a_k or a_{k+1}, depending on whether i reports P_i or P_i'.

Case 2: Suppose that c was exhausted under (P_i, P_{-i}). Then there exists at least one student i'' who ranked c first under P_{-i}^1. We change the first choices of i' and i'' to a_1 to obtain P_{-i}^4. Under (P_i, P_{-i}^4), i and two more students (i''' and i'''', say) are rejected from a_1 in the first round. We change the preference orders of i''' and i'''' to

$$P_{i'''}, P_{i''''} : a_1 \succ \ldots \succ a_{k-1} \succ c \succ a_k \succ a_{k+1} \succ \ldots \tag{211}$$

to obtain P_{-i}^5. Under the preference profile (P_i, P_{-i}^5), the students i, i''', i'''' will all be rejected from $a_1, \ldots, a_{k-1}$. In the kth round, i will get either a_k or a_{k+1}, depending on whether it reports P_i or P_i', and i''' and i'''' will compete for c.

In both cases, we have shown that we can construct preference reports (P_{-i}^3 and P_{-i}^5, respectively), such that i can ensure an assignment of either a_k or a_{k+1}, whichever it claims to prefer.

The argument for the case where i receives a_{k+1} under P_i' and $c \neq a_k$ under P_i follows by symmetry. Note that the construction is independent of the particular (strict) priorities underlying the deterministic naïve Boston mechanism.

Sensitivity of $\text{ABM}^{\boldsymbol{\pi}}$ for arbitrary priority profiles $\boldsymbol{\pi}$ follows by the same construction. The only difference is that instead of applying to (and being rejected from) $a_2, \ldots, a_{k+1}$, the students i, i''', i'''' in case 2 simply skip all these schools. □

3.B.4 Efficiency of the Naïve Boston Mechanism

Proof of Proposition 9

Proof of Proposition 9.

1. *For any priority distribution $\mathbb{P}$, NBM$^{\mathbb{P}}$ is ex-post efficient but may not be ordinally or rank efficient.*

2. *Among all upper invariant mechanisms, NBM$^{\mathbb{U}}$ is not on the efficient frontier with respect to ordinal (or rank) dominance. Equivalently, there exists an upper invariant mechanism that strictly ordinally (and rank) dominates NBM$^{\mathbb{U}}$.*

Ex-post efficiency: We first prove that $\text{NBM}^{\boldsymbol{\pi}}(\boldsymbol{P})$ is ex-post efficient at $\boldsymbol{P}$ for any fixed priority profile $\boldsymbol{\pi}$. Let $N_r \subseteq N, r = 1, \ldots, m$ be the set of students who receive their rth choice school under $x = \text{NBM}^{\boldsymbol{\pi}}(\boldsymbol{\pi})$. Let $\sigma \in \Pi$ be some (any) priority order over the students such that all students in N_1 have priority over students in N_2, who in turn have priority over N_3, etc., i.e., for any $r = 1, \ldots, m-1$ and any $i_r \in N_r, i_{r+1} \in N_{r+1}$ we have that $i_r \ \sigma \ i_{r+1}$. Now, consider the assignment $y = \text{SD}^{\boldsymbol{\sigma}}(\boldsymbol{P})$. This is equivalent to an application of the serial dictatorship mechanism to the preference profile $\boldsymbol{P}$ with picking order σ, and therefore, y is ex-post efficient at $\boldsymbol{P}$.

Under the serial dictatorship mechanism, all the students in N_1 receive their first choices because (1) they are allowed to pick first and (2) there is sufficient capacity at there first choices (otherwise, they could not all have received these schools under the Boston mechanism). When the first student from N_2 gets to pick a school, the situation is exactly the same as at the beginning of the second round of the Boston mechanism: the first choice school of the remaining students have no more seats available. By the same argument as for first choices, all students from N_2 must get their second choice school. Inductively, we get that all students in N_r get their rth choice under the serial dictatorship mechanism with picking order σ, and consequently the assignments x and y are the same. Hence, x is ex-post efficient.

Since $\text{NBM}^{\mathbb{P}}(\boldsymbol{P})$ is simply a lottery over deterministic assignments $\text{NBM}^{\boldsymbol{\pi}}(\boldsymbol{P})$, each of which is ex-post efficient by the previous arguments. Therefore, $\text{NBM}^{\mathbb{P}}$ is ex-post efficient at any preference profile and for any priority distribution.

Ordinal and Rank Inefficiency: Consider the setting with 4 students $N =$

$\{1, \ldots, 4\}$, 4 schools $M = \{a, \ldots, d\}$, each with unit capacities, and the preference profile

$$\begin{aligned} P_1 &: a \succ b \succ c \succ d, \\ P_2 &: a \succ b \succ d \succ c, \\ P_3 &: b \succ a \succ c \succ d, \\ P_4 &: b \succ a \succ d \succ c. \end{aligned}$$

The assignments is

$$\mathrm{NBM}^{\mathbb{U}}(\boldsymbol{P}) = \begin{pmatrix} \frac{1}{2} & 0 & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{2} & 0 & \frac{1}{8} & \frac{3}{8} \\ 0 & \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ 0 & \frac{1}{2} & \frac{1}{8} & \frac{3}{8} \end{pmatrix}, \tag{212}$$

which is ordinally dominated at $\boldsymbol{P}$ by the assignment

$$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \tag{213}$$

Since ordinal dominance implies rank dominance, failure of ordinal efficiency implies failure of rank efficiency.

Failure to be on the Efficient Frontier: We construct a mechanism that is upper invariant and ordinally dominates $\mathrm{NBM}^{\mathbb{U}}$. This mechanism, NBM^{+}, is essentially the same mechanism as $\mathrm{NBM}^{\mathbb{U}}$, except that the assignment is altered at certain preference profiles. Again, consider the setting with 4 students and 4 schools in unit capacity. We say that a preference profile satisfies *separable wants* if the schools and students can be renamed such that

- students 1 and 2 have first choice a,
- students 3 and 4 have first choice b,
- students 1 and 3 prefer c to d,
- and students 2 and 4 prefer d to c.

Formally,

$$\begin{aligned} P_1 &: a \succ \{b, c, d\} \text{ and } c \succ d, \\ P_2 &: a \succ \{b, c, d\} \text{ and } d \succ c, \\ P_3 &: b \succ \{a, c, d\} \text{ and } c \succ d, \\ P_4 &: b \succ \{a, c, d\} \text{ and } d \succ c. \end{aligned}$$

NBM^+ is the same as NBM, except that the outcome is adjusted for preference profiles with separable wants. Let

$$\text{NBM}^+(\boldsymbol{P}) = \begin{cases} \text{PS}(\boldsymbol{P}), & \text{if } \boldsymbol{P} \text{ satisfies separable wants,} \\ \text{NBM}^{\mathbb{U}}(\boldsymbol{P}), & \text{else,} \end{cases} \tag{214}$$

where PS denotes the Probabilistic Serial mechanism (Bogomolnaia and Moulin, 2001). At some preference profile $\boldsymbol{P}$ that satisfies separable wants, the assignment under PS (after appropriately renaming of the students and schools) is

$$\text{PS}(\boldsymbol{P}) = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \tag{215}$$

Observe that under $\text{NBM}^{\mathbb{U}}$, a is split equally between 1 and 2, and b is split equally between 3 and 4. Consequently, students 1 and 2 get no share of b and students 3 and 4 get no share of a, just as under PS. Among all assignments that distribute a and b in this way, student 1 prefers the ones that give her higher probability at c. This is at most $\frac{1}{2}$, since she already receives a with probability $\frac{1}{2}$. Similarly, students $2, 3$, and 4 prefer their respective assignment under PS to any other assignment that splits a and b in the same way as PS and $\text{NBM}^{\mathbb{U}}$. Therefore, NBM^+ weakly ordinally dominates $\text{NBM}^{\mathbb{U}}$, and the dominance is strict for preference profiles $\boldsymbol{P}$ with separable wants.

It remains to be shown that NBM^+ is upper invariant. To verify this, we only need to consider the change in assignment that the mechanism prescribes if some student swaps two adjacent schools in its reported preference ordering. Starting with any preference profile $\boldsymbol{P}$, the swap produces a new preference profile $\boldsymbol{P}'$. If neither $\boldsymbol{P}$ nor $\boldsymbol{P}'$ satisfy separable wants, the mechanism behave like $\text{NBM}^{\mathbb{U}}$.

For swaps where at least one of the preference profiles satisfies separable wants, we

can assume without loss of generality that this is $\boldsymbol{P}$. Such a swap will lead to a new preference profile $\boldsymbol{P}'$ and one of the following three cases:

(i) The new preference profile satisfies separable wants.

(ii) The composition of the first choices has changed.

(iii) The preference profile no longer satisfies separable wants, but the composition of the first choices has not changed.

By symmetry, we can restrict our attention to student 1 whose preference order satisfies

$$P_1 : a \succ c \succ d \text{ and } a \succ b. \tag{216}$$

Case (i) implies that $\boldsymbol{P}'$ still satisfies separable wants with respect to the same mappings μ, ν. Thus, NBM^+ will not change the assignment, i.e., upper invariance is not violated.

In case (ii), student 1 has a new first choice. If the new first choice is b, she will receive b with probability $\frac{1}{3}$ and a with probability 0 under $\text{NBM}^{\mathbb{U}}(\boldsymbol{P}')$. If the new first choice is c or d, the student will receive that that school with certainty under $\text{NBM}^{\mathbb{U}}(\boldsymbol{P}')$. Both changes are consistent with upper invariance.

Finally, in case (iii), the swap must involve c and d, since this is the only way in which separable wants can be violated. Since $\boldsymbol{P}'$ violates separable wants, we have $\text{NBM}^+(\boldsymbol{P}') = \text{NBM}^{\mathbb{U}}(\boldsymbol{P}')$, and therefore, a will still be split equally between the students who rank it first, and the same is true for b. Thus, student 1 will receive $\frac{1}{2}$ of a and 0 of b, which is the same as under $\text{NBM}^+(\boldsymbol{P}) = \text{PS}(\boldsymbol{P})$. The only change can affect the assignment for the schools c and d. This is consistent with upper invariance. □

Proof of Theorem 11

Proof of Theorem 11. $\text{NBM}^{\mathbb{U}}$ strictly comparably rank dominates $\text{DA}^{\mathbb{U}}$:

1. *$\text{NBM}^{\mathbb{U}}(\boldsymbol{P})$ rank dominates $\text{DA}^{\mathbb{U}}(\boldsymbol{P})$ at $\boldsymbol{P}$ for any $\boldsymbol{P} \in \mathcal{P}^N$ where $\text{NBM}^{\mathbb{U}}(\boldsymbol{P})$ and $\text{DA}^{\mathbb{U}}(\boldsymbol{P})$ are comparable by rank dominance,*
2. *there exists a preference profile $\boldsymbol{P} \in \mathcal{P}^N$ such that $\text{NBM}^{\mathbb{U}}(\boldsymbol{P})$ strictly rank dominates $\text{DA}^{\mathbb{U}}(\boldsymbol{P})$ at $\boldsymbol{P}$.*

We first establish the following lemmas about rank dominance. Let $x, x^1, \ldots, x^K \in \Delta(X)$ be assignments such that

$$x = \sum_{k=1}^{K} x^k \cdot \alpha_k \tag{217}$$

for some $\alpha_1, \ldots, \alpha_K > 0$ with $\sum_{k=1}^{K} \alpha_k = 1$, i.e., x is the convex combination of the assignments $x^k, k = 1, \ldots, K$ with coefficients $\alpha_k, k = 1, \ldots, K$.

Lemma 3. *The rank distribution d^x of x (at some preference profile $\boldsymbol{P}$) is equal to the convex combination of the rank distributions d^{x^k} of the x^k with respect to coefficients $\alpha_1, \ldots, \alpha_K$, i.e.,*

$$d^x = \sum_{k=1}^{K} d^{x^k} \cdot \alpha_k. \tag{218}$$

Lemma 3 is obvious from the definition of the rank distribution in Definition 21.

Lemma 4. *Let $y, y^1, \ldots, y^K \in \Delta(X)$ be assignments such that*

$$y = \sum_{k=1}^{K} y^k \cdot \alpha_k \tag{219}$$

for the same coefficients $\alpha_1, \ldots, \alpha_K$, and let there be $rank_k, k = 1, \ldots, K$ such that for all $k = 1, \ldots, K$ and all $rank' < rank_k$ we have

$$d^{x^k}_{rank'} = d^{y^k}_{rank'}. \tag{220}$$

Furthermore, if $rank_k \leqslant m$, then

$$d^{x^k}_{rank_k} > d^{y^k}_{rank_k} \tag{221}$$

(otherwise, $d^{x^k} = d^{y^k}$).

Then y does not even weakly rank dominate x.

Proof of Lemma 4: Let $\text{rank}_{\min} = \{\text{rank}_k | k = 1, \ldots, K\}$ be the lowest rank (i.e., the best choice) at which inequality (221) holds strictly, and let $k_{\min} \in \{1, \ldots, K+1\}$ be an index for which this is the case. Then for all $\text{rank}' < \text{rank}_{\min}$ and all $k = 1, \ldots, K$ we have

$$d^{x^k}_{\text{rank}'} = d^{y^k}_{\text{rank}'}, \tag{222}$$

so that by Lemma 3

$$d^{x}_{\text{rank}'} = d^{y}_{\text{rank}'}. \tag{223}$$

In words, the rank distributions of x and y coincides for all ranks before $\text{rank}_{\min}$. Furthermore,

$$d^{x^k}_{\text{rank}_{\min}} \geqslant d^{y^k}_{\text{rank}_{\min}} \tag{224}$$

for all $k \neq k_{\min}$, and

$$d^{x^{k_{\min}}}_{\text{rank}_{\min}} > d^{y^{k_{\min}}}_{\text{rank}_{\min}}. \tag{225}$$

Thus, by Lemma 3 and the fact that $\alpha_{k_{\min}} > 0$,

$$\begin{aligned}
d^{x}_{\text{rank}_{\min}} &= \sum_{k=1}^{K} d^{x^k}_{\text{rank}_{\min}} \cdot \alpha_k && (226)\\
&= d^{x^{k_{\min}}}_{\text{rank}_{\min}} \cdot \alpha_{k_{\min}} + \sum_{k=1, k\neq k_{\min}}^{K} d^{x^k}_{\text{rank}_{\min}} \cdot \alpha_k && (227)\\
&> d^{y^{k_{\min}}}_{\text{rank}_{\min}} \cdot \alpha_{k_{\min}} + \sum_{k=1, k\neq k_{\min}}^{K} d^{x^k}_{\text{rank}_{\min}} \cdot \alpha_k && (228)\\
&\geqslant \sum_{k=1}^{K} d^{y^k}_{\text{rank}_{\min}} \cdot \alpha_k = d^{y}_{\text{rank}_{\min}}. && (229)
\end{aligned}$$

□

We now proceed to prove the Theorem in two steps.

Step 1 (for any fixed single priority profile): First, we show that for any fixed single priority profile $\boldsymbol{\pi} = (\pi, \ldots, \pi) \in \Pi^M$ and any preference profile $\boldsymbol{P}$, the assignment $y^{\boldsymbol{\pi}} = \text{DA}^{\boldsymbol{\pi}}(\boldsymbol{P})$ never strictly rank dominates the assignment $x^{\boldsymbol{\pi}} = \text{NBM}^{\boldsymbol{\pi}}(\boldsymbol{P})$ at $\boldsymbol{P}$. In fact, we show something stronger, namely that the conditions of Lemma 4 are satisfied for $x^{\boldsymbol{\pi}}$ and $y^{\boldsymbol{\pi}}$, i.e., either $d^{x^{\boldsymbol{\pi}}} = d^{y^{\boldsymbol{\pi}}}$, or there exists some $r \in \{1, \ldots, m\}$ such that $d^{x^{\boldsymbol{\pi}}}_r > d^{y^{\boldsymbol{\pi}}}_r$, and $d^{x^{\boldsymbol{\pi}}}_{r'} = d^{y^{\boldsymbol{\pi}}}_{r'}$ for all $r' < r$.

In the proof, we consider the slightly larger domain, where schools can have zero capacity. Under $\text{DA}^{\boldsymbol{\pi}}$, including additional empty schools does not make a difference for the resulting assignment. Under the $\text{NBM}^{\boldsymbol{\pi}}$, it is easy to see that the assignments can be decomposed into two parts:

1. Run the first round of the mechanism, in which a set of students N_1 receives their first choice schools.
2. Remove the students N_1 from N, and also remove these students from all priority orders in the priority profile $\boldsymbol{\pi}$. Reduce the capacities of the schools they received by the number of students who received each school. Change the preference orders of all *remaining* students by moving their first choice to the end of their ranking. Then, run the mechanism again on the reduced problem (which may include schools of capacity zero).

In the final assignment resulting from $\text{NBM}^{\boldsymbol{\pi}}$, the students N_1 will receive their first choices, and the other students will receive the schools they got in the reduced setting.

Claim 11. *$DA^{\boldsymbol{\pi}}(\boldsymbol{P})$ assigns a weakly lower number of first choices than $NBM^{\boldsymbol{\pi}}(\boldsymbol{P})$.*

The claim is obvious from the observation that $\text{NBM}^{\boldsymbol{\pi}}$ maximizes the number of assigned first choices.

Claim 12. *If $DA^{\boldsymbol{\pi}}(\boldsymbol{P})$ assigns the same number of first choices as $NBM^{\boldsymbol{\pi}}(\boldsymbol{P})$, then the sets of students who get their first choices under both mechanism coincide.*

Proof. By assumption, $d_1^{x^{\boldsymbol{\pi}}} = d_1^{y^{\boldsymbol{\pi}}}$. Suppose towards contradiction that there exists some student $i \in N$, who receives her first choice school $j \in M$ under $y^{\boldsymbol{\pi}}$ but not under $x^{\boldsymbol{\pi}}$. That means that j was exhausted in the first round by other students, all of whom must have had higher priority than i (according to $\boldsymbol{\pi}$). These students as well as i would also apply to j in the first round of $\text{DA}^{\boldsymbol{\pi}}$. But since j was already exhausted by the other students, i will also be rejected from j in the first round of $\text{DA}^{\boldsymbol{\pi}}$, a contradiction. □

Observer that under $\text{DA}^{\boldsymbol{\pi}}$ students can only obtain their first choice school in the first round. By Claim 12, if $d_1^{x^{\boldsymbol{\pi}}} = d_1^{y^{\boldsymbol{\pi}}}$, then $\text{NBM}^{\boldsymbol{\pi}}$ and $\text{DA}^{\boldsymbol{\pi}}$ assign the same students to their first choice schools, and therefore, none of the students who received their first choice school under $\text{DA}^{\boldsymbol{\pi}}$ (tentatively in the first round) was rejected in any subsequent round. Thus, we can also decompose the assignment from $\text{DA}^{\boldsymbol{\pi}}$ into two parts (as before for $\text{NBM}^{\boldsymbol{\pi}}$):

1. The assignment from the first round.
2. The assignments from applying the mechanism to the reduced and altered setting.

We can now apply Claim 12 inductively to the reduced settings to show that $d_r^{x^{\boldsymbol{\pi}}} = d_r^{y^{\boldsymbol{\pi}}}$ implies that the same students also got their rth choice under both mechanisms. Since $d_r^{x^{\boldsymbol{\pi}}} < d_r^{y^{\boldsymbol{\pi}}}$ is impossible by Claim 11, we get that either the assignments from both mechanism coincide entirely, or $d_r^{x^{\boldsymbol{\pi}}} > d_r^{y^{\boldsymbol{\pi}}}$ for some $r \in \{1, \ldots, m\}$, i.e., the Boston mechanism assigns strictly more rth choices than Deferred Acceptance.

Step 2 (for any single priority distributions): For any single priority distribution $\mathbb{P}$, a single priority profile $\boldsymbol{\pi}$ is drawn at random according to $\mathbb{P}$. By construction

$$x = \text{NBM}^{\mathbb{P}}(\boldsymbol{P}) = \sum_{\boldsymbol{\pi}} \text{NBM}^{\boldsymbol{\pi}}(\boldsymbol{P}) \cdot \mathbb{P}[\boldsymbol{\pi}], \tag{230}$$

and

$$y = \mathrm{DA}^{\mathbb{P}}(\boldsymbol{P}) = \sum_{\boldsymbol{\pi}} \mathrm{DA}^{\boldsymbol{\pi}}(\boldsymbol{P}) \cdot \mathbb{P}[\boldsymbol{\pi}], \tag{231}$$

i.e., both x and y can we written as convex combinations of assignments $x^{\boldsymbol{\pi}} = \mathrm{NBM}^{\boldsymbol{\pi}}(\boldsymbol{P})$ and $y^{\boldsymbol{\pi}} = \mathrm{DA}^{\boldsymbol{\pi}}(\boldsymbol{P})$, respectively, with the same coefficients $\alpha_{\boldsymbol{\pi}} = \mathbb{P}[\boldsymbol{\pi}]$. By Step 1, each pair $x^{\boldsymbol{\pi}}, y^{\boldsymbol{\pi}}$ has the property that $d_{r'}^{x^{\boldsymbol{\pi}}} = d_{r'}^{y^{\boldsymbol{\pi}}}$ for $r' < r \leqslant m$ and $d_r^{x^{\boldsymbol{\pi}}} > d_r^{y^{\boldsymbol{\pi}}}$ (or $d^{x^{\boldsymbol{\pi}}} = d^{y^{\boldsymbol{\pi}}}$). Thus, by Lemma 4, $\mathrm{DA}^{\mathbb{P}}(\boldsymbol{P})$ never strictly rank dominates $\mathrm{NBM}^{\mathbb{P}}(\boldsymbol{P})$.

Step 3 (for any priority distribution): Harless (2015) showed that Claim 12 also holds for multiple priority profiles. We can apply the same reasoning as in Step 2 to obtain comparable rank dominance of $\mathrm{NBM}^{\mathbb{P}}$ over $\mathrm{DA}^{\mathbb{P}}$ for any priority distribution $\mathbb{P}$. □

3.B.5 Efficiency of the Adaptive Boston Mechanism

Proof of Proposition 10

Proof of Proposition 10.

1. *For any priority distribution $\mathbb{P}$, $ABM^{\mathbb{P}}$ is ex-post efficient but may not be ordinally or rank efficient.*
2. *Among all partially strategyproof mechanisms, $ABM^{\mathbb{U}}$ is not on the efficient frontier with respect to ordinal (or rank) dominance. Equivalently, there exists a partially strategyproof mechanism that strictly ordinally (and rank) dominates $ABM^{\mathbb{U}}$.*

Ex-post Efficiency & Ordinal and Rank Inefficiency: The proof of ex-post efficiency is completely analogous to the proof of ex-post efficiency for NBM. Furthermore, ordinal and rank inefficiency can be seen using the same example as for NBM because the assignments from NBM and ABM at the particular preference profile coincide.

Failure to be on the Efficient Frontier: We construct the mechanism ABM^{+} in the same way as NBM^{+}, i.e., we take $\mathrm{ABM}^{\mathbb{U}}$ as a baseline mechanism but replace the outcomes for preference profiles with separable wants by the outcomes chosen by the PS mechanism.

As for NBM^{+}, we consider a swap of two adjacent schools in the preference report of student 1, such that $\boldsymbol{P}$ satisfies separable wants.

In case (i), when the new profile also satisfies separable wants, the assignment does not change, which is consistent with upper invariance and swap monotonicity.

In case (ii), when the composition of first choices changes, student 1 must have ranked her second choice first. In this case, she will receive this new first choice with probability $\frac{1}{3}$ and the prior first choice with probability 0. This is also consistent with upper invariance and swap monotonicity.

Finally, in case (iii), the swap must involve c and d. She will still receive her first choice with probability $\frac{1}{2}$ and her second choice with probability 0 (as in the proof for $\text{NBM}^{\mathbb{U}}$). Therefore, her assignment for the school she brought down, can only decrease, and her assignment for the school she brought up can only increase, and both change by the same absolute value. This is consistent with upper invariance and swap monotonicity. □

Examples from Section 3.6.4

Example 11. Consider a setting with 6 students $N = \{1, \dots, 6\}$, 6 schools $M = \{a, \dots, f\}$, each available in unit capacity (i.e., $q_j = 1$ for all $j \in M$), and the preferences

$$\begin{aligned} P_1, \dots, P_4 &: a \succ b \succ c \succ d \succ e \succ f, \\ P_5, P_6 &: e \succ b \succ a \succ d \succ f \succ c. \end{aligned}$$

Consider the single priority profile given by the priority order $1 \; \pi \dots \pi \; 5$ which ranks the students according to their names. The deterministic mechanism $\text{SD}^{\boldsymbol{\pi}}$ will assign schools a through d to students 1 through 4. Students 5 and 6 will get schools e and f. Thus, $\text{SD}^{\boldsymbol{\pi}}$ assigns 2 first, 1 second, 1 third, 1 fourth, and 1 fifth choice. For the same priority profile, $\text{ABM}^{\boldsymbol{\pi}}$ will assign a, b, c to students $1, 2, 3$, respectively. Students 5 and 6 will get e and d, which leaves student 4 with f. Observe that in this case, $\text{ABM}^{\boldsymbol{\pi}}$ assigns 2 first, 1 second, 1 third, 1 fourth, no fifth, and 1 sixth choice. This is strictly rank dominated by the rank efficient assignment chosen by $\text{SD}^{\boldsymbol{\pi}}$ for this fixed priority profile.

For the probabilistic assignments that arise under the single uniform priority distribution we get that

$$\text{ABM}^{\mathbb{U}}(\boldsymbol{P}) = \frac{1}{60} \begin{pmatrix} 15 & 12 & 15 & 3 & 0 & 15 \\ 15 & 12 & 15 & 3 & 0 & 15 \\ 15 & 12 & 15 & 3 & 0 & 15 \\ 15 & 12 & 15 & 3 & 0 & 15 \\ 0 & 6 & 0 & 24 & 30 & 0 \\ 0 & 6 & 0 & 24 & 30 & 0 \end{pmatrix}, \tag{232}$$

and

$$\mathrm{DA}^{\mathbb{U}}(\boldsymbol{P}) = \frac{1}{60}\begin{pmatrix} 15 & 12 & 15 & 8 & 0 & 10 \\ 15 & 12 & 15 & 8 & 0 & 10 \\ 15 & 12 & 15 & 8 & 0 & 10 \\ 15 & 12 & 15 & 8 & 0 & 10 \\ 0 & 6 & 0 & 14 & 30 & 10 \\ 0 & 6 & 0 & 14 & 30 & 10 \end{pmatrix}. \tag{233}$$

Since the rank distribution $d^{\mathrm{DA}^{\mathbb{U}}(\boldsymbol{P})} = (2, 1, 1, 1, 1/3, 2/3)$ strictly stochastically dominates $d^{\mathrm{ABM}^{\mathbb{U}}(\boldsymbol{P})} = (2, 1, 1, 1, 0, 1)$, $\mathrm{DA}^{\mathbb{U}}$ strictly rank dominates $\mathrm{ABM}^{\mathbb{U}}$ at $\boldsymbol{P}$. Note that in this case in fact all assignments chosen by $\mathrm{DA}^{\mathbb{U}}$ are rank inefficient, but some of those chosen by $\mathrm{ABM}^{\mathbb{U}}$ are not.

Example 12. Consider a setting with 5 students $N = \{1, \ldots, 5\}$, 5 schools $M = \{a, \ldots, e\}$, each available in unit capacity (i.e., $q_j = 1$ for all $j \in M$), and the preferences

$$\begin{aligned} P_1, P_2 &: a \succ b \succ c \succ d \succ e, \\ P_3, P_4 &: a \succ d \succ c \succ e \succ b, \\ P_5 &: b \succ \ldots. \end{aligned}$$

For the single priority profile $\boldsymbol{\pi} = (\pi, \ldots, \pi)$ with $1 \; \pi \; 3 \; \pi \; 4 \; \pi \; 2 \; \pi \; 5$, $\mathrm{NBM}^{\boldsymbol{\pi}}$ will assign 1 to a, 5 to b, 3 to d, 4 to c, and 2 to e, which yields 2 first, 1 second, 1 third, and 1 fifth choices. $\mathrm{ABM}^{\boldsymbol{\pi}}$ will also assign 1 to a, 5 to b, and 3 to d. However, 2 will get c and 4 will get e, so that we get 2 first, 1 second, 1 third, and 1 fourth choices, a strictly rank dominant assignment.

The probabilistic assignments under $\mathrm{NBM}^{\mathbb{U}}$ and $\mathrm{ABM}^{\mathbb{U}}$ are

$$\mathrm{NBM}^{\mathbb{U}}(\boldsymbol{P}) = \frac{1}{60}\begin{pmatrix} 15 & 0 & 25 & 0 & 20 \\ 15 & 0 & 25 & 0 & 20 \\ 15 & 0 & 5 & 30 & 10 \\ 15 & 0 & 5 & 30 & 10 \\ 0 & 60 & 0 & 0 & 0 \end{pmatrix}, \tag{234}$$

$$\mathrm{ABM}^{\mathbb{U}}(\boldsymbol{P}) = \frac{1}{60}\begin{pmatrix} 15 & 0 & 30 & 0 & 15 \\ 15 & 0 & 30 & 0 & 15 \\ 15 & 0 & 0 & 30 & 15 \\ 15 & 0 & 0 & 30 & 15 \\ 0 & 60 & 0 & 0 & 0 \end{pmatrix}. \tag{235}$$

The rank distribution under $\mathrm{NBM}^{\mathbb{U}}$ is $d^{\mathrm{NBM}^{\mathbb{U}}(\boldsymbol{P})} = (2, 1, 1, 1/3, 2/3)$, which is dominated by $d^{\mathrm{ABM}^{\mathbb{U}}(\boldsymbol{P})} = (2, 1, 1, 1/2, 1/2)$.

Proof of Theorem 13

Proof of Theorem 13. Let $(N^k, M^k, \mathbf{q}^k)_{k\geqslant 1}$ be a sequence of settings such that

- *the set of schools does not change (i.e., $M^k = M$ for all k),*
- *the capacity of each school increases (i.e., $\min_{j\in M} q_j^k \to \infty$ for $k \to \infty$),*
- *the number of students equals the number of seats (i.e., $|N^k| = \sum_{j\in M} q_j^k$).*

Then the share of preference profiles where $DA^{\mathbb{U}}$ rank dominates $ABM^{\mathbb{U}}$ (even weakly) vanishes in the limit:

$$\lim_{k\to\infty} \frac{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k} : DA^{\mathbb{U}}(\boldsymbol{P}) \text{ rank dominates } ABM^{\mathbb{U}}(\boldsymbol{P}) \text{ at } \boldsymbol{P}\}}{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k}\}} = 0. \tag{236}$$

An assignment x is *first-choice-maximizing* at preference profile $\boldsymbol{P}$ if it can be represented as a lottery over deterministic assignments that give the maximum number of first choices, i.e.,

$$d_1^x = \sum_{i\in N} x_{i,j} \mathbb{1}_{r_i(j)=1} = \max_{y\in\mathcal{X}} d_1^y. \tag{237}$$

Since any ex-post efficient assignment is supported by a serial dictatorship, $\mathrm{DA}^{\mathbb{U}}$ puts positive probability on *all* ex-post efficient, deterministic assignments. In contrast, $\mathrm{ABM}^{\mathbb{U}}$ assigns positive probabilities to only *some* ex-post efficient, deterministic assignments. In particular, $\mathrm{ABM}^{\mathbb{U}}$ is first choice maximizing, i.e., it gives no probability to any assignment that does not yield the maximum possible number of first choices. Consequently, if at some preference profile $\boldsymbol{P}$ there exists at least one ex-post efficient, deterministic assignment that is not first-choice-maximizing, then $\mathrm{DA}^{\mathbb{U}}$ will assign strictly less first choices than $\mathrm{ABM}^{\mathbb{U}}$. At these preference profiles, $\mathrm{DA}^{\mathbb{U}}$ is guaranteed not to rank dominate $\mathrm{ABM}^{\mathbb{U}}$ (even weakly).

Using this observation, we can now prove the following Claim 13 which in turn yields the limit result, Theorem 13.

Claim 13. *For any fixed number of schools $m \geqslant 3$ and any $\epsilon > 0$, there exists $q_{\min} \in \mathbb{N}$, such that for any capacities $q_1, \ldots, q_m$ with $q_j \geqslant q_{\min}$ for all $j \in M$ and $n = \sum_{j \in M} q_j$ students, and for $\boldsymbol{P}$ chosen uniformly at random from $\mathcal{P}^n$, the probability that $DA^{\mathbb{U}}(\boldsymbol{P})$ is first-choice-maximizing is smaller than ϵ.*

For a given preference profile $\boldsymbol{P} \in \mathcal{P}^N$, the *first choice profile* $\mathbf{k}^{\boldsymbol{P}} = (k_j^{\boldsymbol{P}})_{j \in M}$ is the vector of non-negative integers, where $k_j^{\boldsymbol{P}}$ represents the number of students whose first choice is j. For a fixed setting, i.e., the triple $(N, M, \mathbf{q})$, we consider a uniform distribution on the space of preference profiles $\mathcal{P}^n$. As the preference profile $\boldsymbol{P}$ is held fixed, we suppress the index and simply write k_j. We say that a school $j \in M$ is

- *un-demanded* if $k_j = 0$,
- *under-demanded* if $k_j \in \{1, \ldots, q_j - 1\}$,
- *exhaustively demanded* if $k_j = q_j$,
- and *over-demanded* if $k_j > q_j$.

For any first choice profile $\mathbf{k}^{\boldsymbol{P}}$, one of the following cases must hold:

(I) There is at least one un-demanded school.

(II) All schools are exhaustively demanded.

(III) No school is un-demanded, at least one school is over-demanded, and at least one other school is exhaustively demanded.

(IV) No school is un-demanded, but at least two schools are over-demanded.

(V) There is exactly one over-demanded school, and all other schools are under-demanded.

We will show that for fixed m and increasing minimum capacity, the probabilities for cases (I) and (II) become arbitrarily small. We will further show that in cases (III), (IV), and (V), the probabilities that $\mathrm{DA}^{\mathbb{U}}$ assigns the maximum number of first choices become arbitrarily small.

(I) The probability that under a randomly chosen preference profile at least one school is un-demanded is upper-bounded by

$$\frac{\binom{m}{1}(m-1)^n}{m^n} = m\left(\frac{m-1}{m}\right)^n, \tag{238}$$

which converges to 0 as $n = \sum_{j \in M} q_j \geqslant m q_{\min}$ becomes large (where m is fixed).

(II) Let $\tilde{q} = \frac{n}{m}$. Without loss of generality, $\tilde{q}$ can be chosen as a natural number (otherwise, we increase the capacity of the school with least capacity until n is divisible by m). The probability that under a randomly chosen preference profile all schools are exhaustively demanded is

$$\frac{\binom{n}{q_1,\ldots,q_m}}{m^n} \leqslant \frac{\binom{n}{\tilde{q},\ldots,\tilde{q}}}{m^n} \lesssim \frac{(m\tilde{q})^{m\tilde{q}}}{(\tilde{q})^{m\tilde{q}} m^{m\tilde{q}}} \sqrt{\frac{m\tilde{q}}{\tilde{q}^m}} = \sqrt{\frac{m}{\tilde{q}^{m-1}}}, \tag{239}$$

which converges to 0 as $\tilde{q} \geqslant q_{\min}$ becomes large (where m is fixed).

(III) Suppose that $\mathrm{DA}^{\mathbb{U}}$ is first choice maximizing. If one school a is over-demanded and another school b is exhaustively demanded, then no student with first choice a can have b as second choice. Otherwise, there exists an order of the student such that a student with first choice a will get b. In that case, b is not assigned entirely to students with first choice b, and hence, the assignment can not maximize the number of first choices. Thus, the probability that the k_a students who have first choice a all have a second choice different from b (conditional on the first choice profile) is

$$\left(\frac{m-2}{m-1}\right)^{k_1} < \left(\frac{m-2}{m-1}\right)^{q_1} \leqslant \left(\frac{m-2}{m-1}\right)^{q_{\min}}. \tag{240}$$

This becomes arbitrarily small for increasing $q_{\min}$. Thus, the probability that the maximum number of first choices is assigned by $\mathrm{DA}^{\mathbb{U}}$, conditional on case (III) becomes small.

(IV) This case is analogous to (III).

(V) Suppose that for some preference profile consistent with case (V), $\mathrm{DA}^{\mathbb{U}}$ assigns the maximum number of first choices. Let a be the school that is over-demanded and let $j_2, \ldots, j_m$ be the under-demanded schools. Then the maximum number of first choices is assigned if and only if

- q_a students with first choice a receive a, and
- all students with first choices $j_2, \ldots, j_m$ receive their respective first choice.

If $\mathrm{DA}^{\mathbb{U}}$ maximizes the number of first choices, then for any ordering of the students, the maximum number of first choices must be assigned, i.e., the two conditions are true. If the students with first choice a get to pick before all other students, then

they exhaust a and get at most $q_j - k_j$ of the schools $j \neq a$: otherwise, if they got more than $q_j - k_j$ of school j, then some student with first choice j would get a worse choice, which violates first choice maximization.

After any q_a of the k_a students with first choice a consume school a, there are $k_a - q_a$ students left which will consume other schools. Since $n = \sum_{j \in M} q_j = \sum_{j \in M} k_j$, we get that

$$k_a - q_a = n - \sum_{j \neq a} k_j - (n - \sum_{j \neq a} q_j) = \sum_{j \neq a} q_j - k_j. \tag{241}$$

Therefore, the second choice profile of these $k_a - q_a$ students must be $(l_2, \ldots, l_m)$, where $l_r = q_{j_r} - k_{j_r} \geqslant 1$. In addition, some student i' who consumed a has second choice j', and some student i'' with first choice a gets its second choice $j'' \neq j'$. If we exchange the place of i' and i'' in the ordering, i'' will get a and i' will get j'. But then $q_{j'} - k_{j'} + 1$ students with first choice a get their second choice j'. Therefore, when the students with first choice j' get to pick their schools, there are only $k_{j'} - 1$ copies of j' left, which is not sufficient. Thus, we have constructed an ordering of the students under which the number of assigned first choices is not maximized. This implies that for any preference profile with first choice profile satisfying case (V), $\mathrm{DA}^{\mathbb{U}}$ does not assign the maximum number of first choices.

Combining the arguments for all cases, we can find $q_{\min}$ sufficiently high, such that we can estimate the probability that $\mathrm{DA}^{\mathbb{U}}$ maximizes first choices (*$DA^{\mathbb{U}}$ mfc.*) by

$$\begin{aligned}
\mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}] &= \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{I})]\mathbb{Q}[(\mathrm{I})] && (242)\\
&\quad + \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{II})]\mathbb{Q}[(\mathrm{II})] && (243)\\
&\quad + \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{III})]\mathbb{Q}[(\mathrm{III})] && (244)\\
&\quad + \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{IV})]\mathbb{Q}[(\mathrm{IV})] && (245)\\
&\quad + \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{V})]\mathbb{Q}[(\mathrm{V})] && (246)\\
&\leqslant \mathbb{Q}[(\mathrm{I})] + \mathbb{Q}[(\mathrm{II})] + \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{III})] && (247)\\
&\quad + \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{IV})] + \mathbb{Q}[\mathrm{DA}^{\mathbb{U}} \text{ mfc.}|(\mathrm{V})] && (248)\\
&\leqslant \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} + 0 = \epsilon. && (249)
\end{aligned}$$

Here, $\mathbb{Q}$ is the probability measure induced by the random selection of a preference profile. □

Proof of Theorem 14

Proof of Theorem 14. Let $(N^k, M^k, \mathbf{q}^k)_{k\geqslant 1}$ be a sequence of settings such that

- *the number of schools equals the number of students (i.e., $|M^k| = |N^k| = k$),*
- *all schools have unit capacity (i.e., $q_j^k = 1$ for $j \in M^k$).*

Then the share of preference profiles where $DA^{\mathbb{U}}$ rank dominates $ABM^{\mathbb{U}}$ (even weakly) vanishes in the limit:

$$\lim_{k\to\infty} \frac{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k} : DA^{\mathbb{U}}(\boldsymbol{P}) \text{ rank dominates } ABM^{\mathbb{U}}(\boldsymbol{P}) \text{ at } \boldsymbol{P}\}}{\#\{\boldsymbol{P} \in \mathcal{P}^{N^k}\}} = 0. \tag{250}$$

As for the proof of Theorem 13, we establish that $\mathrm{DA}^{\mathbb{U}}$ is almost never first choice maximizing at a randomly selected preference profile.

Claim 14. *For any $\epsilon > 0$, there exists $n \in \mathbb{N}$, such that for any setting $(N, M, \mathbf{q})$ with $\#M = \#N \geqslant n$ and $q_j = 1$ for all $j \in M$, and for $\boldsymbol{P}$ chosen uniformly at random, the probability that $DA^{\mathbb{U}}(\boldsymbol{P})$ is first-choice-maximizing is smaller than ϵ.*

Recall that for a fixed preference profile $\boldsymbol{P}$, $\mathrm{DA}^{\mathbb{U}}$ is first-choice maximizing only if all ex-post efficient assignments are first-choice maximizing. We will introduce *no overlap*, a necessary condition on the preference profile that ensures that $\mathrm{DA}^{\mathbb{U}}$ assigns the maximum number of first choices. Conversely, if a preference profile violates *no overlap*, $\mathrm{DA}^{\mathbb{U}}$ will not assign the maximum number of first choices. To establish Claim 14, we show that the share of preference profiles that exhibit no overlap vanishes as n becomes large.

The proof requires some more formal definitions: for convenience, we will enumerate the set M of schools by the integers $\{1, \ldots, n\}$. As in the proof of Theorem 13, $\mathbf{k}^{\boldsymbol{P}} = (k_1^{\boldsymbol{P}}, \ldots, k_n^{\boldsymbol{P}})$ is called the *first choice profile* of the type profile $\boldsymbol{P}$, where $k_j^{\boldsymbol{P}}$ is the number of students whose first choice is school j. To reduce notation, we suppress the superscript $\boldsymbol{P}$. For some first choice profile $\mathbf{k}$ and school j we define the following indicators:

$$w_{\mathbf{k}}(j) := \begin{cases} 1, & \text{if } k_j \geqslant 1 \\ 0, & \text{else} \end{cases} \quad \text{and } o_{\mathbf{k}}(j) := \begin{cases} 1, & \text{if } k_j \geqslant 2 \\ 0, & \text{else.} \end{cases} \tag{251}$$

w indicates whether j is *demanded*, i.e., it is the first choice of at least one student, and o indicates whether j is *over-demanded*, i.e., it is the first choice of more than one student.

Further, we define

$$W_{\mathbf{k}} := \sum_{j \in M} w_{\mathbf{k}}(j), \quad O_{\mathbf{k}} := \sum_{j \in M} o_{\mathbf{k}}(j), \quad C_{\mathbf{k}} = \sum_{j \in M} k_j \cdot o_{\mathbf{k}}(j). \tag{252}$$

$W_{\mathbf{k}}$ is the number of schools that are demanded by at least one student, $O_{\mathbf{k}}$ is the number of over-demanded schools, and $C_{\mathbf{k}}$ is the number of students competing for over-demanded schools. Finally, a preference profile $\boldsymbol{P}$ *exhibits overlap* if there exists a student $i \in N$ with first choice j_1 and second choice j_2, such that $o_{\mathbf{k}^{\boldsymbol{P}}}(j_1) = 1$ and $w_{\mathbf{k}^{\boldsymbol{P}}}(j_2) = 1$, i.e., student i's first choice is over-demanded and its second choice is demanded as a first choice by at least one other student. As an example consider a setting where three students have preferences

$$P_1 : a \succ \ldots, \quad \succ_2 : a \succ b \succ \ldots, \quad \succ_3 : b \succ \ldots. \tag{253}$$

The maximum number of first choices that can be assigned is 2, e.g., by giving a to 1 and b to 3. But for the priority order $1 \ \pi \ 2 \ \pi \ 3$, student 1 will get a and student 2 will get b. Then student 3 cannot take b, and consequently $\mathrm{DA}^{\mathbb{U}}$ will not assign the maximum number of first choices. If a preference profile exhibits overlap, a situation as in (253) will arise for some priority order, and therefore, $\mathrm{DA}^{\mathbb{U}}$ will not assign the maximum number of first choices. Conversely, no overlap in $\boldsymbol{P}$ is a necessary condition for $\mathrm{DA}^{\mathbb{U}}(\boldsymbol{P})$ to assign the maximum number of first choices. We will show in the following that the share of preference profiles exhibiting no overlap becomes small for increasing n.

Consider a uniform distribution (denoted $\mathbb{Q}$) on the preference profiles, i.e., all students draw their preference order independently and uniformly at random from the space of all possible preference orders. Then the statement that *the share of preference profiles exhibiting no overlap becomes small* is equivalent to the statement that the *probability of selecting a preference profile with no overlap converges to 0.* The proof of the following Claim 15 is technical and requires involved combinatorial and asymptotic arguments.

Claim 15. $\mathbb{Q}[\boldsymbol{P}$ *no overlap*$] \to 0$ *for* $n \to \infty$.

Proof of Claim 15. Using conditional probability, we can write the probability that a preference profile is without overlap as

$$\mathbb{Q}[\boldsymbol{P} \text{ no overlap}] = \sum_{\mathbf{k}} \mathbb{Q}[\mathbf{k} = \mathbf{k}^{\boldsymbol{P}}] \cdot \mathbb{Q}[\boldsymbol{P} \text{ no overlap} \mid \mathbf{k} = \mathbf{k}^{\boldsymbol{P}}]. \tag{254}$$

The number of preference profiles that have first choice profile $\mathbf{k} = (k_1, \ldots, k_n)$ is

proportional to the number of ways to distribute n unique balls (students) across n urns (first choices), such that k_j balls end up in urn j. Thus,

$$\mathbb{Q}[\mathbf{k} = \mathbf{k}^{\boldsymbol{P}}] = \frac{\binom{n}{k_1,\ldots,k_n}(n-1)!^n}{(n!)^n} = \frac{\binom{n}{k_1,\ldots,k_n}}{n^n}. \tag{255}$$

In order to ensure no overlap, a student with an over-demanded first choice cannot have as her second choice a school that is the first choice of any other student. Students whose first choice is not over-demanded can have any school (except for their own first choice) as second choice. Thus, given a first choice profile $\mathbf{k}$, the conditional probability of no overlap is

$$\begin{aligned}
&\mathbb{Q}[\boldsymbol{P} \text{ no overlap} \mid \mathbf{k} = \mathbf{k}^{\boldsymbol{P}}] && (256)\\
=\ &\prod_{j\in M}\left((1-o_{\mathbf{k}}(j)) + o_{\mathbf{k}}(j)\left(\frac{n-W_{\mathbf{k}}}{n-1}\right)^{k_j}\right) && (257)\\
=\ &\left(\frac{n-W_{\mathbf{k}}}{n-1}\right)^{\sum_{j\in M} k_j\cdot o_{\mathbf{k}}(j)} = \left(\frac{n-W_{\mathbf{k}}}{n-1}\right)^{C_{\mathbf{k}}} = \left(\frac{C_{\mathbf{k}}-O_{\mathbf{k}}}{n-1}\right)^{C_{\mathbf{k}}}, && (258)
\end{aligned}$$

where the last equality holds, since $n - W_{\mathbf{k}} = n - (n - C_{\mathbf{k}} + O_{\mathbf{k}}) = C_{\mathbf{k}} - O_{\mathbf{k}}$. Thus, the probability of no overlap can be determined as

$$\mathbb{Q}[\boldsymbol{P} \text{ no overlap}] = \frac{1}{n^n}\sum_{\mathbf{k}}\binom{n}{k_1,\ldots,k_n}\left(\frac{C_{\mathbf{k}}-O_{\mathbf{k}}}{n-1}\right)^{C_{\mathbf{k}}}. \tag{259}$$

$C_{\mathbf{k}}$ is either 0 or $\geqslant 2$, since a single student cannot be in competition. If no students compete ($C_{\mathbf{k}} = 0$), all must have different first choices. Thus, for $\mathbf{k} = (1,\ldots,1)$, the term in the sum in (259) is

$$\binom{n}{k_1,\ldots,k_n}\left(\frac{C_{\mathbf{k}}-O_{\mathbf{k}}}{n-1}\right)^{C_{\mathbf{k}}} = \binom{n}{1,\ldots,1}\cdot 1 = n!. \tag{260}$$

Using this and sorting the terms for summation by c for $C_{\mathbf{k}}$ and o for $O_{\mathbf{k}}$, we get

$$\begin{aligned}
&\mathbb{Q}[\boldsymbol{P} \text{ no overlap}]\\
=\ &\frac{1}{n^n}\left[n! + \sum_{c=2}^{n}\sum_{o=1}^{\lfloor\frac{c}{2}\rfloor}\left(\frac{c-o}{n-1}\right)^{c}\sum_{\mathbf{k}:C_{\mathbf{k}}=c,O_{\mathbf{k}}=o}\binom{n}{k_1,\ldots,k_n}\right]. && (261)
\end{aligned}$$

Consider the inner sum

$$\sum_{\mathbf{k}:C_{\mathbf{k}}=c,O_{\mathbf{k}}=o} \binom{n}{k_1,\ldots,k_n} \tag{262}$$

in (261): with a first choice profile $\mathbf{k}$ that satisfies $C_{\mathbf{k}} = c$ and $O_{\mathbf{k}} = o$ there are exactly o over-demanded schools (i.e., schools j with $k_j \geqslant 2$), $n - c$ singly-demanded schools (with $k_j = 1$), and $c - o$ un-demanded schools (with $k_j = 0$). Therefore,

$$\begin{aligned}
&\sum_{\mathbf{k}:C_{\mathbf{k}}=c,O_{\mathbf{k}}=o} \binom{n}{k_1,\ldots,k_n} && (263)\\
=\ &\binom{n}{c-o} \sum_{\mathbf{k}'=(k'_1,\ldots,k'_{n-c+o}):C_{\mathbf{k}'}=c,O_{\mathbf{k}'}=o} \binom{n}{k'_1,\ldots,k'_{n-c+o}} && (264)\\
=\ &\binom{n}{c-o}\binom{n-c+o}{n-c}\frac{n!}{c!} \sum_{\mathbf{k}''=(k''_1,\ldots,k''_o):k''_j\geqslant 2} \binom{c}{k''_1,\ldots,k''_o}. && (265)
\end{aligned}$$

The first equality holds because we simply choose $c-o$ of the n schools to be un-demanded, and

$$\binom{n}{k_1,\ldots,k_{r-1},0,k_{r+1},\ldots,k_m} = \binom{n}{k_1,\ldots,k_{r-1},k_{r+1},\ldots,k_m}. \tag{266}$$

The second equality holds because we select the $n - c$ singly-demanded schools from the remaining $n - c + o$ schools as well as the $n - c$ students to demand them. The sum (265) is equal to the number of ways to distribute c unique balls to o unique urns such that each urn contains at least 2 balls. This in turn is equal to

$$o! \left\{\!\!\left\{ \begin{matrix} c \\ o \end{matrix} \right\}\!\!\right\}, \tag{267}$$

where $\{\{:\}\}$ denotes the 2-associated Stirling number of the second kind. This number represents the number of ways to partition c unique balls such that each partition contains at least 2 balls. The factor $o!$ in (267) is included to make the partitions unique. $\{\{:\}\}$ is upper-bounded by $\{:\}$, the Stirling number of the second kind, which represents the number of ways to partition c unique balls such that no partition is empty. Furthermore, the Stirling number of the second kind has the upper bound

$$\left\{ \begin{matrix} c \\ o \end{matrix} \right\} \leqslant \binom{c}{o} o^{c-o}. \tag{268}$$

Thus, the sum in (265) can be upper-bounded by

$$\sum_{\mathbf{k}''=(k_1'',\ldots,k_o'')\colon k_j''\geqslant 2} \binom{c}{k_1'',\ldots,k_o''} \leqslant o!\binom{c}{o}o^{c-o}. \tag{269}$$

Combining all the previous observations, we can estimate the probability $\mathbb{Q}[\boldsymbol{P}$ no overlap$]$ from (261) by

$$\begin{aligned} &\mathbb{Q}[\boldsymbol{P} \text{ no overlap}] \qquad (270)\\ \leqslant\ & \frac{1}{n^n}\left[n! + \sum_{c=2}^{n}\sum_{o=1}^{\lfloor\frac{c}{2}\rfloor}\left(\frac{c-o}{n-1}\right)^c\binom{n}{c-o}\binom{n-c+o}{n-c}\binom{c}{o}\frac{n!o!}{c!}o^{c-o}\right]. \end{aligned}$$

The Stirling approximation yields

$$\sqrt{2\pi}e^{\frac{1}{12n+1}} \leqslant \frac{n!}{\sqrt{n}\left(\frac{n}{e}\right)^n} \leqslant \sqrt{2\pi}e^{\frac{1}{12n}}, \tag{271}$$

and therefore $n! \approx \left(\frac{n}{e}\right)^n\sqrt{n}$ up to a constant factor. Using this, we observe that the first term in (270) converges to 0 as n increases, i.e.,

$$\frac{n!}{n^n} \approx \frac{\sqrt{n}}{e^n} \to 0 \text{ for } n\to\infty. \tag{272}$$

Now we need to estimate the double sum in (270):

$$\begin{aligned} &\frac{1}{n^n}\sum_{c=2}^{n}\sum_{o=1}^{\lfloor\frac{c}{2}\rfloor}\left(\frac{c-o}{n-1}\right)^c\binom{n}{c-o}\binom{n-c+o}{n-c}\binom{c}{o}\frac{n!o!}{c!}o^{c-o} \qquad (273)\\ =\ & \frac{n!}{n^n}\sum_{c=2}^{n}\sum_{o=1}^{\lfloor\frac{c}{2}\rfloor}\frac{n!}{(c-o)!(n-c+o)!}\cdot\frac{(n-c+o)!}{o!(n-c)!}\cdot\frac{c!}{o!(c-o)!}\cdot\frac{o!}{c!}\cdot\frac{(c-o)^c o^{c-o}}{(n-1)^c}\\ =\ & \frac{n!}{n^n}\sum_{c=2}^{n}\sum_{o=1}^{\lfloor\frac{c}{2}\rfloor}\binom{n}{c}\binom{c}{o}\left(\frac{n}{n-1}\right)^c\frac{1}{n^c}\cdot\frac{(c-o)^c o^{c-o}}{(c-o)!} \qquad (274)\\ \lesssim\ & \left[\sqrt{n}\left(\frac{n}{n-1}\right)^{n-1}\left(\frac{n-1}{n}\right)\right]\frac{1}{e^n}\sum_{c=2}^{n}\sum_{o=1}^{\lfloor\frac{c}{2}\rfloor}\binom{n}{c}\binom{c}{o}\frac{1}{n^c}\cdot\frac{(c-o)^c o^{c-o}}{(c-o)^{c-o}}e^{c-o}\\ \leqslant\ & [e\sqrt{n}]\frac{1}{e^n}\sum_{c=2}^{n}\sum_{o=1}^{\lfloor\frac{c}{2}\rfloor}\binom{n}{c}\binom{c}{o}\frac{1}{n^c}\cdot(c-o)^o o^{c-o}e^{c-o}, \qquad (275) \end{aligned}$$

where we use that $\left(1+\frac{x}{n}\right)^n \leqslant e^x$. Using the binomial theorem and the fact that the function $o \mapsto (c-o)^o o^{c-o}$ is maximized by $o = \frac{c}{2}$, we can further estimate (275) by

$$\begin{aligned}
& \left[e\sqrt{n}\right] \frac{1}{e^n} \sum_{c=2}^{n} \binom{n}{c} \left(\frac{ec}{2n}\right)^c \sum_{o=1}^{\left\lfloor \frac{c}{2} \right\rfloor} \binom{c}{o} \left(\frac{1}{e}\right)^o && (276)\\
\leqslant\ & \left[e\sqrt{n}\right] \frac{1}{e^n} \sum_{c=2}^{n} \binom{n}{c} \left(\frac{ec}{2n}\right)^c \left(1+\frac{1}{e}\right)^c && (277)\\
\leqslant\ & \left[e\sqrt{n}\right] \frac{1}{e^n} \sum_{c=2}^{n} \binom{n}{c} \left(\alpha \cdot \frac{c}{n}\right)^c && (278)
\end{aligned}$$

with $\alpha = \frac{(1+e)}{2}$. To estimate the sum in (278), we first consider even n and note the following:

- $\alpha = \frac{(1+e)}{2} \approx 1.85914\ldots < e$, and therefore, the last term of the sum for $c = n$ can be ignored as $\binom{n}{n}\left(\frac{\alpha}{e}\right)^n \to 0$ for $n \to \infty$.

- $\binom{n}{c} = \binom{n}{n-c}$, and therefore, both terms $\binom{n}{c}\left(\alpha\frac{c}{n}\right)^c$ and $\binom{n}{c}\left(\alpha\frac{n-c}{n}\right)^{n-c}$ have the same binomial coefficient in the sum.

- The idea is to estimate the sum of both terms by an exponential function of the form $c \mapsto e^{mc+b}$, where m and b depend only on n and α.

- Indeed, the log of the sum, the function $c \mapsto \log\left(\left(\alpha\frac{c}{n}\right)^c + \left(\alpha\frac{n-c}{n}\right)^{n-c}\right)$, is strictly convex and on the interval $\left[1, \frac{n}{2}\right]$ it is upper-bounded by the linear function

$$f(c) = \left(\frac{\log(4)}{n} - \log(2\alpha)\right) c + n\log(\alpha). \tag{279}$$

- Thus,

$$\binom{n}{c}\left(\alpha\frac{c}{n}\right)^c + \binom{n}{n-c}\left(\alpha\frac{n-c}{n}\right)^{n-c} \leqslant \binom{n}{c} e^{f(c)}. \tag{280}$$

We can bound (278) by

$$
\begin{aligned}
& \left[e\sqrt{n}\right] \frac{1}{e^n} \sum_{c=1}^{\frac{n}{2}} \binom{n}{c} e^{f(c)} && (281)\\
= \; & \left[e\sqrt{n}\right] \frac{1}{e^n} \sum_{c=1}^{\frac{n}{2}} \binom{n}{c} 4^{\frac{c}{n}} \alpha^n \left(\frac{1}{2\alpha}\right)^c && (282)\\
\leqslant \; & \left[4e\sqrt{n}\right] \frac{\alpha^n \left(1+\frac{1}{2\alpha}\right)^n}{e^n} && (283)\\
= \; & \left[4e\sqrt{n}\right] \left(\frac{\frac{1}{2}+\alpha}{e}\right)^n \approx \left[4e\sqrt{n}\right] \left(\frac{2.35914\ldots}{e}\right)^n . && (284)
\end{aligned}
$$

Since $2.35914 < e$, the exponential convergence of the last term dominates the divergence of the first terms, which is of the order $\sqrt{n}$, and the expression converges to 0.

For odd n the argument is essentially the same, except that we need to also consider the central term (for $c = \frac{n}{2} + 1$) separately.

$$
\begin{aligned}
& \binom{n}{\frac{n}{2}+1} \left(\alpha \frac{\frac{n}{2}+1}{2}\right)^{\frac{n}{2}+1} && (285)\\
\leqslant \; & 2^n \left(\sqrt{\alpha\left(\frac{1}{2}+\frac{1}{n}\right)}\right)^n \cdot \alpha\left(\frac{1}{2}+\frac{1}{n}\right) && (286)\\
= \; & \left(\sqrt{\alpha\left(2+\frac{4}{n}\right)}\right)^n \cdot \alpha\left(\frac{1}{2}+\frac{1}{n}\right). && (287)
\end{aligned}
$$

With $\sqrt{\alpha\left(2+\frac{4}{n}\right)} \approx 1.92828\ldots < e$, the result follows for odd n as well. □

This completes the proof of Theorem 14. □

4 Hybrid Mechanisms: Trading Off Strategyproofness and Efficiency of Random Assignment Mechanisms

Abstract

Severe impossibility results restrict the design of strategyproof random assignment mechanisms, and trade-offs are necessary when aiming for more demanding efficiency requirements, such as ordinal or rank efficiency. We introduce *hybrid mechanisms*, which are convex combinations of two component mechanisms. We give a set of conditions under which such hybrids facilitate a non-degenerate trade-off between strategyproofness (in terms of *partial strategyproofness*) and efficiency (in terms of *dominance*). This set of conditions is tight in the sense that trade-offs may become degenerate if any of the conditions are dropped. Moreover, we give an algorithm for the mechanism designer's problem of determining a maximal mixing factor. Finally, we prove that our construction can be applied to mix Random Serial Dictatorship with Probabilistic Serial, as well as with the adaptive Boston mechanism, and we illustrate the efficiency gains numerically.

4.1 Introduction

When a set of indivisible goods or resources (called *objects*) has to be assigned to self-interested agents without the use of monetary transfers, we face an *assignment problem*. Examples include the assignment of students to schools, subsidized housing to tenants, and teachers to training programs (Niederle, Roth and Sönmez, 2008). Since the seminal paper of Hylland and Zeckhauser (1979), this problem has attracted much attention from mechanism designers (e.g., Abdulkadiroğlu and Sönmez (1998); Bogomolnaia and Moulin (2001); Abdulkadiroğlu and Sönmez (2003*a*); Featherstone (2011)).

It is often desirable or even required that assignment mechanisms perform well on

multiple dimensions, such as *efficiency*, *fairness*, and *strategyproofness*. However, severe impossibility results prevent the design of mechanisms that excel on all these dimensions simultaneously (Zhou, 1990). This makes the assignment problem a challenge for mechanism designs. The *Random Serial Dictatorship (RSD)* mechanism is *strategyproof* and *anonymous*, but only satisfies the baseline requirement of *ex-post efficiency*. If strategyproofness is relaxed to *weak strategyproofness*, the more demanding requirement of *ordinal efficiency* can be achieved via the *Probabilistic Serial (PS)* mechanism. However, no ordinally efficient, symmetric mechanism can be strategyproof (Bogomolnaia and Moulin, 2001). The even more demanding requirement of *rank efficiency* can be achieved by *Rank Value (RV)* mechanisms, but at the same time, rank efficiency is in conflict with even weak strategyproofness. Obviously, trade-offs are necessary and have been the focus of recent research (e.g., see (Abdulkadiroğlu, Pathak and Roth, 2009; Budish, 2012; Aziz, Brandt and Brill, 2013*a*; Azevedo and Budish, 2015)).

In this paper we investigate a straightforward approach to the problem of trading off strategyproofness and efficiency of random assignment mechanisms. Specifically, we use *partial strategyproofness* (Mennle and Seuken, 2015*b*) to compare mechanisms by their incentive properties, and we use *ordinal dominance* and *rank dominance* to compare them by their efficiency properties. We introduce *hybrid mechanisms*, which are convex combinations of two component mechanisms, and we show that, subject to a set of quite intuitive conditions, hybrid mechanisms behave exactly as one would expect: they facilitate a non-degenerate trade-off between strategyproofness and efficiency. Instantiating this approach with popular assignment mechanisms, such as RSD, PS, RV, and variants of the Boston mechanism, we illustrate that the conditions are not trivial; but when they hold, the efficiency gains (over RSD) can be substantial.

4.1.1 Partially Strategyproof Hybrid Mechanisms

Due to restrictive impossibility results pertaining to strategyproofness, we cannot hope to improve efficiency of mechanisms without relaxing strategyproofness, especially in the presence of additional fairness criteria, such as anonymity. In (Mennle and Seuken, 2015*b*) we have introduced a relaxed incentive requirement for assignment mechanisms: a mechanism is *r-partially strategyproof* if truthful reporting is a dominant strategy for agents who have sufficiently different valuations for different objects. The numerical parameter r controls the extent to which their valuations must vary across objects. r yields a parametric measure for the strength of the incentive properties of non-strategyproof mechanisms. Larger values of $r \in [0, 1]$ imply stronger incentive guarantees, $r = 1$

corresponds to strategyproofness, and $r = 0$ does not yield any incentive guarantees. We use this *degree of strategyproofness* to quantify the performance of mechanisms on the strategyproofness-dimensions.

In this paper, we study how hybrid mechanisms trade off strategyproofness and efficiency. The key idea of *hybrid mechanisms* is to "mix" a mechanism φ that has good incentive properties and another mechanism ψ that has good efficiency properties. Concretely, for two component mechanisms φ and ψ the β-hybrid is given by $h^\beta = (1-\beta)\varphi + \beta\psi$. The parameter $\beta \in [0,1]$ is called the *mixing factor*. Obviously, $h^0 = \varphi$ and $h^1 = \psi$, so that the hybrid mechanisms at the extreme mixing factors $\beta = 0$ and $\beta = 1$ trivially inherit the desirable property of the respective component mechanism. For intermediate mixing factors $\beta \in (0,1)$ hybrids should intuitively inherit a *share* of the desirable properties from both component mechanisms.

Regarding the strategyproofness-dimension, we find that this intuition may not always be justified: as we show in this paper, it can happen that any non-trivial share of ψ (i.e., any $\beta > 0$) causes the degree of strategyproofness of the hybrid to drop to 0 immediately. Our first main result is a set of sufficient conditions that prevent such "degenerate" behavior: a pair (φ, ψ) is *hybrid-admissible* if

(1) φ is strategyproof,

(2) ψ is *upper invariant*: a swap of two adjacent objects in an agent's report does not change that agent's probabilities for obtaining an object it prefers to any of the two,

(3) ψ is *weakly less varying* than φ: whenever a swap leads to a change of an agent's assignment under ψ, then that agent's assignment must also change under φ.

For any hybrid-admissible pair, we show that a non-degenerate trade-off is possible in the sense that for all (even small) relaxations of strategyproofness, the share of ψ that can be included in the hybrid is non-trivial. Furthermore, we show that it is not possible to drop any of the three conditions from hybrid-admissibility and still guarantee that hybrid mechanisms with intermediate degrees of strategyproofness can be constructed.

4.1.2 Harnessing Efficiency Improvements

Towards understanding the efficiency improvements that we can obtain through hybrid mechanisms, we employ the well-known concepts of *ordinal* and *rank dominance*. Our second main result is that if ψ dominates φ, then h^β dominates φ but is dominated by ψ. Thus, hybrids have *intermediate efficiency* in a well-defined sense. One challenge is

that a comparison of φ and ψ by dominance may not be possible at every preference profile. For cases where they are incomparable at some preference profile, we show that h^β is not dominated by φ and h^β does not dominate ψ. In other words, the dominance comparison of h^β, φ, and ψ *points in the right direction* whenever this comparison is possible, and it *never points in the wrong direction* when φ and ψ are not comparable.

This shows that if some mechanism φ offers good incentives, and another mechanism ψ has desirable efficiency, then a mechanism designer can trade off strategyproofness and efficiency systematically by constructing hybrids of φ and ψ. Concretely, she can specify the minimal acceptable degree of strategyproofness $\underline{\rho}$ and then choose the mixing factor β as high as possible. The *maximal mixing factor* $\beta^{\max}$ is the largest value of $\beta \in [0, 1]$ for which h^β is $\underline{\rho}$-partially strategyproof. The parameter $\beta^{\max}$ has an appealing interpretation: it serves as a measure for how far relaxing strategyproofness from "$r = 1$" (i.e., strategyproofness) to "$r \geqslant \underline{\rho}$" (i.e., $\underline{\rho}$-partial strategyproofness) will take us between the baseline efficiency of φ to the more desirable efficiency of ψ.

This "trade-off" approach to the design of random assignment mechanisms gives rise to a computational problem: given a setting (i.e., number of agents, number of objects, and object capacities), as well as a minimal acceptable degree of strategyproofness $\underline{\rho}$, the mechanism designer faces the problem of determining the maximal mixing factor $\beta^{\max}$. In this paper, we show how this problem can be solved algorithmically for any finite setting.

4.1.3 Hybrids of Popular Mechanisms

Finally, we apply our theory of hybrids to pairs of popular mechanisms. First, we show that Random Serial Dictatorship (RSD) and the Probabilistic Serial (PS) mechanism form a hybrid-admissible pair. Since PS is ordinally efficient, it ordinally dominates RSD whenever the two mechanisms are comparable. Therefore, hybrids of RSD and PS can be used to trade off strategyproofness and efficiency in terms of ordinal dominance.

Second, we show two impossibility result: neither the classic Boston mechanism (NBM),[1] nor Rank Value (RV) mechanisms are weakly less varying than RSD. Furthermore, we demonstrate that hybrids of RSD with NBM or RV indeed have degenerate incentive properties (i.e., they have a degree of strategyproofness of 0). These findings illustrate that while hybrid-admissibility is sufficient for non-degenerate trade-offs, it is also close to being necessary. On a broader scale, these impossibility results serve as

[1] We call this *naïve* Boston mechanism because it is "naïve" in the sense that agents apply to exhausted objects in the application process (Mennle and Seuken, 2015*d*).

reminders that straightforward approaches, like the construction of hybrid mechanisms, do not always yield the seemingly obvious outcomes that our intuition may suggest.

Third, we show that the pair of RSD and the adaptive Boston Mechanism (ABM)[2] is also hybrid-admissible. ABM rank dominates RSD whenever the outcomes are comparable, except in a negligibly small number of cases (Mennle and Seuken, 2015*d*). Thus, hybrids of RSD and ABM allow non-degenerate trade-offs between strategyproofness and efficiency in terms of the rank dominance relation (except for the small number of cases). For both pairs (RSD,PS) and (RSD,ABM), we find numerically that efficiency gains (in terms of the maximal mixing factor) from relaxing strategyproofness can be substantial.

Organization of this paper: In Sections 4.2 and 4.3, we discuss related work and introduce our formal model. In Section 4.4, we introduce hybrid-admissibility and show that it enables the construction of partially strategyproof hybrids.[3] In Section 4.5, we show how hybrids trade off strategyproofness and efficiency, and we give an algorithm for the mechanism designer's problem of determining a maximal mixing factor. In Section 4.6, we apply our results to popular assignment mechanisms. Section 4.7 concludes.

4.2 Related Work

Hylland and Zeckhauser (1979) proposed a pseudo-market mechanism for the problem of assigning students to on-campus housing. However, eliciting agents' cardinal preferences is often difficult if not impossible to in settings without money. For this reason, recent work has focused on *ordinal* mechanisms, where agents submit rankings over objects. Carroll (2011), Huesmann and Wambach (2015), and Ehlers et al. (2015) gave systematic arguments for the focus on ordinal mechanisms.

For the deterministic case, strategyproofness of assignment mechanisms has been studied extensively, e.g., in (Pápai, 2000; Ehlers and Klaus, 2006, 2007; Hatfield, 2009; Pycia and Ünver, 2014). For non-deterministic mechanisms, Abdulkadiroğlu and Sönmez (1998) showed that RSD is equivalent to the Core from Random Endowments mechanism for house allocation (if agents' initial houses are drawn uniformly at random). Erdil (2014) showed that when capacity exceeds demand, RSD is not the only strategyproof, ex-post efficient mechanism that satisfies symmetry. On the other hand, Bade (2014) showed that taking any ex-post efficient, strategyproof, non-bossy, deterministic mechanism and

[2] ABM is a variant of the Boston school choice mechanisms in which students automatically skip exhausted schools in the application process; see (Mennle and Seuken, 2015*d*).

[3] We emphasize that the partial strategyproofness concept imported from (Mennle and Seuken, 2015*b*) in Section 4.4.1 should *not* be considered a contribution of the present paper.

assigning agents to roles in the mechanism uniformly at random is equivalent to using RSD. However, when capacity equals demand, it is still an open conjecture whether RSD is the unique mechanism that is strategyproof, ex-post efficient, and symmetric (Lee and Sethuraman, 2011). Despite the fact that this conjecture remains to be proven, is evident that the space of "useful" strategyproof mechanisms is very small.

The research community has also introduced stronger efficiency concepts, such as *ordinal efficiency*. The Probabilistic Serial (PS) mechanism, which achieves ordinal efficiency, was originally introduced by Bogomolnaia and Moulin (2001) for strict preferences and since then it has been studied extensively: Katta and Sethuraman (2006) introduced an extension that allows agents to be indifferent between objects. Hashimoto et al. (2014) showed that PS with equal eating speads is the unique mechanism that is ordinally fair and non-wasteful. In terms of incentives, Bogomolnaia and Moulin (2001) showed that PS is not strategyproof but *weakly strategyproof* in the sense that no agent can obtain a first order-stochastically dominant assignment by misreporting.

While ex-post efficiency and ordinal efficiency are the two most well-studied efficiency concepts for assignment mechanisms, some mechanisms used in practice aim at *rank efficiency*, which is a further refinement of ordinal efficiency (Featherstone, 2011). However, no rank efficient mechanism can even be weakly strategyproof. Other popular mechanisms, like the Boston mechanism (see (Ergin and Sönmez, 2006)), are manipulable but are nevertheless in frequent use. Budish and Cantillon (2012) showed evidence from course allocation at Harvard Business School, suggesting that using a non-strategyproof mechanism may lead to higher social welfare than using a strategyproof mechanism such as RSD. This challenges the view that strategyproofness should be a hard requirement for mechanism design.

Given that strategyproofness is so restrictive, some researchers have considered relaxed incentive requirements. For example, Carroll (2013) used approximate strategyproofness for normalized vNM utilities in the voting domain to quantify the incentives to manipulate under different non-strategyproof voting rules. Budish (2011) proposed the Competitive Equilibrium from Approximately Equal Incomes mechanism for the combinatorial assignment problem. For the random social choice domain, Aziz, Brandt and Brill (2013*a*) considered first order-stochastic dominance (SD) and sure thing (ST) dominance. They showed that while RSD is SD-strategyproof, it is merely ST-efficient; they contrasted this with Strictly Maximal Lotteries, which are SD-efficient but only ST-strategyproof.

The construction of hybrid mechanisms in the present paper differs from these ap-

proaches: rather than comparing discrete points in the design space, we enable a continuous trade-off between strategyproofness and efficiency that can be described in terms of two parameters: the *degree of strategyproofness* (Mennle and Seuken, 2015*b*) for incentive properties and the *mixing factor* for efficiency. Formally, hybrid mechanisms are simply probability mixtures of two component mechanisms. Gibbard (1977) used such mixtures in his seminal characterization of the set of strategyproof random ordinal mechanisms. In (Mennle and Seuken, 2015*c*), we have extended Gibbard's result by giving a structural characterization of the Pareto frontier of *approximately* strategyproof random mechanisms. Hybrid mechanisms play a central role in our characterization. The present paper differs from (Mennle and Seuken, 2015*c*) in that we consider the random assignment domain specifically, and we employ *partial strategyproofness*, which is a more appropriate relaxation of strategyproofness in this domain than approximate strategyproofness.

4.3 Formal Model

A *setting* $(N, M, \boldsymbol{q})$ consists of a set N of n *agents*, a set M of m *objects*, and a vector $\boldsymbol{q} = (q_1, \ldots, q_m)$ of *capacities* (i.e., there are q_j units of object j). We assume that supply satisfies demand (i.e., $n \leqslant \sum_{j \in M} q_j$), since we can always add a dummy object with capacity n. Each agent $i \in N$ has a strict preference order P_i over objects, where $P_i : a \succ b$ means that i strictly prefers object a to object b. We denote the set of all preference orders by $\mathcal{P}$. A *preference profile* $\boldsymbol{P} = (P_1, \ldots, P_n) \in \mathcal{P}^N$ is a collection of preference orders of all agents, where $P_{-i} \in \mathcal{P}^{N \setminus \{i\}}$ are the preference orders of all agents, except i. Agents' preferences over objects are extended to preferences over lotteries via *von Neumann-Morgenstern (vNM) utilities* $u_i : M \to \mathbb{R}^+$. A utility function u_i is *consistent* with preference order P_i if $P_i : a \succ b$ whenever $u_i(a) > u_i(b)$. We denote by U_{P_i} the set of all utility functions that are consistent with P_i.

In the *random assignment* problem, each agent ultimately receives a single object, but we evaluate random mechanisms based on the resulting *interim assignments*. Such assignments are represented by an $n \times m$-matrix $x = (x_{i,j})_{i \in N, j \in M}$ satisfying the *fulfillment constraint* $\sum_{j \in M} x_{i,j} = 1$, the *capacity constraint* $\sum_{i \in N} x_{i,j} \leqslant q_j$, and the *probability constraint* $x_{i,j} \in [0, 1]$ for all $i \in N, j \in M$. The entries of the matrix x are interpreted as probabilities, where $x_{i,j}$ is the probability that i gets j. An assignment is *deterministic* if all agents get exactly one full object, such that $x_{i,j} \in \{0, 1\}$ for all $i \in N, j \in M$. For any agent i, the ith row $x_i = (x_{i,j})_{j \in M}$ of the matrix x is called the *assignment vector*

of i, or *i's assignment* for short. The Birkhoff-von Neumann Theorem (Birkhoff, 1946; von Neumann, 1953) and its extensions (Budish et al., 2013) ensure that, given any probabilistic assignment, we can always find a lottery over deterministic assignments that implements its marginal probabilities. Finally, we denote by X and $\Delta(X)$ the spaces of all deterministic and probabilistic assignments, respectively.

A *mechanism* is a mapping $\varphi : \mathcal{P}^N \to \Delta(X)$ that chooses an assignment based on a profile of reported preference orders. $\varphi_i(P_i, P_{-i})$ is the assignment vector that agent i receives if it reports P_i and the other agents report P_{-i}. Note that mechanisms only receive rank ordered lists as input but no additional cardinal information. Thus, we consider *ordinal* mechanisms, which determine assignments based on the ordinal preference reports alone. The expected utility for i is given by the scalar product

$$\mathbb{E}_{\varphi_i(P_i,P_{-i})}[u_i] = \langle u_i, \varphi_i(P_i, P_{-i})\rangle = \sum_{j\in M} u_i(j) \cdot \varphi_{i,j}(P_i, P_{-i}). \tag{288}$$

Finally, we define hybrid mechanisms, which we study in this paper.

Definition 28 (Hybrid)**.** For mechanisms φ, ψ and a *mixing factor* $\beta \in [0, 1]$, the *β-hybrid of φ and ψ* is given by $h^\beta = (1 - \beta)\varphi + \beta\psi$, where for all preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, the assignment $h^\beta(\boldsymbol{P})$ is the β-convex combination of the assignments of $\varphi(\boldsymbol{P})$ and $\psi(\boldsymbol{P})$.

4.4 Partially Strategyproof Hybrid Mechanisms

In this section, we first provide a short overview of the partial strategyproofness concept, which we have introduced in (Mennle and Seuken, 2015*b*). We then give our first main result, a set of conditions under which the construction of hybrid mechanisms with non-degenerate degrees of strategyproofness is possible. Subsequently, we show that none of the conditions can be dropped without losing this guarantee.

4.4.1 Full and Partial Strategyproofness

Under a strategyproof mechanism, agents have a dominant strategy to report truthfully. For random mechanisms, this means that truthful reporting of ordinal preferences maximizes any agent's expected utility, independent of the reports of the other agents and the particular utility function underlying that agent's preference order.

Definition 29 (Strategyproof)**.** A mechanism φ is *strategyproof* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any utility function $u_i \in U_{P_i}$ that is consistent with P_i, we have

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_j(P_i', P_{-i}) \rangle \geqslant 0 \tag{289}$$

This notion of strategyproofness for random mechanism coincides with the one used by Gibbard (1977) for random voting mechanisms. For deterministic mechanisms, it reduces to the requirement that no agent can obtain a strictly preferred object by misreporting. Furthermore, it is equivalent to *strong stochastic dominance-strategyproofness*, which requires that any agent's assignment from misreporting is first order-stochastically dominated by the assignment that the agent can obtain from reporting truthfully.

Partially strategyproof mechanisms (Mennle and Seuken, 2015*b*) remain strategyproof on a particular domain restriction. The agents can still have any preference order, but their underlying utility functions are constrained.

Definition 30 (Uniformly Relatively Bounded Indifference)**.** For $r \in [0, 1]$, a utility function $u \in U_P$ satisfies *uniformly relatively bounded indifference* with respect to *indifference bound* r (URBI(r)) if for any objects $a, b \in M$ with $P : a \succ b$, we have

$$r \cdot \left(u(a) - \min_{j \in M} u(j) \right) \geqslant u(b) - \min_{j \in M} u(j). \tag{290}$$

To obtain some intuition about this domain restriction, observe that a utility function $u : M \to \mathbb{R}^+$ can be interpreted as a vector in $(\mathbb{R}^+)^m$. The set U_P corresponds to a convex cone containing all the vectors for which the a-component is strictly larger than the b-component (provided $P : a \succ b$). Then the set of utility functions that satisfy URBI(r) and are consistent with P corresponds to a smaller cone inside U_P. This smaller cone is strictly bounded away from the indifference hyperplanes $H_{a,b} = \{u(a) = u(b)\}$ for any two objects $a, b \in M$. Note that the URBI(r) constraint is independent of affine transformations: if u is translated by adding a constant δ (i.e., $\tilde{u}(j) = u(j) + \delta$ for all $j \in M$), then this value will be subtracted again in (290), since $\min_{j \in M} \tilde{u}(j) = \min_{j \in M} u(j) + \delta$. If u is scaled by a factor $\alpha > 0$, then this affects both sides of (290) equally, so that the *relative* bound r is preserved.

For convenience, we denote by URBI(r) the set of all utility functions that satisfy uniformly relatively bounded indifference with respect to r. With this domain restriction, the definition of partial strategyproofness is analogous to the definition of strategyproof-

ness, except that the inequality only needs to hold for agents with utility functions in URBI(r).

Definition 31 (Partially Strategyproof)**.** For a given setting $(N, M, \boldsymbol{q})$ and $r \in [0, 1]$ we say that a mechanism φ is *r-partially strategyproof in* $(N, M, \boldsymbol{q})$ if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any utility function $u_i \in U_{P_i} \cap \text{URBI}(r)$ that is consistent with P_i and satisfies URBI(r), we have

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_j(P_i', P_{-i}) \rangle \geqslant 0. \tag{291}$$

For the remainder of the paper we will fix an arbitrary setting $(N, M, \boldsymbol{q})$. Thus, we will simply say that a mechanism is *r-partially strategyproof*, omitting the setting but keeping in mind that the value of r is specific to the respective setting.

One of the main findings in (Mennle and Seuken, 2015*b*) is that strategyproofness can be decomposed into three simple axioms, and that the set of partially strategyproof mechanisms arises by dropping the least important of these axioms. The three axioms restrict the way in which a mechanism may change an agent's assignment when this agent changes its report. For any preference order $P \in \mathcal{P}$, its *neighborhood* N_P is the set of preference orders that can be obtained by swapping two objects that are adjacent in P. Formally, for $P : a_1 \succ \ldots \succ a_m$ we have

$$N_P = \left\{ P' \in \mathcal{P} \;\middle|\; \begin{array}{l} P' : a_1 \succ \ldots \succ a_{k-1} \succ a_{k+1} \succ a_k \succ a_{k+2} \succ \ldots \succ a_m \\ \text{for some } k \in \{1, \ldots, m-1\} \end{array} \right\}. \tag{292}$$

Definition 32 (Swap Monotonic)**.** A mechanism φ is *swap monotonic* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a \succ b$ and $P_i' : b \succ a$, one of the following holds:

- either $\varphi_i(P_i, P_{-i}) = \varphi_i(P_i', P_{-i})$,
- or $\varphi_{i,a}(P_i, P_{-i}) > \varphi_{i,a}(P_i', P_{-i})$ and $\varphi_{i,b}(P_i, P_{-i}) < \varphi_{i,b}(P_i', P_{-i})$.

Definition 33 (Upper Invariant)**.** A mechanism φ is *upper invariant* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in N_{P_i}$ from the neighborhood of P_i with $P_i : a \succ b$ and $P_i' : b \succ a$, we have that i's assignment for objects from the upper contour set of a does not change (i.e., $\varphi_{i,j}(P_i, P_{-i}) = \varphi_{i,j}(P_i', P_{-i})$ for all $j \in M$ with $P_i : j \succ a$).

Definition 34 (Lower Invariant)**.** A mechanism φ is *lower invariant* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any misreport $P_i' \in N_{P_i}$ from the

neighborhood of P_i with $P_i : a \succ b$ and $P_i' : b \succ a$, we have that i's assignment for objects from the lower contour set of b does not change (i.e., $\varphi_{i,j}(P_i, P_{-i}) = \varphi_{i,j}(P_i', P_{-i})$ for all $j \in M$ with $P_i : b \succ j$).

Swap monotonicity means that if the mechanism changes an agent's assignment after this agent has swapped two adjacent objects in its report, then this change of assignment must be *direct* and *responsive*: if there is any change at all, there must be some change for the objects for which differential preferences have been reported, and this change has to be in the right direction. Upper invariance means that an agent cannot improve its chances at more preferred objects by changing the order of less preferred objects. In the presence of an outside option, this is equivalent to robustness to manipulation by truncation (Hashimoto et al., 2014). Finally, lower invariance is the natural counterpart for upper invariance. Strategyproofness decomposes into these three axioms, and partial strategyproofness arises by dropping lower invariance.

Fact 6 (Mennle and Seuken, 2015*b*)**.** *A mechanism is strategyproof* if and only if *it is swap monotonic, upper invariant, and lower invariant.*

Fact 7 (Mennle and Seuken, 2015*b*)**.** *Given a setting* $(N, M, \boldsymbol{q})$*, a mechanism is* r*-partially strategyproof for some* $r > 0$ if and only if *it is swap monotonic and upper invariant.*

Furthermore, the URBI(r) domain restriction is *maximal* in the sense that for a swap monotonic, upper invariant mechanism, there is no systematically larger set of utility functions for which we can also guarantee that truthful reporting is a dominant strategy. This allows us to define a meaningful, parametric measure for the incentive properties of non-strategyproof mechanisms.

Definition 35 (Degree of Strategyproofness)**.** Given a setting $(N, M, \boldsymbol{q})$ and a mechanism φ, the *degree of strategyproof* of φ is the largest indifference bound $r \in [0, 1]$ for which φ is r-partially strategyproof. Formally,

$$\rho_{(N,M,\boldsymbol{q})}(\varphi) = \max\{r \in [0, 1] \mid \varphi \text{ is } r\text{-partially strategyproof}\}. \tag{293}$$

By virtue of the maximality of the URBI(r) domain restriction, it is meaningful to compare mechanisms by their degree of strategyproofness. This comparison is consistent with (but not equivalent to) the comparison of mechanisms by their vulnerability to manipulation (Pathak and Sönmez, 2013). In this paper, we use the degree of strategyproofness to measure and compare mechanisms on the strategyproofness-dimension.

4.4.2 Construction of Partially Strategyproof Hybrids

To state our first main result, we define what it means for one mechanism ψ to be *weakly less varying* than another mechanism φ. This condition is part of our subsequent definition of *hybrid-admissibility*.

Definition 36 (Weakly Less Varying)**.** For mechanisms φ, ψ, we say that ψ *is weakly less varying than* φ if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, and any report $P_i' \in N_{P_i}$ from the neighborhood of P_i we have that

$$\varphi_i(P_i, P_{-i}) = \varphi_i(P_i', P_{-i}) \;\Rightarrow\; \psi_i(P_i, P_{-i}) = \psi_i(P_i', P_{-i}). \tag{294}$$

Loosely speaking, this means that the mechanism ψ (as a function of preference profiles) must be at least as *coarse* as φ. If φ does not change i's assignment, then a weakly less varying mechanism ψ must not change it either. This is important for the incentive properties of hybrids: suppose that some misreport by some agent is beneficial under ψ. If for the same misreport, φ does not change that agent's assignment, then any share of φ in the hybrid is insufficient to counteract the benefit that the agent obtains from this manipulation.

We are now ready to formulate hybrid-admissibility.

Definition 37 (Hybrid-Admissible)**.** A pair (φ, ψ) is *hybrid-admissible* if

(1) φ is strategyproof,

(2) ψ is upper invariant,

(3) ψ is weakly less varying than φ.

The following Theorem 15 is our first main result. It shows that under hybrid-admissibility, the degree of strategyproofness $\rho(h^\beta)$ of hybrid mechanisms varies in a non-degenerate fashion for varying mixing factors $\beta \in [0, 1]$.

Theorem 15. *Given a setting $(N, M, \boldsymbol{q})$, for any hybrid-admissible pair (φ, ψ) we have:*

1. for any $r < 1$ there exists a non-trivial $\beta > 0$ such that h^β is r-partially strategyproof,

2. the mapping $\beta \mapsto \rho_{(N,M,\boldsymbol{q})}(h^\beta)$ is monotonic and decreasing.

Proof Outline (formal proof in Appendix 4.C.1). Consider agent i with $P_i : a_1 \succ \ldots \succ a_m$ and a misreport $P_i : a_1 \succ \ldots \succ a_K \succ a'_{K+1} \succ \ldots \succ a'_m$, where the positions of the first K objects remain unchanged. The key insight is that we only need to consider

cases where i's assignment of a_{K+1} *strictly decreases* under φ. If i receives less of a_{K+1}, this has a negative effect on i's expected utility from reporting P_i'. We show that for utility functions in URBI(r) and sufficiently small $\beta > 0$, this negative effect suffices to make the misreport P_i' useless. Finally, $\beta > 0$ can be chosen uniformly for all preference profiles and misreports (while it may depend on the mechanisms and the setting). □

Theorem 15 confirms our intuition about the manipulability of hybrids h^β when β varies between 0 and 1. For mechanism designers, this result is good news: given any setting, a hybrid-admissible pair of mechanisms and a minimal acceptable degree of strategyproofness $\underline{\rho} \in [0, 1)$, we can always find a non-trivial hybrid (i.e., h^β with $\beta > 0$) that is $\underline{\rho}$-partially strategyproof. The fact that a *strictly positive* β can be chosen implies that any (even small) relaxation of strategyproofness can enable improvements on the efficiency-dimension.

If ψ is a more efficient mechanism, then a mechanism designer would intuitively like to choose a mixing factor as large as possible because more of the more efficient ψ would be included. In Section 4.5 we give a precise understanding of the way in which mixing affects the efficiency of hybrids, and we show that the mechanism designer's problem of determining a *maximal mixing factor* can be solved algorithmically.

4.4.3 Independence of Hybrid-Admissibility

We have seen that under hybrid-admissibility, the degree of strategyproofness of hybrid mechanisms in fact behave as our intuition suggests. Next, we show that dropping either of the three conditions from hybrid-admissibility will lead to a collapse of this guarantee.

Proposition 12. *If φ is not strategyproof, there exists a mechanism ψ that is upper invariant and weakly less varying than φ, and a bound $r \in (0, 1)$, such that no non-trivial hybrid of the pair (φ, ψ) will be r-partially strategyproof.*

Proof. Consider a constant mechanism ψ that yields the same assignment, independent of the agents' reports. If φ is manipulable by some agent i with utility u_i, we choose r such that $u_i \in$ URBI(r). Then i can manipulate any non-trivial hybrid of φ and ψ. □

Proposition 13. *For any strategyproof φ and any ψ that is weakly less varying than φ, but not upper invariant, no non-trivial hybrid of the pair (φ, ψ) is r-partially strategyproof for any bound $r \in (0, 1]$.*

Proof. Since φ is strategyproof, it must be upper invariant by Fact 6. If ψ is not upper invariant, then neither is any non-trivial hybrid of φ and ψ. Consequently, the hybrid is not r-partially strategyproof for any $r > 0$ by Fact 7. □

Proposition 14. *Let φ be a strategyproof mechanism such that for some agent $i \in N$, some preference profile (P_i, P_{-i}), and some misreport P_i', we have $\varphi_i(P_i, P_{-i}) = \varphi_i(P_i', P_{-i})$. Then there exits an upper invariant mechanism ψ such that no non-trivial hybrid of the pair (φ, ψ) is r-partially strategyproof for any bound $r \in (0, 1]$.*

Proof. Let j be the best choice object under P_i that changes position between P_i and P_i'. Then let ψ be upper invariant with $\psi_{i,j}(P_i, P_{-i}) = 0$ and $\psi_{i,j}(P_i', P_{-i}) = 1$. If $\beta > 0$, then i can manipulate the hybrid h^β in a first order-stochastic dominance sense. Therefore, h^β cannot be partially strategyproof by Proposition 2 in (Mennle and Seuken, 2015*b*). □

In combination, Propositions 12, 13, and 14 show that none of the three requirements for hybrid admissibility can be dropped, or else the relaxed incentive properties of the hybrid mechanisms may be degenerate.

In Section 4.6 we prove hybrid admissibility for pairs of Random Serial Dictatorship and Probabilistic Serial, as well as Random Serial Dictatorship with the adaptive Boston mechanisms. In contrast, for the naïve Boston mechanism and Rank Value mechanisms we show that neither is weakly less varying than Random Serial Dictatorship. Furthermore, hybrids of these mechanisms will have a degree of strategyproofness of 0 (unless $\beta = 0$). This illustrates that while hybrid admissibility is a sufficient condition, it is also close to being necessary for non-degenerate trade-offs.

4.5 Parametric Trade-offs Between Strategyproofness and Efficiency

We have obtained a good understanding of the incentive properties of hybrid mechanisms. However, ultimately, we are interested in the trade-off between strategyproofness and efficiency. To this end, we need to understand the efficiency properties of hybrids. We first review three notions of dominance, namely ex-post, ordinal, and rank dominance, and the corresponding efficiency requirements. We then show that, loosely speaking, hybrid mechanisms inherit a share of the efficiency advantages of the more efficient component, and this share is proportional to the mixing factor β.

4.5.1 Dominance and Efficiency Notions

Ex-post efficiency is ubiquitous in assignment problems. Most assignment mechanisms considered in theory and applications are ex-post efficient, such as Random Serial Dictatorship, Probabilistic Serial, Rank Value mechanisms, and variants of the Boston mechanism. Ex-post efficiency requires that *ex-post*, when every agent finally holds one object, no Pareto improvements are possible by re-assigning objects.

Definition 38 (Ex-post Efficient)**.** Given a preference profile $\boldsymbol{P} \in \mathcal{P}^N$, a deterministic assignment x *ex-post dominates* another deterministic assignment y *at* $\boldsymbol{P}$ if all agents weakly prefer their assigned object under x to their assigned object under y. The dominance is *strict* if at least one agent strictly prefers its assigned object under x. A deterministic assignment x is *ex-post efficient at* $\boldsymbol{P}$ if it is not strictly ex-post dominated by any other deterministic assignment at $\boldsymbol{P}$. Finally, a random assignment is *ex-post efficient at* $\boldsymbol{P}$ if it has a lottery decomposition consisting only of deterministic assignments that are ex-post efficient at $\boldsymbol{P}$.

To compare random assignments by their efficiency, we draw on notions of dominance for random assignments.

Definition 39 (Ordinally Efficient)**.** For a preference order $P : a_1 \succ \ldots \succ a_m$ and two assignment vectors $v = v_{j \in M}$ and $w = w_{j \in M}$, we say that v *first order-stochastically dominates* w at P if for all ranks $k \in \{1, \ldots, m\}$ we have

$$\sum_{j \in M : j \succ a_k} v_j \geqslant \sum_{j \in M : j \succ a_k} w_j. \tag{295}$$

For a preference profile $\boldsymbol{P}$, an assignment x *ordinally dominates* another assignment y *at* $\boldsymbol{P}$ if for all agents $i \in N$, the assignment vector x_i first order-stochastically dominates y_i at P_i. x *strictly* ordinally dominates y at $\boldsymbol{P}$ if in addition inequality (295) is strict for some agent $i \in N$ and some rank $k \in \{1, \ldots, m\}$. Finally, x is *ordinally efficient at* $\boldsymbol{P}$ if it is not strictly ordinally dominated by any other assignment at $\boldsymbol{P}$.

If x ordinally dominates y at $\boldsymbol{P}$ and $\boldsymbol{P}$ is the true preference profile of the agents, then all agents will prefer x to y, independent of their underlying utility functions. Bogomolnaia and Moulin (2001) showed that the Probabilistic Serial mechanism produces ordinally efficient assignments (at the reported preference profiles). Moreover, these assignments may strictly ordinally dominate the assignments obtained from Random Serial Dictatorship at the same preference profiles.

Featherstone (2011) introduced a strict refinement of ordinal efficiency, called *rank efficiency*, and he developed Rank Value mechanisms that produce rank efficient assignments.

Definition 40 (Rank Efficient)**.** For a preference profile $\boldsymbol{P}$ let $\mathrm{ch}_{P_i}(k)$ denote the kth choice object of the agent i with preference order P_i. The *rank distribution* of an assignment x at $\boldsymbol{P}$ is a vector $\boldsymbol{d}^x = (d_1^x, \ldots, d_m^x)$ with

$$d_k^x = \sum_{i \in M} x_{i,\mathrm{ch}_{P_i}(k)} \text{ for } k \in \{1, \ldots, m\}. \tag{296}$$

d_k^x is the expected number of kth choices assigned under x with respect to preference profile $\boldsymbol{P}$. An assignment x *rank dominates* another assignment y at $\boldsymbol{P}$ if $\boldsymbol{d}^x$ first order-stochastically dominates $\boldsymbol{d}^y$ (i.e., $\sum_{k=1}^{r} d_k^x - d_k^y \geqslant 0$ for all $r \in \{1, \ldots, m\}$). x *strictly rank dominates* y at $\boldsymbol{P}$ if this inequality is strict for some rank $r \in \{1, \ldots, m\}$. x is *rank efficient* at $\boldsymbol{P}$ if is not strictly rank dominated by any other assignment at $\boldsymbol{P}$.

Rank dominance captures the intuition that, for society as a whole, assigning two first choices and one second choice is preferable to assigning one first and two second choices. Rank efficient mechanisms in the assignment domain correspond to *positional scoring rules* in the social choice domain (Xia and Conitzer, 2008) because they can be interpreted as maximizing an aggregate score based on ranks (Featherstone, 2011).

4.5.2 Efficiency of Hybrid Mechanisms

Using the notions of ex-post efficiency, ordinal dominance, and rank dominance, we show that hybrids inherit a share of the good efficiency properties from the more efficient component.

Theorem 16. *Given a setting $(N, M, \boldsymbol{q})$, for any mechanisms φ and ψ, any preference profile $\boldsymbol{P} \in \mathcal{P}^N$, and any mixing factor $\beta \in [0, 1]$ the following hold:*

1. *if $\varphi(\boldsymbol{P})$ and $\psi(\boldsymbol{P})$ are ex-post efficient at $\boldsymbol{P}$, then $h^\beta(\boldsymbol{P})$ is ex-post efficient at $\boldsymbol{P}$,*
2. *$\psi(\boldsymbol{P})$ ordinally (or rank) dominates $\varphi(\boldsymbol{P})$ at $\boldsymbol{P}$* if and only if
   - *$h^\beta(\boldsymbol{P})$ ordinally (or rank) dominates $\varphi(\boldsymbol{P})$ at $\boldsymbol{P}$, and*
   - *$\psi(\boldsymbol{P})$ ordinally (or rank) dominates $h^\beta(\boldsymbol{P})$ at $\boldsymbol{P}$.*

The proof is given in Appendix 4.C.2.

Theorem 16 shows that hybrid mechanisms inherit a part of the desirable efficiency properties from their more efficient component. Statement 1 is important to ensure that the baseline requirement of ex-post efficiency is preserved. Statement 2 yields that if the component ψ is more efficient than the component φ in the sense of ordinal or rank dominance, then all hybrids will have intermediate efficiency (i.e., h^β will dominate φ but be dominated by ψ). Furthermore, it is straightforward to see that under these conditions, efficiency improves as β increases: consider two different hybrid mechanisms h^β and $h^{\beta'}$ with $\beta < \beta'$. By setting $\beta^* = \frac{\beta'-\beta}{1-\beta}$, we can write

$$h^{\beta'} = (1-\beta)\cdot\varphi + \beta\cdot\psi = (1-\beta^*)\cdot h^\beta + \beta^*\cdot\psi \tag{297}$$

as a β^*-hybrid with components h^β and ψ. Consequently, hybrids with higher mixing factors dominate hybrids with lower mixing factors.

However, not all mechanisms are comparable everywhere. For example, the Probabilistic Serial mechanism is ordinally efficient, but it does not ordinally dominate the ordinally inefficient Random Serial Dictatorship mechanism at *all* preference profiles. Instead, some assignments resulting under the two mechanisms may not be comparable by ordinal dominance. In these cases, the second direction of the equivalence in statement 2 becomes useful: when dominance does not permit a clear decision between assignments, then the hybrid will not have clearly worse efficiency than either component. Thus, intuitively, efficiency of the hybrid h^β is better than the efficiency of φ *whenever this statement is meaningful.*

4.5.3 A Parametric Measure for Efficiency Gains

Hybrid mechanisms yield a natural measure for efficiency gains, namely the mixing factor β. First, consider a preference profile $\boldsymbol{P}$ and two mechanisms φ, ψ, such that $\psi(\boldsymbol{P})$ ordinally dominates $\varphi(\boldsymbol{P})$ at $\boldsymbol{P}$. Independent of the particular vNM utility functions underlying the agents' ordinal preferences, we know that every agent has (weakly) higher expected utility under $\psi(\boldsymbol{P})$ than under $\varphi(\boldsymbol{P})$. Moreover, the agents' expected utility under the hybrid h^β is a linear function of the mixing factor because

$$\mathbb{E}_{h_i^\beta(\boldsymbol{P})}[u_i] = (1-\beta)\cdot\mathbb{E}_{\varphi_i(\boldsymbol{P})}[u_i] + \beta\cdot\mathbb{E}_{\psi_i(\boldsymbol{P})}[u_i]. \tag{298}$$

Thus, the gain in any agent's expected utility from using h^β rather than φ is exactly the β-share of the gain in the agent's expected utility from using ψ rather than φ.

Second, suppose that $\psi(\boldsymbol{P})$ rank dominates $\varphi(\boldsymbol{P})$ at $\boldsymbol{P}$. A *rank valuation* $v : \{1, \ldots, m\} \to \mathbb{R}$ with $v(k) \geqslant v(k+1)$ is a function that associates a value $v(k)$ with giving some agent its kth choice object. The *v-rank value* of an assignment $x \in \Delta(X)$ is the aggregate expected value from choosing x and it is given by

$$v(x, \boldsymbol{P}) = \sum_{k=1}^{m} d_k^x \cdot v(k). \tag{299}$$

Consequently, the v-rank value of the hybrid h^β is a linear function of the mixing factor:

$$v(h^\beta(\boldsymbol{P}), \boldsymbol{P}) = (1-\beta) \cdot v(\varphi(\boldsymbol{P}), \boldsymbol{P}) + \beta \cdot v(\psi(\boldsymbol{P}), \boldsymbol{P}). \tag{300}$$

The fact that $\psi(\boldsymbol{P})$ rank dominates $\varphi(\boldsymbol{P})$ implies that the v-rank value of $\psi(\boldsymbol{P})$ is (weakly) higher than the v-rank value of $\varphi(\boldsymbol{P})$ for any rank valuation v (Featherstone, 2011). Thus, the gain in v-rank value from using h^β rather than φ is exactly the β-share of the gain in v-rank value from using ψ rather than φ.

In combination, Theorems 15 and 16 show that parametric trade-offs between strategyproofness (measured by the degree of strategyproofness ρ) and efficiency (measured by the mixing factor β) are possible via hybrid mechanisms: when a pair of mechanisms is hybrid-admissible and the second component dominates the first, a higher mixing factor will yield hybrids that are more efficient (whenever such a statement is meaningful) but also have lower degree of strategyproofness.

4.5.4 Computability of Maximal Mixing Factor

Given our understanding of hybrids, the question arises how a mechanism designer can use this construction to perform a trade-off between strategyproofness and efficiency. Suppose that a minimal acceptable degree of strategyproofness $\underline{\rho}$ is given. Then the mechanism designer faces the computational problem of *finding the highest mixing factor β, such that h^β remains $\underline{\rho}$-partially strategyproof.* Formally, she is interested in

$$\beta^{\max}_{(N,M,\boldsymbol{q}),\varphi,\psi}(\underline{\rho}) = \max\left\{\beta \in [0,1] \mid h^\beta \text{ is } \underline{\rho}\text{-partially SP in } (N, M, \boldsymbol{q})\right\} \tag{301}$$

In (Mennle and Seuken, 2015*b*), we have shown that the degree of strategyproofness $\rho_{(N,M,\boldsymbol{q})}(\varphi)$ is computable. Thus, we have a solution to the problem of "finding $\rho(h^\beta)$, given β." However, the mechanism designer's problem is the *inverse* of this problem,

ALGORITHM 3: BETAMAX

Input: setting $(N, M, \boldsymbol{q})$, mechanisms φ, ψ, bound $\underline{\rho}$
Variables: agent i, preference profile (P_i, P_{-i}), misreport P_i', vectors $\delta^\varphi, \delta^\psi$, rank K, choice function ch, real values $\beta^{\max}, p_K^\varphi, p_K^\psi$

```
begin
    β^max ← 1
    for i ∈ N, (P_i, P_{-i}) ∈ 𝒫^N, P_i' ∈ 𝒫 do
        ∀j ∈ M : δ_j^φ ← φ_{i,j}(P_i, P_{-i}) − φ_{i,j}(P_i', P_{-i})
        ∀j ∈ M : δ_j^ψ ← ψ_{i,j}(P_i, P_{-i}) − ψ_{i,j}(P_i', P_{-i})
        for K ∈ {1, ..., m} do
            p_K^φ ← Σ_{k=1}^{K} δ^φ_{ch_{P_i}(k)} · ρ̲^k
            p_K^ψ ← Σ_{k=1}^{K} δ^ψ_{ch_{P_i}(k)} · ρ̲^k
            if p_K^ψ < 0 then
                β^max ← min{β^max, p_K^φ / (p_K^φ − p_K^ψ)}
            end
        end
    end
    return β^max
end
```

namely to "find β, given $\underline{\rho}$." The following algorithm solves this problem.

Algorithm 3 optimistically sets its guess of $\beta^{\max}$ to 1. Then it iterates through all possible preference profiles, all agents, and all misreports that agents may submit. For each of these combinations, it uses the partial dominance interpretation of partial strategyproofness (Theorem 4 in (Mennle and Seuken, 2015*b*)) to determine whether the current guess is too high, and the value is adjusted downward if necessary.

Proposition 15. *Given a setting $(N, M, \boldsymbol{q})$, a hybrid-admissible pair of mechanisms (φ, ψ), and a bound $\underline{\rho} \in [0, 1]$ Algorithm 3 (*BETAMAX*) is complete and correct for the mechanism designer's problem of finding the maximal mixing factor $\beta^{max}_{(N,M,\boldsymbol{q}),\varphi,\psi}(\underline{\rho})$.*

The proof is given in Appendix 4.C.3.

Computational Cost of BetaMax

Note that our main goal is to show *computability*, not computational efficiency. Nonetheless, we can make a statement about the computational cost of running BETAMAX: computing the random assignments from the mechanisms φ and ψ may itself be a

costly operation.[4] Thus, if $O(\varphi)$ and $O(\psi)$ denote the cost of determining φ and ψ for a single preferences profile, respectively, then the overall cost of Algorithm 3 is $O\left(n \cdot m \cdot (m!)^{n+1} \left(O(\varphi) + O(\psi)\right)\right)$.

In the most general case (i.e., without any additional restrictions), a mechanism is specified in terms of a set of assignment matrices $\{\varphi(\boldsymbol{P}), \boldsymbol{P} \in \mathcal{P}^N\}$. This set will contain $(m!)^n$ matrices of dimension $n \times m$. Consequently, the size of the problem is $S = (m!)^n \cdot n \cdot m$. In terms of S, Algorithm 3 has complexity $O\left(S \sqrt[n]{S}\right)$. Thus, for the general case, there is not much room for improvement: since the algorithm must consider each preference profile at least once, *any* correct and complete algorithm has computational cost of at least S.

Reductions of Computational Cost

Reductions of the computational complexity are possible if more information is available about the mechanisms φ and ψ. For anonymous φ and ψ, the identities of the agents is irrelevant. In this case, the computational cost can be reduced to

$$O\left(n \cdot m! \binom{m! + n - 1}{n} \left(O(\varphi) + O(\psi)\right)\right), \tag{302}$$

because only $\binom{m! + n - 1}{n}$ preference profiles must be considered. Moreover, if the mechanisms are also neutral (i.e., the assignment does not depend on the objects' names either), then it suffices to consider only agent 1 with a fixed preference order. With this, the computational cost can be further reduced to

$$O\left(m! \binom{m! + n - 2}{n - 1} \left(O(\varphi) + O(\psi)\right)\right). \tag{303}$$

Even with these reductions, running Algorithm 3 is costly for larger settings. However, it is likely that more efficient algorithms exist for mechanisms with additional restrictions, and bounds may be derived analytically for certain interesting mechanisms, such as Probabilistic Serial. Having shown computability, we leave the design of computationally more efficient algorithms to future research.

[4] Determining the probabilistic assignment of a mechanism may be computationally hard, even if implementing the mechanism is easy (e.g., see (Aziz, Brandt and Brill, 2013*b*)).

4.6 Applications to Popular Mechanisms

So far, we have considered abstract hybrid mechanisms and we have derived general results. In this section, we consider concrete instantiations of our construction. Indeed, it is applicable to some (but not all) well-known mechanisms. $\varphi = \text{RSD}$ is a canonical choice because it is the only known mechanism that is strategyproof, ex-post efficient, and anonymous. In order to apply Theorem 15 (for the construction of partially strategyproof hybrids), we must establish two requirements for the second component: ψ must be upper invariant, and ψ must be weakly less varying than RSD. Furthermore, to obtain efficiency gains, ψ must be more efficient than RSD in some sense. Table 4.1 provides an overview of our results. Trade-offs for ordinal dominance can be achieved via hybrids of RSD and PS, and trade-offs for rank dominance are possible via hybrids of RSD and ABM. However, NBM and RV both violate hybrid-admissibility (in combination with RSD), and we find that in fact they do not admit a non-degenerate trade-off.[5]

φ	ψ	**Dominance**	**UI**	**WLV**	h^β **PSP**
RSD	PS	Ordinal	✓	✓	✓
RSD	RV	Rank	✗	✗	✗
RSD	NBM	Rank	✓	✗	✗
RSD	ABM	Rank (with exceptions)	✓	✓	✓

Table 4.1: Results overview, *UI*: ψ upper invariance, *WLV*: ψ weakly less varying than φ, *PSP*: r-partially strategyproof for some $r > 0$.

4.6.1 Hybrids of RSD and PS

By Theorem 2 of Hashimoto et al. (2014), PS is upper invariant. Since PS is ordinally efficient, it is never ordinally dominated by RSD at any preference profile. Furthermore, PS may (but does not always) ordinally dominate RSD (Bogomolnaia and Moulin, 2001). Thus, PS ordinally dominates RSD whenever the two mechanisms are comparable. To obtain hybrid-admissibility of the pair (RSD, PS), it remains to be shown that PS is weakly less varying than RSD.

Theorem 17. *PS is weakly less varying than RSD.*

[5] We provide short descriptions of the mechanisms RSD, PS, NBM, ABM, and RV in Appendix 4.A.

Proof Outline (formal proof in Appendix 4.C.4). Consider an agent i that swaps two objects in its report from $P_i : a \succ b$ to $P_i' : b \succ a$. First, we show that PS changes the assignment *if and only if* neither a nor b are exhausted when i finishes consuming objects that it strictly prefers to both. Next, we show that RSD changes the assignment *if and only if* there exists an ordering of the agents such that all objects that i prefers strictly to a are assigned before i gets to pick, but neither a nor b are assigned by then. Finally, we show that the first condition (for PS) implies the second condition (for RSD), using an inductive argument. The key idea is to show that if no such ordering of the agents exists for m objects, then we can construct a case with $m - 1$ objects where no such ordering exists either. □

Corollary 5. *The pair (RSD,PS) admits the construction of partially strategyproof hybrids that improve efficiency in terms of ordinal dominance.*

4.6.2 Two Impossibility Results

A mechanism designer may also want to trade off strategyproofness for improvements of the rank distribution. Mechanisms that aim at achieving a good rank distribution are Rank Value mechanisms (Featherstone, 2011) and Boston mechanisms (Mennle and Seuken, 2015*d*). It turns out, however, that neither RV nor the naïve variant of the Boston mechanism (NBM) are suitable second components in combination with RSD.

Impossibility Result for Rank Value Mechanisms

RV rank dominates RSD whenever their outcomes are comparable, since RV is rank efficient, but RSD is not. However, no rank efficient mechanism can be upper invariant, as we demonstrate in Example 13. Therefore, the pair (RSD, RV) violates hybrid-admissibility.[6]

Example 13. Consider a setting with agents $N = \{1, 2, 3\}$ and objects $M = \{a, b, c\}$, each available in unit capacity. If the agents have preferences

$$\begin{aligned} P_1, P_2 &: a \succ b \succ c, \\ P_3 &: c \succ a \succ b, \end{aligned}$$

[6] In addition to violating upper invariance, RV is not weakly less varying than RSD in general; see Example 15 in Appendix 4.B.

then any rank efficient assignment must assign c to agent 3. Therefore, at least one of the agents 1 and 2 has a positive probability for b. Without loss of generality, let $\mathrm{RV}_{1,b}(\boldsymbol{P}) > 0$. If agent 1 reports

$$P_1' \quad : \quad a \succ c \succ b$$

instead, then the unique rank efficient assignment must assign a to agent 1. Since this misreport changes agent 1's assignment of a, RV is not upper invariant.

It follows from Example 13 that any non-trivial hybrid h^β of RSD and RV will violate upper invariance. This means that h^β will not be r-partially strategyproof for any positive $r > 0$ (by Fact 7), or equivalently, h^β will have a degree of strategyproofness of 0. This teaches us that RSD and RV indeed do not admit the construction of hybrid mechanisms that make a non-degenerate trade-off between strategyproofness and efficiency.

Impossibility Result for the Naïve Boston Mechanism

We consider the Boston mechanism with no priorities and single uniform tie-breaking (Miralles, 2008). The "naïve" variant of the Boston mechanism (*NBM*) lets agents apply to their respective next best choices in consecutive rounds, even if the objects to which they apply have no more remaining capacity. The assignments from NBM rank dominate those from RSD whenever they are comparable, and NBM is also upper invariant (Mennle and Seuken, 2015*d*). However, NBM is *not* weakly less varying than RSD, as Example 14 shows. Thus, pairs of RSD and NBM violate hybrid-admissibility.

Example 14. Consider a setting with agents $N = \{1, \ldots, 6\}$, objects $M = \{a, \ldots, f\}$, each available in unit capacity. Let the agents have preferences

$$\begin{aligned} P_1, P_2 \quad &: \quad a \succ b \succ c \succ d \succ e \succ f \\ P_3, P_4, P_5, P_6 \quad &: \quad c \succ b \succ f \succ d \succ a \succ e. \end{aligned}$$

Under RSD, agent 1's assignment is $\left(\frac{1}{2}, \frac{1}{10}, 0, \frac{7}{30}, \frac{1}{6}, 0\right)$ of objects a through f, respectively. Swapping c and d in its report will not change its assignment under RSD. Under NBM and truthful reporting, the assignment is the same as under RSD. But if agent 1 changes its report by swapping c and d, its assignment under NBM changes to $\left(\frac{1}{2}, \frac{1}{10}, 0, \frac{2}{5}, 0, 0\right)$. It strictly prefers this assignment in a first order-stochastic dominance sense.

In fact, Example 14 shows something more, namely that at the particular preference profile, NBM is manipulable in a first order-stochastic dominance sense, while RSD

does not change the assignment at all. Thus, any hybrid of RSD and NBM will also be manipulable in a first order-stochastic dominance sense at this preference profile. Consequently, the hybrid cannot be r-partially strategyproof for any $r > 0$ (by Proposition 2 in (Mennle and Seuken, 2015*b*)). Analogous to the pair (RSD, RV), we learn that the pair (RSD, NBM) does not admit the construction of hybrid mechanisms with non-degenerate degrees of strategyproofness either.

4.6.3 Hybrids of RSD and ABM

In (Mennle and Seuken, 2015*d*), we have formalized an adaptive variant of the Boston mechanism (ABM), which is sensitive to the fact that agents cannot benefit from applying to objects that were already exhausted in previous rounds. Instead, in each round, agents who have not been assigned so far, apply to their most preferred *available* object.

Our analysis of ABM in (Mennle and Seuken, 2015*d*) has revealed two further attributes: first, ABM is upper invariant, one of the conditions we need for hybrid-admissibility. Second, ABM rank dominates RSD whenever the two mechanisms are comparable, except in certain special cases. These exceptions occur rarely, and the probability of encountering them vanishes as markets get large. Thus, if we can show that ABM is also weakly less varying than RSD, then we can use this pair to construct partially strategyproof hybrids that trade off strategyproofness and efficiency in terms of rank dominance (with the exception of a tiny number of preference profiles).

Theorem 18. *ABM is weakly less varying than RSD.*

Proof outline (formal proof in Appendix 4.C.5). Both RSD and ABM are implemented by randomizing over orderings π of agents. Suppose i manipulates by swapping a and b. If ABM changes the assignment, then there exists π such that all objects that i prefers strictly to a and b are assigned in previous rounds to other agents. Then i gets to "pick" between a and b under ABM. Starting with π, we construct an ordering π' such that if the ordering π' is drawn under RSD, i gets to pick between a and b but no object it strictly prefers to a or b under RSD. This is sufficient for RSD to also change the assignment. □

Corollary 6. *The pair (RSD,ABM) admits the construction of partially strategyproof hybrids that improve efficiency in terms of rank dominate (with few exceptions).*

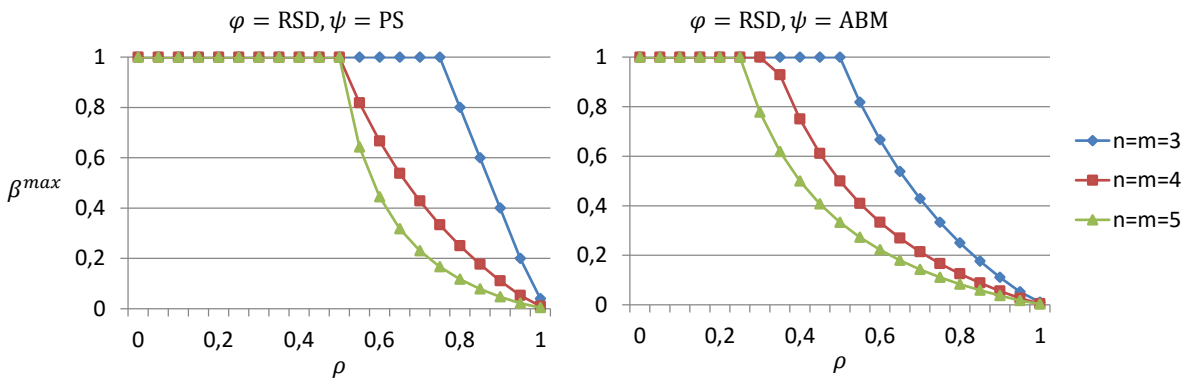

Figure 4.1: Plots of $\beta^{\max}$ by acceptable degree of strategyproofness $\rho \in [0, 1]$, for components $\varphi = \text{RSD}, \psi \in \{\text{PS}, \text{ABM}\}$, and settings with $n = m \in \{3, 4, 5\}$.

4.6.4 Numerical Results

We have shown that we can construct interesting hybrids by combining RSD with PS or ABM. This gives mechanism designers the possibility to trade off strategyproofness for better efficiency. To illustrate the magnitude of these trade-offs, we have computed $\beta^{\max}$ for a variety of settings $(N, M, \boldsymbol{q})$ and acceptable degrees of strategyproofness $\rho \in [0, 1]$.

Figure 4.1 shows plots of the maximal mixing factor $\beta^{\max}$ for settings with unit capacity and different numbers of objects and agents. Observe that as the acceptable degree of strategyproofness for the hybrid increases, the allowable share of ψ decreases and becomes 0 if full strategyproofness is required. We also see that the relationship between ρ and $\beta^{\max}$ is not linear. In particular, the first efficiency improvements (from $\beta^{\max} = 0$ to $\beta^{\max} > 0$) are the most "costly" in terms of a reduction of the degree of strategyproofness ρ. On the other hand, for mild strategyproofness requirements, the share of PS or ABM in the hybrid can be significant, e.g., more than 30% of PS or 17% of ABM for $\rho = 0.75$ and $n = m = 4$.

Figure 4.2 shows plots of $\beta^{\max}$ for hybrids of RSD and PS, where we hold the number of objects constant at $m = 3$ but vary the capacity of the objects $q \in \{2, 3, 4\}$ (with $n = q \cdot m$ agents). We observe that for larger capacities, the hybrids can contain a larger share of PS. This is consistent with findings by Kojima and Manea (2010), who have shown that for a fixed agent and a fixed number of objects, PS makes truthful reporting a dominant strategy if the capacities of the objects are sufficiently high. It is conceivable that the degree of strategyproofness of PS keeps increasing and converges to 1 in the

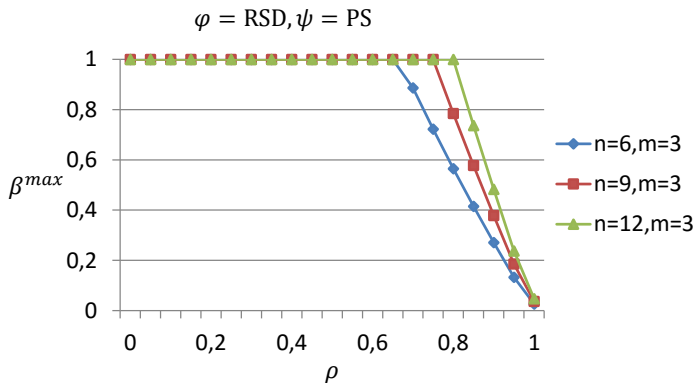

Figure 4.2: Plots of $\beta^{\max}$ by acceptable degree of strategyproofness $\rho \in [0, 1]$, for components $\varphi = \text{RSD}, \psi = \text{PS}$, and settings with $m = 3, n \in \{6, 9, 12\}$.

limit as capacity increases, an interesting question for future research.

4.7 Conclusion

In this paper, we have presented a novel approach to trading off strategyproofness and efficiency for random assignment mechanisms. We have introduced hybrid mechanisms, which are convex combinations of two component mechanisms, as a method to facilitate these trade-offs. Typically, the first component φ introduces better incentives while the second component ψ introduces better efficiency.

For our first result, we have employed partial strategyproofness, a new concept for relaxing strategyproofness in a parametric way that we have introduced in (Mennle and Seuken, 2015*b*). If (1) φ is strategyproof and (2) ψ is upper invariant and (3) weakly less varying than φ, we have shown that partially strategyproof hybrids can be constructed for any desired degree of strategyproofness. At the same time, our hybrid-admissibility requirement is tight in the sense that none of the three conditions can be dropped without risking degenerate trade-offs.

For our second result, we have shown that hybrids inherit ex-post efficiency from their components, and their efficiency (relative to the components) can be understood in terms of ordinal (or rank) dominance. This means that, in line with intuition, hybrid mechanisms in fact trade off strategyproofness for efficiency: as the mixing factor β (i.e., the share of ψ) increases, efficiency of the hybrid increases, but the degree of strategyproofness decreases. This has important consequences for mechanism designers: if φ is a strategyproof mechanism, ψ is a non-strategyproof alternative that is more

appealing due to its efficiency properties, and a certain degree of strategyproofness $\underline{\rho} < 1$ is acceptable, then a hybrid can be used to improve efficiency, subject to the $\underline{\rho}$-partial strategyproofness constraint. As we have shown in Section 4.5.4, the mechanism designer's problem of determining the maximal mixing factor can be solved algorithmically.

Finally, we have presented instantiations of hybrid mechanisms with $\varphi = \text{RSD}$ as the strategyproof component. Using $\psi = \text{PS}$ yields better efficiency in an ordinal dominance sense, and using $\psi = \text{ABM}$, an adaptive variant of the Boston mechanism, yields better efficiency in a rank dominance sense (with few exceptions). Numerically, we have illustrated the connection between the degree of strategyproofness ρ and the maximal mixing factor $\beta^{\max}$, and we have shown that the latter can be significant for even mild reductions of the minimal acceptable degree of strategyproofness.

This paper contributes to an important area of research concerned with trade-offs between strategyproofness and efficiency in the assignment domain. Hybrid mechanisms break new ground because the method is constructive, it enables a parametric trade-off, and the mechanism designer's problem of determining a suitable hybrid is computable. Our hybrids shed light on the frontiers of such trade-offs and can serve as benchmark mechanisms for future research.

Appendix for Chapter 4

4.A Mechanisms

We explain how each mechanism determines the assignment based on a reported profile $\boldsymbol{P}$ of preferences.

4.A.1 Random Serial Dictatorship Mechanism

The *Random Serial Dictatorship* mechanism chooses an agent uniformly at random and assigns this agent its first choice object. Next, it chooses another agent uniformly at random from the remaining agent and assigns this agent the object that it prefers most out of all the objects that have remaining capacity. This continues until all agents have received an object. The random assignment matrix arises from the fact that agents do not know when they will be chosen by the mechanism.

4.A.2 Probabilistic Serial Mechanism

Under the *Probabilistic Serial* mechanism, the objects are treated as if they were divisible. All agents start consuming probability shares of their first choice objects at equal speeds. Once all capacity of an object is completely consumed, all agents who were consuming this object, move on to their next preferred object. If this next object is already exhausted as well, they go directly to the next object, and so on. This process continues until all agents have collected a total of 1 units of some objects. The shares of objects that each agent has collected are the entries in the assignment matrix of the Probabilistic Serial mechanism.

4.A.3 Naïve Boston Mechanism

Under the *naïve Boston* mechanism, all agents report their preferences and then draw a random number. The assignment process occurs in rounds. In the first round, each agent

applies to its most preferred object. Applicants are assigned the objects to which they applied if these have sufficient capacity. If an object has more applicants than remaining capacity, preference is give to agents with higher random numbers. The agents who did not get an object in the first round continue to the second round. In the kth round, each remaining agent applies to its kth choice. Again, objects are assigned to agents until their capacity is exhausted, and the unlucky agents with the lowest random numbers enter the next round. The assignment process ends when all agents have received an object. The random assignment matrix arises from the fact that agents do not know their random numbers.

4.A.4 Adaptive Boston Mechanism

The *adaptive Boston* mechanism works like the naïve Boston mechanism, except that in each round, the remaining agents apply to the object that they prefer most out of all the objects that still have remaining capacity. Again, the random assignment matrix arises from the fact that agents do not know their random numbers.

4.A.5 Rank Value Mechanism

Rank Value mechanisms are a class of mechanisms. Given a rank valuation $v : \{1, \dots, m\} \to \mathbb{R}$ with $v(k) \geqslant v(k+1)$, a v-Rank Value mechanism determines an assignment by solving the following linear program:

$$\begin{aligned}
\text{maximize} \quad & \sum_{i \in N} \sum_{j \in M} v(\text{rank}_{P_i}(j)) \cdot x_{i,j}, \\
\text{subject to} \quad & \sum_{i \in N} x_{i,j} = 1, \text{ for all } j \in M, \\
& \sum_{j \in M} x_{i,j} \leqslant q_j, \text{ for all } i \in N, \\
& x_{i,j} \in [0,1], \text{ for all } i \in N, j \in M,
\end{aligned}$$

where $\text{rank}_{P_i}(j)$ is the rank of j under the preference ranking of agent i, i.e., the number of objects that this agent weakly prefers to j.

4.B Example from Section 4.6.2

Example 15 (RV not Weakly Less Varying than RSD)**.** Consider a setting $N = \{1, \ldots, 3\}, M = \{a, b, c\}, q_a = q_b = q_c = 1$. For the preference profile

$$\begin{aligned} P_1 &: a \succ b \succ c, \\ P_2, P_3 &: c \succ a \succ b, \end{aligned}$$

suppose the rank valuation is $v(1) = 10, v(2) = 6, v(3) = 0$. Then RV will assign b to agent 1 with certainty. To see this suppose that agent 1 gets a instead. Then some other agent i received b. If agent 1 and agent i trade, the objective increases by $6-10+6-0 = 2$. Now suppose that agent 1 gets c. Again some agent i gets object a. If agent 1 and agent i trade, this improves the objective by $10 - 0 + 6 - 10 = 6$. We have argued that agent 1 will get b in any deterministic assignment chosen by RV with rank valuation v. Then by definition, agent 1 must get b with certainty.

Suppose now that agent 1 reports

$$P_1' : a \succ c \succ b$$

instead, i.e., it swaps objects b and c in its report. Then under any rank efficient assignment (with respect to (P_1', P_{-1})), agent 1 will receive object a. This is because whenever agent 1 gets another object in some deterministic assignment, the objective improves if agent 1 trades with the agent who received a (independent of v). Since no rank efficient assignment will give agent 1 any other object than a, swapping b and c in its report is a beneficial manipulation for agent 1. This is independent of its actual utility, as long as the utility is consistent with P_1.

Now consider the outcome of RSD: it is easy to see that for any ordering of the agents, if agent 1 does not receive a when it gets to choose, object c will not be available. Therefore, $\text{RSD}_1(P_1, P_{-1}) = \text{RSD}_1(P_1', P_{-1})$, i.e., RSD does not change the assignment of agent 1. This means that RV with the specific choice of rank valuation v is not weakly less varying than RSD, and agent 1 in the given situation would want to manipulate any non-trivial hybrid of RSD and RV.

4.C Omitted Proofs

4.C.1 Proof of Theorem 15

Proof of Theorem 15. Given a setting $(N, M, \boldsymbol{q})$, for any hybrid-admissible pair (φ, ψ) we have:

1. *for any $r < 1$ there exists a non-trivial $\beta > 0$ such that h^β is r-partially strategyproof,*
2. *the mapping $\beta \mapsto \rho_{(N,M,\boldsymbol{q})}(h^\beta)$ is monotonic and decreasing.*

To see statement 2, fix an agent $i \in N$, a preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, a misreport $P_i' \in \mathcal{P}$, and a utility function $u_i \in U_{P_i}$. If for any $\beta \in [0, 1]$, the hybrid h^β is manipulable for i in this situation, then

$$\left\langle u_i, h_i^\beta(P_i, P_{-i}) - h_i^\beta(P_i', P_{-i}) \right\rangle < 0. \tag{304}$$

By linearity, we can decompose the left side to

$$\begin{aligned} & \left\langle u_i, h_i^\beta(P_i, P_{-i}) - h_i^\beta(P_i', P_{-i}) \right\rangle & (305) \\ = \; & (1-\beta) \left\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_i(P_i', P_{-i}) \right\rangle & (306) \\ & + \beta \left\langle u_i, \psi_i(P_i, P_{-i}) - \psi_i(P_i', P_{-i}) \right\rangle. & (307) \end{aligned}$$

The first part (with factor $(1-\beta)$ must be non-negative by strategyproofness of φ. Thus, $\langle u_i, \psi_i(P_i, P_{-i}) - \psi_i(P_i', P_{-i}) \rangle < 0$. This implies that for any $\beta \in [0, B]$, agent i in this fixed situation will prefer truthful reporting to misreporting P_i', and for any $\beta \in (B, 1]$, it will strictly prefer misreporting P_i'. Consequently, the set of utility functions, for which the hybrid h^β makes truthful reporting a dominant strategy shrinks as β increases. Therefore, the maximal bound r for which we can guarantee truthful reporting to be a dominant strategy for any agent with utility in URBI(r) also shrinks. This implies that the mapping $\beta \mapsto \rho(h^\beta)$ is monotonic and decreasing.

The proof for statement 1 is more challenging. Consider a strategyproof mechanism φ and a weakly less varying, upper invariant mechanism ψ, a fixed setting $(N, M, \boldsymbol{q})$, and a fixed bound $r < 1$. We must find a mixing factor $\beta > 0$ such that no agent with a utility satisfying URBI(r) will find a beneficial manipulation to the hybrid h^β.

Let $\boldsymbol{P} = (P_i, P_{-i}) \in \mathcal{P}^N$ be a preference profile, $u_i \in U_{P_i}$ a utility function for agent i,

and let $P_i' \in \mathcal{P}$ be a potential misreport, where

$$P_i : a_1 \succ \ldots \succ a_m. \tag{308}$$

Suppose that ψ changes the assignment for i (otherwise the incentive constraint for the hybrid mechanism is trivially satisfied for this preference profile and misreport by strategyproofness of φ). By Lemma 5, there exists a rank $L \in \{1, \ldots, m-1\}$ such that the gain in expected utility from reporting P_i' instead of P_i under ψ is upper-bounded by

$$\langle u_i, \psi_i(P_i', P_{-i}) - \psi_i(P_i, P_{-i})\rangle \leqslant u_i(a_L) - u_i(a_m), \tag{309}$$

and the utility gain from reporting P_i truthfully instead of the misreport P_i' under φ is lower-bounded by

$$\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_i(P_i', P_{-i})\rangle \geqslant \varepsilon \cdot (u_i(a_L) - u_i(a_{L+1}))\,, \tag{310}$$

where $\varepsilon > 0$ depends only on the setting and the mechanism φ. Thus, the utility gain from reporting P_i truthfully instead of the misreport P_i' under the hybrid h^β is lower bounded by

$$\begin{aligned}
& \left\langle u_i, h_i^\beta(P_i, P_{-i}) - h_i^\beta(P_i', P_{-i})\right\rangle && (311)\\
= \; & (1-\beta)\langle u_i, \varphi_i(P_i, P_{-i}) - \varphi_i(P_i', P_{-i})\rangle && (312)\\
& +\beta\langle u_i, \psi_i(P_i, P_{-i}) - \psi_i(P_i', P_{-i})\rangle && (313)\\
\geqslant \; & (1-\beta)\varepsilon\,(u_i(a_L) - u_i(a_{L+1})) - \beta\,(u_i(a_L) - u_i(a_m)) && (314)\\
= \; & (u_i(a_L) - u_i(a_m))\,(\varepsilon(1-\beta) - \beta) && (315)\\
& -(u_i(a_{L+1}) - u_i(a_m))\,(\varepsilon(1-\beta))\,. && (316)
\end{aligned}$$

If u_i satisfies URBI(r), we can lower bound the difference $u_i(a_L) - u_i(a_m)$ by $r\,(u_i(a_{L+1}) - u_i(a_m))$ and get

$$\begin{aligned}
& \left\langle u_i, h_i^\beta(P_i, P_{-i}) - h_i^\beta(P_i', P_{-i})\right\rangle && (317)\\
\geqslant \; & \frac{u_i(a_{L+1}) - u_i(a_m)}{r}\,(\varepsilon(1-\beta) - \beta - r\varepsilon(1-\beta))\,. && (318)
\end{aligned}$$

Since $\frac{u_i(a_{L+1})-u_i(a_m)}{r} \geqslant 0$, this is positive if and only if

$$\varepsilon(1-\beta)-\beta-r\varepsilon(1-\beta) \geqslant 0 \quad \Leftrightarrow \quad \beta \leqslant \frac{\varepsilon(1-r)}{\varepsilon(1-r)+1}. \tag{319}$$

This upper bound for β is strictly positive and independent of the specific utility function u_i, the preference profile (P_i, P_{-i}), and the misreport P_i'. Therefore, h^β is r-partially strategyproof if β is chosen to satisfy (319).

Lemma 5. *Consider a setting $(N, M, \boldsymbol{q})$, a strategyproof mechanism φ, a weakly less varying, upper invariant mechanism ψ, an agent $i \in N$, a preference profile $\boldsymbol{P} = (P_i, P_{-i}) \in \mathcal{P}^N$, a misreport $P_i' \in \mathcal{P}$, and a utility function $u_i \in U_{P_i}$. If $\varphi_i(P_i, P_{-i}) \neq \varphi_i(P_i', P_{-i})$, then there exists $L \in \{1, \ldots, m-1\}$ such that the gain in expected utility from reporting P_i' instead of P_i under ψ is upper-bounded by*

$$u_i(a_L) - u_i(a_m), \tag{320}$$

and the gain in expected utility from reporting P_i truthfully instead of P_i' under φ is lower-bounded by

$$\varepsilon\left(u_i(a_L) - u_i(a_{L+1})\right), \tag{321}$$

where $\varepsilon > 0$ depends only on the setting and the mechanism φ.

Proof. We first introduce the auxiliary concept of the canonical transition. Consider two preference orders P and P'. A *transition* from P to P' is a sequence of preference orders $\tau(P, P') = (P^0, \ldots, P^S)$ such that

- $P^0 = P$ and $P' = P^S$,
- $P^{k+1} \in N_{P^k}$ for all $k \in \{0, \ldots, S-1\}$,

where N_P is the neighborhood of preference order P. A transition can be interpreted as a sequence of swaps of adjacent objects that transform one preference order into another if applied in order. Suppose,

$$P' : a_1 \succ a_2 \succ \ldots \succ a_m. \tag{322}$$

Then the *canonical transition* is the transition that results from starting at P and swapping a_1 (which may not be in first position for P) up until it is in first position. Then do the same for a_2, until it is in second position, and so on, until P' is obtained.

Suppose P_i corresponds to the preference ordering

$$P_i \quad : \quad a_1 \succ \ldots \succ a_{L-1} \succ a_L \succ \ldots \succ a_m,$$

and let a_L be the best choice object (under P_i) for which the assignment under φ changes, i.e.,

$$\varphi_{i,a_k}(P_i, P_{-i}) = \varphi_{i,a_k}(P_i', P_{-i}) \text{ for } k < L, \quad \varphi_{i,a_L}(P_i, P_{-i}) \neq \varphi_{i,a_L}(P_i', P_{-i}). \tag{323}$$

Consider the canonical transition from P_i' to P_i. This will bring the objects $a_k, k < L$ into position (as they are under P_i) first. By Theorem 1 in (Mennle and Seuken, 2015*b*) and because φ is strategyproof, the assignment for each of these objects can only weakly increase or weakly decrease. However, by (323) their assignments remain unchanged. Therefore, the assignment does not change for any of the swaps that bring the objects $a_k, k < L$ into position. Using that ψ is weakly less varying than φ, we can assume that

$$P_i' \quad : \quad a_1 \succ \ldots \succ a_{L-1} \succ a_L' \succ \ldots \succ a_m'$$

without loss of generality.

By upper invariance of ψ, the highest gain the agent could obtain from reporting P_i' instead of P_i arises if all probability for its last choice is converted to probability for the best choice for which the assignment can change at all, i.e., a_L. Thus, the utility gain is bounded by

$$u_i(a_L) - u_i(a_m). \tag{324}$$

Let

$$\varepsilon = \min \left\{ |\varphi_{i,j}(P_i, P_{-i}) - \varphi_{i,j}(P_i', P_{-i})| \;\middle|\; \begin{array}{c} j \in M, i \in N, \\ (P_i, P_{-i}) \in \mathcal{P}^N, P_i' \in \mathcal{P} : \\ \varphi_{i,j}(P_i, P_{-i}) \neq \varphi_{i,j}(P_i', P_{-i}) \end{array} \right\} \tag{325}$$

be the smallest positive amount by which the assignment of some object to some agent can change upon a change of report by that agent under φ. In the canonical transition from P_i to P_i', the object a_L will only be swapped downwards, i.e., its assignment can not increase in any step. But since we assumed that it changes, it must strictly decrease. This decrease has at least magnitude ε by definition. Thus, when misreporting, the agent looses at least ε probability for a_L in some swap. From Theorem 1 in (Mennle and Seuken, 2015*b*) we know that the assignment for the other object involved in that swap

must strictly increase by the same amount ε. Since all other swaps reverse the order of objects from "right" (as under P_i) to "wrong" (as under P_i'), the assignment can only get weakly worse for the agent. Therefore, the gain from reporting P_i truthfully instead of P_i' under φ is at least $\varepsilon\,(u_i(a_L) - u_i(a_{L+1}))$. □

This completes the proof of Theorem 15 □

4.C.2 Proof of Theorem 16

Proof of Theorem 16. Given a setting $(N, M, \boldsymbol{q})$, for any mechanisms φ and ψ, any preference profile $\boldsymbol{P} \in \mathcal{P}^N$, and any mixing factor $\beta \in [0,1]$ the following hold:

1. *if $\varphi(\boldsymbol{P})$ and $\psi(\boldsymbol{P})$ are ex-post efficient at $\boldsymbol{P}$, then $h^\beta(\boldsymbol{P})$ is ex-post efficient at $\boldsymbol{P}$,*
2. *$\psi(\boldsymbol{P})$ ordinally (or rank) dominates $\varphi(\boldsymbol{P})$ at $\boldsymbol{P}$* if and only if
   - *$h^\beta(\boldsymbol{P})$ ordinally (or rank) dominates $\varphi(\boldsymbol{P})$ at $\boldsymbol{P}$, and*
   - *$\psi(\boldsymbol{P})$ ordinally (or rank) dominates $h^\beta(\boldsymbol{P})$ at $\boldsymbol{P}$.*

To see statement 1, note that $h^\beta(\boldsymbol{P})$ is a convex combination of (and therefore a lottery over) the assignments $\varphi(\boldsymbol{P})$ and $\psi(\boldsymbol{P})$. Since both are ex-post efficient, each has a lottery decomposition into ex-post efficient, deterministic assignments. Therefore, we can construct a lottery decomposition of $h^\beta(\boldsymbol{P})$ into ex-post efficient, deterministic assignments by combining the two lotteries. This shows ex-post efficiency of $h^\beta(\boldsymbol{P})$ at $\boldsymbol{P}$.

Suppose, $\psi(\boldsymbol{P})$ ordinally dominates $\varphi(\boldsymbol{P})$, i.e., for all $i \in N$ and all $j \in M$ we have

$$\sum_{j' \in M:\ P_i : j' \geq j} \varphi_{i,j}(\boldsymbol{P}) \leqslant \sum_{j' \in M:\ P_i : j' \geq j} \psi_{i,j}(\boldsymbol{P}). \tag{326}$$

With $h^\beta(\boldsymbol{P}) = (1-\beta)\varphi(\boldsymbol{P}) + \beta\psi(\boldsymbol{P})$ it follows directly that for any $\beta \in [0,1]$,

$$\sum_{j' \in M:\ P_i : j' \geq j} \varphi_{i,j}(\boldsymbol{P}) \leqslant \sum_{j' \in M:\ P_i : j' \geq j} h^\beta_{i,j}(\boldsymbol{P}) \leqslant \sum_{j' \in M:\ P_i : j' \geq j} \psi_{i,j}(\boldsymbol{P}), \tag{327}$$

i.e., ψ ordinally dominates h^β at $\boldsymbol{P}$, which in turn dominates φ. Conversely, if $\psi(\boldsymbol{P})$ does not ordinally dominate $\varphi(\boldsymbol{P})$, then there exists some agent $i \in N$ and some $j \in M$, such that

$$\sum_{j' \in M:\ P_i : j' \geq j} \varphi_{i,j}(\boldsymbol{P}) > \sum_{j' \in M:\ P_i : j' \geq j} \psi_{i,j}(\boldsymbol{P}). \tag{328}$$

Again, by linearity, this implies

$$\sum_{j' \in M:\ P_i : j' \geq j} \varphi_{i,j}(\boldsymbol{P}) > \sum_{j' \in M:\ P_i : j' \geq j} h^{\beta}_{i,j}(\boldsymbol{P}) > \sum_{j' \in M:\ P_i : j' \geq j} \psi_{i,j}(\boldsymbol{P}), \tag{329}$$

which means that h^{β} does not ordinally dominate φ and is not ordinally dominated by ψ. This establishes statement 2 for ordinal dominance. For rank dominance, the result is analogous, where we exploit the fact that the rank distribution of h^{β} is the β-convex combination of the rank distributions of φ and ψ. □

4.C.3 Proof of Proposition 15

*Proof of Proposition 15. Given a setting $(N, M, \boldsymbol{q})$, a hybrid-admissible pair of mechanisms (φ, ψ), and a bound $\underline{\rho} \in [0, 1]$ Algorithm 3 (*BetaMax*) is complete and correct for the mechanism designer's problem of finding the maximal mixing factor $\beta^{max}_{(N,M,\boldsymbol{q}),\varphi,\psi}(\underline{\rho})$.*

Since there are only finitely many agents, preference profiles, misreports, and ranks, the loops of the algorithm eventually terminate. Thus, the algorithm terminates on any admissible input parameters (i.e., completeness).

For correctness, we use the fact that by Theorem 4 in (Mennle and Seuken, 2015*b*), r-partial strategyproofness is equivalent to strong r-partial dominance-strategyproofness. Formally, for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P'_i \in \mathcal{P}$, and any $K \in \{1, \ldots, m\}$, we define the following polynomials (in r)

$$p^{\varphi}_K(r) = \sum_{j:\ \mathrm{rank}_{P_i}(j) \leqslant K} r^{\mathrm{rank}_{P_i}(j)} \cdot \left(\varphi_{i,j}(P_i, P_{-i}) - \varphi_{i,j}(P'_i, P_{-i})\right), \tag{330}$$

$$p^{\psi}_K(r) = \sum_{j:\ \mathrm{rank}_{P_i}(j) \leqslant K} r^{\mathrm{rank}_{P_i}(j)} \cdot \left(\psi_{i,j}(P_i, P_{-i}) - \psi_{i,j}(P'_i, P_{-i})\right), \tag{331}$$

where $\mathrm{rank}_{P_i}(j)$ is the rank of j in the preference order of agent i, i.e., the number of objects that i weakly prefers to object j. For the hybrid mechanism, the correspoinding polynomial is

$$p^{h^{\beta}}_K(r) = (1 - \beta) p^{\varphi}_k(r) + \beta p^{\psi}_k(r), \tag{332}$$

and h^{β} is $\underline{\rho}$-partially strategyproof if and only if $p^{h^{\beta}}_K(\underline{\rho}) \geqslant 0$ for all combinations $i, (P_i, P_{-i}), P'_i, K$. Since φ is strategyproof, $p^{\varphi}_K(\underline{\rho}) \geqslant 0$, and therefore, the only way that $p^{h^{\beta}}_K(\underline{\rho})$ can be negative is for $p^{\psi}_K(\underline{\rho})$ to be negative. Conversely, if $p^{h^{\beta}}_K(\underline{\rho}) \geqslant 0$ for some β, then $p^{h^{\beta'}}_K(\underline{\rho}) \geqslant 0$ for any $\beta' \leqslant \beta$ as well, i.e., reducing β will not lead to a violation of

any of the positivity constraints. Finally, the only constraints where β is not arbitrary are those where $p_K^{\psi}(\underline{\rho}) < 0$ strictly. In this case,

$$p_K^{h^\beta}(\underline{\rho}) = (1-\beta)p_K^{\varphi}(\underline{\rho}) + \beta p_K^{\psi}(\underline{\rho}) \geqslant 0 \tag{333}$$

$$\Leftrightarrow \quad \beta \leqslant \frac{p_K^{\varphi}(\underline{\rho})}{p_K^{\varphi}(\underline{\rho}) - p_K^{\psi}(\underline{\rho})} \tag{334}$$

Algorithm BetaMax starts with an optimistic guess of $\beta^{\max} = 1$ and then reduces this value if this is required to establish a positivity constraint. As we observed, subsequent further reductions of $\beta^{\max}$ cannot lead to a renewed violation of a previously checked constraint. Since the algorithm reduces $\beta^{\max}$ only when this is strictly required by some constraint, and this reduction is minimal, the final value of the variable $\beta^{\max}$ will be precisely the maximal mixing factor for which h^β is $\underline{\rho}$-partially strategyproof. □

4.C.4 Proof of Theorem 17

Proof of Theorem 17. PS is weakly less varying than RSD.

Suppose, n agents compete for $m = m_a + 2 + m_b$ objects with capacities given by $\boldsymbol{q}$, and let $M = \{a_1, \ldots, a_{m_a}, x, y, b_1, \ldots, b_{m_b}\}$. Agent 1 is considering the two preference reports

$$\begin{aligned} P_1 &: a_1 \succ \ldots \succ a_{m_a} \succ x \succ y \succ b_1 \succ \ldots \succ b_{m_b}, \\ P_1' &: a_1 \succ \ldots \succ a_{m_a} \succ y \succ x \succ b_1 \succ \ldots \succ b_{m_b}, \end{aligned}$$

where the positions of x and y are reversed in the second report. The reports of the other agents are fixed and given by P_{-1}.

Further suppose that with reports (P_1, P_{-1}), the objects where exhausted at times $0 < \tau_1 \leqslant \tau_2 \leqslant \ldots \leqslant \tau_m = 1$ under PS. Re-label the objects as $j_1, \ldots, j_m$ in increasing order of the times at which they were exhausted. If two objects were exhausted at the same time, re-label them in arbitrary order. Denote by τ_x and τ_y the times at which x and y were exhausted, respectively.

Given these considerations, Claim 16 yields equivalent conditions under which PS changes the assignment, Claim 17 yields similar conditions under which RSD changes the assignment, and Claim 18 shows that the former condition implies the latter.

Claim 16. *In Theorem 17, $PS_1(P_1, P_{-1}) \neq PS_1(P_1', P_{-1})$ if and only if*

1. *there exists* $k \geqslant m_a$ *such that* $\tau_1 \leqslant \ldots \leqslant \tau_k < \min(\tau_x, \tau_y) \leqslant 1$, *and*
2. *for all* $l \in \{1, \ldots, m_a\}$ *there exists* $l' \in \{1, \ldots, k\}$ *with* $a_l = j_{l'}$.

Proof. **"⇒"** Choose k such that j_k is the last of the $a_1, \ldots, a_{m_a}$ to run out. Suppose, $\tau_y \leqslant \tau_k$. Agent 1 is busy consuming shares of other objects until time τ_k, regardless of the reported order of x and y. After τ_k agent 1 consumes shares of x until it is exhausted. Because y was already exhausted before τ_k, agent 1 gets no shares of y. Under report P_1', it would finish consuming other objects at τ_k and find objects y exhausted. Hence, it would begin consuming shares of x immediately, just as it did under report P_1. Thus, the order in which x and y are reported does not matter for the times at which it consumes objects x and y. Because P_1 and P_1' only differ in the order of x and y, the remaining objects are also consumed in the same order and at the same times. Hence, agent 1's assignment does not change.

The case for $\tau_x \leqslant \tau_k$ is analogous.

Because PS is non-bossy (Ekici and Kesten, 2012), we know that if the switch from P_1 to P_1' did not change the assignment for agent 1, it did not change the assignment at all.

"⇐" Suppose the last of the objects $a_1, \ldots, a_{m_a}$ to be exhausted is j_k, and $\tau_k < \tau_y \leqslant \tau_x$. Then agent 1 gets no shares of y. If it switches its report to P_1', it will receive a non-trivial share of y, hence the assignment changes.

Now suppose the opposite, namely $\tau_y > \tau_x$. Agent 1 begins consumption of x at time τ_k and then turns to y at time τ_x. Thus, agent 1 receives $\tau_x - \tau_k$ shares of x and $\tau_y - \tau_x$ shares of y. When it switches its report to P_1', it will consume shares of y between τ_k and τ_y'. We need to show that $\tau_y' - \tau_k > \tau_y - \tau_x$. If $\tau_y' \geqslant \tau_y$, this is clear, because $\tau_k < \tau_x$ by assumption. In the following we assume $\tau_y' < \tau_y$.

Let $n_y(\tau)$ be the number of agents other than agent 1 consuming shares of y at time τ. n_y is integer-valued and increasing in τ, and there must exist a $\delta > 0$ such that $n_y(\tau_y - \delta) \geqslant 1$. This means that agent 1 is not the only agent consuming shares of y before it is exhausted. Otherwise, agent 1 would exhaust y alone, which implies that agent 1 received no shares of x, a contradiction.

If agent 1 reports P_1' instead, let $n_y'(\tau)$ be the corresponding number of agents consuming y at times τ. We observe that x will be exhausted later, because agent 1 is no longer consuming shares of it. This means that agents who prefer x over y will arrive later at y. Agents arriving at y from other objects than x may also

arrive later, because they face less competition from the agents stuck at x, etc. Therefore $n'_y \leqslant n_y$.

Under report P_1 from agent 1, y is exhausted by τ_y, i.e.,

$$q_y = \int_0^{\tau_y} n_y(\tau) + \mathbb{1}_{\{\tau \geqslant \tau_x\}} d\tau, \tag{335}$$

and under report P'_1, y is exhausted by τ'_y, i.e.,

$$q_y = \int_0^{\tau'_y} n'_y(\tau) + \mathbb{1}_{\{\tau \geqslant \tau_k\}} d\tau \leqslant \int_0^{\tau'_y} n_y(\tau) + \mathbb{1}_{\{\tau \geqslant \tau_k\}} d\tau. \tag{336}$$

Equating (335) and (336) gives

$$\int_0^{\tau_y} n_y(\tau) + \mathbb{1}_{\{\tau \geqslant \tau_x\}} d\tau \quad \leqslant \quad \int_0^{\tau'_y} n_y(\tau) + \mathbb{1}_{\{\tau \geqslant \tau_k\}} d\tau. \tag{337}$$

This implies

$$\int_{\tau'_y}^{\tau_y} n_y(\tau) + \mathbb{1}_{\{\tau \geqslant \tau_k\}} d\tau \tag{338}$$

$$\leqslant \int_0^{\tau'_y} \mathbb{1}_{\{\tau \geqslant \tau_k\}} d\tau - \int_0^{\tau_y} \mathbb{1}_{\{\tau \geqslant \tau_x\}} dt + \int_{\tau'_y}^{\tau_y} \mathbb{1}_{\{\tau \geqslant \tau_k\}} d\tau \tag{339}$$

$$= \int_0^{\tau_y} \mathbb{1}_{\{\tau \geqslant \tau_k\}} - \mathbb{1}_{\{\tau \geqslant \tau_x\}} d\tau = \tau_x - \tau_k. \tag{340}$$

We know that j_k is exhausted before τ'_y and hence $n_y(\tau) + \mathbb{1}_{\{\tau \geqslant \tau_k\}} \geqslant 1$ for $\tau \in [\tau'_y, \tau_y]$, and $\geqslant 2$ for $\tau \in [\tau_y - \delta, \tau_y]$. This yields

$$\tau_y - \tau'_y < \tau_x - \tau_k, \tag{341}$$

or equivalently $\tau_y - \tau_x < \tau'_y - \tau_k$.

□

Claim 17. *In Theorem 17, $RSD_1(P'_1, P_{-1}) \neq RSD_1(P_1, P_{-1})$ if and only if there exists a sequence $(c_1, \ldots, c_{k_c})$ of k_c agents such that if RSD chose these agents first and in this order, they remove all objects $a_1, \ldots, a_{m_a}$ (and possibly more), but neither x, nor y.*

Proof. In the RSD mechanism, a permutations of agents is chose amongst all possible permutations with uniform probability. The probability for agent 1 to get some object j

is

$$P[1 \text{ gets } j] = \frac{|\{\pi \text{ permutation of } N : 1 \text{ gets } j \text{ under } \pi\}|}{|\{\pi \text{ permutation of } N\}|}, \tag{342}$$

where the denominator is $n!$, and each permutation under which agent 1 gets j contributes $\frac{1}{n!}$ to the total probability.

For some permutation π consider the turn of agent 1. There are 5 possible cases:

1. Agent 1 faces a choice set including some a_l's. This makes no contribution to its chances of getting x or y.
2. Agent 1 faces a choice set consisting only of b_l's. Again, this makes no contribution to its chances of getting x or y.
3. Agent 1 faces only b_l's and x, but not y. This case contributes $\frac{1}{n!}$ to its chances of getting x. This contribution is independent of the order in which it ranked x and y in its report.
4. Agent 1 faces only b_l's and y, but not x. This case contributes $\frac{1}{n!}$ to its chances of getting y and the contribution is again independent of the ranking of x and y.
5. Agent 1 faces x, y and some b_l's, but no a_l's. This case contributes $\frac{1}{n!}$ to either the probabilities for x or y, depending on the ranking.

"⇒" If changing from P_1 to P_1' influences the assignment, the assignment for agent 1 must have changed. This is because RSD is non-bossy (by Lemma 6). RSD also is strategyproof, hence by Theorem 1 in Mennle and Seuken (2015*b*) the probabilities for objects x and y must have changed. In all but the last case, the chances do not depend on the order in which x and y are reported. Thus, at least one permutation leads to case (5). This means that the sequence of agents chosen prior to agent 1 removes all a_l's, but neither x nor y.

"⇐" Under report P_1, agent 1 will receive x any time case (5) occurs, while under P_1' it will receive y. If a sequence $(c_1, \ldots, c_{k_c})$ as defined in Claim 17 exists, it is also the beginning of at least one permutation. When this permutation is selected, case (5) occurs. Switching from report P_1 to P_1' thus strictly increases agent 1's chances of getting y.

□

Claim 18. *In Theorem 17, 1. and 2. from Claim 16 imply the existence of a sequence as described in Claim 17.*

Proof. We prove the claim by constructing a sequence of agents

$$(c_1, \ldots, c_{k_c}) = (c_1^1, \ldots, c_1^{q_1}, \ldots, c_k^1, \ldots, c_k^{q_k}) \tag{343}$$

inductively. Under RSD this sequence will remove objects $j_1, \ldots, j_k$ in this order.

Selection of $c_k^1, \ldots, c_k^{q_k}$ By assumption j_k was consumed strictly before x, hence $\tau_k < 1$. Then at least $q_k + 1$ agents receive non-trivial shares of j_k. Otherwise, if only q_k agents received shares of j_k, they would get the entire capacity and take time 1 to consume it, a contradiction. Select q_k of these agents other than agent 1 as $c_k^1, \ldots, c_k^{q_k}$.

Because all $c_k^1, \ldots, c_k^{q_k}$ actually received shares of j_k under PS, they must all prefer j_k to all other objects except for possibly $j_1, \ldots, j_{k-1}$. In other words, suppose that $j_1, \ldots, j_{k-1}$ were removed under RSD in previous turns, the selected agents would remove j_k completely if chosen next (in arbitrary order).

Selection of $c_l^1, \ldots, c_l^{q_l}, l < k$ Suppose, $c_{l+1}^1, \ldots, c_k^{q_k}$ have been selected. Suppose further that m_l agents (plus possibly agent 1) receive non-trivial shares of j_l under PS. There are two cases:

Case 1 At least q_l of the m_l agents have not been selected as any of the $c_{l+1}^1, \ldots, c_k^{q_k}$ so far. Then these agents are chosen as $c_l^1, \ldots, c_l^{q_k}$.

Case 2 Only $n_l < q_l$ of the m_l agents have not been selected so far. The rest of the m_l agents have been selected at k' other objects. Let these objects be $j_{\rho(1)}, \ldots, j_{\rho(k')}$ with $\rho(l') \in \{l+1, \ldots, k\}$ for all $l' \in \{1, \ldots, k'\}$. At each of the objects $j_{\rho(l')}$, $q_{\rho(l')}$ agents are selected. Now there must be at least $q_l - n_l + 1$ additional agents (possibly including agent 1) consuming non-trivial shares of the objects $j_{\rho(l')}$, otherwise at most $n_l + q_{\rho(1)} + \ldots q_{\rho(k')} + q_l - n_l$ agents fully consume objects $j_l, j_{\rho(1)}, \ldots, j_{\rho(k')}$. This will take them until time 1, a contradiction.

There are two possible cases for these additional $q_l - n_l$ agents (excluding agent 1).

Case 2.1 All of them are available for selection. Then they are selected for the objects $j_{\rho(l')}$ of which they consume non-trivial shares, and the now free agents can be selected for j_l.

Case 2.1 Some of these agents are selected at some other objects $j_{\rho(k'+1)}, \ldots, j_{\rho(k'+k'')}$. Then we use the free agents as in case 2.1, say $n_{l'}$. Then we still need $q_l - n_l - n_{l'}$ agents for j_l. There must be at least $q_l + q_{\rho(1)} + \ldots + q_{\rho(k'')} + 1$ agents consuming non-trivial shares of the objects $j_l, j_{\rho(1)}, \ldots, j_{\rho(k'')}$. $q_l - n_l - n_{l'}$ are not selected for any of these objects. Again there are two cases.

We repeat this argument inductively until enough agents are found who are still available and can replace agents such that the need at object j_l can be satisfied. This must happen, otherwise all agents selected so far as $c_{l+1}^1, \ldots, c_k^{q_k}$, some $n_l^{'''} < q_l$ agents and possibly agent 1 fully consume objects $j_l, j_{l+1}, \ldots, j_k$ objects, again a contradiction.

The fact that all selected agents $c_l^1, \ldots, c_l^{q_l}, l \in \{1, \ldots, k\}$ receive a non-trivial share in the objects j_l implies that they each prefer j_l to all other objects, except possibly $j_1, \ldots, j_{l-1}$. Thus, the sequence $(c_1^1, \ldots, c_k^{q_k})$ has the properties needed for 17. □

Lemma 6. *For any distribution over orderings, the respective RSD is non-bossy.*

Proof. Fix a distribution over orderings of the agents and let p_π be the probability that ordering π is chosen. Suppose that RSD is bossy, then there exists an agents i, i', preference orders P_i, P_i', and $P_{-i} \in \mathcal{P}^{N-i}$ such that $\text{RSD}_i(P_i, P_{-i}) = \text{RSD}_i(P_i', P_{-i})$, but $\text{RSD}_{i'}(P_i, P_{-i}) \neq \text{RSD}_{i'}(P_i', P_{-i})$. For the sake of brevity, we write P and P' for P_i and P_i', respectively.

Let $\text{Can}(P, P') = (P_0 = P, P_1, \ldots, P_{k-1}, P_k = P')$ be the canonical transition from $P = P_i$ to $P' = P_i'$. As in the proof of Lemma 5, the fact that the assignment is the same at the start and at the end of the transition implies that the assignment never changes during the transition, i.e., $\text{RSD}_i(P_l, P_{-1}) = \text{RSD}_i(P_{l+1}, P_{-1})$ for all $l \in \{0, \ldots, k-1\}$. Recall that under strategyproof mechanisms, the effect of swaps in the canonical transition is never undone by subsequent swaps and that swaps only effect the probabilities for adjacent objects (see Theorem 1 in (Mennle and Seuken, 2015*b*)). Let $\text{Can}(P, P') = (P_0 = P, P_1, \ldots, P_{k-1}, P_k = P')$ be the canonical transition from $P = P_0$ to $P' = P_k$. As in the proof of Lemma 5, the fact that the assignment is the same at the start and at the end of the transition implies that the assignment never changes during the transition, i.e., $\text{RSD}_i(P_l, P_{-1}) = \text{RSD}_i(P_{l+1}, P_{-1})$ for all $l \in \{0, \ldots, k-1\}$. Recall that under strategyproof mechanisms, the effect of swaps in the canonical transition is never undone by subsequent swaps and that swaps only effect the probabilities for adjacent objects (see Theorem 1 in (Mennle and Seuken, 2015*b*)).

But the assignment changed for agent i', hence it must have changed for agent i' at some swap in the transition, say from $P_{l'}$ to $P_{l'+1} \in N_{P_{l'}}$. Let j', j'' be the objects that were swapped in this transition. Consider an ordering of the agents π with $p_\pi > 0$. There are two cases.

- Agent i gets the same object under $P_{l'}$ as under $P_{l'+1}$. Then the swap had no effect on the assignment of any other agent, i.e., under π the swap does not change the assignment of the other agents.
- Agent i receives j' under $P_{l'}$, but j'' under $P_{l'+1}$. Then the swap changes the assignment of the agent that received j'' under $P_{l'}$. The magnitude of the change is $-p_\pi < 0$. This agent can be i' by assumption.

However, the latter case is impossible, because this would also strictly increase agent i's chances of receiving j'' (by $p_\pi > 0$), implying $\mathrm{RSD}_i(P_{l'}, P_{-1}) \neq \mathrm{RSD}_i(P_{l'+1}, P_{-1})$, a contradiction. □

This concludes the proof of Theorem 17 □

4.C.5 Proof of Theorem 18

Proof of Theorem 18. ABM is weakly less varying than RSD.

Suppose the following manipulation by agent i by a swap:

$$\begin{aligned} & P_i : a_1 \succ \ldots \succ a_{m_a} \succ x \succ y \succ b_1 \succ \ldots \succ b_{m_b} \\ \rightsquigarrow \quad & P_i' : a_1 \succ \ldots \succ a_{m_a} \succ y \succ x \succ b_1 \succ \ldots \succ b_{m_b}. \end{aligned}$$

By Lemma 17, RSD changes the assignment ($\mathrm{RSD}_i(P_i, P_{-i}) \neq \mathrm{RSD}_i(P_i', P_{-i})$) if and only if there exists an ordering of the agents π such that i gets to pick between x and y in its turn, but all objects $a_1, \ldots, a_{m_a}$ are exhausted by higher-ranking agents. We show that if ABM changes the assignment, then such an ordering π exists. Thus, a change of assignment under ABM implies a change of assignment under RSD.

Suppose, the change of report by agent i from P_i to P_i' changes the outcome of ABM for i, i.e., $\mathrm{ABM}_i(P_i, P_{-i}) \neq \mathrm{ABM}_i(P_i', P_{-i})$. Then from the proof of swap monotonicity (Mennle and Seuken, 2015*d*) we know that there exists an ordering of the agents π' such that in some round (say L), i has not been assigned an object yet, all $a_1, \ldots, a_{m_a}$ are exhausted, but neither x nor y are exhausted. Let $r(i')$ be the round in which i' is

assigned its object, and let

$$R(r) = \{i' \in N \mid r(i') = r\}$$

be the set of agents who receive their assignment in round r (given ordering π'). If i' is assigned object j in round r, i' has applied to j in that round. Thus, out of all the objects with capacity available at the beginning of round r, i' must prefer j. Facing the same set of choices under RSD, i' would also pick j.

Consider an ordering π that ranks an agent i' before another agent i'' if $r(i') < r(i'')$ and ranks them in arbitrary order if $r(i') = r(i'')$. Additionally, let π rank i after all the agents in the set $R(1) \cup \ldots \cup R(L-1)$. If RSD chooses π as the ordering of the agents, then all agents in $R(1)$ receive their first choice (as under ABM). Next, all agents in $R(2)$ face the choice sets out of which they most prefer the object they were assigned under ABM. This continues until finally i faces a choice set that includes none of the $a_1, \ldots, a_{m_a}$, but both x and y. Hence, π is the ordering we are looking for, and its existence concludes the proof. □

5 The Pareto Frontier for Random Mechanisms

Abstract

We study the trade-offs between strategyproofness and other desiderata, such as efficiency or fairness, that often arise in the design of random ordinal mechanisms. We use ε-approximate strategyproofness to measure the *manipulability* of non-strategyproof mechanisms, and we introduce the *deficit* to quantify the performance of mechanisms with respect to a desideratum. When the desideratum is incompatible with strategyproofness, mechanisms that trade off manipulability and deficit optimally form the *Pareto frontier*. Our main contribution is a structural characterization of this Pareto frontier, and we present algorithms that exploit this structure to compute it. To illustrate its shape, we apply our results for two orthogonal desiderata, namely Plurality and Veto scoring, in a setting with 3 agents and 3 alternatives.

5.1 Introduction

In many situations, a group of individuals has to make a collective decision by selecting an alternative: *who should be the next president? who gets a seat at which public school? or where to build a new stadium? Mechanisms* are systematic procedures to make such decisions. Formally, a mechanism collects the individuals' (or *agents'*) preferences and selects an alternative based on this input. While the goal of the mechanism is to select an alternative that is desirable for society as a whole, it may be possible for individual agents to manipulate the outcome to their own advantage by being insincere about their preferences. However, a mechanism that receives false input from lying agents will have difficulty in determining an outcome that is desirable with respect to the true preferences. Therefore, incentives for truthtelling are a major concern in mechanism design.

In this paper, we consider *ordinal* mechanisms with *random outcomes*. These collect

preference orders and select lotteries over alternatives. We study these mechanisms in the domain where all agents can have arbitrary weak and strict preferences. However, our results continue to hold on any domain restriction that limits the set of preference profiles to a subset. Important examples include the domain of strict preferences, the assignment domain, and the two-sided matching domain.

5.1.1 The Curse of Strategyproofness

Strategyproof mechanisms make truthful reporting a dominant strategy for all agents and it is therefore the "gold standard" among the incentive concepts. However, the seminal impossibility results by Gibbard (1973) and Satterthwaite (1975) established that if there are at least three alternatives and all strict preference orders are possible, then the only unanimous and deterministic mechanisms that are strategyproof are *dictatorships*. Gibbard (1977) extended these insights to mechanisms that involve randomization and showed that all strategyproof random mechanisms are probability mixtures of strategyproof unilateral and strategyproof duple mechanisms. Obviously, these results greatly restrict the mechanism design space when strategyproofness is viewed as an indispensable design requirement. In particular, many common desiderata are incompatible with strategyproofness, such as Condorcet consistency, stability of matchings, or egalitarian fairness.

When a desideratum is incompatible with strategyproofness, designing "ideal" mechanisms is impossible. For example, suppose that our goal is to select an alternative that is the first choice of many agents. The *Plurality* mechanism selects an alternative that is the first choice for a maximal number of agents. Thus, it achieves our goal perfectly (given truthful preference reports). At the same time, any mechanism that achieves this goal perfectly must be manipulable. The *Random Dictatorship* mechanism selects the first choice of a randomly chosen agent. This mechanism is strategyproof, but there is a non-trivial probability that it selects "wrong" alternatives. If Plurality is "too manipulable" and Random Dictatorship is "wrong too often," then trade-offs are necessary. Mechanism designers can make this trade-off by choosing an intermediate mechanism that is *less likely* to select "wrong" alternatives than Random Dictatorship but also has a *lower manipulability* than Plurality. In this paper, we study mechanisms that make this trade-off optimally. Such mechanisms form the *Pareto frontier* because they achieve the desideratum as well as possible, subject to a given limit on manipulability. Our main result is a structural characterization of this Pareto frontier.

5.1.2 Measuring Manipulability and Deficit

In order to understand these trade-offs formally, we need measures for the performance of mechanisms on both dimensions (i.e., manipulability and achieving the desideratum).

Approximate Strategyproofness and Manipulability

A mechanism that is *strategyproof* will not allow agents to obtain any positive gain from misreporting. *Approximate strategyproofness* is a common relaxation of strategyproofness: instead of requiring the gain from misreporting to be weakly negative, ε-approximate strategyproofness imposes a small (albeit positive) upper bound ε on this gain. The economic intuition behind this concept is that if the potential gain is small, then the agents might not be willing to collect the necessary information and deliberate about misreports but stick with truthful reporting instead. To obtain a notion of approximate strategyproofness that is meaningful for ordinal mechanisms, we follow earlier work (e.g., (Birrell and Pass, 2011; Carroll, 2013)) and consider agents with von Neumann-Morgenstern utility functions that are bounded between 0 and 1. This allows the formulation of a parametric measure for the strength of a mechanism's incentive properties: we define the *manipulability* $\varepsilon(\varphi)$ of a mechanism φ as the smallest bound ε for which φ is ε-approximately strategyproof.

Welfare Functions and Deficit

By allowing a higher manipulability, the design space of possible mechanisms grows. This raises the question how the new freedom can be harnessed to improve the performance of mechanisms with respect to other desiderata. To measure the ability of mechanisms to achieve these desiderata, we introduce the *deficit*.

We first illustrate the idea of the deficit using an example: consider the desideratum of selecting an alternative that is the first choice of many agents. It is natural to think of "the share of agents for whom an alternative is the first choice" as the *welfare* that society derives from selecting this alternative, and for lotteries over alternatives the welfare is simply the respective expected welfare. In this sense, Plurality is *welfare maximizing* at all preference profiles but Random Dictatorship is not. By selecting a lottery that is not welfare-maximizing, society incurs a loss; we quantify this loss by the difference between the highest possible welfare and the expected welfare from the particular lottery. Finally, the *deficit* of the Random Dictatorship mechanism is simply the highest of these losses that occurs across all preference profiles from using Random Dictatorship. This yields a

meaningful, numerical measure for "how badly Random Dictatorship fails" at achieving the desideratum of selecting an alternative that is the first choice of many agents.

Analogous to this example, notions of deficit for different desiderata can be constructed. This construction always has two steps: first, we need to express the desideratum as a *welfare function*. Loosely speaking, a welfare function specifies the value from selecting a given alternative at a given preference profile. We use the term "welfare" rather freely: *high* welfare can mean that the agents are simply better off, but it can also mean that an alternative is *desirable for society* more generally. For example, it may be more fair or stable. The essential restriction is that the welfare of an alternative must be determined on the basis of the agents' ordinal preferences (i.e., the information that ordinal mechanisms elicit), and the welfare of lotteries must be given by their expected welfare.[1] The second step of the construction is to define the deficit itself. This can be done in several ways: in the example, we considered *absolute* differences and the *worst-case* deficit across all preference profiles. Alternatively, we can consider *relative* differences, take the *expected* deficit with respect to some prior distribution over preference profiles, or both. For any mechanism φ, we denote its deficit by $\delta(\varphi)$. We can construct many meaningful notions of deficit in this way, and the results in this paper hold for any of them.

5.1.3 Optimal Mechanisms and the Pareto Frontier

Together, the measure $\varepsilon(\varphi)$ for manipulability and the measure $\delta(\varphi)$ for deficit yield a way to compare different mechanisms in terms of their incentive and welfare properties. To this end, we define the *signature* of a mechanism φ, which is the point $(\varepsilon(\varphi), \delta(\varphi))$ in the unit square $[0, 1] \times [0, 1]$. Figure 5.1 illustrates this comparison for three mechanisms φ, ψ, and ϑ. Ideally, a mechanism would be strategyproof *and* welfare maximizing. This corresponds to a signature in the origin $(0, 0)$. However, for desiderata that are incompatible with strategyproofness, designing ideal mechanisms is not possible. Instead, there exist strategyproof mechanisms which have a non-trivial deficit, such as Random Dictatorship; and there exist welfare maximizing mechanisms which have non-trivial

[1] Many important desiderata can be expressed via welfare functions, including binary desiderata such as *unanimity*, *Condorcet consistency*, *egalitarian fairness*, *Pareto optimality*, *v-rank efficiency* of assignments, *stability* of matchings, any desideratum specified via a *target mechanism* (or a *target correspondence*), or any *logical combination* of these. Moreover, it is possible to express quantitative desiderata, such as *maximizing positional score* in voting, *maximizing v-rank value* of assignments, or *minimize the number of blocking pairs* in matching. We formally introduce welfare functions in Section 5.5.1, and we discuss their generality and limitations in Section 5.10.2.

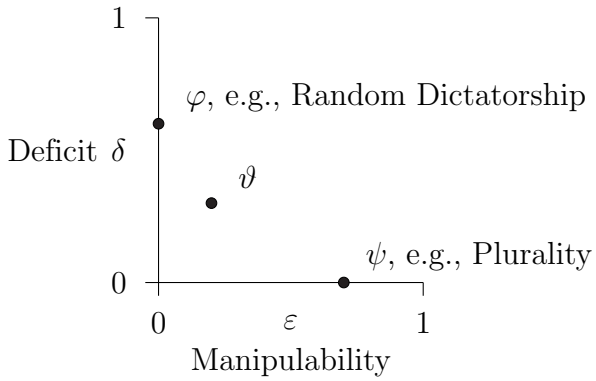

Figure 5.1: Example signatures of mechanisms in a manipulability-deficit-plot.

manipulability, such as Plurality (when the goal is to select an alternative that is the first choice of many agents). Finally, there may exist mechanisms that have intermediate signatures, such as ϑ in Figure 5.1. Choosing between these mechanisms requires trade-offs between manipulability and deficit.

Finding Optimal Mechanisms

Naturally, we want to consider mechanisms that make this trade-off optimally. We say that a mechanism is *optimal at manipulability bound* ε if (1) it has manipulability of at most ε, and (2) it has the lowest deficit among all such mechanisms. For a given ε we denote by $\textsc{Opt}(\varepsilon)$ the set of all mechanisms that are optimal at ε. Given an optimal mechanism, it is not possible to reduce the deficit without increasing manipulability at the same time. In this sense, the set of all optimal mechanisms constitutes the *Pareto frontier*. Our first result in this paper yields a finite set of linear constraints that is equivalent to ε-approximate strategyproofness. This equivalence allows us to formulate a linear program that searches for optimal mechanisms. Precisely, for a fixed bound on manipulability ε, each possible solution to the linear program uniquely identifies one of the mechanisms that are optimal at ε. This enables the computation of optimal mechanisms by solving the respective linear program algorithmically.

Trade-offs via Hybrid Mechanisms

Given two mechanisms, *mixing* them suggests itself as a straightforward approach to creating new mechanisms with intermediate signatures. Formally, the β*-hybrid* of two mechanisms φ and ψ is the β-convex combination of the two mechanisms. Such a hybrid can be implemented by first collecting the agents' preference reports, then randomly

deciding to use ψ or φ with probabilities β and $1-\beta$, respectively. If φ has lower manipulability and ψ has lower deficit, then it is intuitive that their hybrid should inherit a share of the appealing properties on both dimensions. Our second result in this paper formalizes this intuition: we prove that the signature of a β-hybrid is always *weakly preferable* on both dimensions to the β-convex combination of the signatures of the two component mechanisms. This insight teaches us that interesting intermediate mechanisms can indeed be obtained via mixing.

The result has important consequences for the Pareto frontier. We use it to show that the first unit of manipulability that we sacrifice will yield the greatest return in terms of a reduction of deficit. Furthermore, the marginal return on relaxing the incentive requirement further decreases as the mechanisms become more and more manipulable. This is obviously good news for mechanism designers: if welfare gains can be obtained by relaxing strategyproofness, then the most substantial gains will be unlocked by relaxing incentive constraints just "a little bit."

Structural Characterization of the Pareto Frontier

Recall that our first result enables us to compute mechanisms on the Pareto frontier for isolated manipulability bounds ε, and our second result shows that we can use hybrids to create new intermediate (albeit not necessarily optimal) mechanisms. However, to fully understand the possible and necessary trade-offs between manipulability and deficit, we need to identify the whole Pareto frontier across *all* manipulability bounds.

Our main result in this paper is a structural characterization of this Pareto frontier: we show that there exists a finite set $\varepsilon_0 < \ldots < \varepsilon_K$ of *supporting manipulability bounds*, such that between any two of them (between ε_{k-1} and ε_k, say) the Pareto frontier consists precisely of the hybrids of two mechanisms that are optimal at ε_{k-1} and ε_k, respectively. Consequently, the two building blocks of the Pareto frontier are (1) the optimal mechanisms at the supporting manipulability bounds $\{\varepsilon_k : k = 0, \ldots, K\}$, which can be identified by a linear program, and (2) the hybrids of optimal mechanisms at adjacent supporting manipulability bounds for any intermediate $\varepsilon \notin \{\varepsilon_k : k = 0, \ldots, K\}$. Thus, the Pareto frontier can be represented concisely in terms of a finite number of optimal mechanisms and their hybrids. Our characterization can be exploited to compute the whole Pareto frontier algorithmically. This enables the mechanism designer to identify the possible and necessary trade-offs exactly and determine the most desirable compromises.

In summary, we provide a novel perspective on the possible and necessary trade-offs

between manipulability and the performance of mechanisms with respect to a large class of desiderata. Our results unlock the Pareto frontier of random mechanisms to analytic, axiomatic, and algorithmic explorations.

Organization of this paper: In Section 5.2, we discuss related work. In Section 5.3, we introduce our formal model. In Section 5.4, we formalize *manipulability* and prove an equivalence for ε-approximate strategyproofness. In Section 5.5, we introduce welfare functions and the notion of *deficit*. In Section 5.6, we present the linear program FindOpt that identifies optimal mechanisms. In Section 5.7, we study the signatures of hybrid mechanisms. In Section 5.8, we give the structural characterization result for the Pareto frontier. In Section 5.9, we provide the algorithms FindBounds and FindLower, and we present applications to two concrete problems. Section 5.11 concludes.

5.2 Related Work

Severe impossibility results restrict the design of strategyproof ordinal mechanisms. The seminal Gibbard-Satterthwaite Theorem (Gibbard, 1973; Satterthwaite, 1975) established that if all strict preferences over at least 3 alternatives are possible, then the only unanimous, strategyproof, and deterministic mechanisms are dictatorial. Gibbard (1977) extended this result for random mechanisms by showing that any strategyproof random mechanism is a probability mixture of strategyproof unilateral and strategyproof duple mechanisms. Thus, many important desiderata are incompatible with strategyproofness, such as selecting a Condorcet winner or maximizing Borda count (Pacuit, 2012). Similar restrictions persist in other domains, such as the random assignment problem, where strategyproofness is incompatible with v-rank efficiency (Featherstone, 2011), or the two-sided matching problem, where strategyproofness is incompatible with stability (Roth, 1982).

Many research efforts have been made to circumvent these impossibility results to obtain mechanisms with better performance on other dimensions. One way to reconcile strategyproofness with other desiderata is to consider restricted domains: Moulin (1980) showed that in the single-peaked domain, all strategyproof, anonymous, and efficient mechanisms are variants of the Median mechanism with additional virtual agents, and Ehlers, Peters and Storcken (2002) extended this result to random mechanisms. More broadly, Chatterji, Sanver and Sen (2013) showed that a *semi-single-peaked* structure is *almost* the defining characteristic of domains that admit the design of strategyproof deterministic mechanisms with appealing properties; an analogous result for random

mechanisms is outstanding.

An alternative way to circumvent impossibility results pertaining to strategyproofness is to continue working in full domains but to relax the strategyproofness requirement "a little bit." This may allow the design of mechanisms that come closer to achieving a given desideratum but still have appealing (albeit imperfect) incentive properties. In (Mennle and Seuken, 2015*b*) we introduced *partial strategyproofness*, a relaxation of strategyproofness that has particular appeal in the assignment domain. Azevedo and Budish (2015) proposed *strategyproofness in the large*, which requires that as the number of agents in a market grows, the incentives for any individual agent to misreport its preferences should vanish in the limit. Intuitively, this means that the mechanism has good incentives for agents who are "price takers." However, strategyproofness in the large is unsuited for the exact analysis of finite settings which we perform in this paper. Instead, we follow Birrell and Pass (2011) and Carroll (2013), who used *approximate strategyproofness* to quantify manipulability of non-strategyproof ordinal mechanisms. Using approximate strategyproofness for agents whose vNM utilities are bounded between 0 and 1, Carroll (2013) quantified the asymptotic manipulability of various voting mechanisms in different informational environments. In the present paper, we use the same notion of approximate strategyproofness to derive a parametric measure for manipulability; however, we give exact rather than asymptotic results.

Some prior work has also considered the trade-offs between strategyproofness and efficiency explicitly. Using efficiency notions based on dominance relations, Aziz, Brandt and Brill (2013*a*) and Aziz, Brandl and Brandt (2014) proved compatibility and incompatibility of various combinations of incentive and efficiency requirements. Procaccia (2010) considered an *approximation ratio* based on positional scoring and gave bounds on how well strategyproof random mechanisms can approximate optimal positional scores as markets get large. While he found most of these to be inapproximable by strategyproof mechanisms, Birrell and Pass (2011) obtained positive limit results for approximation of deterministic target mechanisms by *approximately* strategyproof random mechanisms. In the present paper, we define the notion of *deficit* to encode a broad spectrum of desiderata. This concept of deficit is novel but sufficiently general to capture desiderata based on positional scoring or on target mechanisms (see Section 5.10 for a discussion of generality and limitations).

5.3 Formal Model

Let N be a set of n *agents* and M be a set of m *alternatives*, where the tuple (N, M) is called a *setting*. Each agent $i \in N$ has a *preference order* P_i over alternatives, where $P_i : a \succeq b$ indicates that agent i weakly prefers alternative a to alternative b. If $P_i : a \succeq b$ and $P_i : b \succeq a$, then i is said to be *indifferent* between a and b, which we denote by $P_i : a \sim b$. Conversely, if $P_i : a \succeq b$ and $P_i : b \not\succeq a$, then i *strictly prefers* a to b, which we denote by $P_i : a \succ b$. We denote the set of all preference orders over alternatives by $\mathcal{P}$. For agent i's preference order P_i, we denote by $r_{P_i}(j) = \#\{j' \in M \mid P_i : j' \succ j\} + 1$ the *rank* of alternative j under P_i, i.e., the number of alternatives that i strictly prefers to j, where 1 is added to ensure that first choices receive rank 1 (not 0).

A collection of preference orders from all agents $\boldsymbol{P} = (P_i, P_{-i})$ is called a *preference profile*, where P_{-i} is the collection of preference orders of all the other agents, except i. A *(random) mechanism* is a mapping $\varphi : \mathcal{P}^N \to \Delta(M)$, where $\Delta(M)$ is the space of lotteries over alternatives. Any lottery $x \in \Delta(M)$ is called an *outcome*.

Agents' preferences over alternatives are extended to preferences over lotteries via von Neumann-Morgenstern utility functions: all agents $i \in N$ are endowed with a *utility function* $u_i : M \to [0, 1]$ that represents their preference order, i.e., $u_i(a) \geqslant u_i(b)$ holds whenever $P_i : a \succeq b$. We denote the set of all utility functions that represent the preference order P_i by U_{P_i}. Utilities are bounded between 0 and 1, so that the model admits a non-degenerate notion of approximate strategyproofness (see Remark 14).

Remark 13. Our results are formulated for the full domain, but they extend naturally to a variety of domains, including the domain of strict preferences, the assignment problem, and the two-sided matching problem. In Section 5.10, we discuss generality and limitations.

5.4 Approximate Strategyproofness and Manipulability

Our goal in this paper is to study mechanisms that trade off manipulability and other desiderata optimally. For this purpose we need measures for the performance of different mechanisms with respect to the two dimensions of this trade-off. In this section, we review the approximate strategyproofness concept, derive a measure for the manipulability of non-strategyproof mechanisms, and present our first main result.

5.4.1 Approximate Strategyproofness

The most demanding incentive concept is strategyproofness, which requires that truthful reporting is a dominant strategy for all agents. For random mechanisms, this means that truthful reporting always maximizes any agent's *expected utility*.

Definition 41 (Strategyproofness)**.** Given a setting (N, M), a mechanism φ is *strategyproof* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any utility $u_i \in U_{P_i}$, and any misreport $P_i' \in \mathcal{P}$, we have

$$\sum_{j \in M} u_i(j) \cdot (\varphi_j(P_i', P_{-i}) - \varphi_j(P_i, P_{-i})) \leqslant 0. \tag{344}$$

The left side of (344) is the change in its own expected utility that agent i can affect by falsely reporting P_i' instead of reporting P_i truthfully. For later use, we denote this difference by

$$\varepsilon(u_i, (P_i, P_{-i}), P_i', \varphi) = \sum_{j \in M} u_i(j) \cdot (\varphi_j(P_i', P_{-i}) - \varphi_j(P_i, P_{-i}))\,. \tag{345}$$

The fact that $\varepsilon(u_i, (P_i, P_{-i}), P_i', \varphi)$ is upper bounded by 0 for strategyproof mechanisms means that deviating from the true preference report yields weakly lower expected utility for any agent in any situation, independent of the other agents' reports.

Conversely, if a mechanism φ is *not* strategyproof, there necessarily exists at least one situation in which $\varepsilon(u_i, (P_i, P_{-i}), P_i', \varphi)$ is strictly positive. To relax strategyproofness we follow earlier work: Birrell and Pass (2011) and Carroll (2013) employed bounded vNM utility functions to obtain a meaningful notion of approximate strategyproofness for ordinal mechanisms.

Definition 42 (ε-Approximately Strategyproof)**.** Given a setting (N, M) and $\varepsilon \in [0, 1]$, a mechanism φ is *ε-approximately strategyproof* if for any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any utility $u_i \in U_{P_i}$, and any misreport $P_i' \in \mathcal{P}$, we have

$$\varepsilon(u_i, (P_i, P_{-i}), P_i', \varphi) = \sum_{j \in M} u_i(j) \cdot (\varphi_j(P_i', P_{-i}) - \varphi_j(P_i, P_{-i})) \leqslant \varepsilon. \tag{346}$$

Note that this definition is analogous to Definition 41 of strategyproofness, except that the upper bound in (346) is ε (instead of 0). Thus, 0-approximate strategyproofness coincides with strategyproofness. Furthermore, the gain for agents whose utilities are bounded

between 0 and 1 will never exceed 1, which makes 1-approximate strategyproofness a void constraint that is trivially satisfied by any mechanism.

The interpretation of intermediate values of $\varepsilon \in (0,1)$ is more challenging. Unlike the utilities in quasi-linear domains, vNM utilities are not comparable across agents. Thus, we cannot simply think of ε as the "value" (e.g., in dollars) that an agent can gain by misreporting. Nonetheless, we can interpret the bound ε as follows: since u_i is bounded between 0 and 1, a change of magnitude 1 in expected utility corresponds to the selection of some agent's first choice alternative instead of that agent's last choice. Thus, "1" corresponds to the *maximal gain* from misreporting that *any* agent could obtain under an *arbitrary* mechanism. Relative to this value 1, the parameter ε is the *share of this maximal gain* by which any agent can at most improve its expected utility under an ε-approximately strategyproof mechanism.

Remark 14. The bounds on utilities are essential for ε-approximate strategyproofness to be a useful relaxation of strategyproofness for ordinal mechanisms. To see this, suppose that a mechanism φ was manipulable and allowed a gain of $\varepsilon(u_i, (P_i, P_{-i}), P_i', \varphi) > 0$. Then, scaling the utility function u_i by a factor $\alpha > 1$ would result in a linear increase of this gain (i.e., $\varepsilon(\alpha u_i, (P_i, P_{-i}), P_i', \varphi) = \alpha\varepsilon(u_i, (P_i, P_{-i}), P_i', \varphi)$). Since this value could become arbitrarily large for large α, ε-approximate strategyproofness for unbounded utilities would coincide with strategyproofness.

5.4.2 Manipulability

If φ is ε-approximately strategyproof, then it is also ε'-approximately strategyproof for any $\varepsilon' \geqslant \varepsilon$. Thus, lower values of ε correspond to stronger incentive properties. With this in mind, we define the manipulability of a mechanism.

Definition 43 (Manipulability)**.** Given a setting (N, M), the *manipulability* of a mechanism φ (in the setting (N, M)) is given by

$$\varepsilon(\varphi) = \min\{\varepsilon' \in [0,1] : \varphi \text{ is } \varepsilon'\text{-approximately SP in } (N, M)\}. \tag{347}$$

Intuitively, $\varepsilon(\varphi)$ is the lowest bound ε' for which φ is ε'-approximately strategyproof. This minimum is in fact attained because all inequalities from (346) are weak. Note that in a different setting (N', M'), the manipulability of a mechanism may vary. However, for all statements in this paper a setting is held fixed. Therefore, the value $\varepsilon(\varphi)$ should be understood as the manipulability of the mechanism φ in the fixed setting from the respective context.

5.4.3 An Equivalent Condition for Approximate Strategyproofness

Recall that the definition of ε-approximate strategyproofness imposes an upper bound on the gain that agents can obtain by misreporting. In particular, inequality (346) must hold for all vNM utility functions that represent the manipulating agent's preference order. Since there are infinitely many such utility functions, a naïve approach to verifying ε-approximate strategyproofness of a given mechanism would involve checking an infinite number of constraints. This is somewhat unattractive from an axiomatic perspective and even prohibitive from an algorithmic perspective. Fortunately, we can bypass this issue, as the next Theorem 19 shows.

Theorem 19. *Given a setting (N, M), a bound $\varepsilon \in [0, 1]$, and a mechanism φ, the following are equivalent:*

1. *φ is ε-approximately strategyproof in (N, M).*
2. *For any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any rank $r \in \{1, \ldots, m\}$, we have*

$$\sum_{j \in M : r_{P_i}(j) \leqslant r} \varphi_j(P_i', P_{-i}) - \varphi_j(P_i, P_{-i}) \leqslant \varepsilon. \tag{348}$$

Proof Outline (formal proof in Appendix 5.E.1). The key is a representation of any utility function as an element of the convex hull of a certain set of *extreme* utility functions. For any combination of agents, preference profiles, misreports, and ranks, the inequalities in statement 2 are precisely the ε-approximate strategyproofness constraints for these extreme utility functions. □

Theorem 19 yields that ε-approximate strategyproofness can be equivalently expressed as a *finite* set of *weak*, *linear* inequalities. This has far-reaching consequences. In general, it unlocks approximate strategyproofness for use under the automated mechanism design paradigm (Sandholm, 2003). Specifically, it enables our identification of optimal mechanisms as solutions to a particular linear program (Section 5.6).

5.5 Welfare Functions and Deficit

While it is important to construct mechanisms that elicit truthful preferences, good incentives alone do not make a mechanism attractive. Instead, a mechanism should

ultimately select alternatives which are *desirable*, where the desirability of an alternative depends on the agents' preferences. In this section, we introduce a formal method to quantify the ability of mechanisms to achieve a given desideratum.

5.5.1 Welfare Functions

To express a desideratum formally, we define *welfare functions*. These reflect the value that society derives from selecting a particular alternative when the agents have a particular profile of preferences.

Definition 44 (Welfare Function)**.** An *welfare function* $w : M \times \mathcal{P}^N \to [0, 1]$ is a mapping that determines the *w-welfare* $w(j, \boldsymbol{P})$ associated with selecting alternative $j \in M$ when the agents have preferences $\boldsymbol{P}$.

We illustrate with two examples how welfare functions capture different desiderata.

Example 16. Suppose that our goal is to select alternatives that are the first choice of many agents. We can define the corresponding welfare function by setting $w^{\text{PLU}}(j, \boldsymbol{P}) = n_j^1/n$, where n_j^1 is the number of agents whose first choice under $\boldsymbol{P}$ is j. Note that $w^{\text{PLU}}(j, \boldsymbol{P})$ is proportional to the *Plurality score* of j under $\boldsymbol{P}$.

Example 17. Alternatively, we may wish to reflect a binary desideratum, like Condorcet consistency. Given a preference profile $\boldsymbol{P}$, an alternative j is a *Condorcet winner* if it dominates all other alternatives in a pairwise majority comparison. If our goal is to select Condorcet winners whenever they exist, we can express this desideratum by setting $w^{\text{CON}}(j, \boldsymbol{P}) = 1$ for any alternative j that is a Condorcet winner at $\boldsymbol{P}$, and $w^{\text{CON}}(j, \boldsymbol{P}) = 0$ for all other alternatives.

Since we consider random mechanisms whose outcomes are lotteries over alternatives, we extend welfare functions from alternatives to random outcomes by taking expectations.

Definition 45 (Expected Welfare)**.** Given a welfare function w, an outcome $x \in \Delta(M)$, and a preference profile $\boldsymbol{P} \in \mathcal{P}^N$, the *expected w-welfare of x at $\boldsymbol{P}$* is given by

$$w(x, \boldsymbol{P}) = \sum_{j \in M} x_j \cdot w(j, \boldsymbol{P}). \tag{349}$$

If the welfare function w and the preference profile $\boldsymbol{P}$ are clear from the context, we refer to $w(x, \boldsymbol{P})$ simply as the *welfare of x*. The interpretation of this value is straightforward: if the welfare function w quantifies the value of different alternatives

(like the Plurality score in Example 16), then $w(x, \boldsymbol{P})$ is the *expectation* of the value that society to derives from randomly selecting an alternative according to x. If w reflects a binary desideratum (like Condorcet consistency in Example 17), then $w(x, \boldsymbol{P})$ is the *probability* that an alternative with the desirable property will ultimately be selected.

Remark 15. By taking expectations, the welfare from selecting a particular lottery is determined by the welfare of the individual alternatives. This linear structure is a key ingredient to our results. In Section 5.10, we show that many (but not all) popular desiderata admit such a representation, and we also discuss the limitations.

Ideally, mechanisms would always select outcomes that maximize welfare.

Definition 46 (Welfare Maximizing)**.** Given a welfare function w and a preference profile $\boldsymbol{P} \in \mathcal{P}^N$, an outcome $x \in \Delta(M)$ is *w-welfare maximizing at $\boldsymbol{P}$* if

$$w(x, \boldsymbol{P}) = \max_{j \in M} w(j, \boldsymbol{P}). \tag{350}$$

A mechanism φ is *w-welfare maximizing* if for any preference profile $\boldsymbol{P} \in \mathcal{P}^N$, the outcome $\varphi(\boldsymbol{P})$ is w-welfare maximizing at $\boldsymbol{P}$.

By construction, for any preference profile $\boldsymbol{P}$, there always exists at least one alternative that is w-welfare maximizing at $\boldsymbol{P}$. Furthermore, any w-welfare maximizing random outcome must be a lottery over alternatives that are all w-welfare maximizing at $\boldsymbol{P}$.

Example 16, continued. Recall the welfare function $w^{\text{Plu}}(j, \boldsymbol{P}) = n_j^1/n$, where n_j^1 is the number of agents who ranked j as their first choice under $\boldsymbol{P}$. Observe that a mechanism is w^{Plu}-welfare maximizing if and only if it is a Plurality mechanism (modulo tie-breaking because there may be multiple w^{Plu}-welfare maximizing alternatives).

Example 17, continued. Recall the welfare function

$$w^{\text{Con}}(j, \boldsymbol{P}) = \begin{cases} 1, & \text{if } j \text{ is a Condorcet winner at } \boldsymbol{P}, \\ 0, & \text{else,} \end{cases} \tag{351}$$

that expresses the desideratum to select Condorcet winners when they exist. In this case, a w^{Con}-welfare maximizing mechanism selects Condorcet winners whenever they exist. Moreover, at any preference profile $\boldsymbol{P}$ where no Condorcet winner exists, the maximal achievable w-welfare is zero. At these preference profiles, w^{Con}-welfare maximizing mechanisms are free to choose any alternative. Consequently, maximizing w^{Con}-welfare is *equivalent* to achieving Condorcet consistency.

5.5.2 Worst-case and Ex-ante Deficit of Mechanisms

For an outcome that is not welfare maximizing at a given preference profile, we define its deficit as the difference between the maximum achievable welfare and the welfare achieved by the outcome. Intuitively, this value is the loss that society incurs from choosing that outcome (instead of an outcome that maximizes welfare).

Definition 47 (Deficit of Outcomes)**.** Given a welfare function w, an outcome $x \in \Delta(M)$, and a preference profile $\boldsymbol{P} \in \mathcal{P}^N$, the *$w$-deficit of x at $\boldsymbol{P}$* is

$$\delta_w(x, \boldsymbol{P}) = \max_{j \in M} w(j, \boldsymbol{P}) - w(x, \boldsymbol{P}). \tag{352}$$

Remark 16 (Relative Deficit)*.* In Definition 47, the difference is *absolute*. However, in some situations, it may be more natural to consider *relative* differences, such as the ratio between the achieved and the maximal achievable welfare. All our results hold for relative and absolute deficits: in Appendix 5.A we show that without loss of generality we can restrict our attention to the absolute version because any relative deficit can be represented as an absolute deficit by scaling the corresponding welfare function.

Equipped with the notion of deficit for outcomes, we define the deficit for mechanisms. This measure is the welfare counterpart to the measure $\varepsilon(\varphi)$ for manipulability. There are two ways to arrive at such a measure for deficit: the first notion is a *worst-case* deficit, where the deficit of the mechanism φ is determined by the most severe violation of the desideratum across all possible preference profiles.

Definition 48 (Worst-case Deficit)**.** Given a setting (N, M), a welfare function w, and a mechanism φ, the *worst-case w-deficit of φ (in (N, M))* is the highest w-deficit incurred by φ across all preference profiles; formally,

$$\delta_w^{\max}(\varphi) = \max_{\boldsymbol{P} \in \mathcal{P}^N} \delta_w(\varphi(\boldsymbol{P}), \boldsymbol{P}). \tag{353}$$

For the second deficit notion, suppose that the agents' preference profiles are drawn from a distribution $\mathbb{P}$ and that this distribution is known to the mechanism designer. In this case, she may prefer a mechanism that induces high *expected* welfare under $\mathbb{P}$.

Definition 49 (Ex-ante Deficit)**.** Given a setting (N, M), a welfare function w, a probability distribution $\mathbb{P}$ over preference profiles, and a mechanism φ, the *ex-ante*

w-deficit of φ with respect to $\mathbb{P}$ (in (N, M)) is

$$\delta_w^{\mathbb{P}}(\varphi) = \sum_{\boldsymbol{P} \in \mathcal{P}^N} \mathbb{P}[\boldsymbol{P}] \cdot \delta_w(\varphi(\boldsymbol{P}), \boldsymbol{P}). \tag{354}$$

Minimizing $\delta_w^{\mathbb{P}}(\varphi)$ corresponds to minimizing the expected w-deficit from applying φ *ex-ante* (i.e., before the agents' preferences are known). This approach is attractive in situations where the same mechanism is applied repeatedly for different groups of agents, so that the outcomes are attractive *on average* across all the repetitions.

For the remainder of this paper, we assume that a setting (N, M) is fixed. Furthermore, we fix a desideratum, expressed via a welfare function w, and we consider a notion of deficit δ that is derived from w either as a worst-case deficit or as an ex-ante deficit. Fixing the triple (N, M, δ), called a *problem*, we denote by $\delta(\varphi)$ the respective deficit of the mechanism φ, keeping in mind that this value may depend on the setting (N, M). Unless explicitly stated otherwise, our results should be understood to hold for each fixed problem (N, M, δ) separately.

5.6 Optimal Mechanisms

We have constructed the measure $\varepsilon(\varphi)$ for manipulability and the measure $\delta(\varphi)$ for deficit, where the desideratum is specified via a welfare function. With these in hand, we now formalize and study optimal mechanisms, which are mechanisms that trade off manipulability and deficit optimally.

5.6.1 Signatures of Mechanisms

To compare different mechanisms, we introduce signatures.

Definition 50 (Signature)**.** Given a problem (N, M, δ) and a mechanism φ, the tuple $(\varepsilon(\varphi), \delta(\varphi)) \in [0, 1] \times [0, 1]$ is called the *signature of φ (in the problem (N, M, δ))*.[2]

Signatures allow a convenient graphical representation of the performance of any mechanism in terms of manipulability and deficit. Figure 5.2 gives examples of such signatures: since 0-approximate strategyproofness is equivalent to strategyproofness, the signature of any strategyproof mechanism must have value 0 in the first component. On

[2]Since we fix a problem (N, M, δ), we refer to the tuple $(\varepsilon(\varphi), \delta(\varphi))$ simply as the *signature of φ*, keeping in mind that a mechanism's signature may be different for different problems.

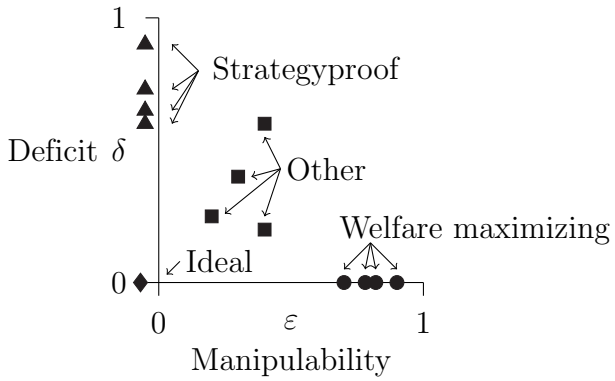

Figure 5.2: Manipulability-deficit-plot with mechanism signatures: strategyproof (▲), welfare maximizing (●), ideal (♦), others (■).

the other hand, any welfare maximizing mechanism must have a signature with a second component of 0. If an ideal mechanism exists, it has a signature in the origin $(0, 0)$. Mechanisms that are neither strategyproof, nor efficient have signatures in the half-open unit square $(0, 1] \times (0, 1]$.

Remark 17. Constant mechanisms are strategyproof. Therefore, at least one signature lies on the δ-axis. Similarly, at each preference profile at least one alternative is welfare maximizing, and a mechanism that selects these alternatives is at least 1-approximate strategyproofness. Therefore, there exists at least one signature on the ε-axis.

5.6.2 Definition and Existence of Optimal Mechanisms

When impossibility results prohibit the design of ideal mechanisms, the decision in favor of any mechanism necessarily involves a trade-off between manipulability and deficit. To make an informed decision about this trade-off, a mechanism designer must be aware of the different design options. One straightforward approach to this problem is to decide on a maximal acceptable manipulability $\varepsilon \in [0, 1]$ up front and use a mechanism that minimizes the deficit among all ε-approximately strategyproof mechanisms. We define

$$\delta(\varepsilon) = \min \{\delta(\varphi) \mid \varphi \ \varepsilon\text{-approximately strategyproof mechanism}\} \tag{355}$$

as the lowest deficit that is achievable by any ε-approximately strategyproof mechanism.

Definition 51 (Optimal Mechanism)**.** Given a problem (N, M, δ) and a bound $\varepsilon \in [0, 1]$, a mechanism φ is *optimal at* ε if φ is ε-approximately strategyproof and $\delta(\varphi) = \delta(\varepsilon)$. We denote by $\textsc{Opt}(\varepsilon)$ the set of all mechanisms that are optimal at ε. Any optimal

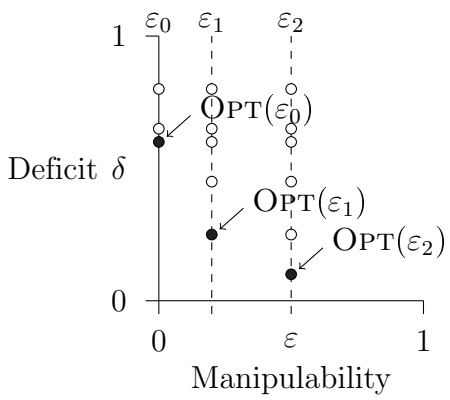

Figure 5.3: Example signatures of optimal mechanisms (•) and non-optimal mechanisms (○) at manipulability bounds $\varepsilon_0, \varepsilon_1, \varepsilon_2$.

mechanism $\varphi \in \textsc{Opt}(\varepsilon)$ is called a *representative of* $\textsc{Opt}(\varepsilon)$.

Proposition 16 shows that optimal mechanisms always exist.

Proposition 16. *Given a problem* (N, M, δ) *and a manipulability bound* $\varepsilon \in [0, 1]$*, there exists at least one mechanism that is optimal at* ε*.*

Existence follows via a compactness argument (see Appendix 5.E.2 for a proof).

Proposition 16 yields the existence of optimal mechanisms for any manipulability bound ε. Thus, it justifies the use of the minimum (rather than the infimum) in the definition of $\delta(\varepsilon)$, since the deficit $\delta(\varepsilon) = \delta(\varphi)$ is in fact attained by some mechanism (namely the optimal one that exists by Proposition 16). Figure 5.3 illustrates optimal mechanisms in terms of their signatures. On the vertical lines at each of the manipulability bounds $\varepsilon_0, \varepsilon_1, \varepsilon_2$, the circles correspond to signatures of non-optimal mechanisms which incur an "unnecessary" deficit, given the respective manipulability bound. The signatures of optimal mechanisms from $\textsc{Opt}(\varepsilon_k), k = 0, 1, 2$ take the lowest positions and are represented by black circles.

5.6.3 Identifying Optimal Mechanisms

The existence proof for optimal mechanisms is implicit and does not provide a way of actually determining optimal mechanisms. Our next result characterizes the set $\textsc{Opt}(\varepsilon)$ as the set of solutions to a linear optimization problem. Using the following linear program, we can solve this problem algorithmically to find representatives of $\textsc{Opt}(\varepsilon)$ and compute the value $\delta(\varepsilon)$.

Linear Program 1 (FindOpt).

$$\text{minimize } \mathtt{d} \qquad \textit{(Objective)}$$

subject to

$$\sum_{j \in M : r_{P_i}(j) \leqslant k} \mathtt{f_j}(\mathtt{P'_i}, \mathtt{P_{-i}}) - \mathtt{f_j}(\mathtt{P_i}, \mathtt{P_{-i}}) \leqslant \varepsilon, \qquad \textit{(}\varepsilon\textit{-approximate SP)}$$

$$\forall i \in N, (P_i, P_{-i}) \in \mathcal{P}^N, P'_i \in \mathcal{P}, k \in \{1, \ldots, m\}$$

$$\mathtt{d} \geqslant \delta(\mathtt{f}), \qquad \textit{(Deficit)}$$

$$\sum_{j \in M} \mathtt{f_j}(\mathtt{P}) = 1, \qquad \forall \boldsymbol{P} \in \mathcal{P}^N \qquad \textit{(Probability)}$$

$$\mathtt{f_j}(\mathtt{P}) \in [0, 1], \qquad \forall \boldsymbol{P} \in \mathcal{P}^N, j \in M \qquad \textit{(Outcome variables)}$$

$$\mathtt{d} \in [0, 1], \qquad \textit{(Deficit variable)}$$

Each variable $\mathtt{f_j}(\mathtt{P})$ corresponds to the probability with which the mechanism φ selects alternative j if the agents report preference profile $\boldsymbol{P}$. Consequently, any assignment of the variables $\{\mathtt{f_j}(\mathtt{P}) : j \in M, \boldsymbol{P} \in \mathcal{P}^N\}$ corresponds to a unique mapping $\varphi : M \times \mathcal{P}^N \to \mathbb{R}^M$. The two constraints labeled *(Probability)* and *(Outcome variables)* ensure that the variable assignment in fact corresponds to a *mechanism* (rather than just a mapping). The variable $\mathtt{d}$ represents this mechanism's deficit and the objective is to minimize its value.

The set of constraints labeled *(ε-approximate SP)* reflects the equivalent conditions for ε-approximate strategyproofness that we obtained from Theorem 19. In combination, these constraints ensure that the mechanisms corresponding to the feasible variable assignments of FindOpt are *exactly* the ε-approximately strategyproof mechanisms. The constraint labeled *(Deficit)* is a placeholder and takes different forms, depending on whether the worst-case or the ex-ante notion of deficit is considered. To minimize worst-case deficit, we need constraints that make $\mathtt{d}$ an upper bound for the deficit of φ at any preference profile $\boldsymbol{P} \in \mathcal{P}^N$. This is achieved by replacing "$\mathtt{d} \geqslant \delta(\mathtt{f})$" with

$$\mathtt{d} \geqslant \max_{j \in M} w(j, \boldsymbol{P}) - \sum_{j \in M} \mathtt{f_j}(\mathtt{P}) \cdot w(j, \boldsymbol{P}), \qquad \forall \boldsymbol{P} \in \mathcal{P}^N. \qquad (\textit{Worst-case deficit})$$

The ex-ante deficit $\delta^{\mathbb{P}}(\varphi)$ (with respect to the distribution $\mathbb{P}$ over preference profiles) is a linear function of the mechanism φ, and we can replace "$\mathtt{d} \geqslant \delta(\mathtt{f})$" with the single

constraint

$$\mathtt{d} \geqslant \sum_{\boldsymbol{P} \in \mathcal{P}^N} \mathbb{P}[\boldsymbol{P}] \cdot \left(\max_{j \in M} w(j, \boldsymbol{P}) - \sum_{j \in M} \mathtt{f_j}(\mathtt{P}) \cdot w(j, \boldsymbol{P}) \right). \qquad \textit{(Ex-ante deficit)}$$

The following Proposition 17 formalizes the correspondence between optimal mechanisms and the solutions to the linear program FindOpt.

Proposition 17. *Given a problem* (N, M, δ) *and a bound* $\varepsilon \in [0, 1]$*, an assignment of the variables* $\{\mathtt{f_j}(\mathtt{P}) : j \in M, \boldsymbol{P} \in \mathcal{P}^N\}$ *is a solution to* FindOpt *at* ε if and only if *the mechanism* φ *defined by* $\varphi_j(\boldsymbol{P}) = \mathtt{f_j}(\mathtt{P})$ *for all* $j \in M, \boldsymbol{P} \in \mathcal{P}^N$ *is optimal at* ε*.*

The proof follows directly from the discussion above. One important consequence of Proposition 17 is that we can now compute optimal mechanisms for any given problem (N, M, δ) and any manipulability bound $\varepsilon \in [0, 1]$. Going back to the mechanism designer's problem of choosing a mechanism that makes attractive trade-offs between manipulability and deficit, we now have a way of determining optimal mechanisms for particular manipulability bounds ε. The linear program FindOpt allows the mechanism designer to evaluate algorithmically what deficits she must accept when she wants to ensure that the manipulability does not exceed some fixed limit ε.

Shifting the burden of design to a computer by encoding good mechanisms in optimization problems is the central idea of automated mechanism design (Sandholm, 2003). A common challenge with this approach is that the optimization problem can become large and difficult to solve; and naïve implementations of FindOpt will face this issue as well. Substantial run-time improvements are possible, e.g., by exploiting additional axioms such as anonymity and neutrality (Mennle, Abaecherli and Seuken, 2015). Nonetheless, determining optimal mechanisms remains a computationally expensive operation.

Computational considerations aside, Proposition 17 provides a new understanding of optimal mechanisms: since the set $\textsc{Opt}(\varepsilon)$ corresponds to the solutions of the linear program FindOpt, $\textsc{Opt}(\varepsilon)$ can be interpreted as a convex polytope. Equipped with this understanding, we will use methods from convex geometry to derive our structural characterization of the Pareto frontier in Section 5.8. The representation of optimal mechanisms as solutions to FindOpt constitutes the first building block of this characterization.

5.7 Hybrid Mechanisms

In this section, we introduce a simple method for creating mechanisms with intermediate signatures. *Hybrid* mechanisms are convex combinations of two *component mechanisms*. Intuitively, by mixing one mechanism with low manipulability and another mechanism with low deficit, we may hope to obtain new mechanisms with intermediate properties on both dimensions. We now formalize this intuition. The construction of hybrid mechanisms is initially independent of the study of optimal mechanisms. However, in Section 5.8, we will see that they constitute the second building block for our structural characterization.

Definition 52 (Hybrid)**.** For $\beta \in [0,1]$ and mechanisms φ and ψ, the *β-hybrid* h_β is given by

$$h_\beta(\boldsymbol{P}) = (1-\beta)\varphi(\boldsymbol{P}) + \beta\psi(\boldsymbol{P}) \tag{356}$$

for any preference profile $\boldsymbol{P} \in \mathcal{P}^N$.

In practice, "running" a hybrid mechanism is straightforward: first, collect the preference reports. Second, toss a β-coin to determine whether to use ψ (probability β) or φ (probability $1-\beta$). Third, apply this mechanism to the reported preference profile.

Our next result shows that the signatures of β-hybrids are always weakly better than the β-convex combination of the signatures of the two component mechanisms.

Theorem 20. *Given a problem (N, M, δ), two mechanisms φ, ψ, and $\beta \in [0,1]$, we have*

$$\varepsilon(h_\beta) \leqslant (1-\beta)\varepsilon(\varphi) + \beta\varepsilon(\psi), \tag{357}$$

$$\delta(h_\beta) \leqslant (1-\beta)\delta(\varphi) + \beta\delta(\psi). \tag{358}$$

Proof Outline (formal proof in Appendix 5.E.3). We write out the definitions of $\varepsilon(h_\beta)$ and $\delta(h_\beta)$, each of which may involve taking a maximum. The two inequalities are then obtained with the help of the triangle inequality. □

The following example shows that the inequalities in Theorem 20 can indeed be strict.

Example 18. Consider a problem with one agent and three alternatives a, b, c, where δ is the worst-case deficit that arises from Plurality scoring. Let φ and ψ be two mechanisms whose outcomes depend only on the agent's relative ranking of b and c.

	φ			ψ		
Report	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$
If $P: b \succeq c$	0	2/3	1/3	5/9	1/9	1/3
If $P: c \succ b$	1/3	1/3	1/3	1/9	5/9	1/3

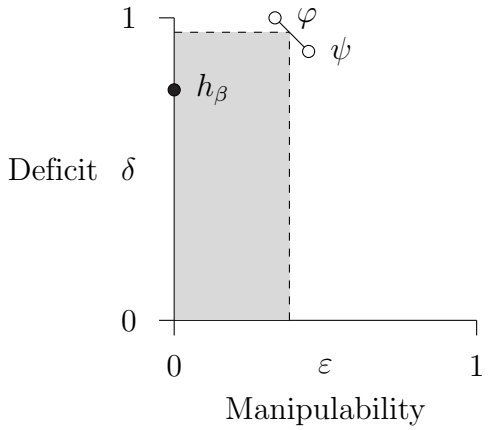

Figure 5.4: Example signature of hybrid h_β ($\bullet$) must lie in the shaded area, weakly to the left and below the β-convex combinations of the signatures of φ and ψ ($\circ$).

It is a simple exercise to verify that the mechanisms' signatures are $(\varepsilon(\varphi), \delta(\varphi)) = (1/3, 1)$ and $(\varepsilon(\psi), \delta(\psi)) = (4/9, 8/9)$. Furthermore, for $\beta = 3/7$, the hybrid h_β is constant and therefore strategyproof, and it has a deficit of $\delta(h_\beta) = 16/21$. Figure 5.4 illustrates these signatures. The guarantees from Theorem 20 are represented by the shaded area, where the signature of h_β can lie anywhere below and to the left of the β-convex combination of the signatures of φ and ψ.

There are two important takeaways from Theorem 20. First, it yields a strong argument in favor of randomization: given two mechanisms with attractive manipulability and deficit, randomizing between the two always yields a mechanism with a signature that is at least as attractive as the β-convex combination of the signatures of the two mechanisms. As Example 18 shows, randomizing in this way can sometimes yield *strictly* preferable hybrid mechanisms that have lower manipulability *and* lower deficit than either of the original mechanisms.

The second takeaway is that the common fairness requirement of *anonymity* comes "for free" in terms of manipulability and deficit (provided that the deficit measure δ is itself anonymous): intuitively, given any mechanism φ, an anonymous mechanism can be constructed by randomly assigning the agents to new roles. This yields a hybrid mechanism with many components, each of which corresponds to the original mechanism with agents assigned to new positions. Under an anonymous measure of deficit, every component will have the same signature as φ. If follows from Theorem 20 that this new, anonymous mechanism has a weakly better signature than φ. Similarly, we can impose

neutrality without having to accept higher manipulability or more deficit (if δ is itself neutral). We formalize these insights in Appendix 5.B.

5.8 The Pareto Frontier

Recall that optimal mechanisms are those mechanisms that trade off manipulability and deficit optimally. These mechanisms form the Pareto frontier because with them, one cannot achieve a strictly lower deficit without accepting strictly higher manipulability.

Definition 53 (Pareto Frontier)**.** Given a problem (N, M, δ), let $\bar{\varepsilon}$ be the smallest manipulability bound that is compatible with welfare maximization; formally,

$$\bar{\varepsilon} = \min\{\varepsilon \in [0,1] \mid \exists \varphi : \varphi \text{ welfare maximizing } \& \ \varepsilon\text{-approximately SP}\}. \tag{359}$$

The *Pareto frontier* is the union of all mechanisms that are optimal for some manipulability bound $\varepsilon \in [0, \bar{\varepsilon}]$; formally,

$$\textsc{Pf} = \bigcup_{\varepsilon \in [0,\bar{\varepsilon}]} \textsc{Opt}(\varepsilon). \tag{360}$$

The special manipulability bound $\bar{\varepsilon}$ is chosen such that maximal welfare can be achieved with an $\bar{\varepsilon}$-approximately strategyproof mechanisms, $\bar{\varphi}$ say, but not with any strictly less manipulable mechanism. Since $\bar{\varphi}$ has deficit 0, any mechanism that is optimal at some larger manipulability bound $\varepsilon > \bar{\varepsilon}$ may be more manipulable than $\bar{\varphi}$, but it cannot have a strictly lower deficit. This motivates restricting our attention to manipulability bounds between 0 and $\bar{\varepsilon}$ (instead of 0 and 1). From the perspective of the mechanism designer, mechanisms on the Pareto frontier are the only mechanisms that should be considered; if a mechanism is *not* on the Pareto frontier, we can find another mechanism that is a Pareto-improvement in the sense that it has strictly lower manipulability, strictly lower deficit, or both.

5.8.1 Monotonicity and Convexity

Recall that we have defined $\delta(\varepsilon)$ as the smallest deficit that can be attained by any ε-approximately strategyproof mechanism. Thus, the signatures of mechanisms on the Pareto frontier are described by the mapping $\varepsilon \mapsto \delta(\varepsilon)$ that associates each manipulability bound $\varepsilon \in [0, \bar{\varepsilon}]$ with the deficit $\delta(\varepsilon)$of the mechanisms that are optimal mechanisms at this manipulability bound. Based on our results so far, we can make interesting observations about the Pareto frontier by analyzing this mapping.

Corollary 7. *Given a problem (N, M, δ), the mapping $\varepsilon \mapsto \delta(\varepsilon)$ is monotonically decreasing and convex.*

Monotonicity follows from the definition of optimal mechanisms, and convexity is a consequence of Theorem 20 (see Appendix 5.E.4 for a formal proof). While monotonicity is merely reassuring, convexity is non-trivial and very important. It means that when we relax strategyproofness, the first unit of manipulability that we give up allows the largest reduction of deficit. For any additional unit of manipulability that we sacrifice, the deficit will be reduced at a lower rate, which means *decreasing marginal returns*. Thus, we can expect to capture most welfare gains from relaxing strategyproofness early on. Moreover, convexity and monotonicity together imply *continuity*. This means that trade-offs along the Pareto frontier are *smooth* in the sense that a tiny reduction of the manipulability bound ε does not require accepting a vastly higher deficit.

For mechanism designers, these observations deliver a very important lesson: if welfare gains can be obtained by relaxing strategyproofness, then the most substantial gains will be unlocked by relaxing incentive constraints just "a little bit." This provides encouragement to investigate gains from accepting even small amounts of manipulability. On the other hand, if initial gains are not worth the sacrifice, then gains from accepting even more manipulability will not be "surprisingly" attractive either.

5.8.2 Structural Characterization of the Pareto Frontier

In Section 5.6, we have shown that we can identify isolated mechanisms on the Pareto frontier by solving the linear program FindOpt for a given manipulability bound ε. In Section 5.7, we have introduced hybrids, and we have shown how mixing two mechanisms can result in new mechanisms with intermediate or even superior signatures. We now give our main result, a structural characterization of the Pareto frontier in terms of these two building blocks, namely *optimal mechanisms* and *hybrids*.

Theorem 21. *Given a problem (N, M, δ), there exists a finite set of* supporting manipulability bounds

$$\varepsilon_0 = 0 < \varepsilon_1 < \ldots < \varepsilon_{K-1} < \varepsilon_K = \bar{\varepsilon}, \tag{361}$$

such that for any $k \in \{1, \ldots, K\}$ and any $\varepsilon \in [\varepsilon_{k-1}, \varepsilon_k]$ with $\beta = \frac{\varepsilon - \varepsilon_{k-1}}{\varepsilon_k - \varepsilon_{k-1}}$ we have that

$$\textsc{Opt}(\varepsilon) = (1-\beta)\textsc{Opt}(\varepsilon_{k-1}) + \beta\textsc{Opt}(\varepsilon_k), \tag{362}$$

$$\delta(\varepsilon) = (1-\beta)\delta(\varepsilon_{k-1}) + \beta\delta(\varepsilon_k). \tag{363}$$

Proof Outline (formal proof in Appendix 5.E.5). Our proof exploits that $\textsc{Opt}(\varepsilon)$ corresponds to the solutions of the linear program $\textsc{FindOpt}$ (Section 5.6.3) with feasible set $F_\varepsilon = \{x \,|\, Dx \leqslant d, Ax \leqslant \varepsilon\}$, where neither D, nor d, nor A depend on ε. First, we show that if a set of constraints is binding for F_ε, then it is binding for $F_{\varepsilon'}$ for ε' within a compact interval $[\varepsilon^-, \varepsilon^+]$ that contains ε and not binding for any $\varepsilon'' \notin [\varepsilon^-, \varepsilon^+]$. With finiteness of the number of constraints of the LP, this yields a finite segmentation of $[0, \bar{\varepsilon}]$. The vertex-representations (Grünbaum, 2003) can then be used to show that on each segment $[\varepsilon_{k-1}, \varepsilon_k]$, the *solution sets* $S_\varepsilon = \operatorname{argmin}_{F_\varepsilon} \mathtt{d}$ are exactly the β-convex combinations of $S_{\varepsilon_{k-1}}$ and S_{ε_k} with $\beta = \frac{\varepsilon - \varepsilon_{k-1}}{\varepsilon_k - \varepsilon_{k-1}}$. □

It would be particularly simple if the optimal mechanisms at some manipulability bound ε were just the β-hybrids of optimal mechanisms at 0 and $\bar{\varepsilon}$ with $\beta = \varepsilon/\bar{\varepsilon}$. While this straw-man hypothesis is not true in general, Theorem 21 shows that the Pareto frontier has a linear structure over each element of a finite segmentation of the interval $[0, \bar{\varepsilon}]$. It is completely identified by two building blocks: (1) the *sets of optimal mechanisms* $\textsc{Opt}(\varepsilon_k)$ for finitely many $\varepsilon_k, k \in \{0, \ldots, K\}$, and (2) *hybrid mechanisms*, which provide the missing link for $\varepsilon \neq \varepsilon_k, k \in \{0, \ldots, K\}$. Representatives of $\textsc{Opt}(\varepsilon_k)$ can be obtained by solving the corresponding linear program $\textsc{FindOpt}$ at the supporting manipulability bounds ε_k; and for intermediate bounds $\varepsilon = (1-\beta)\varepsilon_{k-1} + \beta\varepsilon_k$, we have shown that a mechanism φ is optimal at ε *if and only if* it is a β-hybrid of two mechanisms $\varphi_{k-1} \in \textsc{Opt}(\varepsilon_{k-1})$ and $\varphi_k \in \textsc{Opt}(\varepsilon_k)$.

Theorem 21 allows an additional insight about the mapping $\varepsilon \mapsto \delta(\varepsilon)$ (where $\delta(\varepsilon) = \min\{\delta(\varphi) \mid \varphi\ \varepsilon\text{-approximately strategyproof mechanism}\}$). We have already observed that this mapping must be monotonic, decreasing, convex, and continuous (see Corollary 7). In addition, we obtain piecewise linearity.

Corollary 8. *Given a problem (N, M, δ), the mapping $\varepsilon \mapsto \delta(\varepsilon)$ is piecewise linear.*

Figure 5.5 illustrates the results of Theorem 21 and Corollary 8: first, the mapping $\varepsilon \mapsto \delta(\varepsilon)$ is monotonic, decreasing, convex, continuous, and piecewise linear, and, second, optimal mechanisms with intermediate manipulability can be obtained by mixing two mechanisms that are optimal at the two adjacent supporting manipulability bounds, respectively. Both Theorem 21 and Corollary 8 are especially interesting from a computational perspective because they yield a finite representation of the Pareto frontier. In the following Section 5.9, we give algorithms that exploits this structure to identify all supporting manipulability bounds and therefore the whole Pareto frontier.

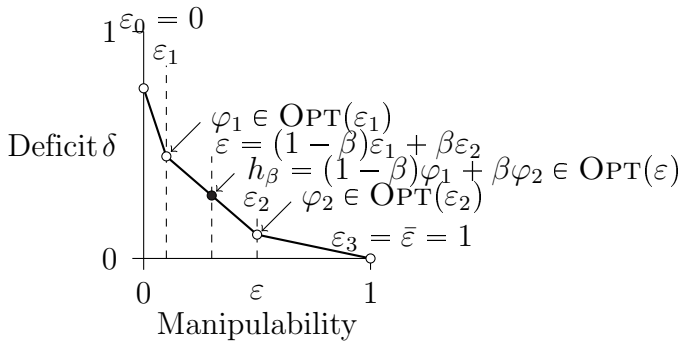

Figure 5.5: Illustrative example of the signatures of the Pareto frontier.

5.9 Computing the Pareto Frontier: Algorithms and Examples

In this section, we give algorithms to compute all supporting manipulability bounds, and we then apply our findings to two example problems.

5.9.1 The Algorithm FindBounds

Recall that the linear program FindOpt can be used to determine a representative of the set of optimal mechanisms $\text{Opt}(\varepsilon)$ at any manipulability bound ε. One naïve approach to finding the Pareto frontier would be to run FindOpt for many different manipulability bounds to obtain optimal mechanisms at each of these bounds, and then consider these mechanisms and their hybrids. However, this method has two drawbacks: first, and most importantly, it would not yield the correct Pareto frontier. The result can, at best, be viewed as a *conservative estimate* of the Pareto frontier, since choosing fixed manipulability bounds is not guaranteed to identify any actual supporting manipulability bounds exactly. Second, from a computational perspective, executing FindOpt is expensive, which is why we would like to keep the number of executions as low as possible.

Knowing the structure of the Pareto frontier, we can take a more refined approach. The algorithm FindBounds (Algorithm 4 in Appendix 5.C) applies FindOpt sequentially to determine the signatures of optimal mechanisms at different manipulability bounds. However, instead of exploring a fixed or random set of bounds, it uses the information obtained in each step to interpolate the most promising candidates for supporting manipulability bounds.

Before we state our formal result about the correctness, completeness, and runtime of FindBounds, we provide an outline of how the algorithm works (see Appendix 5.C for a pseudocode implementation). Observe that by Theorem 21, the problem of computing all supporting manipulability bounds is equivalent to identifying the path of the piecewise monotonic, convex, linear function $\varepsilon \mapsto \delta(\varepsilon)$. FindBounds keeps an inventory of "known signatures" and "verified segments"on this path. It uses these to interpolate and verify new segments.

Interpolation: Suppose that we know four points $s_0 = (\varepsilon_0, \delta_0), s_1 = (\varepsilon_1, \delta_1), s_2 = (\varepsilon_2, \delta_2), s_3 = (\varepsilon_3, \delta_3)$ with $\varepsilon_0 < \varepsilon_1 < \varepsilon_2 < \varepsilon_3$ on the path. If the two segments $\overline{s_0\ s_1}$ and $\overline{s_2\ s_3}$ are linearly independent, their affine hulls have a unique point of intersection ε' over the interval $[\varepsilon_1, \varepsilon_2]$. This point of intersection is the candidate for a new supporting manipulability bound that FindBounds considers. The left plot in Figure 5.6 illustrates the geometry of this step.

Verification: Once we have identified the candidate ε', we use FindOpt to compute the deficit $\delta' = \delta(\varepsilon')$. The signature (ε', δ') is either on or below the straight line connecting the signatures $(\varepsilon_1, \delta_2)$ and $(\varepsilon_2, \delta_2)$. In the first case, if all three signatures lie on a single straight line segment, then we can infer that this line segment must be part of the path of $\varepsilon \mapsto \delta(\varepsilon)$. We can mark this segment as "verified," because there are no supporting manipulability bounds in the open interval $(\varepsilon_1, \varepsilon_2)$ (but possibly at its limits). In the second case, if (ε', δ') lies strictly below the line segment, then there must exist at least one supporting manipulability bound in the open interval $(\varepsilon_1, \varepsilon_2)$ at which the path has a real "kink." The right plot in Figure 5.6 illustrates the geometry of this step. The signature (ε', δ') must lie somewhere on the gray vertical line at ε'.

In this way FindBounds repeatedly identifies and verifies line segments on the path of $\varepsilon \mapsto \delta(\varepsilon)$. If no additional supporting manipulability bounds exist on a segment, then this segment is verified. Otherwise, one more point (ε', δ') is added to the collection of known signatures, and the algorithm continues to interpolate.

Proposition 18. *Given a problem (N, M, δ), the algorithms* FindLower *and* FindBounds *require at most $4K + \log_2(1/\varepsilon_1) - 1$ executions of the linear program* FindOpt *to determine all supporting manipulability bounds of the Pareto frontier, where K is the number of supporting manipulability bounds and ε_1 is the smallest non-trivial supporting*

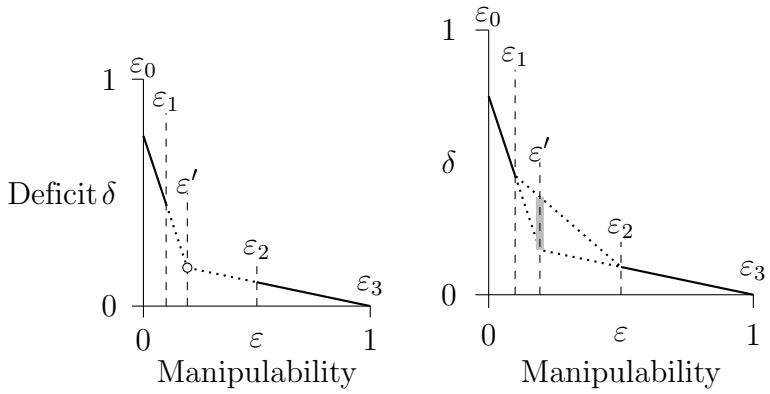

Figure 5.6: Illustration of interpolation (left) and verification step (right).

manipulability bound.[3]

Proof Outline (formal proof in Appendix 5.E.6). The challenge is showing that the algorithm makes at most 3 "wrong guesses" before it correctly identifies a new supporting manipulability bound. Intuitively, this is possible because once the algorithm has found three points that are linearly dependent, by Theorem 21, there cannot be any supporting manipulability bound between the two outer points. Furthermore, there must be exactly two supporting manipulability bounds in their affine hull, a fact which Algorithm 4 exploits in the interpolation. □

The interpolation process in FindBounds is set up in such a way that at most 4 executions of FindOpt are required for finding each supporting manipulability bound. Thus, assuming that representatives of $\text{Opt}(\varepsilon_k)$ must be determined by executing FindOpt at each ε_k, FindBounds is at most a constant factor 4 slower than an algorithm that correctly "guesses" all supporting manipulability bounds and uses FindOpt for these only.

5.9.2 Examples: Plurality vs. Veto Scoring

In this section, we consider two concrete problems and derive the respective Pareto frontiers. The two examples highlight the different shapes that the Pareto frontier can take. Moreover, they uncover an interesting connection to the characterization of

[3] For technical reasons, FindBounds assumes knowledge of some value $\underline{\varepsilon} \in (0, \varepsilon_1]$, i.e., some strictly positive lower bound for the smallest non-trivial supporting manipulability bound. This is needed to initialize the interpolation process. In Appendix 5.C we also give the algorithm FindLower (Algorithm 5) that identifies a suitable value of $\underline{\varepsilon}$.

strategyproof mechanisms by Gibbard (1977). Both settings contain 3 alternatives and 3 agents with strict preferences over the alternatives, but they differ in the desideratum.

Plurality and Random Dictatorships

Consider a setting with agents $N = \{1, 2, 3\}$, alternatives $M = \{a, b, c\}$, and where agents' preferences are strict. Suppose further that the desideratum is to select alternatives that are the first choice of many agents. From Example 16 we know that the corresponding welfare function is $w^{\text{PLU}}(j, \boldsymbol{P}) = n_j^1/n$, where n_j^1 is the number of agents whose first choice under $\boldsymbol{P}$ is j. Finally, suppose that the deficit δ is the worst-case w^{PLU}-deficit; formally,

$$\delta(\varphi) = \max_{\boldsymbol{P} \in \mathcal{P}^N} \left(\max_{j \in M} \left(w^{\text{PLU}}(j, \boldsymbol{P}) \right) - w^{\text{PLU}}(\varphi(\boldsymbol{P}), \boldsymbol{P}) \right). \tag{364}$$

Using FindBounds, we find that in this problem, the Pareto frontier has only the two extreme supporting manipulability bounds $\varepsilon_0 = 0$ and $\bar{\varepsilon} = \varepsilon_1 = 1/3$. Thus, it consists precisely of the hybrids of mechanisms that are optimal at ε_0 and ε_1 by Theorem 21. Formally,

$$\text{Opt}(\beta/3) = (1 - \beta)\text{Opt}(0) + \beta\text{Opt}(1/3) \tag{365}$$

for any $\beta \in [0, 1]$. Moreover, we can use FindOpt to obtain representatives of Opt(0) and Opt(1/3). Interestingly, the representatives determined by the algorithm have a familiar structure: first, the representative of Opt(0) corresponds to Random Dictatorship with a uniform choice of the dictator. Thus, in this problem, no strategyproof mechanism has lower deficit than Random Dictatorship. Second, the representative of Opt(1/3) corresponds to *Uniform Plurality*, a mechanism that determines the set of welfare maximizing alternatives and selects one of these uniformly at random. Thus, in this problem, no welfare maximizing mechanism has lower manipulability than Uniform Plurality. Figure 5.7 depicts the signatures of the Pareto frontier.

In addition, we prove these insights analytically.

Proposition 19. *In a problem (N, M, δ) with $n = 3$ agents, $m = 3$ alternatives, strict preferences, and where δ is the worst-case w^{PLU}-deficit, the following hold:*

1. *The Pareto frontier has two supporting manipulability bounds $\{0, 1/3\}$.*
2. *Random Dictatorship is a representative of* Opt(0).
3. *Uniform Plurality is a representative of* Opt(1/3).

The formal proof is given in Appendix 5.E.7.

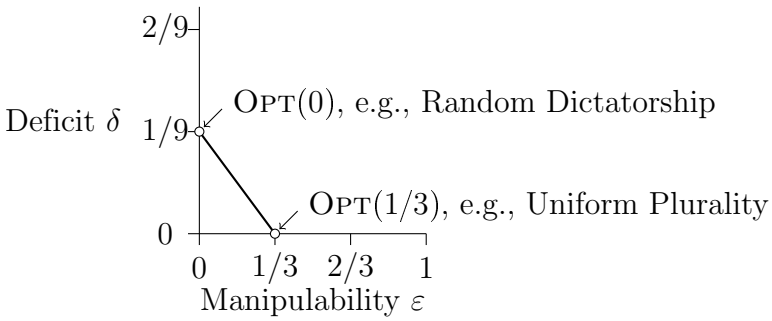

Figure 5.7: Plot of signatures of the Pareto frontier for $n = 3$ agents, $m = 3$ alternatives, strict preferences, and worst-case w^{Plu}-deficit δ.

Veto and Random Duple

The second problem again involves a setting with agents $N = \{1, 2, 3\}$, alternatives $M = \{a, b, c\}$, and where agents' preferences are strict. The difference of this problem to the previous one is the different desideratum: this time, our goal is to select an alternatives that is the *last* choice of *few* agents. This desideratum is reflected by the Veto scoring function, and the corresponding welfare function is given by

$$w^{\text{Veto}}(j, \boldsymbol{P}) = \frac{n - n_j^m}{n}, \tag{366}$$

where n_j^m is the number of agents who ranked j as their last choice under $\boldsymbol{P}$. We use this welfare function to define the worst-case w^{Veto}-deficit δ; formally,

$$\delta(\varphi) = \max_{\boldsymbol{P} \in \mathcal{P}^N} \left(\max_{j \in M} \left(w^{\text{Veto}}(j, \boldsymbol{P}) \right) - w^{\text{Veto}}(\varphi(\boldsymbol{P}), \boldsymbol{P}) \right). \tag{367}$$

Again, we use FindBounds to determine the supporting manipulability bounds of the Pareto frontier in this problem. These are

$$\varepsilon_0 = 0, \varepsilon_1 = 1/21, \varepsilon_2 = 1/12, \bar{\varepsilon} = \varepsilon_3 = 1/2. \tag{368}$$

As in the previous example, we can compute representatives of the optimal mechanisms at each of these bounds. For the extreme bounds $\varepsilon_0 = 0$ and $\bar{\varepsilon} = 1/2$, these representatives again have a familiar structure: first, the representative of Opt(0) corresponds to *Random Duple*, a mechanism that picks two alternatives uniformly at random and then selects the one that is preferred to the other by a majority of the agents (breaking ties randomly). Second, the representative of Opt(1/2) corresponds to *Uniform Veto*, a mechanism that

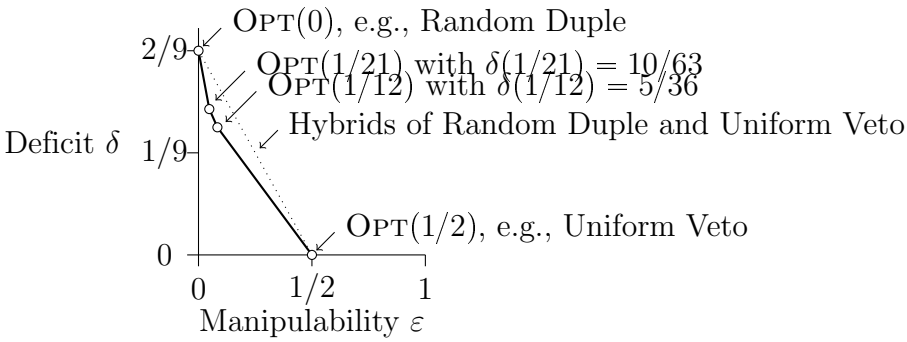

Figure 5.8: Plot of signatures of the Pareto frontier for $n = 3$ agents, $m = 3$ alternatives, strict preferences, and worst-case w^{Veto}-deficit δ.

finds all alternatives that are the last choice of a minimal number of agents and selects one of these alternatives uniformly at random. Figure 5.8 depicts the signatures of the Pareto frontier in this problem.

We also created representatives of Opt(1/21) and Opt(1/12) via FindOpt (they are given in Appendix 5.E.8 in their numerical form). While the interpretation of these two mechanisms is not straightforward, it is clear that neither of them is a hybrid Random Duple and Uniform Veto. Indeed, in order to generate mechanisms on the Pareto frontier, we cannot simply consider hybrids of optimal mechanisms from the extreme supporting manipulability bounds. Instead, we can (and have to) exploit the additional design freedom in the particular problem by separately designing optimal mechanisms for the two intermediate supporting manipulability bounds $\varepsilon_1 = 1/21$ and $\varepsilon_2 = 1/12$ specifically.

As in the previous example, we convince ourselves of the correctness of these assertions by proving them analytically.

Proposition 20. *In a problem (N, M, δ) with $n = 3$ agents, $m = 3$ alternatives, strict preferences, and where δ is the worst-case w^{Veto}-deficit, the following hold:*

1. *The Pareto frontier has four supporting manipulability bounds $\{0, 1/21, 1/12, 1/2\}$.*
2. *Random Duple is a representative of* Opt(0).
3. *Uniform Veto is a representative of* Opt(1/2).
4. *Hybrids of Random Duple and Uniform Veto are not on the Pareto frontier, except for $\beta \in \{0, 1\}$.*

The formal proof is given in Appendix 5.E.8.

Connection to Gibbard's Characterization of Strategyproof Mechanisms

Juxtaposing the two examples shows an interesting connection to earlier work that characterized strategyproof random mechanisms: Gibbard (1977) showed that in a domain with three or more alternatives and strict preferences, any mechanism that is strategyproof must be a probability mixture (i.e., a hybrid with an arbitrary number of components) of strategyproof *unilateral* mechanisms (i.e., whose random outcomes depend only on the preference order of a single agent) and strategyproof *duple* mechanisms (i.e., which assign positive probability to at most two alternatives).

Our examples show that the optimal strategyproof mechanisms for the particular problems are exactly the two extremes of this representation: when the desideratum is based on Plurality scoring, optimal strategyproof mechanism arise by randomizing over unilateral mechanisms only, namely dictatorships. Conversely, when the desideratum is based on Veto scoring, optimal strategyproof mechanisms arise by randomizing over duple mechanisms only, namely the majority vote between two alternatives. Thus, our two examples teach us in what sense unilateral and duple mechanisms can be understood to reside on "opposite ends" of the spectrum of strategyproof mechanisms.

To prove Propositions 19 and 20, we have used a *symmetric decomposition* theorem, which is a refinement of Gibbard's strong characterization for *anonymous, neutral, strategyproof* mechanisms. This theorem may be of independent interest and is given in Appendix 5.B.4.

5.10 Alternative Domains and Desiderata

With the exception of the examples in Sections 5.9.2, we have formulated our results for the general domain with weak preferences. In the following Section 5.10.1, we explain how our findings continue to hold in many other interesting domains. Subsequently, in Section 5.10.2, we discuss the expressiveness of welfare functions for the purpose of encoding desiderata.

5.10.1 Domain Restrictions

We have proven our results (Theorems 19 through 21 and Proposition 17) for the unrestricted preference domain with indifferences. However, all these results continue to hold when the domain is restricted by only admitting a subset of the possible preference profiles. Formally, let $\boldsymbol{\mathcal{P}} \subseteq \mathcal{P}^N$ be *any* subset of the set of preference profiles; then all

our results still hold, even if we admit only preference profiles from $\boldsymbol{\mathcal{P}}$. This extension is possible because none of our proofs make use of the richness of the full domain.

Important domains that can be expressed as domain restrictions $\boldsymbol{\mathcal{P}} \subset \mathcal{P}^N$ are:

- The *full domain of strict ordinal preferences*, which is the basis for the classical impossibility results pertaining to strategyproofness of deterministic mechanisms (Gibbard, 1973; Satterthwaite, 1975) and random mechanisms (Gibbard, 1977).
- The *assignment domain*, where each agent must be assigned to a single, indivisible object (e.g., a seat at a public school) (Abdulkadiroğlu and Sönmez, 2003*b*).
- The *two-sided matching domain*, where agents of two different types must be matched to each other to form pairs (e.g., men and women, doctors and hospitals, mentors and mentees), and variations of this domain, such as *matching with couples* (Roth, 1984).
- The *combinatorial assignment domain*, where agents receive bundles of indivisible objects (e.g., course schedules) (Budish, 2011)

Of course, the fact that our results continue to hold in these domains does not mean that the actual Pareto frontiers will be equal across the different domains. On the contrary, one would expect optimal mechanisms to be highly adapted to the particular domain in which they are designed to operate.

Remark 18 (Full Utility Assumption). One essential ingredient to our results is the fact that ε-approximate strategyproofness can be equivalently expressed via a finite set of linear constraints. Theorem 19, which yields this equivalence, relies on the assumption of *full utility*: given a preference order P_i, the agent i can have vNM utility function $u_i \in U_{P_i}$ that represents this preference order (and is bounded between 0 and 1). A domain restriction that excludes *subsets of utility functions* would violate this condition. For example, suppose that we imposed the additional restriction that the agent's utility are *normalized* (i.e., $\|u_i\|_2 = \sqrt{\sum_{j \in M} u_i(j)^2} = 1$). This restriction limits the gain that an agent can possibly obtain from misreporting in a non-linear way. The linear constraints from Theorem 19 would be sufficient for ε-approximate strategyproofness. However, they would no longer be necessary, and we would lose *equivalent*.

5.10.2 Encoding Desiderata via Welfare Functions

Welfare functions are a very versatile means to express different design goals that the mechanism designer may have. We have already seen how they can be used to reflect

desiderata based on Plurality scoring (Example 16), Condorcet consistency (Examples 17), or Veto scoring (Section 5.9.2). We now give additional examples that illustrate how welfare functions can be used to encode common desiderata.

Binary Desiderata

Many desirable properties of mechanisms are simply properties of the alternatives they select at any given preference profile. For instance, *Pareto optimality* requires the mechanism to choose alternatives that are not Pareto dominated by any other alternative at the respective preference profile. In general, let Θ be a desirable property that an alternative can have at some preference profile (e.g., Pareto optimality). Welfare functions can capture Θ by setting

$$w(j, \boldsymbol{P}) = \begin{cases} 1, & \text{if alternative } j \text{ has property } \Theta \text{ at } \boldsymbol{P}, \\ 0, & \text{else.} \end{cases} \tag{369}$$

The w-deficit of a mechanism at a given preference profile,

$$\delta_w(\varphi(\boldsymbol{P}), \boldsymbol{P}) = \max_{j \in M} w(j, \boldsymbol{P}) - \sum_{j \in M} \varphi_j(\boldsymbol{P}) \cdot w(j, \boldsymbol{P}), \tag{370}$$

has a straightforward interpretation: it is simply the "probability that an alternative selected by φ at $\boldsymbol{P}$ violates property Θ at $\boldsymbol{P}$." Consequently, if δ is the worst-case deficit, then the probability that φ selects an alternative that violates Θ is at most $\delta(\varphi)$ across all preference profiles. Alternatively, if $\delta^{\mathbb{P}}$ is the ex-ante deficit, then $\delta^{\mathbb{P}}(\varphi)$ is the ex-ante probability (i.e., before the preferences have been reported) that φ selects an alternative that violates Θ, given the prior distribution $\mathbb{P}$ over agents' preference profiles.

Common binary desiderata include

- *unanimity*: if all agents agree on a first choice, this alternative should be selected,
- *Condorcet consistency*: if some alternatives weakly dominate any other alternatives in a pairwise majority comparison, one of these alternatives should be selected,
- *Pareto optimality*: only undominated alternatives should be selected,
- *egalitarian fairness*: let $R(j) = \max_{i \in N} r_{P_i}(j)$ be the rank of alternative j in the preference order of the agent who likes it least, then only alternatives with minimal $R(j)$ should be selected.

While not all of these desiderata are in conflict with strategyproofness individually, their combinations may be. It is possible to encode combinations of binary properties: given two welfare functions w_1 and w_2, the minimum $w = \min(w_1, w_2)$ is again a welfare function, and it expresses the desideratum that both properties (encoded by w_1 and w_2) should hold simultaneously; and the maximum $w = \max(w_1, w_2)$ expresses that at least one of the properties should hold.

Even in the absence of a clear-cut objective, a desideratum may be *implicitly* specified via a *target mechanism*. A target mechanism ϕ is a mechanism that one would like to use if there were no concerns about manipulability (Birrell and Pass, 2011). The welfare function induced by ϕ takes value 1 at the alternatives that are selected by ϕ at the respective preference profiles, or equivalently, $w(j, \boldsymbol{P}) = \phi_j(\boldsymbol{P})$. An analogous construction is possible for a *target correspondence*, which is a mechanism Φ that selects a *set of alternatives*. To reflect a target correspondence, the welfare function can be chosen as $w(j, \boldsymbol{P}) = 1$ if $j \in \Phi(\boldsymbol{P})$ and $w(j, \boldsymbol{P}) = 0$ otherwise, where we denote by $\Phi(\boldsymbol{P})$ the set of alternatives selected by the correspondence Φ at $\boldsymbol{P}$.

In the assignment domain, our notion of Pareto optimality turns into *ex-post efficiency*, which is an important baseline requirement for many assignment mechanisms. Featherstone (2011) introduced *v-rank efficiency*, which is a refinement of ex-post efficiency (and implies ordinal efficiency). He showed that for any *rank valuation* v, the set of v-rank efficient assignments is equal to the convex hull of the set of deterministic v-rank efficient assignments. Thus, v-rank efficiency is representable by a welfare function by setting $w(a, \boldsymbol{P}) = 1$ for all alternatives a that correspond to a v-rank efficient deterministic assignment. In two-sided matching, it is often important to select *stable matchings* to prevent blocking pairs from matching outside the mechanism and thereby causing unraveling of the market. Since stability is a property of deterministic matchings, we can encode this desideratum by a welfare function.

With any such welfare function in hand, we can then define a notion of deficit and use the results in this paper to understand the respective Pareto frontier.

Remark 19 (Linearity Assumption). One example of a desideratum that cannot be represented by a welfare function is the intermediate efficiency requirement of *ordinal efficiency* for assignment mechanisms. To see this, note that ordinal efficiency coincides with ex-post efficiency on deterministic assignments. However, the convex combination of two ordinally efficient assignments is not necessarily again ordinally efficient. In general, to be able to encode a desideratum via a welfare function, the welfare of any random outcome x must depend in a linear fashion on the welfare of the individual alternatives

(see Definition 45). Ordinal efficiency violates this requirement.

Quantifiable Desiderata

In some situations, there exists a quantified notion of the desirability of different alternatives at different preference profiles. Straightforward examples are welfare measures based on positional scoring. Formally, a *scoring function* $v : \{1, \dots, m\} \to \mathbb{R}$ is a mapping that associates a *score* $v(r)$ with selecting some agent's rth choice alternative. Typical examples include *Borda* (i.e., $v(r) = m - r$ for all $r \in \{1, \dots, m\}$), *Plurality* (i.e., $v(1) = 1$ and $v(r) = 0$ for $r \neq 1$), and *Veto* (i.e., $v(r) = 0$ for all $r \neq m$ and $v(m) = -1$). The score of an alternative at a preference profile is then determined by summing the scores of this alternative under the different preference orders of the agents. Formally,

$$\mathrm{sc}_v(j, \boldsymbol{P}) = \sum_{i \in N} v(r_{P_i}(j)), \tag{371}$$

where $r_{P_i}(j)$ is the rank of alternative j in preference order P_i. For any scoring-based desideratum, the respective welfare function arises by scaling sc_v as follows:

$$w(j, \boldsymbol{P}) = \frac{\mathrm{sc}_v(j, \boldsymbol{P}) - \min_{j' \in M} \mathrm{sc}_v(j', \boldsymbol{P})}{\max_{j' \in M} \mathrm{sc}_v(j', \boldsymbol{P}) - \min_{j' \in M} \mathrm{sc}_v(j', \boldsymbol{P})} \tag{372}$$

if $\max_{j' \in M} \mathrm{sc}_v(j', \boldsymbol{P}) > \min_{j' \in M} \mathrm{sc}_v(j', \boldsymbol{P})$ and $w(j, \boldsymbol{P}) = 0$ otherwise.

In the assignment domain, the rank valuation v is used to determine the v-rank value of any assignment. This induces a scoring function on the corresponding alternatives. Consequently, instead of viewing v-rank efficiency as a binary desideratum, the quantified desideratum could be to maximize the v-rank value. This natural measure of welfare was put forward by Featherstone (2011).

In quasi-linear domains, a common objective is to maximize the sum of the agents' cardinal utilities. However, by definition, ordinal mechanisms do not incorporate preference intensities, because they only elicit ordinal preference information. Thus, it is not possible to maximize the agents' aggregate utility directly. For situations where this aggregate value is interesting, Procaccia and Rosenschein (2006) proposed to maximize a *conservative estimate* of this value instead. This (slightly unorthodox) desideratum is also representable by a welfare function (see Appendix 5.D for details).

5.11 Conclusion

In this paper, we have presented a structural characterization of the Pareto frontier for random ordinal mechanisms. Loosely speaking, the *Pareto frontier* consists of those mechanisms that make optimal trade-offs between incentive properties and other desiderata, such as Condorcet consistency or the goal of choosing an alternative that is the first choice of many agents. We have derived an intuitive measure for *manipulability* from the approximate strategyproofness concept; and we have formalized the *deficit*, a very general way of measuring the performance of mechanisms with respect to any desideratum specified by a *welfare function*. This has allowed us to assess mechanisms φ by their *signatures* $(\varepsilon(\varphi), \delta(\varphi))$. All mechanisms on the Pareto frontier are *optimal* in the sense that it is not possible to improve on their performance with respect to the desideratum (i.e., reduce the deficit) without accepting additional manipulability.

We have achieved our main result in three distinct steps: first, we have shown that ε-approximate strategyproofness can be equivalently expressed by a finite set of linear constraints. This has enabled us to define the linear program FindOpt to identify the set of all mechanisms that have minimal deficit but satisfy ε-approximate strategyproofness for a given manipulability bound ε. Second, we have shown how hybrid mechanisms, which are convex combinations of two component mechanisms, trade off manipulability against deficit. In particular, we have given a guarantee that the signature of a β-hybrid h_β is always at least as good as the β-convex combination of the signatures of the two component mechanisms. Third, we have shown that the Pareto frontier consists of two building blocks: (1) there exists a finite set of *supporting manipulability bounds* $\varepsilon_0, \ldots, \varepsilon_K$ such that we can characterize the set of optimal mechanisms at each of the bounds ε_k as the set of solutions to the linear program FindOpt at ε_k, and (2) for any intermediate manipulability bound $\varepsilon = (1-\beta)\varepsilon_{k-1} + \beta\varepsilon_k$, the set of optimal mechanisms at ε is precisely the set of β-hybrids with components taken from the optimal mechanisms at each of the two adjacent supporting manipulability bounds ε_{k-1} and ε_k.

Our results have a number of interesting consequences (beyond their relevance in this paper): first, Theorem 19 gives a finite set of linear constraints that is equivalent to ε-approximate strategyproofness. This makes ε-approximate strategyproofness accessible to algorithmic analysis. In particular, it enables the use of this incentive requirement under the automated mechanism design paradigm.

Second, the performance guarantees for hybrid mechanisms from Theorem 20 yield convincing arguments in favor of randomization. In particular, we learn that the important

requirements of anonymity and neutrality come "for free;" mechanism designers do not have to accept a less desirable signature when imposing either or both (provided that the deficit measure is anonymous, neutral, or both).

Third, our main result, Theorem 21, has provided a *structural understanding* of the whole Pareto frontier. Knowledge of the Pareto frontier enables mechanism designers to make a completely informed decision about trade-offs between manipulability and deficit. In particular, we now have a way to determine precisely by how much the performance of mechanisms (with respect to a given desideratum) can be improved when allowing additional manipulability. An important learning for mechanism designers is that the mapping $\varepsilon \mapsto \delta(\varepsilon)$, which associates each manipulability bound with the lowest achievable deficit at this manipulability bound, is *monotonic*, *decreasing*, *convex*, and *continuous*. This means that when trading off manipulability and deficit along the Pareto frontier, the trade-offs are smooth, and the earliest sacrifices yield the greatest benefit. Thus, it can be worthwhile for mechanism designers to consider even small bounds $\varepsilon > 0$ in order to obtain substantial improvements. Conversely, if the initial improvements are not worth the sacrifice, then we can confidently abort the search because we know that the marginal return on any further relaxation of strategyproofness is decreasing.

Finally, we have illustrated our results by considering two concrete problems. In both problems, three agents had strict preferences over three alternatives. In the first problem, the desideratum was to choose an alternative that is the *first choice* of *many agents* (corresponding to Plurality scoring), and in the second problem, the desideratum was to choose an alternative that is the *last choice* of a *small number* of agents (corresponding to Veto scoring). In both problems, we have computed the Pareto frontier and verified the resulting structure analytically. The examples showed that the Pareto frontier may be completely linear (first problem) but truly non-linear (second problem).

In summary, we have given novel insights about the Pareto frontier for random ordinal mechanisms. We have proven our results for the full ordinal domain that includes indifferences, but they continue to hold for many other interesting domains that arise by restricting the space of preference profiles, such as the assignment domain and the two-sided matching domain. When impossibility results restrict the design of strategyproof mechanisms, we have provided a new perspective on the unavoidable trade-off between strategyproofness and other desiderata along the Pareto frontier.

Appendix for Chapter 5

5.A Relative Deficit

The deficit for outcomes (Definition 47) is defined as the absolute difference between the achievable and the achieved welfare. However, in some situations, it may be more natural to consider a *relative* difference, e.g., the ratio between the achieved and the maximal achievable welfare. As we will show next, it is without loss of generality that in this paper we have restricted our attention to absolute differences. This is because the relative w-deficit can always be expressed as an absolute $\tilde{w}$-deficit, where the adjusted welfare function $\tilde{w}$ is obtained from w by scaling. Proposition 21 makes this argument precise.

To state this equivalence formally, we need to define the relative deficit: for any preference profile $\boldsymbol{P} \in \mathcal{P}^N$, the *$w$-welfare margin at $\boldsymbol{P}$* is the difference between the highest and the lowest w-welfare achievable by any alternative at $\boldsymbol{P}$. We set $w^{\max}(\boldsymbol{P}) = \max_{j \in M} w(j, \boldsymbol{P})$, $w^{\min}(\boldsymbol{P}) = \min_{j \in M} w(j, \boldsymbol{P})$, and $w^{\text{margin}}(\boldsymbol{P}) = w^{\max}(\boldsymbol{P}) - w^{\min}(\boldsymbol{P})$. Note that for the special case where $w^{\text{margin}}(\boldsymbol{P}) = 0$, all alternatives (and therefore all outcomes) have the same welfare. In this case, *any* alternative is w-welfare maximizing at $\boldsymbol{P}$. For an outcome $x \in \Delta(M)$, the *relative w-deficit of x at $\boldsymbol{P}$* is the w-deficit of x at $\boldsymbol{P}$, normalized by the w-welfare margin at $\boldsymbol{P}$; formally,

$$\delta_w^{\text{relative}}(x, \boldsymbol{P}) = \begin{cases} \frac{w^{\max}(\boldsymbol{P}) - w(x, \boldsymbol{P})}{w^{\text{margin}}(\boldsymbol{P})}, & \text{if } w^{\text{margin}}(\boldsymbol{P}) > 0, \\ 0, & \text{else.} \end{cases} \tag{373}$$

Proposition 21. *For any welfare function w, there exists a welfare function $\tilde{w}$ such that the relative w-deficit coincides with the absolute $\tilde{w}$-deficit, such that for all outcomes $x \in \Delta(M)$ and all preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, we have*

$$\delta_w^{relative}(x, \boldsymbol{P}) = \delta_{\tilde{w}}(x, \boldsymbol{P}). \tag{374}$$

The proof follows immediately by setting $\tilde{w}(j, \boldsymbol{P}) = \frac{w^{\max}(\boldsymbol{P}) - w(j, \boldsymbol{P})}{w^{\text{margin}}(\boldsymbol{P})}$, whenever $w^{\text{margin}}(\boldsymbol{P}) >$

0, and $\tilde{w}(j, \boldsymbol{P}) = 0$ otherwise. Proposition 21 shows that including relative deficit does not enrich the space of possible welfare criteria but that the space of welfare functions is rich enough to cover relative deficits implicitly. Thus, it is without loss of generality that we have restricted attention to absolute deficits in this paper.

5.B Anonymity, Neutrality, and a Symmetric Decomposition

Two common requirements in the design of ordinal mechanisms are anonymity and neutrality. Anonymity captures the intuition that a mechanism should not discriminate between agents; instead, the influence of any agent on the outcome should be the same and independent of the agent's name. Neutrality requires that the mechanism is not biased towards particular alternatives; the decision should depend only on the agents' preferences but not on the names of the alternatives.

In this section we describe the implications of these two requirements for the design of optimal mechanisms: after providing definitions, we show how they can be incorporated in the linear program FindOpt. Then we formally prove the second takeaway from Theorem 20 that (under certain conditions) both requirements come for free in terms of the signatures of optimal mechanisms. Finally, we prove a new symmetric decomposition result for strategyproof, anonymous, neutral mechanisms, that extends the characterization of strategyproof random mechanisms in (Gibbard, 1977).

5.B.1 Definition of Anonymity and Neutrality

First, we define anonymity: for any *renaming of the agents* (i.e., any bijection $\pi : N \to N$) and any preference profile $\boldsymbol{P} = (P_1, \ldots, P_n) \in \mathcal{P}^N$, let $\boldsymbol{P}^\pi = (P_{\pi(1)}, \ldots, P_{\pi(n)})$ be the preference profile where the agents have exchanged their roles according to π. Agent i is now reporting the preference order $P_{\pi(i)}$ that was reported by agent $\pi(i)$ under the original preference profile $\boldsymbol{P}$. For any mechanism φ let φ^π be the mechanism under which the agents trade roles according to π; formally, let $\varphi^\pi(\boldsymbol{P}) = \varphi(\boldsymbol{P}^\pi)$ for any preference profile $\boldsymbol{P} \in \mathcal{P}^N$.

Definition 54 (Anonymity).

- A welfare function w is *anonymous* if for all renamings $\pi : N \to N$, preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, and alternatives $j \in M$, we have $w(j, \boldsymbol{P}) = w(j, \boldsymbol{P}^\pi)$ (i.e., the welfare of alternative is independent of the order in which the agents submit their preferences).

- A probability distribution $\mathbb{P}$ over preference profiles is *anonymous* if for all renamings $\pi : N \to N$ and preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, we have $\mathbb{P}[\boldsymbol{P}] = \mathbb{P}[\boldsymbol{P}^\pi]$ (i.e., the probability of a preference profile does not depend on the order in which the different preference orders appear).
- The worst-case deficit δ_w is *anonymous* if the underlying welfare function w is anonymous. The ex-ante deficit $\delta_w^{\mathbb{P}}$ is *anonymous* if the underlying welfare function and the probability distribution $\mathbb{P}$ are anonymous.
- A mechanism φ is *anonymous* if for all renamings $\pi : N \to N$ and preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, we have $\varphi(\boldsymbol{P}) = \varphi^\pi(\boldsymbol{P})$ (i.e., the outcome of the mechanism is independent of the order in which the agents submit their preferences).

Next, we define neutrality: for any *renaming of the alternatives* (i.e., any bijection $\varpi : M \to M$) and any preference order $P_i \in \mathcal{P}$, let P_i^ϖ be the preference order under which $P_i^\varpi : \varpi(j) \geq \varpi(j')$ whenever $P_i : j \geq j'$ for any alternatives $j, j' \in M$. This means that P_i^ϖ corresponds to P_i, except that the all alternatives have been renamed according to ϖ. For any preference profile $\boldsymbol{P} = (P_1, \ldots, P_n) \in \mathcal{P}^N$, let $\boldsymbol{P}^\varpi = (P_1^\varpi, \ldots, P_n^\varpi)$ be the preference profile where the alternatives inside the agents' preference orders have been renamed according to ϖ. For any mechanism φ let φ^ϖ be the mechanism under which the alternatives are renamed according to ϖ; formally, let $\varphi_j^\varpi(\boldsymbol{P}) = \varphi_{\varpi(j)}(\boldsymbol{P}^\varpi)$ for all preference profile $\boldsymbol{P} \in \mathcal{P}^N$ and alternative $j \in M$.

Definition 55 (Neutrality)**.** We define the following:

- A welfare function w is *neutral* if for all renamings $\varpi : M \to M$, preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, and alternatives $j \in M$, we have $w(j, \boldsymbol{P}) = w(\varpi(j), \boldsymbol{P}^\varpi)$ (i.e., the value of any alternative is independent of its name).
- A probability distribution $\mathbb{P}$ over preference profiles is *neutral* if for all renamings $\varpi : M \to M$ and preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, we have $\mathbb{P}[\boldsymbol{P}] = \mathbb{P}[\boldsymbol{P}^\varpi]$ (i.e., the probability of a preference profile does not depend on the names of the alternatives).
- The worst-case deficit δ_w is *neutral* if the underlying welfare function w is neutral. The ex-ante deficit $\delta_w^{\mathbb{P}}$ is *neutral* if the underlying welfare function and the probability distribution $\mathbb{P}$ are neutral.
- A mechanism φ is *neutral* if for all renamings $\varpi : M \to M$, preference profiles $\boldsymbol{P} \in \mathcal{P}^N$, and alternatives $j \in M$, we have $\varphi_j(\boldsymbol{P}) = \varphi_j^\varpi(\boldsymbol{P}) = \varphi_{\varpi(j)}(\boldsymbol{P}^\varpi)$ (i.e., the outcomes of the mechanism are independent of the names of the alternatives).

Incorporating anonymity and neutrality as additional constraints in the linear program FindOpt is straightforward and can be done as follows.

Linear Program 1, extended (FindOpt)**.**

$$
\begin{aligned}
&\dots \\
&\mathtt{f_j}(\mathtt{P}) = \mathtt{f_j}(\mathtt{P}^{\pi}), && \forall \boldsymbol{P} \in \mathcal{P}^N, j \in M, \pi : N \to N \text{ bij.} && \textit{(Anonymity)} \\
&\mathtt{f_j}(\mathtt{P}) = \mathtt{f}_{\varpi(\mathtt{j})}(\mathtt{P}^{\varpi}), && \forall \boldsymbol{P} \in \mathcal{P}^N, j \in M, \varpi : M \to M \text{ bij.} && \textit{(Neutrality)}
\end{aligned}
$$

5.B.2 Costlessness of Requiring Anonymity and Neutrality

With these notions of anonymity and neutrality in mind, observe that for any given mechanism it is possible to construct an anonymous mechanism or a neutral mechanism by randomizing over the roles of the agents or the roles of the alternatives in the mechanism, respectively: let

$$\varphi_j^{\text{anon}}(\boldsymbol{P}) = \frac{1}{n!} \sum_{\substack{\pi : N \to N \\ \text{bijection}}} \varphi_j^{\pi}(\boldsymbol{P}) \tag{375}$$

and

$$\varphi_j^{\text{neut}}(\boldsymbol{P}) = \frac{1}{m!} \sum_{\substack{\varpi : M \to M \\ \text{bijection}}} \varphi_j^{\varpi}(\boldsymbol{P}) \tag{376}$$

for all preference profiles $\boldsymbol{P} \in \mathcal{P}^N$ and alternatives $j \in M$. Observe that φ^{anon} and φ^{neut} are simply hybrid mechanisms with many components, where each component is used with a probability of $\frac{1}{n!}$ or $\frac{1}{m!}$, respectively. With these definitions we can formally derive the second main takeaway from Theorem 20.

Corollary 9. *Given a problem (N, M, δ), where the deficit is anonymous/neutral/both, for any mechanism φ there exists a mechanism $\tilde{\varphi}$ that is anonymous/neutral/both and has a weakly better signature than φ; formally, $\varepsilon(\tilde{\varphi}) \leqslant \varepsilon(\varphi)$ and $\delta(\tilde{\varphi}) \leqslant \delta(\varphi)$*

Proof. Observe that if the deficit δ is anonymous, then φ and φ^{π} have the same signature for any renaming of the agents $\pi : N \to N$. Consequently, φ^{anon} is a hybrid of $n!$ mechanisms which all have the same signature, and therefore it has a weakly better signature by Theorem 20. Similarly, if δ is neutral, then φ and φ^{ϖ} have the same signature for any renaming of the alternatives $\varpi : M \to M$, and the result follows analogously. □

Intuitively, Corollary 9 means that, given the right welfare desideratum, the two requirements anonymity and neutrality are "free" in terms of manipulability and deficit.

We do not have to accept a less attractive signatures in order to achieve either property.

5.B.3 Non-trivial Signature Cost of Other Desiderata

In contrast to anonymity and neutrality, other common desiderata do not come for free, such as Condorcet consistency, Pareto optimality, or even the rather weak requirement of unanimity.

Definition 56 (Unanimity, Pareto Optimality, Condorcet Consistency)**.** For a given preference profile $\boldsymbol{P} \in \mathcal{P}^N$, define the following:

- An alternative $j \in M$ is a *unanimity winner at* $\boldsymbol{P}$ if for all agents $i \in N$ and all other alternatives $j' \in M$ we have $P_i : j \succeq j'$. Let $M^{\text{unan}}(\boldsymbol{P})$ be the set of all unanimity winners at $\boldsymbol{P}$ and $M^{\neg\text{unan}}(\boldsymbol{P}) = M \backslash M^{\text{unan}}(\boldsymbol{P})$ the set of non-winners.
- An alternative $j \in M$ *Pareto dominates* another alternative $j' \in M$ at $\boldsymbol{P}$ if for all agents $i \in N$ we have $P_i : j \succeq j'$ and for some agent i' we have $P_{i'} : j \succ j'$. j is *Pareto optimal at* $\boldsymbol{P}$ if there exists no other alternative j' that Pareto dominates j. Let $M^{\text{Pareto}}(\boldsymbol{P})$ be the set of Pareto optimal alternatives at $\boldsymbol{P}$, and let $M^{\neg\text{Pareto}}(\boldsymbol{P}) = M \backslash M^{\text{Pareto}}(\boldsymbol{P})$ be the set of alternatives that are Pareto dominated by another alternative at $\boldsymbol{P}$.
- For any two alternatives $a, b \in M$ let $n_{a \succ b}(\boldsymbol{P}) = \#\{i \in N \mid P_i : a \succ b\}$ be the number of agents who strictly prefer a to b at $\boldsymbol{P}$. An alternative $j \in M$ is a *Condorcet winner at* $\boldsymbol{P}$ if for all other alternatives $j' \in M$ we have $n_{j \succ j'}(\boldsymbol{P}) \geqslant n_{j' \succ j}(\boldsymbol{P})$. $M^{\text{Condorcet}}(\boldsymbol{P})$ is the set of Condorcet winners at $\boldsymbol{P}$, and the set of non-Condorcet winners is $M^{\neg\text{Condorcet}}(\boldsymbol{P}) = M \backslash M^{\text{Condorcet}}(\boldsymbol{P})$.

A mechanism φ satisfies unanimity, Pareto optimality, or Condorcet consistency, if it only selects alternatives that have the respective property whenever they exist. One way to incorporate these desiderata into the linear program FindOpt is to include them in the objective function by using a welfare function that assigns higher value to alternatives that have the respective property.

However, this is not sufficient if the optimal mechanisms must satisfy the requirement *completely*, independent of the resulting increase in manipulability. For example, the goal could be to find a unanimous, strategyproof mechanism that minimizes the worst-case deficit based on Veto scoring. In this case, it would not suffice to include unanimity in the welfare function, because the only way to guarantee unanimity in this case would be to impose full welfare maximization. Alternatively, we can incorporate the

property as additional constraints in the linear program FindOpt directly. The following linear constraints can be used to require unanimity, Pareto optimality, and Condorcet consistency, respectively.

Linear Program 1, extended (FindOpt)**.**

$$
\begin{aligned}
&\dots \\
&\mathtt{f_j}(\mathtt{P}) = 0, && \text{(Unanimity)} \\
&\quad \forall \boldsymbol{P} \in \mathcal{P}^N \text{ such that } M^{\text{unan}}(\boldsymbol{P}) \neq \varnothing \text{ and } j \in M^{\neg\text{unan}}(\boldsymbol{P}) \\
&\mathtt{f_j}(\mathtt{P}) = 0, && \text{(Pareto)} \\
&\quad \forall \boldsymbol{P} \in \mathcal{P}^N, j \in M^{\neg\text{Pareto}}(\boldsymbol{P}) \\
&\mathtt{f_j}(\mathtt{P}) = 0, && \text{(Condorcet)} \\
&\quad \forall \boldsymbol{P} \in \mathcal{P}^N \text{ such that } M^{\text{Condorcet}}(\boldsymbol{P}) \neq \varnothing \text{ and } j \in M^{\neg\text{Condorcet}}(\boldsymbol{P})
\end{aligned}
$$

The next example illustrates that, unlike anonymity and neutrality, unanimity does not come for free in terms of the mechanisms' signatures. Since Pareto optimality and Condorcet consistency imply unanimity, the same is true for both other desiderata.

Example 19. Consider the same problem as in Section 5.9.2, where $n = m = 3$, agents have strict preferences over alternatives a, b, c, δ is the worst-case deficit based on Veto scoring. Let φ be a strategyproof mechanism that is also unanimous.

By the characterization in (Gibbard, 1977), φ must be a hybrid of strategyproof unilateral and strategyproof duple mechanisms. However, for $m \geqslant 3$ alternatives, no duple mechanism is unanimous, and unilateral mechanisms are unanimous only if they always select the first choice of the dictating agent. But as soon as a single component of a hybrid is not unanimous, the hybrid is not unanimous either. Consequently, φ must be a hybrid of dictatorships. Since unanimity is an anonymous and neutral constraint (i.e., the constraint is invariant to renamings of agents or alternatives), we obtain from Corollary 9 that the signature of φ is at most as good as the signature of Random Dictatorship, where the dictating agent is chosen *uniformly* at random. However, Random Dictatorship has a deficit of $\frac{4}{9}$ at the preference profile

$$P_1, P_2 \;:\; a \succ b \succ c, \tag{377}$$

$$P_3 \;:\; c \succ b \succ a. \tag{378}$$

But we already observed that Random Duple is strategyproof and optimal (but not unanimous) in this problem with strictly lower deficit of $\frac{2}{9}$. This means that requiring

unanimity in this problem leads to a strict increase in the lowest deficit that is achievable with strategyproof, optimal mechanisms.

5.B.4 Symmetric Decomposition of Strategyproof, Anonymous, and Neutral Mechanisms

We present a refinement of Gibbard's strong characterization of strategyproof mechanisms (Gibbard, 1977). Our *symmetric decomposition* characterizes mechanisms that are *strategyproof*, *anonymous*, and *neutral*. We use this result to establish the shapes of Pareto frontiers in Sections 5.9.2 analytically. Beyond this application, the symmetric decomposition may be of independent interest.

In the full domain of *strict* preferences, Gibbard (1977) showed that any strategyproof mechanism is a hybrid of multiple "simple" mechanisms, namely strategyproof *unilateral* and *duple* mechanisms.

Definition 57 (Gibbard, 1977)**.** A mechanism uni is *unilateral* if the outcome only depends on the report of a single agent; formally, there exists $i \in N$ such that for all preference profiles $\boldsymbol{P}, \boldsymbol{P}' \in \mathcal{P}^N$ we have that $P_i = P_i'$ implies $\text{uni}(\boldsymbol{P}) = \text{uni}(\boldsymbol{P}')$.

Definition 58 (Gibbard, 1977)**.** A mechanism dup is *duple* if only two alternatives are possible; formally, there exist $a, b \in M$ such that for all preference profiles $\boldsymbol{P} \in \mathcal{P}^N$ we have $\text{dup}_j(\boldsymbol{P}) = 0$ for all $j \neq a, b$.

The strong characterization result is the following.

Fact 8 (Gibbard, 1977)**.** *A mechanism φ is strategyproof* if and only if *it can be written as a hybrid of mechanisms $\varphi^1, \ldots, \varphi^K$, and each component φ^k is strategyproof and either unilateral or duple.*[4]

Obviously, duple mechanisms cannot satisfy neutrality (unless $m = 2$) and unilateral mechanisms cannot satisfy anonymity (unless the mechanism is constant or $n = 1$). This means that anonymity and neutrality of strategyproof mechanisms are the result of mixing the unilateral and duple components "correctly." This intuition gives rise to the following more refined decomposition of strategyproof, anonymous, neutral mechanism.

Theorem 22 (Symmetric Decomposition)**.** *A mechanism φ is strategyproof, anonymous, and neutral if and only if there exist*

[4] Gibbard further refined this result by replacing strategyproofness with *localized* and *non-perverse*.

1. *strategyproof, neutral, unilateral mechanisms* $uni_k, k \in \{1, \dots, K^{uni}\}$,
2. *strategyproof, anonymous, duple mechanisms* $dup_k, k \in \{K^{uni}+1, \dots, K^{uni}+K^{dup}\}$,
3. *coefficients* $\beta_k \geqslant 0, k \in \{1, \dots, K^{uni}+K^{dup}\}$ *with* $\sum_{k=1}^{K^{uni}+K^{dup}} \beta_k = 1$,

such that

$$\varphi = \sum_{\substack{\pi : N \to N \\ \textit{bijection}}} \sum_{k=1}^{K^{uni}} \left(\frac{\beta_k}{n!}\right) uni_k^{\pi} + \sum_{\substack{\varpi : M \to M \\ \textit{bijection}}} \sum_{k=K^{uni}+1}^{K^{uni}+K^{dup}} \left(\frac{\beta_k}{m!}\right) dup_k^{\varpi}. \tag{379}$$

Proof. By anonymity and neutrality of φ we get that

$$\varphi = (\varphi^{\pi})^{\varpi} = (\varphi^{\varpi})^{\pi} = \varphi^{\pi,\varpi} \tag{380}$$

for all bijections $\pi : N \to N$ and $\varpi : M \to M$, which implies

$$\varphi = \sum_{\pi,\varpi} \frac{1}{n!m!} \varphi^{\pi,\varpi}. \tag{381}$$

Since $\varphi^{\pi,\varpi}$ is strategyproof, we can use Fact 8 to decompose it into $K^{\pi,\varpi}$ strategyproof unilateral and duple mechanisms, i.e.,

$$\varphi^{\pi,\varpi} = \sum_{k=1}^{K^{\pi,\varpi}} \beta_k^{\pi,\varpi} \varphi^{k,\pi,\varpi}. \tag{382}$$

By symmetry, the decomposition can be chosen such that for any pair of renamings π, ϖ and $k \in \{1, \dots, K^{\pi,\varpi}\}$ we have

- $K^{\text{uni}} + K^{\text{dup}} = K^{\pi,\varpi}$,
- $\beta_k = \beta_k^{\pi,\varpi}$,
- if $\varphi^k = \varphi^{k,\text{id},\text{id}}$ is a strategyproof unilateral (or duple) mechanism, than $\varphi^{k,\pi,\varpi}$ is a strategyproof unilateral (or duple) mechanism with $\varphi^k_{\varpi(j)}(\boldsymbol{P}^{\pi,\varpi}) = \varphi^{k,\pi,\varpi}_j(\boldsymbol{P})$,
- without loss of generality, φ^k is unilateral for $k \in \{1, \dots, K^{\text{uni}}\}$ and duple for $k \in \{K^{\text{uni}}+1, \dots, K^{\text{uni}}+K^{\text{dup}}\}$.

Averaging over all renamings π, ϖ we get

$$\varphi = \sum_{\pi,\varpi} \left(\frac{1}{n!m!}\right) \sum_{k=1}^{K^{\pi,\varpi}} \beta_k^{\pi,\varpi} \varphi^{k,\pi,\varpi} = \sum_{k=1}^{K^{\text{uni}}+K^{\text{dup}}} \sum_{\pi,\varpi} \left(\frac{\beta_k}{n!m!}\right) \varphi^{k,\pi,\varpi}. \tag{383}$$

If $\varphi^{k,\mathrm{id},\varpi}$ is strategyproof and duple, then

$$\mathrm{dup}^{k,\varpi} = \sum_{\pi} \left(\frac{1}{n!}\right) \varphi^{k,\pi,\varpi} \tag{384}$$

is strategyproof and duple as well. Similarly, if $\varphi^{k,\pi,\mathrm{id}}$ is strategyproof and unilateral, then

$$\mathrm{uni}^{k,\pi} = \sum_{\varpi} \left(\frac{1}{m!}\right) \varphi^{k,\pi,\varpi} \tag{385}$$

is strategyproof and unilateral as well. With this we can rewrite (383) as

$$\varphi = \sum_{k=1}^{K^{\mathrm{uni}}} \sum_{\pi} \frac{\beta_k}{n!} \mathrm{uni}^{k,\pi} + \sum_{k=K^{\mathrm{uni}}+1}^{K^{\mathrm{uni}}+K^{\mathrm{dup}}} \sum_{\varpi} \frac{\beta_k}{m!} \mathrm{dup}^{k,\varpi}, \tag{386}$$

which is precisely the symmetric decomposition (379). □

The symmetric decomposition (379) is a consequence of Gibbard's strong characterization and the fact that for any anonymous, neutral mechanism we have $\varphi = \varphi^{\pi,\varpi}$. It is *symmetric* in the sense that for any component uni^k (or dup^k) that occur with coefficient β_k, the corresponding component $\mathrm{uni}^{k,\pi}$ (or $\mathrm{dup}^{k,\varpi}$) occur with the same coefficient. (379) decomposes φ into two parts: a neutral part on the left, that gets "anonymized" by randomization, and an anonymous part on the right that gets "neutralized" by randomization.

5.C Algorithms

ALGORITHM 4: FINDBOUNDS

Input: bound $\underline{\varepsilon} > 0$
Variables: set of supporting manipulability bounds `supp.bnds`, stacks of *unverified*, *verified*, and *outer* segments `segments.u`, `segments.v`, `segments.o`

```
begin
    supp.bnds ← {0}
    segments.u ← {(signature(ε̲), signature(1))}
    segments.v ← ∅
    segments.o ← {(signature(0), signature(ε̲)), (signature(1), signature(2))}
    while segments.u ≠ ∅ do
        (P⁻, P⁺) ← pop(segments.u)
        (P⁻⁻, P⁻), (P⁺, P⁺⁺) ∈ segments.v ∪ segments.u ∪ segments.o
        e ← (affine.hull({P⁻⁻, P⁻}) ∩ affine.hull({P⁻⁻, P⁻}))_ε
        P ← signature(e)
        if P ∈ (affine.hull({P⁻⁻, P⁻}) ∩ affine.hull({P⁺, P⁺⁺})) then
            supp.bnds ← supp.bnds ∪ {P}
            segments.v ← segments.v ∪ {(P⁻⁻, P⁻), (P⁺, P⁺⁺), (P⁻, P), (P, P⁺)}
        end
        else if P ∈ affine.hull(P⁻, P⁺) then
            segments.v ← segments.v ∪ {(P⁻, P⁺)}
        end
        else
            segments.u ← segments.u ∪ {(P⁻, P), (P, P⁺)}
        end
    end
    return supp.bnds
end
```

ALGORITHM 5: FINDLOWER

Variables: signatures $\mathtt{sign}^0, \mathtt{sign}, \mathtt{sign}^+$, bound $\underline{\varepsilon}$

```
begin
    ε̲ ← 1/2
    sign⁰ ← signature(0), sign⁺ ← signature(1), sign ← signature(ε̲)
    while sign ∉ affine.hull(sign⁰, sign⁺) do
        sign⁺ ← sign, ε̲ ← ε̲/2, sign ← signature(ε̲)
    end
    return ε̲
end
```

5.D Conservative Lower Bound on Aggregate Utility

Even though the vNM utilities in our model are not comparable across agents, one may choose to maximize the sum of these utilities. However, since agents only report ordinal preferences, we cannot maximize their utilities directly. Instead, we can maximize a *conservative estimate* for this sum. Formally, the respective desideratum would be to maximize

$$w(j, \boldsymbol{P}) = \inf_{u_1 \in U_{P_1}, \ldots, u_n \in U_{P_n}} \sum_{i \in N} u_i(j). \tag{387}$$

Recall that we bounded utilities between 0 and 1, and let us assume that $\max_{j \in M} u_i(j) = 1$. Then the aggregate utility from any outcome $x \in \Delta(M)$ is lowest if agents have utility (close to) 0 for all except their first choices. Thus, maximizing the conservative lower bound for aggregate utility in our model corresponds to selecting an alternative that is the first choice of many agents. This is equivalent to the quantitative measure of welfare induced by Plurality scoring.

5.E Omitted Proofs

5.E.1 Proof of Theorem 19

Proof of Theorem 19. Given a setting (N, M), a bound $\varepsilon \in [0, 1]$, and a mechanism φ, the following are equivalent:

1. *φ is ε-approximately strategyproof in (N, M).*
2. *For any agent $i \in N$, any preference profile $(P_i, P_{-i}) \in \mathcal{P}^N$, any misreport $P_i' \in \mathcal{P}$, and any rank $r \in \{1, \ldots, m\}$, we have*

$$\sum_{j \in M : r_{P_i}(j) \leqslant r} \varphi_j(P_i', P_{-i}) - \varphi_j(P_i, P_{-i}) \leqslant \varepsilon. \tag{388}$$

Fix an agent i, a preference profile (P_i, P_{-i}), and a misreport P_i'. The admissible set of utility functions for agent i is U_{P_i}, i.e., all the utilities $u_i : M \to [0, 1]$ for which $u_i(j) \geqslant u_i(j')$ whenever $P_i : j \succeq j'$. Let $B^{\{0,1\}}(P_i)$ denote the set of *binary utilities associated with* P_i, i.e.,

$$B^{\{0,1\}}(P_i) = \{u : M \to \{0, 1\} \mid u(j) \geqslant u(j') \text{ whenever } P_i : j \succeq j'\}. \tag{389}$$

We first show the direction "⇐," i.e., the condition 2 implies ε-approximate strategyproofness (1). Let $u \in B^{\{0,1\}}(P_i)$, then the incentive constraint (346) for this particular utility function has the form

$$\begin{aligned} & \varepsilon\left(u,\left(P_i,P_{-i}\right),P_i',\varphi\right) && (390)\\ = & \sum_{j\in M} u(j)\cdot\left(\varphi_j(P_i',P_{-i})-\varphi_j(P_i,P_{-i})\right) && (391)\\ = & \sum_{j\in M:u(j)=1} \varphi_j(P_i',P_{-i})-\varphi_j(P_i,P_{-i}) && (392)\\ = & \sum_{j\in M:r_{P_i}(j)\leqslant r} \varphi_j(P_i',P_{-i})-\varphi_j(P_i,P_{-i}) && (393)\end{aligned}$$

for some $k \in \{1,\dots,m\}$. By (388) from the condition 2, this term is always upper bounded by ε. By Lemma 7, $U_{P_i} \subseteq Conv\left(B^{\{0,1\}}(P_i)\right)$, which means that any $u_i \in U_{P_i}$ that represents the preference order P_i can be written as a convex combination of utility functions in $B^{\{0,1\}}(P_i)$, i.e.,

$$u_i = \sum_{l=1}^{L} \alpha_l u^l \tag{394}$$

for $u^l \in B^{\{0,1\}}(P_i)$ and $\alpha_l \geqslant 0$ for all $l \in \{1,\dots,L\}$ and $\sum_{l=1}^{L}\alpha_l = 1$. By linearity of the incentive constraint (346) we get that

$$\begin{aligned} \varepsilon\left(u_i,\left(P_i,P_{-i}\right),P_i',\varphi\right) &= \sum_{l=1}^{L}\alpha_l\varepsilon\left(u^l,\left(P_i,P_{-i}\right),P_i',\varphi\right) && (395)\\ &\leqslant \sum_{l=1}^{L}\alpha_l\varepsilon = \varepsilon. && (396)\end{aligned}$$

This proves the direction "⇐."

Next, we prove the direction "⇒". Towards contradiction, assume that the constraint (388) is violated for some $k \in \{1,\dots,m\}$, i.e.,

$$\sum_{j\in M:r_{P_i}(j)\leqslant k} \varphi_j(P_i',P_{-i})-\varphi_j(P_i,P_{-i}) = \varepsilon+\delta \tag{397}$$

with $\delta > 0$. Let $u \in B^{\{0,1\}}(P_i)$ be the binary utility function with

$$u(j) = \begin{cases} 1, & \text{if } r_{P_i}(j) \leqslant k,\\ 0, & \text{else.}\end{cases} \tag{398}$$

Then

$$\varepsilon\left(u,\left(P_i,P_{-i}\right),P_i',\varphi\right)=\varepsilon+\delta. \tag{399}$$

Choose any utility function $u' \in U_{P_i}$ and let $\beta = \frac{\delta/2}{\varepsilon+\delta+1}$. The utility function constructed by $\tilde{u} = (1-\beta)u + \beta u'$ represents P_i and we have

$$\begin{aligned}
& \varepsilon\left(\tilde{u},\left(P_i,P_{-i}\right),P_i',\varphi\right) && (400)\\
= \;& (1-\beta)\varepsilon\left(u,\left(P_i,P_{-i}\right),P_i',\varphi\right)+\beta\varepsilon\left(u',\left(P_i,P_{-i}\right),P_i',\varphi\right) && (401)\\
\geqslant \;& (1-\beta)(\varepsilon+\delta)-\beta && (402)\\
= \;& -\beta(\varepsilon+\delta+1)+(\varepsilon+\delta)=\varepsilon+\delta/2, && (403)
\end{aligned}$$

since the change in utility from manipulation is lower bounded by -1. Thus, the ε-approximate strategyproofness constraint is violated (for the utility function $\tilde{u}$, a contradiction.

Lemma 7. *For any preference order $P \in \mathcal{P}$ define the set of* binary utilities associated with P *by*

$$B^{\{0,1\}}(P)=\{u:M\to\{0,1\}\mid u(j)\geqslant u(j')\text{ whenever } P:j\succeq j'\}. \tag{404}$$

Then $U_P \subseteq Conv\left(B^{\{0,1\}}(P)\right)$, where $Conv(\cdot)$ denotes the convex hull of a set.

Proof of Lemma 7. First, suppose that the preference order P is strict, i.e., $P : j \succ j'$ or $P : j' \succ j$ for all $j \neq j'$, and without loss of generality,

$$P: j_1 \succ j_2 \succ \ldots \succ j_m. \tag{405}$$

In this case, $B^{\{0,1\}}(P)$ consists of all the functions $u^r : M \to \{0,1\}$ with

$$u^k(j_r)=\begin{cases}1, & \text{if } r\leqslant k,\\ 0, & \text{else.}\end{cases} \tag{406}$$

With $k \in 0,\ldots,m$. Let $\Delta u(0) = 1 - u(j_1)$, $\Delta u(k) = u(j_k) - u(j_{k+1})$ for all $k \in \{1,\ldots,m-1\}$, and $\Delta u(m) = u(j_m)$. Then represent u by

$$u(j_r)=\sum_{k=r}^{m}\Delta u(k). \tag{407}$$

Now we construct the utility function

$$\tilde{u} = \sum_{k=0}^{m} \Delta u(k) \cdot u^k. \tag{408}$$

First note that $\sum_{k=0}^{m} \Delta u(k) = 1$ and $\Delta u(k) \geqslant 0$ for all k, so that $\tilde{u}$ is a convex combination of elements of $B^{\{0,1\}}(P)$. Furthermore, for any $j_r \in M$, we have that

$$\begin{aligned} \tilde{u}(j_r) &= \sum_{k=0}^{m} \Delta u(k) \cdot u^k(j_r) && (409) \\ &= \sum_{k=r}^{m} \Delta u(k) = u(j_r), && (410) \\ & && (411) \end{aligned}$$

i.e., $u = \tilde{u}$. This establishes the Lemma for strict preference orders.

For arbitrary preference orders (i.e., with indifferences) the proof can be easily extended by combining all alternatives about which an agent with preference order P is indifferent into a single virtual alternative. Then we apply the proof for the strict case. The utility functions in $B^{\{0,1\}}(P)$ will be exactly those that put equal value on the groups of alternatives between which an agent with preference order P is indifferent. This concludes the proof of the Lemma. □

This concludes the proof of Theorem 19. □

5.E.2 Proof of Proposition 16

Proof of Proposition 16. Given a problem (N, M, δ) and a manipulability bound $\varepsilon \in [0, 1]$, there exists at least one mechanism that is optimal at ε.

A strategyproof mechanism always exists (e.g., the constant mechanism), and any strategyproof mechanism is also ε-approximately strategyproof. Thus, the set of candidates for $\textsc{Opt}(\varepsilon)$ is never empty. Since the deficit of any mechanism is upper bounded by 1, we get $\delta(\varepsilon) \leqslant 1$ for all $\varepsilon \in [0, 1]$. Next, for some $\varepsilon \in [0, 1]$ let

$$\delta(\varepsilon) = \inf \{\delta(\varphi) \mid \varphi \ \varepsilon\text{-approximately strategyproof}\}. \tag{412}$$

By definition (of the infimum), there exists a sequence of mechanisms $(\varphi_k)_{k \geqslant 1}$ such that $\delta(\varphi^k) \to \delta(\varepsilon)$ as $k \to \infty$. Since all φ_k are mechanisms, they are uniformly bounded

functions from the finite set $\mathbb{P}^N$ of preference profiles to the compact set $\Delta(M)$. Therefore, some sub-sequence of $(\varphi_k)_{k\geqslant 1}$ converges uniformly to a limit $\tilde{\varphi}$, which must itself be a mechanism.

By Theorem 19, ε-approximate strategyproofness is equivalent to a finite set of weak, linear inequalities. Since these are all satisfied by the elements of the sequence $(\varphi_k)_{k\geqslant 1}$, they are also satisfied by $\tilde{\varphi}$, i.e., $\tilde{\varphi}$ is an ε-approximately strategyproof mechanism. The deficit of a mechanism φ is either a maximum (worst-case deficit) or a weighted average (ex-ante deficit) of a finite set of values, but in any case, it is a uniformly continuous projection of φ onto the compact interval $[0,1]$. Thus, by uniform convergence, the deficit of $\delta(\tilde{\varphi})$ must be the limit of the sequence of deficits $(\delta(\varphi_k))_{k\geqslant 1}$, i.e., $\delta(\varepsilon)$, which yields $\tilde{\varphi} \in \textsc{Opt}(\varepsilon)$. □

5.E.3 Proof of Theorem 20

Proof of Theorem 20. Given a problem (N, M, δ), two mechanisms φ, ψ, and $\beta \in [0,1]$, we have

$$\varepsilon(h_\beta) \leqslant (1-\beta)\varepsilon(\varphi) + \beta\varepsilon(\psi), \tag{413}$$

$$\delta(h_\beta) \leqslant (1-\beta)\delta(\varphi) + \beta\delta(\psi). \tag{414}$$

The manipulability of a mechanism is determined by the maximum manipulability across all agents $i \in N$, preference profiles $(P_i, P_{-i}) \in \mathcal{P}^N$, and misreports $P_i' \in \mathcal{P}$. Using the triangle inequality for this max-operator, we get that

$$\begin{aligned}
\varepsilon(h_\beta) &= \max \varepsilon\left(u, (P_i, P_{-i}), P_i', h_\beta\right) && (415)\\
&= \max \sum_{j\in M} u_i(j) \cdot ((h_\beta)_j(P_i', P_{-i}) - (h_\beta)_j(P_i, P_{-i})) && (416)\\
&\leqslant (1-\beta) \max \sum_{j\in M} u_i(j) \cdot (\varphi_j(P_i', P_{-i}) - \varphi_j(P_i, P_{-i})) && (417)\\
&\quad +\beta \max \sum_{j\in M} u_i(j) \cdot (\psi_j(P_i'; P_{-i}) - \psi_j(P_i, P_{-i})) && (418)\\
&= (1-\beta)\varepsilon(\varphi) + \beta\varepsilon(\psi). && (419)
\end{aligned}$$

For worst-case deficit, observe that the extended welfare function for random outcomes is linear in the first argument, i.e., $w((1-\beta)x + \beta y, \boldsymbol{P}) = (1-\beta)w(x, \boldsymbol{P}) + \beta w(y, \boldsymbol{P})$

for any outcomes x, y preference profile $\boldsymbol{P}$. Thus,

$$\delta(h_\beta) = \max \delta\left(h_\beta(\boldsymbol{P}), \boldsymbol{P}\right) \leqslant (1-\beta)\varepsilon(\varphi) + \beta\varepsilon(\psi) \tag{420}$$

follows analogously. Since the ex-ante deficit is simply a single linear function of the mechanism, we get exact equality, i.e., $\delta(h_\beta) = (1-\beta)\varepsilon(\varphi) + \beta\varepsilon(\psi)$. □

5.E.4 Proof of Corollary 7

Proof of Corollary 7. Given a problem (N, M, δ), the mapping $\varepsilon \mapsto \delta(\varepsilon)$ is monotonically decreasing and convex.

Monotonicity follows (almost) by definition: observe a mechanism $\varphi \in \textsc{Opt}(\varepsilon)$ is also a candidate for $\textsc{Opt}(\varepsilon')$ for any larger manipulability bound $\varepsilon' \geqslant \varepsilon$. The only reason for φ to not be optimal at ε' is that some other ε'-approximately strategyproof mechanism has strictly lower deficit. Thus, the mapping $\varepsilon \mapsto \delta(\varepsilon)$ is weakly monotonic and decreasing.

To see convexity, assume towards contradiction that the mapping is not convex. Then there must exist bounds $\varepsilon, \varepsilon' \in [0,1]$ and $\beta \in [0,1]$, such that for $\varepsilon_\beta = (1-\beta)\varepsilon + \beta\varepsilon'$, we have

$$\delta(\varepsilon_\beta) > (1-\beta)\delta(\varepsilon) + \beta\delta(\varepsilon'). \tag{421}$$

Let $\varphi \in \textsc{Opt}(\varepsilon)$ and $\varphi' \in \textsc{Opt}(\varepsilon')$ and consider the hybrid $h_\beta = (1-\beta)\varphi + \beta\varphi'$. By Theorem 20, this hybrid has a manipulability of at most ε_β and a deficit of at most $(1-\beta)\delta(\varepsilon) + \beta\delta(\varepsilon')$. Thus,

$$\delta(\varepsilon_\beta) \leqslant \delta(h_\beta) \leqslant (1-\beta)\delta(\varepsilon) + \beta\delta(\varepsilon') < \delta(\varepsilon_\beta), \tag{422}$$

a contradiction. □

5.E.5 Proof of Theorem 21

Proof of Theorem 21. Given a problem (N, M, δ), there exists a finite set of supporting manipulability bounds

$$\varepsilon_0 = 0 < \varepsilon_1 < \ldots < \varepsilon_{K-1} < \varepsilon_K = \bar{\varepsilon}, \tag{423}$$

such that for any $k \in \{1, \dots, K\}$ *and any* $\varepsilon \in [\varepsilon_{k-1}, \varepsilon_k]$ *with* $\beta = \frac{\varepsilon - \varepsilon_{k-1}}{\varepsilon_k - \varepsilon_{k-1}}$ *we have that*

$$\textsc{Opt}(\varepsilon) = (1-\beta)\textsc{Opt}(\varepsilon_{k-1}) + \beta\textsc{Opt}(\varepsilon_k), \tag{424}$$

$$\delta(\varepsilon) = (1-\beta)\delta(\varepsilon_{k-1}) + \beta\delta(\varepsilon_k). \tag{425}$$

From Section 5.6 we know that for each $\varepsilon \in [0, \bar{\varepsilon}]$ we can write $\textsc{Opt}(\varepsilon)$ as the set of solutions to a linear program, i.e.,

$$\textsc{Opt}(\varepsilon) = \quad \operatorname*{argmin}_x \langle v, x \rangle \tag{426}$$

$$\text{s.t.} \quad Dx \leqslant d, \tag{427}$$

$$Ax \leqslant \varepsilon, \tag{428}$$

where D and A are matrices, v and d are vectors, ε is a vector with all entries equal to ε, and x is a vector of variables of dimension L. Observe that ε enters the constraints only as the upper bound in a number of linear inequalities. The proof utilizes this characterization of $\textsc{Opt}(\varepsilon)$.

Before we proceed with the proof of Theorem 21, we require a number of definitions. We denote by F_ε the set of *feasible points at* ε, i.e.,

$$F_\varepsilon = \{x \mid Dx \leqslant d, Ax \leqslant \varepsilon\}, \tag{429}$$

and we denote by S_ε the set of *solutions at* ε, i.e.,

$$S_\varepsilon = \operatorname*{argmin}_{x \in F_\varepsilon} \langle v, x \rangle. \tag{430}$$

A constraint is a row C_l of either the matrix A or the matrix D with the corresponding bound c_l equal to respective entry of d or ε. A feasible point $x \in F_\varepsilon$ is an *extreme point of* F_ε if there exist L independent constraints $C_1, \dots, C_L$ such that

$$C_l x = c_l \tag{431}$$

for all $l \in \{1, \dots, L\}$, i.e., the constraints are satisfied with equality at x. x is then said to be an extreme point of F_ε *with respect to* $(C_1, \dots, C_L)$. We say that the set of constraints $(C_1, \dots, C_L)$ is *restrictive at* ε if they are independent and there exists an extreme point

in F_ε with respect to these constraints. Let

$$\mathcal{R}(\varepsilon) = \{(C_1, \ldots, C_L) \mid (C_1, \ldots, C_L) \text{ is restrictive at } \varepsilon\} \tag{432}$$

be the set of all sets of constraints that are restrictive at ε. A set of restrictive constraints $\boldsymbol{C} \in \mathcal{R}(\varepsilon)$ is *binding at* ε if the extreme point $x \in F_\varepsilon$ where the constraints of $\boldsymbol{C}$ are satisfied with equality is a solution, i.e., $x \in S_\varepsilon$. Let

$$\mathcal{B}(\varepsilon) = \{\boldsymbol{C} \in \mathcal{R}(\varepsilon) \mid \boldsymbol{C} \text{ is binding at } \varepsilon\} \tag{433}$$

be the set of all sets of constraints that are binding at ε. We denote by $E(F_\varepsilon)$ and $E(S_\varepsilon)$ the extreme points of F_ε and S_ε, respectively.

Observe that since F_ε and S_ε are polytopes and bounded by finitely many hyperplanes, the extreme points $E(F_\varepsilon)$ and $E(S_\varepsilon)$ form minimal $\mathscr{V}$-representations of F_ε and S_ε (see, e.g., p.51ff in (Grünbaum, 2003)). Thus, $Conv(E(F_\varepsilon)) = F_\varepsilon$ and $Conv(E(S_\varepsilon)) = S_\varepsilon$. Furthermore, since $S_\varepsilon \subseteq F_\varepsilon$, any extreme point of S_ε is also an extreme point of F_ε, i.e., $E(S_\varepsilon) \subseteq E(F_\varepsilon)$. Finally, each extreme point is uniquely determined by the set of constraints with respect to which it is extreme, i.e., if there exists an extreme point with respect to a set of constraints $\boldsymbol{C} \in \mathcal{R}(\varepsilon)$, then this point is unique.

Claim 19. *For $\varepsilon_0, \varepsilon_1 \in [0, \bar{\varepsilon}]$ with $\varepsilon_0 < \varepsilon_1$, if $x_0 \in F_{\varepsilon_0}$ and $x_1 \in F_{\varepsilon_1}$, then for any $\gamma \in [0, 1]$ and $\varepsilon = (1-\gamma)\varepsilon_0 + \gamma\varepsilon_1$ we have that*

$$x = (1-\gamma)x_0 + \gamma x_1 \in F_\varepsilon. \tag{434}$$

Proof. By assumption, $D_k x_0 \leqslant d_k$ and $D_k x_1 \leqslant d_k$ for all k. Thus,

$$D_k((1-\gamma)x_0 + \gamma x_1) = (1-\gamma)D_k x_0 + \gamma D_k x_1 \leqslant d_k. \tag{435}$$

Furthermore, $A_k x_0 \leqslant \varepsilon_0$ and $A_k x_1 \leqslant \varepsilon_1$ for all k, which implies

$$A_k((1-\gamma)x_0 + \gamma x_1) = (1-\gamma)A_k x_0 + \gamma A_k x_1 \leqslant (1-\gamma)\varepsilon_0 + \gamma\varepsilon_1 = \varepsilon. \tag{436}$$

□

Claim 20. *For $\varepsilon_0, \varepsilon_1 \in [0, \bar{\varepsilon}]$ with $\varepsilon_0 < \varepsilon_1$, if $\boldsymbol{C} \in \mathcal{R}(\varepsilon_0)$ and $\boldsymbol{C} \in \mathcal{R}(\varepsilon_1)$, then for any $\gamma \in [0, 1]$ and $\varepsilon = (1-\gamma)\varepsilon_0 + \gamma\varepsilon_1$ we have that $\boldsymbol{C} \in \mathcal{R}(\varepsilon)$, and the γ-convex combination of the extreme points at ε_0 and ε_1 with respect to $\boldsymbol{C}$ is the unique extreme point at ε with*

respect to $\boldsymbol{C}$.

Proof. Since $\boldsymbol{C} \in \mathcal{R}(\varepsilon_0)$ and $\boldsymbol{C} \in \mathcal{R}(\varepsilon_1)$, there exist unique extreme points $x_0 \in E(F_{\varepsilon_0})$ and $x_1 \in E(F_{\varepsilon_0})$ with respect to $\boldsymbol{C}$. By Claim 19, the point $x = (1-\gamma)x_0 + \gamma x_1$ is feasible at ε. For any $l \in \{1, \ldots, L\}$ if $C_l = D_k$ for some k we have that

$$C_l x = (1-\gamma)D_k x_0 + \gamma D_k x_1 = (1-\gamma)d_k + \gamma d_k = d_k = c_l, \tag{437}$$

i.e., x satisfies the constraint C_l with equality. If $C_l = A_k$, then the constraint is also satisfied with equality, since

$$C_l x = (1-\gamma)A_k x_0 + \gamma A_k x_1 = (1-\gamma)\varepsilon_0 + \gamma \varepsilon_1 = \varepsilon. \tag{438}$$

Consequently, x is the unique extreme point at ε with respect to $\boldsymbol{C}$. This in turn implies that $\boldsymbol{C}$ is restrictive at ε. □

Claim 21. *There exists a finite decomposition*

$$0 = \varepsilon_0 < \varepsilon_1 < \ldots < \varepsilon_K = \bar{\varepsilon} \tag{439}$$

of the interval $[0, \bar{\varepsilon}]$*, such that on each interval* $[\varepsilon_{k-1}, \varepsilon_k]$ *we have that* $\mathcal{R}(\varepsilon) = \mathcal{R}(\varepsilon')$ *for all* $\varepsilon, \varepsilon' \in [\varepsilon_{k-1}, \varepsilon_k]$.

Proof. By Claim 20, if some set of L independent constraints $\boldsymbol{C}$ is restrictive at some $\varepsilon \in [0, \bar{\varepsilon}]$, then the set of $\varepsilon' \in [0, \bar{\varepsilon}]$ where $\boldsymbol{C}$ is also restrictive must be compact interval $[\varepsilon^-_{\boldsymbol{C}}, \varepsilon^+_{\boldsymbol{C}}] \subseteq [0, \bar{\varepsilon}]$ with $\varepsilon \in [\varepsilon^-_{\boldsymbol{C}}, \varepsilon^+_{\boldsymbol{C}}]$. Since there is a finite number of constraints, there is also a finite number of constraint sets $\boldsymbol{C}$. Consider the set

$$\{\varepsilon_0, \ldots, \varepsilon_K\} = \bigcup_{\boldsymbol{C} \text{ set of } L \text{ indep. constraints}} \{\varepsilon^-_{\boldsymbol{C}}, \varepsilon^+_{\boldsymbol{C}}\}. \tag{440}$$

Observe that by construction, a set of L independent constraints $\boldsymbol{C}$ becomes restrictive or un-restrictive only at one of the finitely many ε_{k-1}. This proves the claim. □

Claim 22. *On each interval* $[\varepsilon_{k-1}, \varepsilon_k]$ *from Claim 21 and for any* $\varepsilon \in [\varepsilon_{k-1}, \varepsilon_k]$*, if* $\boldsymbol{C} \in \mathcal{R}(\varepsilon)$*, then* $\boldsymbol{C} \in \mathcal{R}(\varepsilon_{k-1}) \cap \mathcal{R}(\varepsilon_k)$.

Proof. Assume towards contradiction that $\boldsymbol{C} \in \mathcal{R}(\varepsilon)$, but $\boldsymbol{C} \notin \mathcal{R}(\varepsilon_{k-1})$. Then there exists an $\varepsilon' \in (\varepsilon_{k-1}, \varepsilon) \subseteq (\varepsilon_{k-1}, \varepsilon_k)$, where $\boldsymbol{C}$ become restrictive for the first time. Then

$\varepsilon' \in \{\varepsilon_0, \varepsilon_{k-1}\}$, and therefore $[\varepsilon_{k-1}, \varepsilon_k]$ cannot be one of the intervals in the decomposition. Instead, it would be split by ε', a contradiction. □

Claim 23. *On each interval $[\varepsilon_{k-1}, \varepsilon_k]$ from Claim 21 and for any $\gamma \in [0,1]$ with $\varepsilon = (1-\gamma)\varepsilon_{k-1} + \gamma\varepsilon_k$ we have that*

$$F_\varepsilon = (1-\gamma)F_{\varepsilon_{k-1}} + \gamma F_{\varepsilon_k}, \tag{441}$$

i.e., the set of feasible points at ε is equal to the γ-convex combination of feasible points at ε_{k-1} and ε_k.

Proof. By Claim 19 we have

$$F_\varepsilon \supseteq (1-\gamma)F_{\varepsilon_{k-1}} + \gamma F_{\varepsilon_k}. \tag{442}$$

By Claim 22 the extreme points of F_ε are the γ-convex combinations of extreme points of $F_{\varepsilon_{k-1}}$ and F_{ε_k}. Since $F_\varepsilon = Conv(E(F_\varepsilon))$, this implies

$$F_\varepsilon \subseteq (1-\gamma)F_{\varepsilon_{k-1}} + \gamma F_{\varepsilon_k}, \tag{443}$$

which concludes the proof of the claim. □

Claim 24. *On each interval $[\varepsilon_{k-1}, \varepsilon_k]$ from Claim 21, if $\boldsymbol{C} \in \mathcal{B}(\varepsilon)$ for some $\varepsilon \in (\varepsilon_{k-1}, \varepsilon_k)$, then $\boldsymbol{C} \in \mathcal{B}(\varepsilon_{k-1}) \cap \mathcal{B}(\varepsilon_k)$. Furthermore, the extreme point of S_ε with respect to $\boldsymbol{C}$ is the γ-convex combination of the extreme points of $S_{\varepsilon_{k-1}}$ and S_{ε_k} with respect to $\boldsymbol{C}$ with $\gamma = \frac{\varepsilon - \varepsilon_{k-1}}{\varepsilon_k - \varepsilon_{k-1}}$.*

Proof. Since $\boldsymbol{C} \in \mathcal{B}(\varepsilon)$, there exists a unique extreme point $x \in E(S_\varepsilon)$ such that x is extreme at ε with respect to $\boldsymbol{C}$. By Claim 23, we can represent $x = (1-\gamma)x_0 + \gamma x_1$ with $x_0 \in F_{\varepsilon_{k-1}}$, $x_1 \in F_{\varepsilon_k}$. By Claim 20, x_0 and x_1 are extreme points of $F_{\varepsilon_{k-1}}$ and F_{ε_k}, respectively.

Suppose towards contradiction that $x_0 \notin S_{\varepsilon_{k-1}}$. Then there exists $x_0' \in S_{\varepsilon_{k-1}}$ such that

$$\langle v, x_0' \rangle < \langle v, x_0 \rangle. \tag{444}$$

By Claim 19, $x' = (1-\gamma)x'_0 + \gamma x_1$ is in F_ε and

$$\begin{aligned} \langle v, x' \rangle &= (1-\gamma)\langle v, x'_0 \rangle + \gamma \langle v, x_1 \rangle && (445) \\ &< (1-\gamma)\langle v, x_0 \rangle + \gamma \langle v, x_1 \rangle && (446) \\ &= \langle v, x \rangle, && (447) \end{aligned}$$

i.e., x' is feasible at ε and has lower objective than x. This contradicts the assumption that $x \in S_\varepsilon$. A similar argument yields $x_1 \in S_{\varepsilon_k}$, which concludes the proof of the claim. □

Claim 25. *On each interval $[\varepsilon_{k-1}, \varepsilon_k]$ from Claim 21 and any $\gamma \in [0,1]$ with $\varepsilon = (1-\gamma)\varepsilon_{k-1} + \gamma\varepsilon_k$ we have that*

$$S_\varepsilon = (1-\gamma)S_{\varepsilon_{k-1}} + \gamma S_{\varepsilon_k}, \tag{448}$$

i.e., the set of solutions at ε is equal to the γ-convex combination of solutions at ε_{k-1} and ε_k.

Proof. By Claim 19 we have

$$S_\varepsilon \supseteq (1-\gamma)S_{\varepsilon_{k-1}} + \gamma S_{\varepsilon_k}. \tag{449}$$

By Claim 24 the extreme points of S_ε are the γ-convex combinations of extreme points of $S_{\varepsilon_{k-1}}$ and S_{ε_k}. Since $S_\varepsilon = Conv(E(S_\varepsilon))$, this implies

$$S_\varepsilon \subseteq (1-\gamma)S_{\varepsilon_{k-1}} + \gamma S_{\varepsilon_k}, \tag{450}$$

which concludes the proof of the claim. □

Claim 25 is the main step in the proof of Theorem 21: every solution $x \in S_\varepsilon$ corresponds to some optimal mechanism $\varphi \in \textsc{Opt}(\varepsilon)$. Furthermore, by the nature of the representation of mechanisms in the linear program, the convex combination of two solutions corresponds to the hybrid mechanism of the two mechanisms. Thus,

$$S_\varepsilon = (1-\gamma)S_{\varepsilon_{k-1}} + \gamma S_{\varepsilon_k}, \tag{451}$$

implies

$$\textsc{Opt}(\varepsilon) = (1-\gamma)\textsc{Opt}(\varepsilon_{k-1}) + \gamma\textsc{Opt}(\varepsilon_k). \tag{452}$$

Since the objective value d is a variable in the solution, we get

$$\delta(\varepsilon) = (1-\gamma)\delta(\varepsilon_{k-1}) + \gamma\delta(\varepsilon_k). \tag{453}$$

□

5.E.6 Proof of Proposition 18

Proof of Proposition 18. Given a problem (N, M, δ), the algorithms FindLower *and* FindBounds *require at most $4K + \log_2(1/\varepsilon_1) - 1$ executions of the linear program* FindOpt *to determine all supporting manipulability bounds of the Pareto frontier, where K is the number of supporting manipulability bounds and ε_1 is the smallest non-trivial supporting manipulability bound.*

We defined *signature*(ε) (or *sign*(ε) for short) as a function that uses the linear program FindOpt to determine the signature of some mechanism that is optimal at ε, i.e., $(\varepsilon, \delta(\varepsilon))$. Algorithm FindLower initially calls to the function *sign* 3 times, for 0, 1, and 1/2. Now suppose that $\varepsilon_1 \in \left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right)$ for some $n \geqslant 1$. Then *sign* will be called for $\frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^{n+1}}$ until $\underline{\varepsilon} = \frac{1}{2^{n+1}}$ is returned. Note that while $\underline{\varepsilon} = \frac{1}{2^n}$ would have been sufficiently small, the algorithm needs to try $\underline{\varepsilon} = \frac{1}{2^{n+1}}$ to verify this. Thus, it takes $n = \log_2(1/\varepsilon_1)$ calls to *sign*.

The remainder of the proof is concerned with the algorithm FindBounds. Any *segment* consists of two points $(sign(\varepsilon), sign(\varepsilon'))$. Initially, there are two *outer* segments $(sign(0), sign(\underline{\varepsilon}))$ and $(sign(1), sign(2))$, which are needed to help the algorithm get started. The algorithm maintains a decomposition of the interval $[\underline{\varepsilon}, 1]$, which initially consists of a single *unverified* segment. In every execution of the `while`-loop, FindBounds selects an unverified segment $(sign(\varepsilon^-), sign(\varepsilon^+))$. Then it uses the segments to the left and right $(sign(\varepsilon^{--}), sign(\varepsilon^-))$ and $(sign(\varepsilon^+), sign(\varepsilon^{++}))$ to "guess" the position of a new supporting manipulability bound between ε^- and ε^+. This guess e is the ε-value of the point of intersection of the affine hulls of the two segments, i.e.,

$$\mathtt{e} = \left(affine.hull\left(\left\{sign(\varepsilon^{--}), sign(\varepsilon^-)\right\}\right) \cap affine.hull\left(\left\{sign(\varepsilon^+), sign(\varepsilon^{++})\right\}\right)\right)_\varepsilon, \tag{454}$$

where *affine.hull* denotes the affine hull. This value is unique and lies inside the open interval $(\varepsilon^-, \varepsilon^+)$ (by Claim 26). Now $\mathtt{P} = sign(\mathtt{e})$ is computed using the linear program FindOpt, and one of three cases can occur:

1. $\mathtt{P}$ may be equal to the unique point of intersection

$$\mathit{affine.hull}\left(\left\{\mathit{sign}(\varepsilon^{--}), \mathit{sign}(\varepsilon^{-})\right\}\right) \cap \mathit{affine.hull}\left(\left\{\mathit{sign}(\varepsilon^{+}), \mathit{sign}(\varepsilon^{++})\right\}\right). \quad (455)$$

   In this case, $\mathtt{e}$ is a supporting manipulability bound. Furthermore, the segments $(\mathit{sign}(\varepsilon^{--}), \mathit{sign}(\varepsilon^{-}))$, $(\mathit{sign}(\varepsilon^{-}), \mathtt{P})$, $(\mathtt{P}, \mathit{sign}(\varepsilon^{+}))$, and $(\mathit{sign}(\varepsilon^{+}), \mathit{sign}(\varepsilon^{++}))$. are all part of the signature of the Pareto frontier (by Claim 28), and there are no other supporting manipulability bound in the interval $(\varepsilon^{--}, \varepsilon^{++})$. FindBounds marks the four segments as *verified* and includes $\mathtt{e}$ in the collection of supporting manipulability bounds.

2. $\mathtt{P}$ lies in the affine hull of the segment $(\mathit{sign}(\varepsilon^{-}), \mathit{sign}(\varepsilon^{+}))$. Then by Claim 27,

$$\mathit{Conv}\left(\left\{\mathit{sign}(\varepsilon^{-}), \mathtt{P}\right\}\right) \cup \mathit{Conv}\left(\left\{\mathtt{P}, \mathit{sign}(\varepsilon^{+})\right\}\right) \quad (456)$$

   is part of the signature of the Pareto frontier, and there are no supporting manipulability bounds inside the interval $(\varepsilon^{-}, \varepsilon^{+})$. FindBounds marks the segment $(\mathit{sign}(\varepsilon^{-}), \mathit{sign}(\varepsilon^{+}))$ as verified.

3. In any other case, FindBounds splits the segment $(\mathit{sign}(\varepsilon^{-}), \mathit{sign}(\varepsilon^{+}))$ by creating two new unverified segments

$$\left(\mathit{sign}(\varepsilon^{-}), \mathtt{P}\right) \text{ and } \left(\mathtt{P}, \mathit{sign}(\varepsilon^{+})\right). \quad (457)$$

We first show correctness of FindBounds, then completeness:

Correctness FindBounds stops running when there are no more unverified segments. Assume towards contradiction that there exists a supporting manipulability bound $\varepsilon \in (\underline{\varepsilon}, 1)$ that has not been identified. Then this supporting manipulability bound lies in some segment $[\varepsilon^{-}, \varepsilon^{+}]$ that was verified.

If the verification happened in a case 1, Claim 28 ensures that there is no other supporting manipulability bound, i.e., the supporting manipulability bound would have been added to the collection during the analysis of the interval $(\varepsilon^{--}, \varepsilon^{++})$.

If the verification happened in a case 2, Claim 27 ensures that $\varepsilon \notin (\varepsilon^{-}, \varepsilon^{+})$, so that $\varepsilon = \varepsilon^{-}$ (without loss of generality we can assume $\varepsilon = \varepsilon^{-}$, the case $\varepsilon = \varepsilon^{+}$ is analogous). The segment $(\mathit{sign}(\varepsilon^{--}), \mathit{sign}(\varepsilon^{-}))$ was not a verified segment at this time, otherwise this would be a case 1. Thus, at some future step some segment

$(sign(\tilde{\varepsilon}), sign(\varepsilon^-))$ with a right end-point in $sign(\varepsilon^-)$ was verified. But at this step $sign(\varepsilon^-)$ was on the intersection of the affine hulls of the segments $\big(sign(\tilde{\tilde{\varepsilon}}), sign(\tilde{\varepsilon})\big)$ and $(sign(\varepsilon^-), sign(\varepsilon^+))$. This creates a case 1, and thus ε was identified as a supporting manipulability bound in this step.

Completeness It remains to be shown that FindBounds stops at some point. By Claim 29, for ever two adjacent supporting manipulability bounds $\varepsilon_k, \varepsilon_{k+1}$, FindBounds computes at most three points $sign(\varepsilon')$, $sign(\varepsilon'')$, $sign(\varepsilon''')$ with $\varepsilon', \varepsilon'', \varepsilon''' \in (\varepsilon_k, \varepsilon_{k+1})$. Since there is a finite number of supporting manipulability bounds, FindBounds loops at most $3 + 1 = 4$ times per interval, which establishes completeness. The run-time bound follows by observing that there exist $K - 2$ intervals between the smallest non-trivial bound ε_1 and the largest bound $\bar{\varepsilon}$. However, we may need to check the interval $(\hat{\varepsilon}, 1)$ if $\hat{\varepsilon} < 1$. Thus, using $\underline{\varepsilon}$ from FindLower we require at most $4K - 4$ executions of FindOpt to run FindBounds.

Claim 26. $\mathsf{e} \in (\varepsilon^-, \varepsilon^+)$

Proof. By convexity of $\varepsilon \mapsto \delta(\varepsilon)$, we get that $\mathsf{e} \in [\varepsilon^-, \varepsilon^+]$.

Now suppose that $\mathsf{e} = \varepsilon^-$. Then $\varepsilon^- \in affine.hull(\{sign(\varepsilon^+), sign(\varepsilon^{++})\})$. Since $\varepsilon^- < \varepsilon^+$, the segments $(sign(\varepsilon^-), sign(\varepsilon^+))$ and $(sign(\varepsilon^+), sign(\varepsilon^{++}))$ would have been verified in a previous step. But this is a contradiction to the assumption that $(sign(\varepsilon^-), sign(\varepsilon^+))$ was an unverified segment. □

Claim 27. *For $0 \leqslant \varepsilon_{-1} < \varepsilon_0 < \varepsilon_1 \leqslant 1$, if*

$$sign(\varepsilon_0) \in Conv(\{sign(\varepsilon_{-1}), sign(\varepsilon_1)\}), \tag{458}$$

then

$$Conv(\{sign(\varepsilon_{-1}), sign(\varepsilon_1)\}) \tag{459}$$

is part of the signature of the Pareto frontier.

Proof. Assume towards contradiction that $Conv(\{sign(\varepsilon_{-1}), sign(\varepsilon_1)\})$ is not part of the signature of the Pareto frontier. Then by convexity of $\varepsilon \mapsto \delta(\varepsilon)$ there exists $\gamma \in (0, 1)$ with

$$\delta(\varphi_{(1-\gamma)\varepsilon_{-1}+\gamma\varepsilon_1}) < (1-\gamma)\delta(\varphi_{\varepsilon_{-1}}) + \gamma\delta(\varphi_{\varepsilon_1}). \tag{460}$$

If $\varepsilon' = (1-\gamma)\varepsilon_{-1} + \gamma\varepsilon_1 > \varepsilon_0$, then for $\gamma' = \frac{\varepsilon_0 - \varepsilon_{-1}}{\varepsilon' - \varepsilon_{-1}}$ we get

$$(1-\gamma')\delta(\varphi_{\varepsilon_{-1}}) + \gamma'\delta(\varphi_{\varepsilon'}) < \delta(\varphi_{\varepsilon_0}) = \delta(\varphi_{(1-\gamma')\varepsilon_{-1}+\gamma'\varepsilon'}), \tag{461}$$

a contradiction to convexity of $\varepsilon \mapsto \delta(\varepsilon)$. The argument for $\varepsilon' < \varepsilon_0$ is analogous. □

Claim 28. *For* $0 \leqslant \varepsilon_{-2} < \varepsilon_{-1} < \varepsilon_0 < \varepsilon_1 < \varepsilon_2 \leqslant 1$, *if*

$$\begin{aligned} & \{sign(\varepsilon)\} \\ = \; & affine.hull(\{sign(\varepsilon_{-2}), sign(\varepsilon_{-1})\}) \cap affine.hull(\{sign(\varepsilon_1), sign(\varepsilon_2)\}), \end{aligned} \tag{462}$$

then

$$Conv(\{sign(\varepsilon_0), sign(\varepsilon_{\pm 2})\}) \tag{463}$$

is part of the signature of the Pareto frontier.

Proof. The claim follows by applying Claim 27 twice. □

Claim 29. *For any two adjacent supporting manipulability bounds* $\varepsilon_k, \varepsilon_{k+1}$, FindBounds *computes at most three points* $sign(\varepsilon')$, $sign(\varepsilon'')$, $sign(\varepsilon''')$ *with* $\varepsilon', \varepsilon'', \varepsilon''' \in (\varepsilon_k, \varepsilon_{k+1})$.

Proof. Suppose that $sign(\varepsilon')$, $sign(\varepsilon'')$, and $sign(\varepsilon''')$ are computed in this order. If $\varepsilon''' < \min(\varepsilon', \varepsilon'')$, then ε' must be a supporting manipulability bound by convexity of $\varepsilon \mapsto \delta(\varepsilon)$, which is a contradiction. The same holds if $\varepsilon''' > \max(\varepsilon', \varepsilon'')$. If $\varepsilon''' \in (\min(\varepsilon', \varepsilon''), \max(\varepsilon', \varepsilon''))$, the segment

$$(sign(\min(\varepsilon', \varepsilon'')), sign(\max(\varepsilon', \varepsilon''))) \tag{464}$$

is verified (case (2)). Another guess ε'''' that lies within $[\varepsilon_k, \varepsilon_{k+1}]$ involve the segment $(sign(\varepsilon'), sign(\varepsilon''))$. Thus, by convexity of $\varepsilon \mapsto \delta(\varepsilon)$, ε'''' is a supporting manipulability bound. □

This concludes the proof of Proposition 18 □

5.E.7 Proof of Proposition 19

Proof of Proposition 19. In a problem (N, M, δ) *with* $n = 3$ *agents,* $m = 3$ *alternatives, strict preferences, and where* δ *is the worst-case* w^{Plu}*-deficit, the following hold:*

1. *The Pareto frontier has two supporting manipulability bounds* $\{0, 1/3\}$.
2. *Random Dictatorship is a representative of* Opt(0).
3. *Uniform Plurality is a representative of* Opt(1/3).

We first prove that Random Dictatorship is optimal at $\varepsilon_0 = 0$. Since Random Dictatorship is by construction a lottery of unilateral, strategyproof mechanisms, it is obviously strategyproof. At any preference profile where all agents have the same first choice, Random Dictatorship will select this alternative and achieve zero deficit. At any preference profile where all agents have different first choices, all alternatives have the same welfare and therefore, any outcome has zero deficit. Finally, consider the case where two agents agree on a first choice, a say, but the third agent has a different first choice, b say. In this case, selecting a would yield a maximal welfare of $\frac{2}{3}$. However, Random Dictatorship will only select alternative a with probability $\frac{2}{3}$ and b otherwise. This leads to an outcome with welfare of $\frac{2}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} = \frac{5}{9}$, and therefore, the deficit of Random Dictatorship is $\frac{1}{9}$.

It remains to be shown that any strategyproof mechanism will have a deficit of at least $\frac{1}{9}$. Following the discussion of the second takeaway from Theorem 20 in Section 5.7, we can restrict our attention to mechanisms φ that are strategyproof, anonymous, and neutral. By Theorem 22, φ has a symmetric decomposition, i.e., it can be represented as a lottery over neutral, strategyproof unilaterals and anonymous, strategyproof duples. Suppose, it contains a some anonymous, strategyproof duple $d_{a,b}$. By the characterization of strategyproofness via swap monotonicity, upper invariance, and lower invariance (Theorem 1 in (Mennle and Seuken, 2015*b*)), it follows that the outcome of $d_{a,b}$ can only depend on the relative rankings of a and b, so that $d_{a,b}$ has the form

$$d_{a,b}(\boldsymbol{P}) = \begin{cases} (p_3, 1-p_3, 0)\,, & \text{if } P_i : a \succ b \text{ for all agents } i, \\ (p_2, 1-p_2, 0)\,, & \text{if } P_i : a \succ b \text{ for two agents } i, \\ (p_1, 1-p_1, 0)\,, & \text{if } P_i : a \succ b \text{ for one agents } i, \\ (p_0, 1-p_0, 0)\,, & \text{if } P_i : a \succ b \text{ for one agents } i, \end{cases} \tag{465}$$

where the vector $(p, 1-p, 0) = (\varphi_a(\boldsymbol{P}), \varphi_b(\boldsymbol{P}), \varphi_c(\boldsymbol{P}))$ denotes the outcome and $p_3 \geqslant p_2 \geqslant \frac{1}{2}$ and $p_0 \leqslant p_1 \leqslant \frac{1}{2}$. Again by symmetry of the symmetric decomposition, it must also contain the anonymous duple $d^{\varpi}_{a,b}$ for any permutation of the alternatives $\varpi : M \to M$. Consider the preference profile

$$P_1 : \quad a \succ b \succ c, \tag{466}$$

$$P_2 : \quad a \succ c \succ b, \tag{467}$$

$$P_3 : \quad b \succ c \succ a. \tag{468}$$

The following table shows what outcomes of the different duples $\{d^{\varpi}_{a,b} \mid \varpi : M \to$

M permutation} at this preference profile.

Duple	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$
$d_{a,b}$	p_2	$1-p_2$	0
$d_{a,c}$	p_2	0	$1-p_2$
$d_{b,c}$	0	p_2	$1-p_2$
$d_{b,a}$	$1-p_1$	p_1	0
$d_{c,a}$	$1-p_1$	0	p_1
$d_{c,b}$	0	$1-p_1$	p_1

A uniform lottery over these duples assigns probability $2 \cdot \frac{1}{6} = \frac{1}{3}$ to alternative b. Since $p_1 \geqslant p_2$, this mechanism selects an outcome with welfare at most $\frac{2}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} = \frac{5}{9}$. Since the best alternative a has welfare $\frac{2}{3}$, the mechanism must have a deficit of at least $\frac{1}{9}$ at this particular preference profile. This is the same deficit that Random Dictatorship has at this profile, which means that including any strategyproof duples in the symmetric decomposition will not improve the deficit of $\varphi(\boldsymbol{P})$.

Suppose now that the symmetric decomposition of φ contains a neutral, strategyproof unilateral u_i. As before, it follows that u_i must pick an outcome $(p_1, p_2, 1-p_1-p_2)$ where $p_1 \geqslant p_2 \geqslant 1-p1-p_2$. Again by symmetry of the symmetric decomposition, it must also contain the neutral unilateral u_i^π for any permutation of the agents $\pi : N \to N$. Consider the same preference profile as before, with

$$P_1 : \quad a \succ b \succ c, \tag{469}$$

$$P_2 : \quad a \succ c \succ b, \tag{470}$$

$$P_3 : \quad b \succ c \succ a. \tag{471}$$

The following table shows what outcomes the different unilaterals $\{u_i^\pi \mid \pi : N \to N$ permutation$\}$ will select.

Unilateral	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$
u_1	p_1	p_2	$1-p_1-p_2$
u_2	p_1	$1-p_1-p_2$	p_2
u_3	$1-p_1-p_2$	p_1	p_2

A uniform lottery over these unilaterals assigns probability $\frac{1}{3}$ to alternative b, and consequently, this mechanism has a deficit of at least $\frac{1}{9}$ at this profile. This is the same deficit that Random Dictatorship has at this profile, which means that including any

strategyproof unilaterals in the symmetric decomposition will not improve the deficit of $\varphi(\boldsymbol{P})$.

Since for all mechanisms the worst-case deficit was attained at the same preference profile, there exists no strategyproof, anonymous, neutral mechanism that has a lower score deficit at $\boldsymbol{P}$ than $\frac{1}{9}$, and therefore, Random Dictatorship has minimal deficit among all strategyproof mechanisms.

Next, we show that the manipulability of Uniform Plurality is minimal among all welfare maximizing mechanisms. Without loss of generality, we restrict attention to anonymous, neutral, and welfare maximizing mechanisms. At any preference profile where an alternative is ranked first by two agents or more, this alternative is implemented with certainty. Thus, if two agents have the same first choice, the third agent has no opportunity to manipulate, because it cannot change the outcome. The two agents with the same first choice are already receiving their favorite outcome, which makes manipulation useless for them as well.

Thus, any manipulability will arise at a preference profile where all agents have different first choices. Consider the preference profile $\boldsymbol{P}$ with

$$P_1: \quad a \succ b \succ c, \tag{472}$$

$$P_2: \quad b \succ c \succ a, \tag{473}$$

$$P_3: \quad c \succ a \succ b. \tag{474}$$

Renaming the alternatives is equivalent to renaming the agents. Thus, an anonymous and neutral mechanisms has to treat all alternatives equally and must therefore select each alternative with probability $\frac{1}{3}$. Now suppose that agent 1 is almost indifferent between a and b, but strongly dislikes c, i.e., its utility function is "close to" the binary utility $u(a) = u(b) = 1, u(c) = 0$. Then, by swapping a and b, agent 1 can enforce the implementation b with certainty. Its gain from this manipulation is close to

$$1 \cdot u(b) - \frac{1}{3}\left(u(a) + u(b) + u(c)\right) = \frac{1}{3}. \tag{475}$$

Thus, any welfare maximizing mechanism has manipulability $\varepsilon(\varphi) \geqslant \frac{1}{3}$.

Now consider the Uniform Plurality mechanism. At the above preference profile agents cannot change the outcome, unless they change their first choice. By anonymity and neutrality, it suffices to show that agent 1 cannot do any better than $\frac{1}{3}$ by manipulating. However, the only other possible misreport that has any effect on the outcome is to

bring c forward and enforce it as the outcome, which would yield no benefit for agent 1. Consequently, Uniform Plurality has minimal manipulability of $\frac{1}{3}$.

Finally, we show that there are no additional supporting manipulability bounds besides 0 and 1/3. So far we have that Random Dictatorship is in $\textsc{Opt}(0)$ with a deficit of $\delta(0) = \frac{1}{9}$. Furthermore, Uniform Plurality is in $\textsc{Opt}(1/3)$ and no welfare maximizing mechanism has strictly lower manipulability. To complete the proof, we will show that for $\varepsilon = \frac{1}{6}$, all optimal mechanisms have deficit $\frac{1}{18}$. By convexity of the mapping $\varepsilon \mapsto \delta(\varepsilon)$, the signature of the Pareto frontier must therefore be a straight line between $(0, 1/9)$ and $(1/3, 0)$. Considering the performance guarantees for hybrid mechanisms (Theorem 20), this implies optimality of the hybrids of Random Dictatorship and Uniform Plurality.

Suppose that φ is $\frac{1}{6}$-approximately strategyproof, anonymous, and neutral. It follows from anonymity and neutrality that at the preference profile

$$P_1 : \quad a \succ b \succ c, \tag{476}$$

$$P_2 : \quad b \succ c \succ a, \tag{477}$$

$$P_3 : \quad c \succ a \succ b, \tag{478}$$

the outcome must be $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ for a, b, c, respectively. If agent 1 changes its report to

$$P_1' : \quad b \succ a \succ c, \tag{479}$$

the outcome changes to $\left(\frac{1}{3} - \underline{\varepsilon}_a, \frac{1}{3} + \underline{\varepsilon}_a + \underline{\varepsilon}_c, \frac{1}{3} - \underline{\varepsilon}_c\right)$ for some values $\underline{\varepsilon}_a \leqslant \frac{1}{3}, \underline{\varepsilon}_c \leqslant \frac{1}{3}$. Suppose now that agent 1 has a utility close to indifference between a and b, i.e., close to $u(a) = u(b) = 1, u(c) = 0$. Then the utility gain for agent 1 from misreporting P_1' is (arbitrarily close to)

$$1 \cdot (-\underline{\varepsilon}_a) + 1 \cdot (\underline{\varepsilon}_a + \underline{\varepsilon}_c) + 0 \cdot (-\underline{\varepsilon}_c) \approx \underline{\varepsilon}_c. \tag{480}$$

If the mechanism is $\frac{1}{6}$-approximately strategyproof, then $\underline{\varepsilon}_c \leqslant \frac{1}{6}$. The welfare of $\varphi(\boldsymbol{P}')$ at

$\boldsymbol{P}' = (P_1', P_2, P_3)$ is

$$\begin{aligned}
\left\langle \varphi(\boldsymbol{P}'), w^{\text{PLU}}(\cdot, \boldsymbol{P}') \right\rangle &= \varphi_a(\boldsymbol{P}') \cdot 0 + \varphi_b(\boldsymbol{P}') \cdot \frac{2}{3} + \varphi_c(\boldsymbol{P}') \cdot \frac{1}{3} && (481)\\
&= \frac{2}{3}\left(\frac{1}{3} + \underline{\varepsilon}_a + \underline{\varepsilon}_c\right) + \frac{1}{3}\left(\frac{1}{3} - \underline{\varepsilon}_c\right) && (482)\\
&= \frac{6}{18} + \frac{4}{6}\underline{\varepsilon}_a + \frac{1}{3}\underline{\varepsilon}_c && (483)\\
&\leqslant \frac{6}{18} + \frac{4}{18} + \frac{1}{18} = \frac{11}{18}. && (484)
\end{aligned}$$

However, the only welfare maximizing alternative at $\boldsymbol{P}'$ is b with $w^{\text{PLU}}(b, \boldsymbol{P}') = \frac{2}{3} = \frac{12}{18}$. Thus, any $\frac{1}{6}$-approximately strategyproof mechanism must incur a deficit of at least $\frac{1}{18}$. □

5.E.8 Proof of Proposition 20

Proof of Proposition 20. In a problem (N, M, δ) with $n = 3$ agents, $m = 3$ alternatives, strict preferences, and where δ is the worst-case w^{Veto}-deficit, the following hold:

1. *The Pareto frontier has four supporting manipulability bounds $\{0, 1/21, 1/12, 1/2\}$.*
2. *Random Duple is a representative of* $\textsc{Opt}(0)$.
3. *Uniform Veto is a representative of* $\textsc{Opt}(1/2)$.
4. *Hybrids of Random Duple and Uniform Veto are not on the Pareto frontier, except for $\beta \in \{0, 1\}$.*

First, we prove that Random Duple is optimal at $\varepsilon_0 = 0$. Since it is a lottery over strategyproof duple mechanisms, it is obviously strategyproof.

At any preference profile where all agents agree on the last choice, Random Duple selects one of the other alternatives, each of which gives maximal welfare. At any preference profile where all agents have different last choices, any outcome has zero deficit. Finally, consider a preference profile with

$$\begin{aligned}
P_1, P_2 : &\quad \ldots \succ c, && (485)\\
P_3 : &\quad \ldots \succ b. && (486)
\end{aligned}$$

The welfare of a is 1 and the welfare of b is $\frac{2}{3}$, so that the maximum welfare is 1. The worst case for Random Duple is that agents 1 and 2 rank b first, in which case a will be

selected with probability $\frac{1}{3}$ and b with probability $\frac{2}{3}$. Thus, the deficit of Random Duple at $\boldsymbol{P}$ is

$$1 - \frac{1}{3} \cdot 1 - \frac{2}{3} \cdot \frac{2}{3} = \frac{2}{9}. \tag{487}$$

In particular, this deficit is attained by Random Duple at the preference profile $\boldsymbol{P}$ with

$$P_1, P_2 : \quad b \succ a \succ c, \tag{488}$$

$$P_3 : \quad c \succ a \succ b. \tag{489}$$

It remains to be proven whether any strategyproof, anonymous, neutral mechanism φ can achieve a lower deficit (where anonymity and neutrality are without loss of generality, similar to the proof of Proposition 19). Consider the preference profile $\boldsymbol{P}$ and let u_i be a strategyproof and neutral unilateral component in the symmetric decomposition of φ. u_i must pick an outcome $(p_1, p_2, 1 - p_1 - p_2)$ with $p_1 \geqslant p_2 \geqslant 1 - p_1 - p_2$, where p_k denotes the probability of agent i's kth choice. The symmetric decomposition implies that u_1, u_2, and u_3 are equally likely to be chosen. Analogous to the proof of Proposition 19, we get

Unilateral	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$
u_1 or u_2	p_1	p_2	$1 - p_1 - p_2$
u_3	p_2	$1 - p_1 - p2$	p_1

Thus, alternative b is selected with probably at least $\frac{1}{3}$, which means that the deficit of φ at the preference profile $\boldsymbol{P}$ is not reduced by including any unilaterals in the symmetric decomposition.

Similarly, if $d_{a,b}$ is a strategyproof, anonymous duple in the symmetric decomposition of φ, it has the form

$$d_{a,b}(\boldsymbol{P}) = \begin{cases} (p_3, 1 - p_3, 0)\,, & \text{if } P_i : a \succ b \text{ for all agents } i, \\ (p_2, 1 - p_2, 0)\,, & \text{if } P_i : a \succ b \text{ for two agents } i, \\ (p_1, 1 - p_1, 0)\,, & \text{if } P_i : a \succ b \text{ for one agents } i, \\ (p_0, 1 - p_0, 0)\,, & \text{if } P_i : a \succ b \text{ for one agents } i, \end{cases} \tag{490}$$

where $p_3 \geqslant p_2 \geqslant \frac{1}{2}$ and $p_0 \leqslant p_1 \leqslant \frac{1}{2}$. Again by symmetry of the symmetric decomposition, it must also contain the anonymous duple $d^{\varpi}_{a,b}$ for any permutation of the alternatives $\varpi : M \to M$. The following table shows what outcomes the different duples $\{d^{\varpi}_{a,b} \mid \varpi : M \to M \text{ permutation}\}$ will select at $\boldsymbol{P}$.

Duple	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$
$d_{a,b}$	p_1	$1-p_1$	0
$d_{a,c}$	p_2	0	$1-p_2$
$d_{b,c}$	0	p_2	$1-p_2$
$d_{b,a}$	$1-p_2$	p_2	0
$d_{c,a}$	$1-p_1$	0	p_1
$d_{c,b}$	0	$1-p_1$	p_1

Thus, since $p_2 \geqslant p_1$, b is selected with probability of at least $\frac{1}{3}$, and therefore, including any other duples in the symmetric decomposition of φ will not improve the deficit at $\boldsymbol{P}$. Consequently, there exists no strategyproof mechanism with a lower deficit than Random Duple.

Next, we show that Uniform Veto has minimal manipulability of among all welfare maximizing mechanisms. First, observe that Uniform Veto is $\frac{1}{2}$-approximately strategyproof. To see this, consider the preference profile $\boldsymbol{P}$

$$P_1: \quad a \succ b \succ c, \tag{491}$$

$$P_2, P_3: \quad \ldots \succ c. \tag{492}$$

Uniform Veto selects a and b with probability $\frac{1}{2}$ each. To manipulate, agent 1 can rank b last and obtain a with certainty. Its gain from this manipulation would be

$$u_1(a) - \frac{1}{2}\left(u_1(a) + u_1(b)\right) = \frac{1}{2}\left(u_1(a) - u_1(b)\right), \tag{493}$$

which is at most $\frac{1}{2}$ for $u_1(a) = 1$ and $u_1(b)$ close to 0. Thus, the manipulability of Uniform Veto is at least $\frac{1}{2}$.

Suppose now that all agents have different last choices. In this case, Uniform Veto selects any of the alternatives with probability $\frac{1}{3}$. By ranking another alternative last, an agent could only ensure the implementation of its third choice with certainty, which is not a beneficial manipulation.

Finally, suppose that two agents have the same last choice, while a third agent has a different last choice. We have the following cases from the perspective of agent 1.

- Case I:

$$P_1: \quad a \succ b \succ c, \tag{494}$$

$$P_2, P_3: \quad \ldots \succ a. \tag{495}$$

In this case, Uniform Veto implements b with certainty. Agent 1 can only enforce c by ranking b last, or rank a last and obtain b and c with probabilities $\frac{1}{2}$ each. Neither of these moves will make agent 1 better off.

- Case II:

$$P_1: \quad a \succ b \succ c, \tag{496}$$
$$P_2, P_3: \quad \ldots \succ b. \tag{497}$$

In this case, Uniform Veto implements a with certainty, which is already agent 1's first choice.

- Case III:

$$P_1: \quad a \succ b \succ c, \tag{498}$$
$$P_2: \quad \ldots \succ b, \tag{499}$$
$$P_3: \quad \ldots \succ c. \tag{500}$$

In this case, Uniform Veto implements a with certainty, which is already agent 1's first choice.

- Case IV:

$$P_1: \quad a \succ b \succ c, \tag{501}$$
$$P_2: \quad \ldots \succ a, \tag{502}$$
$$P_3: \quad \ldots \succ c. \tag{503}$$

In this case, Uniform Veto implements b with certainty. By ranking b last, agent 1 could obtain a probability of $\frac{1}{3}$ for each alternative instead. Its gain from this manipulation is

$$\frac{1}{3}\left(u_1(a) + u_1(b) + u_1(c)\right) - u_1(b) \leqslant \frac{1}{3} - \frac{2}{3}u_1(b) \leqslant \frac{1}{3}. \tag{504}$$

Further observe that by renaming agents and alternatives, the above cases I through IV cover all possible constellations with 2 different alternatives ranked as last choices from the perspective of any agent. Thus, Uniform Veto is $\frac{1}{2}$-approximately strategyproof.

Having shown that the manipulability of Uniform Veto is exactly $\frac{1}{2}$, we now must

show that any other welfare maximizing mechanism φ has manipulability $\varepsilon(\varphi) \geqslant \frac{1}{2}$. We assume without loss of generality that φ is also anonymous and neutral, and we consider the preference profile

$$P_1, P_2 : \quad a \succ b \succ c, \tag{505}$$

$$P_3 : \quad b \succ a \succ c. \tag{506}$$

At this profile, φ has to select a with some probability p_a and b with probability $1 - p_a$. If $p_a \geqslant \frac{1}{2}$, agent 3 can rank a last and enforce selection of b, the only remaining alternative with full welfare. If $u_3(b) = 1$, $u_3(c) = 0$, and $u_3(a)$ is close to 0, its gain will be

$$1 - (1 - p_a) = p_a \geqslant \frac{1}{2}. \tag{507}$$

If $p_a < \frac{1}{2}$, agent 1 can enforce a by ranking b last and obtain a gain of

$$1 - p_a > \frac{1}{2} \tag{508}$$

with a similar utility function. Thus, any welfare maximizing mechanism has manipulability of at least $\frac{1}{2}$.

Last, we show that the hybrids of Random Duples and Uniform Veto do not lie on the Pareto frontier, except for the extreme cases. In fact, the signatures of their hybrids form a straight line: observe that the deficit of Random Duple is attained at the preference profile $\boldsymbol{P}$ with

$$P_1, P_2, P_3 : \quad a \succ b \succ c. \tag{509}$$

Since Uniform Veto is welfare maximizing, its deficit is zero at all preference profiles. Consequently, by linearity of the welfare $w(x, \boldsymbol{P})$ in the outcome x, the deficit of any hybrid h_β of Random Duple and Uniform Veto is determined by the deficit at $\boldsymbol{P}$. Furthermore, the manipulability of Uniform Veto is highest at the same preference profile if agent 1 swaps b and c to enforce a and has a utility close to $u_1(a) = 1$, $u_1(c) = 0$, and $u_1(b)$ close to 0. This misreport leaves the outcome of Random Duple unchanged. Therefore, the manipulability of any hybrid will also be determined by this preference profile and this potential misreport. By linearity of the incentive constraints from Theorem 19 it is evident that the signatures of the hybrids of Random Duple and Uniform Veto for a straight line between the signatures $(0, 2/9)$ and $(1/2, 0)$ of the

Algorithm	ε	$\delta(\varepsilon)$	**Comment**
FindLower	0	2/9	found s.m.b. at $\varepsilon = 0$
FindLower	1/2	0	
FindLower	1/4	1/12	
FindLower	1/8	1/8	
FindLower	1/16	65/432	
FindLower	1/32	13/72	
FindLower	1/64	29/144	found $\underline{\varepsilon} = 1/64 < \varepsilon_1$
FindBounds	1/6	1/9	
FindBounds	1/18	25/162	
FindBounds	3/47	28/187	
FindBounds	1/12	5/36	found s.m.b. at $\varepsilon = 1/12$ and $\varepsilon = 1/2$
FindBounds	1/21	10/63	found s.m.b. at $\varepsilon = 1/21$

Table 5.1: Executions of the linear program FindOpt when using FindLower and FindBounds to determine the Pareto frontier.

respective component mechanism. Consequently, if the Pareto frontier is not linear, then these hybrids will not be on the Pareto frontier for any $\beta \neq 0, 1$.

To find the supporting manipulability bounds of the Pareto frontier, we used the algorithm FindLower to determine a lower bound for the smallest non-zero supporting manipulability bound and then applied FindBounds with this value of $\underline{\varepsilon}$. Table 5.1 gives the signatures on the Pareto frontier that were determined using the *signature*-function in the order in which they were computed.

Tables 5.2 and 5.3 give two mechanisms that are optimal at $\varepsilon_1 = 1/21$ and $\varepsilon_2 = 1/12$, respectively. For preference profiles that are not listed, rename the agents and alternatives to obtain one of the listed preference profiles, and select the respective outcome (renaming the alternatives again).

□

Preference Profile			$\boldsymbol{\varepsilon_1 = 1/21}$		
$\boldsymbol{P_1}$	$\boldsymbol{P_2}$	$\boldsymbol{P_3}$	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$
$a \succ b \succ c$	$a \succ b \succ c$	$a \succ b \succ c$	11/21	3/7	1/21
$a \succ b \succ c$	$a \succ b \succ c$	$a \succ c \succ b$	4/7	3/7	0
$a \succ b \succ c$	$a \succ b \succ c$	$b \succ a \succ c$	11/21	10/21	0
$a \succ b \succ c$	$a \succ b \succ c$	$b \succ c \succ a$	10/21	11/21	0
$a \succ b \succ c$	$a \succ b \succ c$	$c \succ a \succ b$	4/7	8/21	1/21
$a \succ b \succ c$	$a \succ b \succ c$	$c \succ b \succ a$	10/21	11/21	0
$a \succ b \succ c$	$a \succ c \succ b$	$a \succ c \succ b$	4/7	0	3/7
$a \succ b \succ c$	$a \succ c \succ b$	$b \succ a \succ c$	4/7	3/7	0
$a \succ b \succ c$	$a \succ c \succ b$	$b \succ c \succ a$	3/7	8/21	4/21
$a \succ b \succ c$	$a \succ c \succ b$	$c \succ a \succ b$	4/7	0	3/7
$a \succ b \succ c$	$a \succ c \succ b$	$c \succ b \succ a$	3/7	4/21	8/21
$a \succ b \succ c$	$b \succ a \succ c$	$b \succ a \succ c$	10/21	11/21	0
$a \succ b \succ c$	$b \succ a \succ c$	$b \succ c \succ a$	3/7	4/7	0
$a \succ b \succ c$	$b \succ a \succ c$	$c \succ a \succ b$	4/7	8/21	1/21
$a \succ b \succ c$	$b \succ a \succ c$	$c \succ b \succ a$	8/21	4/7	1/21
$a \succ b \succ c$	$b \succ c \succ a$	$b \succ c \succ a$	1/21	4/7	8/21
$a \succ b \succ c$	$b \succ c \succ a$	$c \succ a \succ b$	1/3	1/3	1/3
$a \succ b \succ c$	$b \succ c \succ a$	$c \succ b \succ a$	1/21	4/7	8/21
$a \succ b \succ c$	$c \succ a \succ b$	$c \succ a \succ b$	11/21	0	10/21
$a \succ b \succ c$	$c \succ a \succ b$	$c \succ b \succ a$	8/21	4/21	3/7
$a \succ b \succ c$	$c \succ b \succ a$	$c \succ b \succ a$	0	11/21	10/21

Table 5.2: Optimal mechanism at $\varepsilon_1 = 1/21$ (extended to other preference profiles via the anonymity and neutrality extension).

Preference Profile			$\boldsymbol{\varepsilon_2 = 1/12}$		
$\boldsymbol{P_1}$	$\boldsymbol{P_2}$	$\boldsymbol{P_3}$	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$
$a \succ b \succ c$	$a \succ b \succ c$	$a \succ b \succ c$	7/12	1/3	1/12
$a \succ b \succ c$	$a \succ b \succ c$	$a \succ c \succ b$	2/3	1/4	1/12
$a \succ b \succ c$	$a \succ b \succ c$	$b \succ a \succ c$	1/2	1/2	0
$a \succ b \succ c$	$a \succ b \succ c$	$b \succ c \succ a$	5/12	7/12	0
$a \succ b \succ c$	$a \succ b \succ c$	$c \succ a \succ b$	2/3	1/3	0
$a \succ b \succ c$	$a \succ b \succ c$	$c \succ b \succ a$	5/12	7/12	0
$a \succ b \succ c$	$a \succ c \succ b$	$a \succ c \succ b$	2/3	1/12	1/4
$a \succ b \succ c$	$a \succ c \succ b$	$b \succ a \succ c$	7/12	5/12	0
$a \succ b \succ c$	$a \succ c \succ b$	$b \succ c \succ a$	1/3	5/12	1/4
$a \succ b \succ c$	$a \succ c \succ b$	$c \succ a \succ b$	7/12	0	5/12
$a \succ b \succ c$	$a \succ c \succ b$	$c \succ b \succ a$	1/3	1/4	5/12
$a \succ b \succ c$	$b \succ a \succ c$	$b \succ a \succ c$	1/2	1/2	0
$a \succ b \succ c$	$b \succ a \succ c$	$b \succ c \succ a$	5/12	7/12	0
$a \succ b \succ c$	$b \succ a \succ c$	$c \succ a \succ b$	7/12	5/12	0
$a \succ b \succ c$	$b \succ a \succ c$	$c \succ b \succ a$	5/12	7/12	0
$a \succ b \succ c$	$b \succ c \succ a$	$b \succ c \succ a$	0	2/3	1/3
$a \succ b \succ c$	$b \succ c \succ a$	$c \succ a \succ b$	1/3	1/3	1/3
$a \succ b \succ c$	$b \succ c \succ a$	$c \succ b \succ a$	0	7/12	5/12
$a \succ b \succ c$	$c \succ a \succ b$	$c \succ a \succ b$	7/12	0	5/12
$a \succ b \succ c$	$c \succ a \succ b$	$c \succ b \succ a$	5/12	1/4	1/3
$a \succ b \succ c$	$c \succ b \succ a$	$c \succ b \succ a$	0	7/12	5/12

Table 5.3: Optimal mechanism at $\varepsilon_2 = 1/12$ (extended to other preference profiles via the anonymity and neutrality extension).

References

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*a*. "Ordinal Efficiency and Dominated Sets of Assignments." *Journal of Economic Theory*, 112(1): 157–172.

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

Abdulkadiroğlu, Atila, Parag Pathak, and Alvin Roth. 2005. "The New York City High School Match." *American Economic Review*, 95(2): 364–367.

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.

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda. 2015. "Expanding 'Choice' in School Choice." *American Economic Journal: Microeconomics*, 7(1): 1–42.

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

Azevedo, Eduardo, and Eric Budish. 2015. "Strategy-proofness in the Large." Working Paper.

Aziz, Haris, Felix Brandt, and Markus Brill. 2013*a*. "On the Tradeoff Between Economic Efficiency and Strategy Proofness in Randomized Social Choice." In *Pro-*

ceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems (AAMAS).

Aziz, Haris, Felix Brandt, and Markus Brill. 2013*b*. "The Computational Complexity of Random Serial Dictatorship." *Economics Letters*, 121(3): 341–345.

Aziz, Haris, Florian Brandl, and Felix Brandt. 2014. "On the Incompatibility of Efficiency and Strategyproofness in Randomized Social Choice." In *Proceedings of the 28th Conference on Artificial Intelligence (AAAI).*

Bade, Sophie. 2014. "Random Serial Dictatorship: The One and Only." Mimeo.

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

Barbera, Salvador, Dolors Berga, and Bernardo Moreno. 2012. "Two Necessary Conditions for Strategy-proofness: On What Domains Are They Also Sufficient?" *Games and Economic Behavior*, 75(2): 490–509.

Bartholdi, John, Craig Tovey, and Michael Trick. 1989. "The Computational Difficulty of Manipulating an Election." *Social Choice and Welfare*, 6(3): 227–241.

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

Birkhoff, Garrett. 1946. "Three Observations on Linear Algebra." *Univ. Nac. Tucumán*, Revista A(5): 147–151.

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. 2011. "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes." *Journal of Political Economy*, 119(6): 1061–1103.

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.

Calsamiglia, Caterina, and Maia Güell. 2014. "The Illusion of School Choice: Empirical Evidence from Barcelona." Working Paper.

Carroll, Gabriel. 2011. "On Mechanisms Eliciting Ordinal Preferences." Working Paper.

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.

Chatterji, Shurojit, Remzi Sanver, and Arunava Sen. 2013. "On Domains that Admit Well-behaved Strategy-proof Social Choice Functions." *Journal of Economic Theory*, 148(3): 1050–1073.

Chen, Yan, and Tayfun Sönmez. 2006. "School Choice: an Experimental Study." *Journal of Economic Theory*, 127(1): 202–231.

Cho, Wonki. 2012. "Probabilistic Assignment: A Two-fold Axiomatic Approach." Working Paper.

de Haan, Monique, Pieter Gautier, Hessel Oosterbeek, and Bas van der Klaauw. 2015. "The Performance of School Assignment Mechanisms in Practice." Working Paper.

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

Dur, Umut, Scott Kominers, Parag Pathak, and Tayfun Sönmez. 2014. "The Demise of Walk Zones in Boston: Priorities vs. Precedence in School Choice." Working Paper.

Ehlers, Lars, and Bettina Klaus. 2006. "Efficient Priority Rules." *Games and Economic Behavior*, 55(2): 372–384. Mini Special Issue: Electronic Market Design.

Ehlers, Lars, and Bettina Klaus. 2007. "Consistent House Allocation." *Economic Theory*, 30(3): 561–574.

Ehlers, Lars, Dipiyoti Majumdar, Debasis Mishra, and Arunava Sen. 2015. "Continuous Cardinal Incentive Compatible Mechanisms are Ordinal." Working Paper.

Ehlers, Lars, Hans Peters, and Ton Storcken. 2002. "Strategy-Proof Probabilistic Decision Schemes for One-Dimensional Single-Peaked Preferences." *Journal of Economic Theory*, 105(2): 408–434.

Ekici, Özgün, and Onur Kesten. 2012. "An Equilibrium Analysis of the Probabilistic Serial Mechanism." Working Paper.

Erdil, Aytek. 2014. "Strategy-proof Stochastic Assignment." *Journal of Economic Theory*, 151: 146–162.

Ergin, Haluk, and Tayfun Sönmez. 2006. "Games of School Choice Under the Boston Mechanism." *Journal of Public Economics*, 90(1-2): 215–237.

Fack, Gabrielle, and Julien Grenet. 2010. "When Do Better Schools Raise Housing Prices? Evidence from Paris Public and Private Schools." *Journal of Public Economics*, 94(1-2): 59–77.

Featherstone, Clayton. 2011. "A Rank-based Refinement of Ordinal Efficiency and a new (but Familiar) Class of Ordinal Assignment Mechanisms." Working Paper.

Gale, David, and Lloyd Shapley. 1962. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly*, 69(1): 9–15.

Gibbard, Allan. 1973. "Manipulation of Voting Schemes: a General Result." *Econometrica*, 41(4): 587–601.

Gibbard, Allan. 1977. "Manipulation of Schemes That Mix Voting with Chance." *Econometrica*, 45(3): 665–81.

Grünbaum, Branko. 2003. *Convex Polytopes. Graduate Texts in Mathematics.* 2 ed., Springer.

Harless, Patrick. 2015. "A School Choice Compromise: Between Immediate and Deferred Acceptance." 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.

Hatfield, John. 2009. "Strategy-proof, Efficient, and Nonbossy Quota Allocations." *Social Choice and Welfare*, 33(3): 505–515.

Heo, Eun. 2014. "Probabilistic Assignment Problem with Multi-unit Demands: A Generalization of the Serial Rule and its Characterization." *Journal of Mathematical Economics*, 54: 40–47.

Huesmann, Katharina, and Martin Wambach. 2015. "Constraints on Matching Markets Based on Moral Concerns." Working Paper.

Hylland, Aanund, and Richard Zeckhauser. 1979. "The Efficient Allocation of Individuals to Positions." *The Journal of Political Economy*, 87(2): 293–314.

IOC. 2015. "Official Webseite of the Olympic Movement - IOC Members." http://www.olympic.org/ioc-members-list, accessed September 26, 2015.

Katta, Akshay-Kumar, and Jay Sethuraman. 2006. "A Solution to the Random Assignment Problem on the Full Preference Domain." *Journal of Economic Theory*, 131(1): 231–250.

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 Pöder, and Andre Veski. 2014. "Matching Practices for Elementary Schools - Estonia." accessed May 16, 2015.

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

Lenstra, Arjen, Hendrik Lenstra, and László Lovász. 1982. "Factoring Polynomials with Rational Coefficients." *Mathematische Annalen*, 261(4): 515–534.

Lubin, Benjamin, and David Parkes. 2012. "Approximate Strategyproofness." *Current Science*, 103(9): 1021–1032.

Mennle, Timo, and Sven Seuken. 2014. "An Axiomatic Characterization of Strategyproof Ordinal Mechanisms with Indifferences." Research Note.

Mennle, Timo, and Sven Seuken. 2015*a*. "Hybrid Mechanisms: Trading Off Strategyproofness and Efficiency of Random Assignment Mechanisms." Working Paper.

Mennle, Timo, and Sven Seuken. 2015*b*. "Partial Strategyproofness: An Axiomatic Approach to Relaxing Strategyproofness for Assignment Mechanisms." Working Paper.

Mennle, Timo, and Sven Seuken. 2015*c*. "The Pareto Frontier for Random Mechanisms." Working Paper.

Mennle, Timo, and Sven Seuken. 2015*d*. "Trade-offs in School Choice: Comparing Deferred Acceptance, the Naïve and the Adaptive Boston Mechanism." Working Paper.

Mennle, Timo, Daniel Abaecherli, and Sven Seuken. 2015. "Computing Pareto Frontier for Randomized Mechanisms." Mimeo.

Mennle, Timo, Michael Weiss, Basil Philipp, and Sven Seuken. 2015. "The Power of Local Manipulation Strategies in Assignment Mechanisms." In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI).*

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

Morrill, Thayer. 2013. "An Alternative Characterization of the Deferred Acceptance Algorithm." *International Journal of Game Theory*, 42(1): 19–28.

Mossel, Elchanan, and Miklós Rácz. 2014. "A Quantitative Gibbard-Satterthwaite Theorem Without Neutrality." *Combinatorica*, 1–71.

Moulin, Hervé. 1980. "On Strategy-Proofness and Single Peakedness." *Public Choice*, 35(4): 437–455.

Niederle, Muriel, Alvin Roth, and Tayfun Sönmez. 2008. "Matching and Market Design." In *The New Palgrave Dictionary of Economics.* Palgrace Macmillan.

Pacuit, Eric. 2012. "Voting Methods." In *The Stanford Encyclopedia of Philosophy.* Winter 2012 ed., ed. Edward Zalta. Stanford University.

Pais, Joana, and Agnes Pinter. 2008. "School Choice and Information: an Experimental Study on Matching Mechanisms." *Games and Economic Behavior*, 64: 303–328.

Pápai, Szilvia. 2000. "Strategyproof Assignment by Hierarchical Exchange." *Econometrica*, 68(6): 1403–1433.

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.

Procaccia, Ariel. 2010. "Can Approximation Circumvent Gibbard-Satterthwaite?" In *Proceedings of the 24th Conference on Artificial Intelligence (AAAI).*

Procaccia, Ariel, and Jeffrey Rosenschein. 2006. "The Distortion of Cardinal Preferences in Voting." In *Cooperative Information Agents X.* Vol. 4149 of *Lecture Notes in Computer Science*, ed. Matthias Klusch, Michael Rovatsos and Terry Payne, 317–331. Springer.

Pycia, Marek, and Utku Ünver. 2014. "Incentive Compatible Allocation and Exchange of Discrete Resources." Working Paper.

Roth, Alvin. 1982. "The Economics of Matching: Stability and Incentives." *Mathematical Operations Research*, 7: 617–628.

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.

Roth, Alvin. 2012. "The Theory and Practice of Market Design." *Prize Lecture for The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2012.*

Roth, Alvin, and Marilda Sotomayor. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis.* Cambridge University Press.

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).*

Satterthwaite, Mark. 1975. "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions." *Journal of Economic Theory*, 10(2): 187–217.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations. History of Economic Thought Books*, McMaster University Archive for the History of Economic Thought.

von Neumann, John. 1953. "Contributions to the Theory of Games." , ed. Harold Kuhn Albert Tucker Vol. 2. Princeton University Press, Princeton, New Jersey.

Xia, Lirong. 2011. "Computational Voting Theory: Game-Theoretic and Combinatorial Aspects." PhD diss. Computer Science Department, Duke University, Durham, NC.

Xia, Lirong, and Vincent Conitzer. 2008. "Generalized Scoring Rules and the Frequency of Coalitional Manipulability." In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC).*

Zhou, Lin. 1990. "On a Conjecture by Gale about One-sided Matching Problems." *Journal of Economic Theory*, 52(1): 123–135.

Curriculum Vitae

Personal Information

Name	Timo Mennle né Vollmer
Date of Birth	September 28, 1982
Place of Birth	Neustadt am Rübenberge, Germany
Nationality	German

Education

2012 – present	*Ph.D. Candidate*
	Computation and Economics Research Group
	University of Zurich
2002 – 2009	*Diplom Mathematics* (equiv. Master)
	Financial Mathematics, minor: Computer Science
	University of Freiburg
2001	*Abitur*, IGS Garbsen

Professional Experience

2010 – 2011	*Management Consultant*, McKinsey & Company, Inc.
2003 – 2008	*Teaching Assistant*, Mathematical Institute,
	University of Freiburg
2008	*Intern*, Advisory, PricewaterhouseCoopers AG
2001 – 2002	*Civil Service*, gGIS mbH Hannover

International Experience

2007	*Exchange student*, 2 semesters
	Australian National University, Canberra, Australia
1998-1999	*Exchange student*, 1 school year
	Newman Smith High School, Dallas, Texas, USA