Optimization in Strategic Environments

Itai Feigenbaum

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

ABSTRACT

Optimization in Strategic Environments

Itai Feigenbaum

This work considers the problem faced by a decision maker (planner) trying to optimize over incomplete data. The missing data is privately held by agents whose objectives are different from the planner's, and who can falsely report it in order to advance their objectives. The goal is to design optimization mechanisms (algorithms) that achieve "good" results when agents' reports follow a game-theoretic equilibrium.

In the first part of this work, the goal is to design mechanisms that provide a small worst-case approximation ratio (guarantee a large fraction of the optimal value in all instances) at equilibrium. The emphasis is on strategyproof mechanisms—where truthfulness is a dominant strategy equilibrium—and on the approximation ratio at that equilibrium. Two problems are considered—variants of knapsack and facility location problems. In the knapsack problem, items are privately owned by agents, who can hide items or report fake ones; each agent's utility equals the total value of their own items included in the knapsack, while the planner wishes to choose the items that maximize the sum of utilities. In the facility location problem, agents have private linear single sinked/peaked preferences regarding the location of a facility on an interval, while the planner wishes to locate the facility in a way that maximizes one of several objectives. A variety of mechanisms and lower bounds are provided for these problems.

The second part of this work explores the problem of reassigning students to schools. Students have privately known preferences over the schools. After an initial assignment is made, the students' preferences change, get reported again, and a reassignment must be obtained. The goal is to design a reassignment mechanism that incentivizes truthfulness, provides high student welfare, transfers relatively few students from their initial assignment, and respects student priorities at schools. The class of mechanisms considered is permuted lottery deferred acceptance (PLDA) mechanisms, which is a natural class of mechanisms based on permuting the lottery numbers students initially draw to decide the initial assignment. Both theoretical and experimental evidence is provided to support the use of a PLDA mechanism called reversed lottery deferred acceptance (RLDA). The evidence

suggests that under some conditions, all PLDA mechanisms generate roughly equal welfare, and that RLDA minimizes transfers among PLDA mechanisms.

Table of Contents

Acknowledgments

I would like to thank my adviser, Jay Sethuraman, for his guidance, support, and friendship. Jay was the one who introduced me to the field of Algorithmic Game Theory, and taught me the tools of the trade. On our many joint research projects, Jay regularly provided excellent insights that were crucial to solving the problems we faced—but even more importantly, as only the best of mentors do, helped me learn how to come up with such insights on my own. In addition, I have greatly benefited from his ability to find—and teach me how to find—good research questions to tackle. Jay has helped me immeasurably in my development as a researcher, and I am very grateful for all he has done for me.

I would also like to thank the other members my dissertation committee—Cliff Stein, Dan Bienstock, Yash Kanoria and Matt P. Johnson—for offering their expertise, and for their valuable suggestions and corrections to the earlier versions of this dissertation. In addition, I would like to thank my committee members (all of whom have helped shape my work as teachers or collaborators or both) and the IEOR faculty for their guidance inside and outside of the classroom. My research has also benefited much from collaborating with the brilliant Chun Ye and Irene Lo, and my time at Columbia has been made even more pleasant by their company and the company of my other fellow Ph.D. students.

Finally, I would like to thank my family for their unconditional love and support: my parents- Pua and Arieh, my brothers- Ohad and Yarin, my uncles- Marty and Herb, my aunt Marge and my sister-in-law Leehi. A special thanks goes to Olivia, for her love and support throughout these past few years. My success in completing this dissertation is in very large part due to my family and Olivia, and I would not be where I am without them.

Chapter 1

Introduction

1.1 A Bit on Mechanism Design

Decisions often need to be made under incomplete information. In many cases, the uncertainty regarding the missing data is due simply to random phenomena, such as the outcome of a future event or measurement errors. In contrast, it could be that some or all of the missing data is known—not to the decision maker (who we shall call the *planner* from now on), but rather to strategic agents, who can manipulate it in order to affect the planner's decision in a way that serves their own agendas. As an example, consider the National Resident Matching Program [Roth and Sotomayor, 1992], where a planner attempts to match medical residents with positions at hospitals. Each resident knows her true preferences over the hospitals (where they would like to work), and each hospital knows its true preferences over residents (whom they would like to hire), but that information is unknown to the planner. The planner must rely on the participants' reports of their preferences in order to compute a matching. However, depending on the algorithm used by the planner, it might not be in the participants' best interest to report their preferences truthfully: for example, a resident might be averse to ranking highly a hospital that ranks her low, for fear of "wasting" a top choice on a lost cause. How do agents behave in such an environment? And, since their reports are not necessarily truthful, can the planner still rely on those reports to make decisions?

The field of mechanism design (e.g. [Borgers *et al.*, 2015; Vulkan *et al.*, 2013]) uses the tools of game theory to model the way agents report in such environments. The planner's choice of mechanism (algorithm) induces a game among the agents, where each agent chooses their report (input) to the mechanism in an attempt to get a favorable output. The agents' inputs are assumed to follow some notion of game-theoretic equilibrium, meaning (informally speaking) that no agent can benefit from changing their report. Thus, the planner's goal becomes designing a mechanism that performs well when agents' reports follow an equilibrium.

Much of the mechanism design literature deals with transferable utilities, where monetary compensation can be used to incentivize agents' behavior. The crown jewel is the Vickrey-Clarke-Groves (VCG) mechanism [Groves, 1973], which tackles the problem of social welfare maximization. In this problem, given a set of outcomes, and agents' private valuations of these outcomes, a planner wishes to choose an outcome that maximizes the total valuation. The VCG mechanism cleverly uses money transfer to allow a social welfare-maximizing outcome to be chosen at (a dominant strategy) equilibrium.[1] There is also much literature where money plays a more significant role than just incentivizing: see, for example, Myerson's profit-maximizing mechanism [Myerson, 1981]. These ideas and others have guided the design of a variety of real world mechanisms, such as spectrum auctions [Cramton, 2013] and sponsored search auctions [Edelman *et al.*, 2007] .

In many strategic environments (such as those explored in this work) agents' utilities are non-transferable, due to financial compensation being considered unethical or infeasible. An example is the school choice problem—the assignment of students to public schools—in which students can report false preferences over the schools in an attempt to improve their assignment [Abdulkadiroglu and Sönmez, 2003]. Allowing students to pay for improved assignments is often unlawful, due to the disadvantage suffered by students from less affluent backgrounds. Another example is the organization of kidney exchange programs, where incompatible patient-donor kidney transplant pairs may exchange kidneys, and hospitals/transplant centers may avoid reporting patients to the program in a bid to increase the number of their own patients getting a kidney or increase the number of operations done locally [Ashlagi and Roth, 2014; Ashlagi *et al.*, 2013; Hajaj *et al.*, 2015]. Most nations forbid monetary compensation for organ donation, so again utilities are non-transferable. Non-transferable utilities is a strong restriction in mechanism design; for example, without money, there is usually no way to choose an optimal outcome at equilibrium in social welfare maximization problems. Nevertheless, the field of mechanism design has accomplished much under this restriction. For example, both scenarios described above are the subject of rich and active research, that has lead to the reform of school choice programs throughout the world,

[1] Note—the actual optimal solution, not just the best achievable at equilibrium.

as well as the creation and improvement of kidney exchange programs.

1.2 An Overview of This Dissertation

This work focuses entirely on mechanism design without money, that is with non-transferable utilities. It consists of two parts. The first (and major) part is theoretical in nature, while the second part has a more applied flavor.

1.2.1 Approximate Mechanism Design without Money

In the first part of this work (Chapters 2-5), we contribute to a stream of literature on "Approximate Mechanism Design without Money", which originated in [Procaccia and Tennenholtz, 2009]. Consider an optimization problem in a strategic environment. We say that a mechanism is *strategyproof* if truthfulness is a dominant strategy equilibrium under it (meaning that it is always optimal for each agent to report their private information truthfully, regardless of the other agents' private information and reports). Also, we say that a mechanism has a worst-case approximation ratio α if α is the smallest number for which, on every instance, the mechanism provides a value of at least $\frac{1}{\alpha}$ times the optimal value (when agents are truthful); we say that a mechanism is β-approximate if $\beta \geq \alpha$. In their paper, Procaccia and Tennenholtz suggest to focus on strategyproof mechanisms, and use the worst-case approximation ratio as a benchmark for the quality of mechanisms. In other words, the goal of their approach is to design strategyproof mechanisms with as small as possible worst-case approximation ratios. In our work, we use their approach to tackle strategic variants of the knapsack problem and one-dimensional facility location problems.

For one of the mechanisms in Chapter 2, we slightly deviate from (or more precisely, generalize) Procaccia's and Tennenholtz's approach, and consider a mechanism that is not generally strategyproof, but that still guarantees good approximation as long as the agents play according to a Bayes-Nash or coarse correlated equilibrium. This approach is related to the notion of the "Price of Anarchy," and the techniques used in analyzing that mechanism—to smoothness based proofs [Roughgarden, 2015].

1.2.1.1 Selfish Knapsack

In Chapter 2 we consider a variant of the knapsack problem: there are n agents, each owning a (mutually disjoint) set of items, where every item has a given value and size. Each agent's item set is known only to that agent. Agents may hide items from the planner (understate), or report fake ones that they do not actually own (overstate). The planner must choose, from the reported items, which items to include in a knapsack with a fixed capacity. Each agent gets a utility equal to the total value of their (non-fake) items included in the knapsack, while the planner's objective is to maximize social welfare (the sum of utilities). Our results are:

1. We provide a randomized mechanism, called HALF-GREEDY, satisfying the following properties:

   (a) When agents can only overstate, HALF-GREEDY is strategyproof and 2-approximate. We show that HALF-GREEDY is best possible by providing a matching lower bound (no randomized strategyproof mechanism can beat 2-approximation). In addition, randomization is shown to be necessary: no deterministic strategyproof mechanism can provide any constant approximation ratio.

   (b) When agents can only understate, HALF-GREEDY is not strategyproof, but as long as the agents' reports follow a Bayes-Nash equilibrium or a coarse correlated equilibrium, the mechanism still provides at least $\frac{1}{2}$ of the optimal value on every instance. The same is true when agents are not limited to understating, under a very mild additional assumption.

2. For the special case of a duopoly, namely 2 understating-only agents, we provide a randomized strategyproof $\frac{5+4\sqrt{2}}{7} \approx 1.522$-approximate mechanism. We also provide a lower bound of $\frac{5\sqrt{5}-9}{2} \approx 1.09$ on the worst-case approximation ratio attainable by randomized strategyproof mechanisms.

3. For the special case of one-bad-apple, namely when all agents but one are honest, and the manipulating agent is understating-only, we provide a lower bound of $\frac{1+\sqrt{5}}{2} \approx 1.618$ (the golden ratio) on the worst-case approximation ratio attainable by deterministic strategyproof mechanisms, as well as a deterministic strategyproof mechanism that attains it.

4. We also consider a different model, where each agent owns a single item, whose value-to-size ratio is publicly known, but whose actual value and size are private information. In this model, we show that a small modification of HALF-GREEDY leads to a mechanism that is strategyproof and 2-approximate. We also show that it is best possible (no randomized strategyproof mechanism can beat this approximation ratio) and that randomization is necessary (no deterministic strategyproof mechanism can provide any constant approximation ratio).

All results appear in [Feigenbaum and Johnson, 2015].

1.2.1.2 One-Dimensional Facility Location

In Chapters 3-5, we consider variants of one-dimensional facility location problems. In these problems, the planner wishes to locate a facility on an interval I. There are n agents located throughout that interval, where each agent's location is their own private information. We consider two types of agents: agents who want the facility to be located as far from them as possible (type 1 agents), and get a utility equal to their distance from their facility; and agents with the opposite preference (type 2 agents), who get a disutility equal to their distance from the facility. The planner, of course, attempts to optimize a certain objective function. We consider the following cases of this problem:

1. The desirable facility model, where all agents are type 2, and the planner wishes to minimize the L_p measure of the distances.

2. The obnoxious facility model, where all agents are type 1, and the planner's objective is to maximize the sum of the distances (maxisum) or to maximize the minimum distance (egalitarian).

3. The hybrid model, where we may have both types of agents simultaneously. In this model, in order to make the preferences of both types comparable, we say that type 2 agents get a utility equal to the length of I minus their distance from the facility.[2] The planner wishes to maximize the sum of utilities.

Our results are:

[2]See discussion in Chapter 5.

1. In the desirable facility model:
   (a) We show that the median mechanism, which is deterministic and strategyproof, is a $2^{1-\frac{1}{p}}$ approximate mechanism, along with a matching lower bound on deterministic strategyproof mechanisms. We also show that no randomized strategyproof mechanism, from a class covering all mechanisms currently known in literature, can beat the median's approximation ratio for integer $2 < p < \infty$.
   (b) We show a lower bound of $\frac{1}{2}(2^{1-\frac{1}{p}}+1)$ on the worst-case approximation ratio attainable by randomized strategyproof mechanisms for all p (subject to a couple of very mild assumptions), even when only 2 agents are present.
2. In the obnoxious facility model:
   (a) We characterize all deterministic strategyproof mechanisms. For the maxisum objective, we use our characterization to show a lower bound of 3 on the worst-case approximation ratio attainable by deterministic strategyproof mechanisms (which matches a known mechanism). For the egalitarian objective, we use our characterization to prove that no deterministic strategyproof mechanism can provide any bounded approximation ratio.
   (b) For the maxisum objective, we design a randomized strategyproof $\frac{3}{2}$-approximate mechanism, as well as prove a lower bound of $\frac{2}{\sqrt{3}}$ on the worst-case approximation ratio of randomized strategyproof mechanisms. For the egalitarian objective, we provide a lower bound of $\frac{3}{2}$.
3. In the hybrid model, we provide a deterministic, strategyproof, 3-approximate mechanism (note that the matching lower bound is carried from above), and a randomized, strategyproof, $\frac{23}{13}$-approximate mechanism.
4. We extend some of our results for the obnoxious model to a generalized model where each agent may control multiple locations.

The current state of knowledge regarding the problems discussed is summarized in table 1.1. All results appear in [Feigenbaum and Sethuraman, 2014; Feigenbaum *et al.*, 2013a].

[3] Under some conditions.

[4] Known result from [Cheng *et al.*, 2013a].

	Det. UB	Det. LB	Rand. UB	Rand. LB
Desirable (L_p)	$2^{1-\frac{1}{p}}$	$2^{1-\frac{1}{p}}$	?	$2^{1-\frac{1}{p}}, \frac{1}{2}(2^{1-\frac{1}{p}}+1)...$[3]
Obnoxious (maxisum)	3...[4]	3	$\frac{3}{2}$	$\frac{2}{\sqrt{3}}$
Obnoxious (egalitarian)	∞	∞	?	$\frac{3}{2}$
Hybrid (maxisum)	3	3	$\frac{23}{13}$	$\frac{2}{\sqrt{3}}$

Table 1.1: Facility location results. The first and second columns specify the upper and lower bounds on the worst-case approximation ratio attainable by deterministic strategyproof mechanisms; the third and fourth column do the same for randomized mechanisms.

1.2.2 Reassignment in School Choice

In the second part of this work (Chapter 6) we consider the problem of reassigning students to schools. The problem of assigning students to schools, also known as the school choice problem, has received much attention in the mechanism design literature, starting with [Abdulkadiroglu and Sönmez, 2003]. In school choice problems, the planner is given a set of students and a set of public schools with given capacities. The students have privately known strict preferences over the schools, while the schools have publicly known weak preferences over students, called priorities. The students report their preferences to the planner, who then outputs a match, which must respect capacities and priorities. In New York City, this is done via the Deferred Acceptance (DA) mechanism [Roth, 2008].

However, in practice, a significant number of students do not show up at their assigned school, for a variety of reasons: a superior private school admission, moving to a different city, or simply a change of mind are a few. Thus, additional vacancies are created, and some reassignment is possible. Students may resubmit their preferences, and the planner can compute a reassignment. A reassignment mechanism has seemingly opposite goals. On the one hand, utilizing the vacancies can improve overall welfare; on the other hand, the reassignment should not transfer too many students from their original assignments, since it is difficult for schools to cope with a dramatic change in their student population (note that the reassignment can happen after the school year has begun!).

We consider a class of reassignment mechanisms, which we call *permuted lottery deferred ac-*

ceptance (PLDA). We assume the initial assignment (the one before students change their preferences) is obtained via DA, where a single lottery is used to break ties in school priorities. Once students' submit their revised preferences, PLDA simply re-runs DA, but with (generally) different tie-breaking, which is correlated with the initial lottery and initial assignments. PLDA respects capacities and priorities, and does not transfer students from their initial assignment against their will. It also satisfies a relaxed variant of strategyproofness when the number of students is "very large". We provide both theoretical and experimental evidence to support using a particular PLDA mechanism, called *reversed lottery deferred acceptance* (RLDA). Our results are as follows:

1. We consider a theoretical model with a continuum of students and no school priorities. Within that model, we show that if a certain technical condition is satisfied, which essentially requires the relative demand for schools to remain unchanged after students resubmit their preferences, all PLDA mechanisms provide identical welfare, and RLDA minimizes transfers among all PLDA mechanisms.

2. We use data from New York City's school choice program to show, via simulation, that even with a finite number of students and with priorities, our analysis essentially holds: PLDA mechanisms perform very similarly on welfare measures, and RLDA minimizes transfers among them.

Part I

Worst-Case Approximation and Mechanism Design

Chapter 2

Selfish Knapsack

2.1 Introduction

We study a strategic variant of the knapsack problem, in which there are n agents, each agent owns a set of items, and every item has a given value and size. A social planner must design a mechanism to choose which items to include in a knapsack of a certain capacity, where the total size of the chosen items cannot exceed the capacity. Each agent gets a utility equal to the total value of her own items included in the knapsack, while the designer wishes to maximize social welfare (the sum of the utilities of the agents). However, the set of items each agent owns is private information, and an agent may choose not to disclose all of her items (may report any subset of them). We call this the *understating* model (UM). Revealing all items might not be in an agent's best interest:

Example 2.1. *Assume the knapsack's capacity is* 1*. Consider a mechanism which always chooses an optimal (social welfare maximizing) solution based on the reported items. Assume agent* 1*'s true set of items is* $\{a, b\}$ *and agent* 2*'s true set of items is* $\{c\}$*, where* a*,* b *and* c *have values* 1*,* $\frac{3}{4}$ *and* $\frac{3}{4}$ *and sizes* 1*,* $\frac{1}{2}$ *and* $\frac{1}{2}$ *respectively. If the agents report truthfully, the mechanism chooses* $\{b, c\}$ *as the solution; however, if agent 1 hides item* b *and reports her set of items to be* $\{a\}$*, the chosen solution becomes* $\{a\}$*, increasing agent* 1*'s utility from* $\frac{3}{4}$ *to* 1 *while decreasing social welfare from* $\frac{3}{2}$ *to* 1*.*

To incentivize truthful reporting, we look for *strategyproof* mechanisms, where truth-telling is a dominant strategy equilibrium (no agent can benefit from misreporting). As Example 2.1 suggests, such mechanisms cannot always achieve optimality, so we seek mechanisms that approximate optimality well. Specifically, we try to design strategyproof mechanisms with small worst-case approximation ratios (a mechanism is α-approximate if it provides, on every instance, social welfare value of at least $\frac{1}{\alpha}$ times the optimal welfare; the worst-case approximation ratio is the smallest such α). We note that one of the mechanisms we design actually fails to be strategyproof, but still has a nice strategic property: every induced Bayes-Nash equilibrium (BNE) and coarse correlated equilibrium (CCE) is guaranteed to have a small approximation ratio.

We emphasize that agents can misreport the *existence* of items, but not their *properties*– their size and value; that is, the planner has the power to verify the size and value of the reported items. One example of such a scenario is the allocation of a scientific resource, like time on a particle accelerator or NSF funding. Scientists submit research proposals, each requesting a certain amount of resource and has a certain expected scientific value. This expected scientific value is evaluated/confirmed by an impartial expert. A scientist can avoid submitting some of their proposals, in an attempt to increase the total amount of resource she receives via proposals she does submit (by avoiding "complementing" other scientists' proposals well, similarly to Example 2.1). Thus, the problem of choosing which proposals to accept in order to maximize total expected scientific value falls within our model.

While our focus is on UM, we also consider the *overstating* model (OM) where an agent is allowed to report fake items, which she does not actually own (report supersets of her true set of items); the planner is assumed to be unable to differentiate between real and fake items. Despite the fact that the agent derives no value from the inclusion of fake items in the knapsack, she can use them to indirectly increase her utility:

Example 2.2. *Consider Example 2.1, only now agent* 1*'s true set of items is* $\{a\}$ *and agent* 2*'s true set of items is* $\{c\}$*. If the agents report truthfully, the mechanism chooses* $\{a\}$ *as the solution. However, if agent* 2 *reports* $\{c, d\}$*, where* d *is a fake item of value* $\frac{3}{4}$ *and size* $\frac{1}{2}$*, the chosen solution becomes* $\{c, d\}$*. Agent* 2 *does not derive any benefit from the inclusion of* d *in the knapsack, but she does benefit from the inclusion of* c*; thus this manipulation increases her utility from* 0 *to* $\frac{3}{4}$*, while decreasing social welfare from* 1 *to* $\frac{3}{4}$*.*

In the allocation of scientific resources, fake items are proposals that the scientist has no intention to seriously pursue, submitted with the intention of creating a seemingly high-valued package along with some of her real proposals, potentially inducing the planner to choose this package– and hence increasing the chance of the real proposals within the package to be chosen (similarly to Example 2.2). Finally, in addition to UM and OM, we also consider their joint generalization, the *full* model (FM), where agents can simultaneously hide items and report fake ones.

Our work is part of a growing literature on the subject of approximate mechanism design without money [Procaccia and Tennenholtz, 2013a] (as well as the price of anarchy [Roughgarden, 2015]). This approach has been applied to many types of problems, such as matching [Dughmi and Ghosh, 2010], facility location [Alon *et al.*, 2010a; Feldman and Wilf, 2013a; Feigenbaum *et al.*, 2013b], and kidney exchange [Ashlagi *et al.*, 2013]. The most relevant paper we could find is [Chen *et al.*, 2011], which (among other results) provides a randomized strategyproof mechanism for UM with a large constant approximation ratio; there is no overlap between our results and theirs. Also related is the “Funding Games” model of Bar-Noy et al. [Bar-Noy *et al.*, 2012], where agents wish to maximize the size of their chosen items. In addition, there are other papers that consider manipulation involving the existence of objects rather than their properties. In the context of exchange markets, such manipulation is considered in [Atlamaz and Klaus, 2007; Postlewaite, 1979]; in connection with approximation, similar manipulation is considered in [Ashlagi *et al.*, 2013; Dughmi and Ghosh, 2010; Chen *et al.*, 2011]. Overstating bears similarity to the notion of “slot destruction” in [Schummer and Vohra, 2013], where airlines withhold information regarding cancellation of flights (equivalent to reporting fake flights) in order to manipulate a mechanism assigning landing times. Finally, examples of price of anarchy analysis for non-SP mechanisms can be found in [Caragiannis *et al.*, 2015; Bhawalkar and Roughgarden, 2011].

Contributions. We provide a randomized mechanism, called HALF-GREEDY, with good strategic properties. In OM, it is strategyproof, 2-approximate, no randomized strategyproof mechanism can beat this approximation guarantee, and no deterministic strategyproof mechanism can provide any constant approximation. In UM it is not strategyproof, but every BNE and CCE induced by it is 2-approximate; this remains true in FM, under a mild assumption. In addition, we design mechanisms for two specialized environments in UM. One is the case of a duopoly, namely $n = 2$ agents; there, we design a randomized strategyproof $\frac{5+4\sqrt{2}}{7} \approx 1.522$-approximate mechanism, and provide a lower bound of $\frac{5\sqrt{5}-9}{2} \approx 1.09$ on the approximation ratio attainable by randomized strategyproof mechanisms. The other case is of one-bad-apple, where only one agent is manipulative among an otherwise honest population. For this environment, we provide a deterministic strategyproof $\frac{1+\sqrt{5}}{2} \approx 1.618$-approximate mechanism, along with a matching lower bound. Finally, we consider a different environment, called Known-Quality-Unknown-Quantity (KQUQ), where every agent owns exactly one item, whose value-to-size ratio is known, but whose actual value and size are not known; we show that a simple modification of HALF-GREEDY yields a strategyproof mechanism which is 2-approximate, while no randomized strategyproof mechanism can beat this approximation guarantee, and no deterministic strategyproof mechanism can provide any constant approximation.

The rest of this chapter is organized as follows. In Section 2.2 we formalize our model. In Section 2.3, we discuss HALF-GREEDY and its strategic properties. Sections 2.4 and 2.5 are dedicated to the specialized environments of a duopoly and one-bad-apple, respectively. In Section 2.6 we discuss the KQUQ environment.

2.2 Model

We assume without loss of generality that the knapsack's capacity is 1. Let $N = \{1, 2, \ldots, n\}$ be a set of agents, $n \geq 2$. Each agent i has a ground set of items G_i (informally, the set of items agent i can *potentially* own), and a finite true set of items $X_i \in \widehat{G_i}$, where, for a set A, $\widehat{A}$ is the collection of all finite subsets of A. Each item $a \in G_i$ has size $s(a) \in (0, 1]$ and value $v(a) \in (0, \infty)$; for $A \in \widehat{\cup_{i\in N} G_i}$, we define $s(A) = \sum_{a\in A} s(a)$ and $v(A) = \sum_{a\in A} v(a)$.[1] We assume that $G_i \cap G_j = \varnothing$ for $i \neq j$.

For each agent $i \in N$, let $R_i^*(X_i) \subseteq \widehat{G_i}$ be her report space when her true set of items is X_i. In the *understating* (UM), *overstating* (OM) and *full* (FM) models, $R_i^*(X_i)$ equals $\widehat{X_i}$, $\{A \in \widehat{G_i} : A \supseteq X_i\}$ and $\widehat{G_i}$ respectively. Each agent i reports some $R_i \in R_i^*(X_i)$. A deterministic mechanism is a function $f : \prod_{i\in N} \widehat{G_i} \to \widehat{\cup_{i\in N} G_i}$ which maps the agents' reports to a set of items to include in the knapsack; a randomized mechanism is a function from $\prod_{i\in N} \widehat{G_i}$ to all random variables over

[1] The fact that our formulation allows us to distinguish between items with identical size, value and owner is a mere convenience. All of our results translate to a model where such items are indistinguishable; in such a model, the planner allots each agent room in the knapsack for an item with a certain value and size, but the agent may choose which actual item (with that value and size) to include.

$\widehat{\cup_{i\in N} G_i}$.[2] We restrict our attention to *feasible* mechanisms; a deterministic (resp. randomized) mechanism f is feasible iff, for all $\mathbf{R} \in \prod_{i\in N} \widehat{G_i}$:

1. f only uses the reported items: $f(\mathbf{R}) \subseteq \cup_{i\in N} R_i$ (resp. surely, meaning with probability 1).

2. f doesn't violate the knapsack's capacity: $s(f(\mathbf{R})) \leq 1$ (resp. surely).

The utility that agent i derives from a chosen solution $S \in \widehat{\cup_{i\in N} G_i}$ when her true set is X_i is defined as $u(X_i, S) = v(X_i \cap S)$. A mechanism is *strategyproof* if truthfulness is a dominant strategy equilibrium. For a deterministic (resp. randomized) mechanism f, this means that for all $i \in N$, $\mathbf{X} \in \prod_{j\in N} \widehat{G_j}$, $R_i \in R_i^*(X_i)$, we have $u(X_i, f(\mathbf{X})) \geq u(X_i, f(\mathbf{X}_{-i}, R_i))$ (resp. $\mathbb{E}[u(X_i, f(\mathbf{X}))] \geq \mathbb{E}[u(X_i, f(\mathbf{X}_{-i}, R_i))]$).[3] We emphasize that for randomized mechanisms, the requirement is that truthful reporting maximizes an agent's *expected* utility (independently of all other agents' reports).

Informally speaking, the planner wants to choose a solution S which maximizes social welfare: $\sum_{i\in N} u(X_i, S)$. However, as we saw in Examples 2.1 and 2.2, strategyproof mechanisms cannot always choose the optimal solution. Thus, we settle for an approximation to optimality: we attempt to design strategyproof mechanisms with small *worst-case approximation ratios.* We define the worst-case approximation ratio of a deterministic (resp. randomized) mechanism f to be $\max_{\mathbf{X}\in\prod_{i\in N}\widehat{G_i}} \frac{\sum_{i=1}^n u(X_i, OPT(\cup_{i\in N} X_i))}{\sum_{i=1}^n u(X_i, f(\mathbf{X}))}$ (resp. $\max_{\mathbf{X}\in\prod_{i\in N}\widehat{G_i}} \frac{\sum_{i=1}^n u(X_i, OPT(\cup_{i\in N} X_i))}{\sum_{i=1}^n \mathbb{E}[u(X_i, f(\mathbf{X}))]}$), where $OPT(A)$ is an optimal solution to the knapsack problem when the set of available items is A.[4] A mechanism whose worst-case approximation ratio is at most α is called α-approximate.

Finally, one of our mechanisms is not always strategyproof, but has another strategic property: every *Bayes-Nash equilibrium* (BNE) and *coarse correlated equilibrium* (CCE) induced by it has a small approximation ratio. BNE is a solution concept suitable for when agents have distributional knowledge of each other's items; CCE is suitable for when that knowledge is exact. We briefly remind the reader of the relevant definitions; for a complete discussion, see [Roughgarden, 2009; Roughgarden, 2015].

---

[2] $\prod_{i\in N} \widehat{G_i} = \widehat{G_1} \times \cdots \times \widehat{G_n}$.

[3] We define the notation $(\mathbf{z}_{-i}, z_i') = (z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_n)$ for every $i \in N$ and n-dimensional vector $\mathbf{z}$.

[4] We consider the worst-case approximation ratio to be ∞ if the denominator is 0 and the numerator is not; in case they are both 0, we consider the ratio to be 1. Note that in the randomized case, the only source of randomization is f, thus there is no need to take expectation of the numerator.

1. Let $\dot{\mathbf{X}}$ be a random variable over $\prod_{i\in N}\widehat{G_i}$ with distribution $\mathcal{F}$. A strategy S_i is a function mapping $X_i \in \widehat{G_i}$ to a random variable over $R_i^*(X_i)$; we also define $\dot{S}_i = S_i(\dot{X}_i)$. A strategy profile $\mathbf{S}$ is a BNE w.r.t. mechanism f and distribution $\mathcal{F}$ iff, for every $i \in N$ and strategy S_i', $\mathbb{E}[u(\dot{X}_i, f(\dot{\mathbf{S}}))] \geq \mathbb{E}[u(\dot{X}_i, f(\dot{\mathbf{S}}_{-i}, \dot{S}_i'))]$. $\mathbf{S}$ is α-approximate iff $\frac{\sum_{i=1}^n \mathbb{E}[u(\dot{X}_i, OPT(\cup_{i\in N}\dot{X}_i))]}{\sum_{i=1}^n \mathbb{E}[u(\dot{X}_i, f(\dot{\mathbf{S}}))]} \leq \alpha$.

2. For a given $\mathbf{X} \in \prod_{i\in N}\widehat{G_i}$, a random variable $\dot{\mathbf{R}}$ over $\prod_{i\in N} R_i^*(X_i)$ is a CCE under mechanism f if for every $i \in N$, $R_i' \in R_i^*(X_i)$, we have $\mathbb{E}[u(X_i, f(\dot{\mathbf{R}})] \geq \mathbb{E}[u(X_i, f(\dot{\mathbf{R}}_{-i}, R_i')]$. $\dot{\mathbf{R}}$ is α-approximate iff $\frac{\sum_{i=1}^n u(X_i, OPT(\cup_{i\in N}X_i))}{\sum_{i=1}^n \mathbb{E}[u(X_i, f(\dot{\mathbf{R}}))]} \leq \alpha$.

2.3 The HALF-GREEDY Mechanism

In this section, we analyze the strategic properties of a randomized mechanism we call HALF-GREEDY. In OM, we show that HALF-GREEDY is strategyproof and 2-approximate. We also show that no randomized strategyproof mechanism can beat this approximation guarantee, and no deterministic strategyproof mechanism can provide a constant worst-case approximation ratio. In UM, we show that while HALF-GREEDY is not strategyproof, every BNE and CCE it induces is 2-approximate; in FM, we can preserve this result, under a mild additional assumption.

For technical convenience, we assume the existence of a given total order $\succeq$ on $\cup_{i\in N}G_i$. To define HALF-GREEDY, we need two auxiliary mechanisms. The first one is the GREEDY mechanism, which adds items to the knapsack by decreasing value-to-size ratio, breaking ties according to $\succeq$. It is convenient to define $\succeq'$ to be a total order on $\cup_{i\in N}G_i$, where for $a, b \in \cup_{i\in N}G_i$, if $\frac{v(a)}{s(a)} > \frac{v(b)}{s(b)}$, then $a \succ' b$, and if $\frac{v(a)}{s(a)} = \frac{v(b)}{s(b)}$, then $\succeq'$ agrees with $\succeq$.

Definition 2.1. *For every $A \in \widehat{\cup_{i\in N}G_i}$, and every $b \in A$, let $L_A(b) = \{a \in A : a \succ' b\}$; if $s(A) > 1$, let $o_A = \max_{\succeq'}\{b \in A : s(L_A(b) \cup \{b\}) > 1\}$. Let the reported items be $\mathbf{R} \in \prod_{i\in N}\widehat{G_i}$. Define the GREEDY mechanism GR as follows: if $s(\cup_{i\in N}R_i) \leq 1$ then $GR(\mathbf{R}) = \cup_{i\in N}R_i$, otherwise $GR(\mathbf{R}) = L_{\cup_{i\in N}R_i}(o_{\cup_{i\in N}R_i})$.*

The second auxiliary mechanism is MAXIMUM-VALUE, which returns a single item with the maximum value possible, breaking ties according to $\succeq$:

Definition 2.2. *Let the reported items be $\mathbf{R} \in \prod_{i\in N}\widehat{G_i}$. Define the MAXIMUM-VALUE mechanism MV as follows: if $\cup_{i\in N}R_i = \varnothing$ then $MV(\mathbf{R}) = \varnothing$, otherwise $MV(\mathbf{R}) = \max_{\succeq}\{a \in \cup_{i\in N}R_i : v(a) \geq v(b) \ \ \forall b \in \cup_{i\in N}R_i\}$.*

Now we can define HALF-GREEDY, which is well known to be 2-approximate (see [Burke and Kendall, 2005]):

Definition 2.3. *The HALF-GREEDY mechanism HG runs GR with probability $\frac{1}{2}$ and MV with probability $\frac{1}{2}$ (probabilities chosen independently of the input).*

Before we continue, we show a simple lemma that will be helpful in the analysis of HALF-GREEDY. We define a bit of helpful notation: for every set $A \in \widehat{\cup_{i \in N} G_i}$, we define $\mathbf{X}^A$ to be the unique set profile where $A = \cup_{i \in N} X_i^A$.

Lemma 2.1. *Let $A, B, C \in \widehat{\cup_{i \in N} G_i}$, and assume $B \subseteq A$, $C \cap A \subseteq B$. Then, $C \cap GR(\mathbf{X}^A) \subseteq GR(\mathbf{X}^B)$ and $C \cap MV(\mathbf{X}^A) \subseteq MV(\mathbf{X}^B)$ (that is, every item in C that is included in the knapsack when we run MV/GR on $\mathbf{X}^A$ remains in the knapsack when we run them on $\mathbf{X}^B$).*

Proof. Let $c \in C \cap MV(\mathbf{X}^A)$; note that by feasibility of MV, $c \in A$ and hence $c \in C \cap A$, implying $c \in B$. Now, by definition of MV, we have that $c = \max_{\succeq}\{a \in A : v(a) \geq v(b) \;\; \forall b \in A\}$. Since $B \subseteq A$ and $c \in B$, we have that $c = \max_{\succeq}\{a \in B : v(a) \geq v(b) \;\; \forall b \in B\}$ (as c is maximal in the larger set, it remains maximal in the smaller set), so $c \in MV(\mathbf{X}^B)$.

Now, let $d \in C \cap GR(\mathbf{X}^A)$; by feasibility of GR, $d \in A$, hence $d \in C \cap A$, hence $d \in B$. If $s(B) \leq 1$, then $GR(\mathbf{X}^B) = B$, and in particular $d \in GR(\mathbf{X}^B)$. So assume $s(B) > 1$, and note that this implies $s(A) > 1$. Clearly, $o_A \succeq' o_B$; since $d \in GR(\mathbf{X}^A) = L_A(o_A)$, $d \succ' o_A \succeq' o_B$. Since $d \in B$ and $d \succ' o_B$, then $d \in L_B(o_B)$, so $d \in GR(\mathbf{X}^B)$. □

2.3.1 Overstating Model

We begin with analyzing HALF-GREEDY in OM. Strategyproofness follows from a very simple fact: under both GR and MV, every real item that is included in the knapsack when agent i reports R_i, remains in the knapsack when agent i reports $R_i \cap X_i$, namely avoids reporting the fake items within R_i:

Lemma 2.2. *Let $i \in N$, $\mathbf{X} \in \prod_{j \in N} \widehat{G_j}$ and $R_i \in \widehat{G_i}$. Then $X_i \cap GR(\mathbf{X}_{-i}, R_i) \subseteq GR(\mathbf{X}_{-i}, R_i \cap X_i)$ and $X_i \cap MV(\mathbf{X}_{-i}, R_i) \subseteq MV(\mathbf{X}_{-i}, R_i \cap X_i)$.*

Proof. Apply Lemma 2.1 with $C = X_i$, $B = (R_i \cap X_i) \cup (\cup_{j \in N \setminus \{i\}} X_j)$ and $A = R_i \cup (\cup_{j \in N \setminus \{i\}} X_j)$. □

In other words, an agent never loses from restricting her report to the real items within that report. This immediately implies:

Corollary 2.1. *In OM, HALF-GREEDY is strategyproof and* 2*-approximate.*

Proof. Strategyproofness is immediate from Lemma 2.2, since for every agent i and for every possible report $R_i \supseteq X_i$ in the overstating model, $R_i \cap X_i = X_i$. The fact that the mechanism is 2-approximate is already known, as noted earlier. □

It is important to note that once GREEDY first fails to add an item to the knapsack (namely, it attempts to pick up an item that does not fit in the remaining space), it stops and returns the items currently in the knapsack; it does *not* try to add the next item that fits in the remaining space. This seemingly trivial choice is actually crucial for maintaining strategyproofness, as the following example shows. Thus, one must be careful about choices that are seemingly unimportant for approximation, as they may be important for strategic properties.

Example 2.3. *Consider the BAD-GREEDY mechanism BG, defined as Algorithm 2.1. Consider the case of* $n = 2$*, with* $X_1 = \{a\}$*,* $X_2 = \{b\}$*,* $v(a) = s(a) = 1$*,* $v(b) = \frac{1}{4}$*,* $s(a) = \frac{1}{2}$*; on this instance,* $BG(\mathbf{X}) = \{a\}$*, and the utility of agent* 2 *is* $v(X_2 \cap \{a\}) = 0$*. However, if agent* 2 *reports* $R_2 = \{b, c\}$*, where* $v(c) = 1$*,* $s(c) = \frac{1}{2}$ *(that is, agent* 2 *reports a fake item* c *in addition to her true item* b*), then* $BG(X_1, R_2) = \{b, c\}$*, and agent* 2*'s utility is* $v(X_2 \cap \{b, c\}) = \frac{1}{4}$*. Thus, BAD-GREEDY is not strategyproof in OM (assuming* $a \in G_1$*,* $b, c \in G_2$*).*

We also provide matching lower bounds, which essentially complete the picture for OM. They show that (1) HALF-GREEDY is best possible in OM: no randomized strategyproof mechanism can beat 2-approximation, and (2) randomization is necessary: no deterministic strategyproof mechanism can provide any constant approximation. For lower bounds, we make the assumption that the ground sets are unrestricted: G_i is called *unrestricted* if it contains infinitely many items of size s and value v for every $s \in (0, 1]$ and $v \in (0, \infty)$.

Theorem 2.1. *In OM, if the ground sets are unrestricted, there is no randomized strategyproof mechanism with a worst-case approximation ratio strictly smaller than* 2*. Also, there is no deterministic strategyproof mechanism with a constant worst-case approximation ratio.*

Algorithm 2.1 BAD-GREEDY

Require: $A \in \widehat{\cup_{i \in N} G_i}$

$S \leftarrow \varnothing$, $T \leftarrow A$
while $T \neq \varnothing$ **do**
  $next \leftarrow \max_{\succeq'} T$
  $T \leftarrow T \setminus \{next\}$
  if $s(S \cup \{next\}) \leq 1$ **then**
    $S \leftarrow S \cup \{next\}$
  end if
end while
return S

Proof. We begin with the bound for randomized mechanisms. Let f be a randomized strategyproof mechanism with worst-case approximation ratio $1 < r < 2$ (the case of $r = 1$ has been covered in Example 2.2, as $r = 1$ can only be achieved by the mechanism that always chooses the optimal solution). Consider $\mathbf{X}' \in \prod_{i \in N} \widehat{G_i}$. In this instance, $X_1' = \{a_1, \ldots, a_{M^2}\}$, where $v(a_j) = \frac{1}{M}$ and $s(a_j) = \frac{1}{M^2}$ for all $j = 1, \ldots, M^2$; $X_2' = \{b\}$ where $v(b) = s(b) = 1$; $X_i' = \varnothing$ for all $i > 2$, where M is some very large integer. The optimal solution for this instance is X_1', with optimal value M. If no item in X_1' is chosen with probability strictly more than p, the approximation ratio is at least $\frac{M}{pM+(1-p)}$ (note that choosing any items from X_1' excludes choosing b and vice versa), and thus we must have $\frac{M}{pM+(1-p)} \leq r$, namely $p \geq \frac{M-r}{Mr-r}$. Thus it must be the case that some item a_j is chosen with probability at least $q = \frac{M-r}{Mr-r}$.

Next, consider $\mathbf{X}$, which is identical to $\mathbf{X}'$ except $X_1 = \{a_j\}$. In this instance, the optimal solution is X_2, with optimal value 1. Due to strategyproofness, it must be the case that item a_j is chosen with probability at least q (since otherwise agent 1 has an incentive to report X_1' instead of X_1). Thus, the approximation ratio on this instance is at least $\frac{1}{\frac{q}{M}+(1-q)} = \frac{1}{\frac{M-r}{M^2r-Mr}+\frac{Mr-M}{Mr-r}}$. Sending $M \to \infty$, this becomes $\frac{r}{r-1} > r$ for $r \in (1, 2)$. Contradiction.

Next, we show the bound for deterministic mechanisms. Let g be a deterministic strategyproof mechanism with worst-case approximation ratio $r \in [1, \infty)$. Assume $M > r$, and consider profile X_1'. If the mechanism doesn't choose any item in X_1', the approximation ratio is at least $\frac{M}{1} > r$.

Thus, the mechanism chooses at least one item in X'_1, say a_j. Now, consider $\mathbf{X}$. On that profile, the mechanism must choose a_j due to strategyproofness, leading to an approximation ratio of $\frac{1}{\frac{1}{M}} = M > r$. Thus, we have arrived at a contradiction. □

2.3.2 Understating and Full Models

In UM, HALF-GREEDY is no longer strategyproof—it is sometimes beneficial for an agent to hide items.

Example 2.4. *Consider the case of* $n = 2$, *where* $X_1 = \{a, b\}$, $X_2 = \{c, d\}$, $v(a) = 2$, $v(c) = 2 - \epsilon$, $s(a) = s(c) = \frac{1}{4} + \epsilon$, $v(b) = 3 + \epsilon$, $v(d) = 3$, $s(b) = s(d) = \frac{1}{2}$, *where* $\epsilon > 0$ *is very small. It is easy to check that there are no dominant strategies:*

1. *For agent* 1*:* $\emptyset$ *is strictly dominated by* $\{a\}$, $\{a\}$ *is a worse response than* $\{b\}$ *which is worse than* $\{a, b\}$ *when agent* 2 *reports* $\emptyset$, *and* $\{a, b\}$ *is a worse response than* $\{b\}$ *when agent* 2 *reports* $\{c\}$.

2. *For agent* 2*:* $\emptyset$ *is strictly dominated by* $\{c\}$, $\{c\}$ *is a worse response than* $\{d\}$ *which is worse than* $\{c, d\}$ *when agent* 1 *reports* $\emptyset$, *and* $\{c, d\}$ *is a worse response than* $\{d\}$ *when agent* 1 *reports* $\{a\}$.

However, that is not necessarily bad news. Both GR and MV satisfy the following property: fix some agent $i \in N$. For all *other* agents $j \neq i$, every item of j which is included in the knapsack when agent i reports R_i, remains in the knapsack when agent i reports $R'_i \subseteq R_i$.

Lemma 2.3. *Let* $i \in N$, $R_i, R'_i \in \widehat{G_i}$, $R'_i \subseteq R_i$. *Let* $(R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n)$, $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \in \prod_{j \in N \setminus \{i\}} \widehat{G_j}$. *For all* $j \in N \setminus \{i\}$, $X_j \cap GR(\mathbf{R}) \subseteq GR(\mathbf{R}_{-i}, R'_i)$ *and* $X_j \cap MV(\mathbf{R}) \subseteq MV(\mathbf{R}_{-i}, R'_i)$.

Proof. Apply Lemma 2.1 with $C = X_j$, $B = R'_i \cup (\cup_{k \in N \setminus \{i\}} R_k)$, $A = R_i \cup (\cup_{k \in N \setminus \{i\}} R_k)$. □

Thus, when an agent hides items, all *other* agents weakly benefit. Therefore, if an agent benefits from hiding items, social welfare increases, since then all other agents benefit as well. Following this observation, it is intuitive to expect that the agents' manipulations would result in a higher social welfare than when they are truthful (although one needs to verify that nothing goes wrong when *all* agents simultaneously engage in hiding items). To model how agents hide items, we consider two options regarding the knowledge agents have of each other's items. The first option is when that knowledge is distributional (a.k.a. Bayesian game): agents know a joint distribution from which the real items are drawn. For this option, we assume that agents behave according to a BNE. The second option is when the knowledge is exact (a.k.a complete information game): agents can see their peers' items. For this option, we merely assume that agents behave according to a CCE.[5] Note that HALF-GREEDY is prior independent, namely we do not assume that the planner has any knowledge (apart from the reports) of the agents' items, distributional or exact. We use a smoothness-based argument ([Roughgarden, 2015]) to show that every BNE and CCE under HALF-GREEDY results in social welfare weakly greater than when the agents are truthful; since truthfulness results in 2-approximation, this implies:[6]

Theorem 2.2. *In UM, for every prior $\mathcal{F}$ over $\prod_{i\in N}\widehat{G_i}$, every BNE w.r.t. HALF-GREEDY and $\mathcal{F}$ is 2-approximate. Similarly, for every $\mathbf{X}\in\prod_{i\in N}\widehat{G_i}$, every CCE w.r.t. HALF-GREEDY and $\mathbf{X}$ is 2-approximate.*[7]

Proof. We show the proof for BNE; the proof for CCE is similar. Fix agent $i\in N$. Let $\dot{\mathbf{X}}$ be a random variable over $\prod_{i\in N}\widehat{G_i}$ with probability distribution $\mathcal{F}$. Let $\mathbf{S}$ be a BNE under mechanism HG and probability distribution $\mathcal{F}$. Note that since we are in the understating model, $\dot{S}_j\subseteq\dot{X}_j$ surely for every $j\in N$ (reminder: $\dot{S}_j = S_j(\dot{X}_j)$). Thus, we can apply Lemma 2.3 $n-1$ times to deduce

$$\mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{X}}))] \leq \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}_{-i}, \dot{X}_i))];$$

[5]A fairly weak assumption– CCE is a generalization of mixed Nash equilibrium.

[6]Note that Theorem 2.2 implies that the game induced by HALF-GREEDY has a price of anarchy of 2 w.r.t. these equilibria concepts.

[7]Readers familiar with Rougharden's work on smooth games ([Roughgarden, 2009; Roughgarden, 2015]) might notice that our proof essentially takes the following form: we show that the game induced by HG is $(1,0)$-smooth with respect to the truthful choice function. Then, (a trivial adaptation of) Roughgarden's extension theorems imply that the social welfare obtained at BNE/CCE is weakly larger than the social welfare obtained when agents are truthful. Since truthfulness results in 2-approximation to optimality, the BNE/CCE is also 2-approximate. However, our proof is just as easily described from first principles, so we avoid the smoothness terminology in favor of accessibility.

that is, when all agents other than i hide items, agent i's utility weakly increases. Now, by definition of BNE,

$$\mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}_{-i}, \dot{X}_i))] \leq \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}))]$$

(when all agents other than i play according to $\mathbf{S}$, i's optimal response is to play S_i). Thus, we have that

$$\mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{X}}))] \leq \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}))].$$

As i was chosen arbitrarily, this holds for all agents, and so we have that

$$\sum_{i=1}^{n} \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{X}}))] \leq \sum_{i=1}^{n} \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}))].$$

Now, as we noted, HG is 2-approximate, that is

$$\frac{\sum_{i=1}^{n} u(X_i, OPT(\cup_{j\in N} X_j))}{\sum_{i=1}^{n} \mathbb{E}[u(X_i, HG(\mathbf{X}))]} \leq 2$$

for any fixed $\mathbf{X}$, which implies

$$\frac{\sum_{i=1}^{n} \mathbb{E}[u(\dot{X}_i, OPT(\cup_{j\in N} \dot{X}_j))]}{\sum_{i=1}^{n} \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{X}}))]} \leq 2.$$

Thus, we also have that

$$\frac{\sum_{i=1}^{n} \mathbb{E}[u(\dot{X}_i, OPT(\cup_{j\in N} \dot{X}_j))]}{\sum_{i=1}^{n} \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}))]} \leq 2.$$

□

Next, we consider FM. In FM, Theorem 2.2 almost holds. The reason we say "almost" is indifference. As Lemma 2.2 shows, no agent can benefit from reporting fake items. However, an agent might report fake items in a way that does not change her utility, but decreases other agents' utilities. Let us give an example of such a Nash equilibrium, which is a special case of both BNE and CCE:

Example 2.5. *Consider the case of* $n = 2$ *agents,* $X_1 = \{a\}$, $X_2 = \{b\}$, *where* $v(a) = 1$, $s(a) = \frac{1}{M}$, $v(b) = M - 2$, $s(b) = \frac{M-1}{M}$, *where* M *is some large integer,* $M >> 2$. *Truthful reporting is a Nash equilibrium. Note that when agents report truthfully, HALF-GREEDY chooses a with probability*

$\frac{1}{2}$ *and* b *with probability* 1*. However, if agent 1 reports* $R_1 = \{a, c\}$ *where* $v(c) = M - 1$ *and* $s(c) = \frac{M-1}{M}$*, and agent* 2 *reports truthfully, we still get a Nash equilibrium, in which HALF-GREEDY still chooses* a *with probability* $\frac{1}{2}$*, but* b *is chosen with probability* 0 *(*c *is chosen with probability* 1*, but since it is a fake item, it does not add to the agents' utilities or to the objective function value). In the latter Nash equilibrium, the approximation ratio is* $2M - 2$*.*

We show that problematic equilibria such as the one above cannot occur if agents are not deliberately malicious.

Definition 2.4. *Let* $i \in N$*,* $X_i, R_i \in \widehat{G_i}$*.* R_i *is called a malicious report for agent* i *when her true set of items is* X_i *if there exists* $R_i' \in \widehat{G_i}$ *where for all* $j \in N$ *and all* $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$*,* $(R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n) \in \prod_{k \in N \setminus \{i\}} \widehat{G_k}$*, we have*
$\mathbb{E}[u(X_j, HG(\mathbf{R}))] \le \mathbb{E}[u(X_j, HG(\mathbf{R}_{-i}, R_i'))]$*, with the inequality being strict for at least one agent in at least one instance.*

In other words, a malicious report is a report that can never benefit any agent (including the agent reporting it), and can sometimes hurt an agent. Thus, if the agents are even very mildly altruistic, they would not report maliciously. Also, in a Bayesian game, we say that a strategy S_i is malicious if $S_i(X_i)$ is malicious w.r.t. X_i, for some $X_i \in \widehat{G_i}$, with positive probability.[8] Non-malicious reports satisfy an important property– fake items included in those reports have no impact on the real items included in the solution. We first prove a restricted case of it, where all agents but one are honest:

Lemma 2.4. *Let* $i \in N$*,* $X_i, R_i \in \widehat{G_i}$*. If* R_i *is not malicious for agent* i *with true set of items* X_i*, then for every choice of* $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \in \prod_{k \in N \setminus \{i\}} \widehat{G_k}$*, and for every* $j \in N$*, we have that* $X_j \cap GR(\mathbf{X}_{-i}, R_i) = X_j \cap GR(\mathbf{X}_{-i}, R_i \cap X_i)$ *and* $X_j \cap MV(\mathbf{X}_{-i}, R_i) = X_j \cap MV(\mathbf{X}_{-i}, R_i \cap X_i)$*.*

Proof. Assume that there exists $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \in \prod_{k \in N \setminus \{i\}} \widehat{G_k}$ such that, for some $j \in N$, $X_j \cap MV(\mathbf{X}_{-i}, R_i) \neq X_j \cap MV(\mathbf{X}_{-i}, R_i \cap X_i)$ (the proof for GR instead of MV is identical).

[8] Assuming non-maliciousness is weaker and easier to justify than assuming no fake items are reported. Assuming no fake items reported is similar in spirit to the assumption of "no overbidding" in generalized second price auctions [Caragiannis *et al.*, 2015]. Reporting fake items in our mechanism, much like overbidding in GSP, is a weakly dominated strategy; however, both can lead to unreasonable and bad equilibria due to indifference, and are thus ruled out by assumption.

First, note that for every instance where the true set of items for agent i is X_i, and for any fixed reports for the agents in $N \backslash \{i\}$, reporting $R_i \cap X_i$ instead of R_i weakly increases all agents' utilities (agent i by Lemma 2.2, and the rest by Lemma 2.3). Thus, it is enough to show that on our instance, when all $k \in N \backslash \{i\}$ own and report X_k, reporting $R_i \cap X_i$ instead of R_i strictly increases some agent's utility (since then R_i would be malicious). Lemmas 2.2 and 2.3 imply that, for all $k \in N$,

$$u(X_k, MV(\mathbf{X}_{-i}, R_i \cap X_i)) \geq u(X_k, MV(\mathbf{X}_{-i}, R_i)).$$

and

$$u(X_k, GR(\mathbf{X}_{-i}, R_i \cap X_i)) \geq u(X_k, GR(\mathbf{X}_{-i}, R_i)).$$

By Lemma 2.3 if $j \neq i$ and Lemma 2.2 if $j = i$, $X_j \cap MV(\mathbf{X}_{-i}, R_i) \neq X_j \cap MV(\mathbf{X}_{-i}, R_i \cap X_i)$ implies

$$X_j \cap MV(\mathbf{X}_{-i}, R_i) \subset X_j \cap MV(\mathbf{X}_{-i}, R_i \cap X_i),$$

thus

$$u(X_j, MV(\mathbf{X}_{-i}, R_i \cap X_i)) > u(X_j, MV(\mathbf{X}_{-i}, R_i)),$$

which, in combination with the already established inequality

$$u(X_j, GR(\mathbf{X}_{-i}, R_i \cap X_i)) \geq u(X_j, GR(\mathbf{X}_{-i}, R_i))$$

proves our result. □

We can now get rid of the honesty assumption:

Lemma 2.5. *Let $i \in N$, $X_i, R_i \in \widehat{G_i}$. If R_i is not malicious for agent i with true set of items X_i, then for every choice of $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, $(R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n) \in \prod_{k \in N \backslash \{i\}} \widehat{G_k}$, and for every $j \in N$, we have that $X_j \cap GR(\mathbf{R}) = X_j \cap GR(\mathbf{R}_{-i}, R_i \cap X_i)$ and $X_j \cap MV(\mathbf{R}) = X_j \cap MV(\mathbf{R}_{-i}, R_i \cap X_i)$.*

Proof. We prove for GR; the proof for MV is identical. Let us define the reports and true sets of the agents other than i as $(R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n), (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \in \prod_{k \in N \backslash \{i\}} \widehat{G_k}$. Assume R_i is non malicious for agent i with true set X_i. For $j \neq i$, by Lemma 2.4,

$$R_j \cap GR(\mathbf{R}) = R_j \cap GR(\mathbf{R}_{-i}, R_i \cap X_i) \text{ surely.}$$

Intersecting both sides with X_j yields

$$X_j \cap R_j \cap GR(\mathbf{R}) = X_j \cap R_j \cap GR(\mathbf{R}_{-i}, R_i \cap X_i) \text{ surely.}$$

By feasibility of GR, we have $X_j \cap GR(\mathbf{R}) \subseteq R_j$ and $X_j \cap GR(\mathbf{R}_{-i}, R_i \cap X_i) \subseteq R_j$; thus it follows that

$$X_j \cap GR(\mathbf{R}) = X_j \cap GR(\mathbf{R}_{-i}, R_i \cap X_i).$$

For $j = i$, $X_i \cap GR(\mathbf{R}) = X_i \cap GR(\mathbf{R}_{-i}, R_i \cap X_i)$ is immediate from Lemma 2.4. □

Therefore, non-maliciousness rules out equilibria like the one in Example 2.5. Fake items might be reported at equilibria, but they would have no impact on welfare. This, in addition to Theorem 2.2, leads to the following result:

Theorem 2.3. *In FM, for every prior $\mathcal{F}$ over $\prod_{i \in N} \widehat{G_i}$, every BNE w.r.t. HALF-GREEDY and $\mathcal{F}$ in which no malicious strategy is used is 2-approximate. Similarly, for every $\mathbf{X} \in \prod_{i \in N} \widehat{G_i}$, every CCE w.r.t. HALF-GREEDY and $\mathbf{X}$ in which no malicious report is used with positive probability is 2-approximate.*

Proof. We prove the theorem for BNE; the proof for CCE is similar. Let $\dot{\mathbf{X}}$ be a random variable over $\prod_{i \in N} \widehat{G_i}$ with probability distribution $\mathcal{F}$. Let $\mathbf{S}$ be a BNE under mechanism HG and probability distribution $\mathcal{F}$, and assume that no malicious strategies are being played in $\mathbf{S}$. We define S'_i to be the strategy that maps each set A to $S_i(A) \cap A$. Applying Lemma 2.5 n times in succession, once for each agent, we can immediately conclude that every agent gets the exact same utility under $\mathbf{S}$ and $\mathbf{S}'$. If $\mathbf{S}'$ is a BNE in the full model, then since no fake items are reported, it is clearly also a BNE in the understating model, and hence by Theorem 2.2 it is 2-approximate, and therefore $\mathbf{S}$ is 2-approximate as well. Thus, it is enough to show that $\mathbf{S}'$ is a BNE in the full model.

Fix agent i. Apply Lemma 2.5 $n-1$ times to get that for all $Y_i \in \widehat{G_i}$, $\mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}_{-\mathbf{i}}, Y_i))] = \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}'_{-i}, Y_i))]$; thus, it follows that since S_i is a best response by agent i when all other agents j play S_j, it remains a best response when they all play S'_j instead. Also, by Lemma 2.2, agent i weakly benefits from playing S'_i instead of S_i, so S'_i is also a best response to all other agents j playing S'_j. Thus, $\mathbf{S}'$ is a BNE in the full model. □

Finally, we make a note regarding the existence of BNE: when moving from UM to FM, HG doesn't lose any BNEs.

Theorem 2.4. *Let* $\mathbf{S}$ *be a BNE in under HG and distribution* $\mathcal{F}$ *in the understating model; then it remains such a BNE in the full model.*

Proof. Assume $\mathbf{S}$ is not a BNE in the full model. Then there must exist an agent i and strategy S_i' where $\mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}))] < \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}_{-i}, \dot{S}_i'))]$. But, by Lemma 2.2, this implies that $\mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}))] < \mathbb{E}[u(\dot{X}_i, HG(\dot{\mathbf{S}}_{-i}, \dot{S}_i' \cap \dot{X}_i))]$, and since $\dot{S}_i' \cap \dot{X}_i \subseteq \dot{X}_i$ surely, it follows that $\mathbf{S}$ is not a BNE in the understating model. □

2.4 The EQUAL-UTILITY Mechanism

In this section, we consider the special case of $n = 2$ in UM. We design a specialized randomized mechanism, called EQUAL-UTILITY for this environment, which is strategyproof and $\frac{5+4\sqrt{2}}{7} \approx 1.522$-approximate. We hope that the ideas behind EQUAL-UTILITY lead to generalizations for larger numbers of agents. In the next section, we show that no deterministic strategyproof mechanism can beat EQUAL-UTILITY's approximation ratio, thus randomization leads to strict improvement. We also provide a lower bound of $\frac{5\sqrt{5}-9}{2} \approx 1.09$ on the approximation ratio attainable by randomized strategyproof mechanisms, showing the necessity of some approximation gap.

The idea behind EQUAL-UTILITY, shown as Algorithm 2.2, is to solve the knapsack problem optimally, with one additional constraint: that the agents' utilities are *exactly* equal. Since in general, apart from $\varnothing$, there might not be a deterministic solution that satisfies this additional constraint, we allow for randomized solutions instead. Formally, we want to solve the following mathematical program (PROGRAM), where A is a random decision variable (set of items):[9] maximize $\mathbb{E}[v(A)]$ subject to: (1) $A \subseteq X_1 \cup X_2$ and $s(A) \leq 1$ surely and (2) $\mathbb{E}[v(A \cap X_1)] = \mathbb{E}[v(A \cap X_2)]$. As every agent gets exactly half of A's expected welfare, it is in each agent's best interest to maximize that welfare. Therefore agents have no incentive to restrict the feasible region of PROGRAM by hiding items, and thus PROGRAM alone is a strategyproof mechanism.

However, PROGRAM does not always lead to good approximation. It fails to do so on instances where one agent's items are much superior to the other's. For example, if one agent has one item

[9] PROGRAM can be stated as a linear programming problem with exponentially many variables. Let $T = \{S \subseteq X_1 \cup X_2 : s(S) \leq 1\}$. Then PROGRAM can be stated as: maximize $\sum_{S \in T} v(S) p_S$ subject to $\sum_{S \in T} v(S \cap X_1) p_S = \sum_{S \in T} v(S \cap X_2) p_S$, $\sum_{S \in T} p_S = 1$ and $p_S \geq 0$ for all $S \in T$ (where the p_S's are our decision variables).

of value M, the other agent has one item of value ϵ, and $M >> \epsilon$, the equal utility constraint dictates that the M-valued item is almost never chosen. Thus, we add a preliminary check meant to catch such instances before we turn to PROGRAM. The preliminary check is as follows: say we wish for our mechanism to be α-approximate. Consider $OPT(X_1)$ and $OPT(X_2)$, namely the optimal solutions using just a single agent's items. If $OPT(X_i)$ is significantly bigger than $OPT(X_j)$, to the extent where $OPT(X_i)$ is guaranteed to be an α-approximation on its own to the optimal value, then we simply return $OPT(X_i)$. This is checked via the condition $v(OPT(X_i)) \geq \frac{1}{\alpha}(v(OPT(X_1)) + v(OPT(X_2)))$.[10] If neither agent can satisfy the specified condition, we turn to PROGRAM.

Algorithm 2.2 EQUAL-UTILITY

Require: Sets of items X_1, X_2, where $X_i \in \widehat{G_i}$; parameter $\alpha \in [1, 2)$

1: $Z_1 \leftarrow OPT(X_1)$, $Z_2 \leftarrow OPT(X_2)$
2: **if** $v(Z_i) \geq \frac{1}{\alpha}(v(Z_1) + v(Z_2))$ for some $i \in \{1, 2\}$ **then**
3:     **return** Z_i ▷ option 1
4: **else**
5:     **return** optimal solution to PROGRAM with input $\mathbf{X}$ ▷ option 2
6: **end if**

Theorem 2.5. *In UM, for $\alpha \geq \frac{5+4\sqrt{2}}{7} \approx 1.522$, EQUAL-UTILITY is strategyproof and α-approximate.*

Proof. First, we prove strategyproofness. We break the proof into cases:

1. Assume the mechanism ends at option 1. In that case, agent i clearly gets the best utility she can possibly get, and hence has no incentive to misreport. Consider agent $j \neq i$. Assume agent j reports $X'_j \subset X_j$. Let us denote $Z'_j = OPT(X'_j)$. Note that $v(Z'_j) \leq v(Z_j)$. Since $v(Z_i) \geq \frac{1}{\alpha}(v(Z_i) + v(Z_j))$, trivially $v(Z_i) \geq \frac{1}{\alpha}(v(Z_i) + v(Z'_j))$. Thus, agent j cannot prevent the mechanism from ending at option 1 by misreporting. Furthermore, since at option 1 the mechanism returns Z_i, agent j has no influence on what is returned. Thus agent j cannot change the outcome of the mechanism by misreporting.

---

[10] We use $v(OPT(X_1)) + v(OPT(X_2))$ instead of $v(OPT(X_1 \cup X_2))$ to maintain strategyproofness.

2. Assume the mechanism ends at option 2. Consider agent 1 (the proof for agent 2 is identical). Assume agent 1 reports $X_1' \subset X_1$. Let us denote $Z_1' = OPT(X_1')$. Note that $v(Z_1') \leq v(Z_1)$. Since the mechanism did not end at option 1, we have that

$$\begin{aligned}
v(Z_1) &< \frac{1}{\alpha}(v(Z_1) + v(Z_2)) \\
(\alpha - 1)v(Z_1) &< v(Z_2) \\
(\alpha - 1)v(Z_1') &< v(Z_2) && \text{(since } v(Z_1')) \leq v(Z_1)) \\
v(Z_1') &< \frac{1}{\alpha}(v(Z_1') + v(Z_2)).
\end{aligned}$$

Thus, agent 1 cannot make the mechanism stop at option 1 and return Z_1', and therefore agent 1 cannot benefit from making the mechanism stop at option 1 (since if Z_2 is returned, her payoff is 0). At option 2, agent 1 would like to report all of her items: her utility is exactly half of the optimal solution to PROGRAM, and so enlarging the feasible region of PROGRAM weakly increases her own utility.

Next, we prove the mechanism is α-approximate. Since we know that $v(Z_1) + V(Z_2) \geq v(OPT(\cup_{i\in\{1,2\}} X_i))$, if the mechanism ends at option 1 clearly it provides an α-approximation. So we just need to prove this for the case the mechanism ends at option 2. Let $O = OPT(X_1 \cup X_2)$, $O_1 = O \cap X_1$, $O_2 = O \cap X_2$. In addition, let $a = \text{argmin}_{i\in\{1,2\}} v(O_i)$, $b = \text{argmax}_{i\in\{1,2\}} v(O_i)$ (if $v(O_1) = v(O_2)$, set $a = 1$ and $b = 2$), and let $p = \frac{v(O_b)-v(O_a)}{v(O_b)-v(O_a)+v(Z_a)}$. Consider the random variable A which returns Z_a with probability p and O with probability $(1 - p)$. p was chosen precisely so that A becomes a feasible solution to our program:

$$\begin{aligned}
\mathbb{E}[v(A \cap X_b)] &= (1 - p)v(O_b) \\
&= \frac{v(O_b)v(Z_a)}{v(O_b) - v(O_a) + v(Z_a)} \\
&= \frac{(v(O_b) - v(O_a))v(Z_a)}{v(O_b) - v(O_a) + v(Z_a)} + \frac{v(O_a)v(Z_a)}{v(O_b) - v(O_a) + v(Z_a)} \\
&= p \cdot v(Z_a) + (1 - p)v(O_a) \\
&= \mathbb{E}[v(A \cap X_a)].
\end{aligned}$$

It is therefore enough to show that $\frac{v(O)}{\mathbb{E}[v(A)]} \leq \alpha$. Note that

$$\begin{aligned}\frac{v(O)}{\mathbb{E}[v(A)]} &= \frac{v(O_a)+v(O_b)}{(1-p)(v(O_a)+v(O_b))+pv(Z_a)} \\ &= \frac{(v(O_a)+v(O_b))(v(O_b)-v(O_a)+v(Z_a))}{2v(Z_a)v(O_b)}.\end{aligned}$$

When $v(Z_a)$ and $v(O_b)$ are fixed values, and $v(O_a)$ is a variable, this is a parabola with a maximum at $v(O_a) = \frac{v(Z_a)}{2}$. Plugging that in, we have the following upper bound on the approximation ratio:

$$\frac{(\frac{v(Z_a)}{2}+v(O_b))^2}{2v(Z_a)v(O_b)} = \frac{1}{2} + \frac{v(O_b)}{2v(Z_a)} + \frac{v(Z_a)}{8v(O_b)}.$$

Let us denote $x = \frac{v(O_b)}{v(Z_a)}$; then our upper bound is $\frac{1}{2} + \frac{x}{2} + \frac{1}{8x}$. Since the mechanism did not end at option 1, we know that

$$\begin{aligned}x &= \frac{v(O_b)}{v(Z_a)} \\ &\leq \frac{v(Z_b)}{v(Z_a)} \\ &< \frac{1}{\alpha-1}.\end{aligned}$$

We also know that

$$\begin{aligned}x &= \frac{v(O_b)}{v(Z_a)} \\ &= \frac{2v(O_b)}{2v(Z_a)} \\ &\geq \frac{v(O_b)+v(O_a)}{2v(Z_a)} \\ &= \frac{v(O)}{2v(Z_a)} \\ &\geq \frac{1}{2}.\end{aligned}$$

So, to see how bad our upper bound can be, we maximize $\frac{1}{2} + \frac{x}{2} + \frac{1}{8x}$ over $x \in [\frac{1}{2}, \frac{1}{\alpha-1}]$. Simple analysis shows that for $\alpha > 1$, the maximum is $\frac{1}{2} + \frac{1}{2(\alpha-1)} + \frac{\alpha-1}{8}$. We are therefore guaranteed approximation ratio α from our mechanism as long as $\frac{1}{2} + \frac{1}{2(\alpha-1)} + \frac{\alpha-1}{8} \leq \alpha$, which is easily seen to hold as long as $\alpha \geq \frac{5+4\sqrt{2}}{7}$. □

Our bound on EQUAL-UTILITY's performance is tight:

Theorem 2.6. *Let $r = \frac{5+4\sqrt{2}}{7}$. Assume the ground sets are unrestricted. For every $\delta > 0$, there exists an instance where EQUAL-UTILITY (with $\alpha = r$) provides an approximation ratio strictly larger than $r - \delta$.*

Proof. Consider the profile where $X_1 = \{a, b\}$ with $v(a) = s(a) = 1$ and $v(b) = s(b) = \frac{1}{2}$, and $X_2 = \{c\}$ where $v(c) = \frac{1}{r-1} - \epsilon$ where $\epsilon > 0$ is small, and $s(c) = \frac{1}{2}$. EQUAL-UTILITY will reach option 2 on this instance. In this case, item a is chosen with probability $p = \frac{\frac{1}{r-1}-\epsilon-\frac{1}{2}}{\frac{1}{r-1}-\epsilon+\frac{1}{2}}$. Items b and c are chosen with probability $1-p$. The mechanism's approximation ratio is $\frac{\frac{1}{r-1}-\epsilon+\frac{1}{2}}{p+(1-p)(\frac{1}{r-1}-\epsilon+\frac{1}{2})}$, which, as $\epsilon \to 0$, goes to $\frac{5+4\sqrt{2}}{7}$. □

We next show that no randomized strategyproof mechanism can be arbitrarily close to optimality– some separation is required:

Theorem 2.7. *In UM, if the ground sets are unrestricted, no randomized strategyproof mechanism can provide a worst-case approximation ratio strictly better than $\frac{5\sqrt{5}-9}{2} \approx 1.09$.*

Proof. Let f be a randomized strategyproof mechanism which provides a worst-case approximation ratio $r < \frac{5\sqrt{5}-9}{2}$. Consider the profile where $X_1 = \{a\}$ with $s(a) = 1$ and $v(a) = \phi = \frac{1+\sqrt{5}}{2}$ (the golden ratio), $X_2 = \{b\}$ with $s(b) = \frac{1}{2}$ and $v(b) = 1$, and $X_i = \varnothing$ for all $i \geq 3$. Let $p = \mathcal{P}(a \in f(\mathbf{X}))$. To maintain approximation ratio r, we must have (I) $p\phi + (1-p) \geq \frac{1}{r}\phi$. Now, consider profile $\mathbf{X}' = (\mathbf{X}_{-1}, X_1')$ where $X_1' = \{a, c\}$ and $v(c) = 1$, $s(c) = \frac{1}{2}$. Let $p' = \mathcal{P}(a \in f(\mathbf{X}'))$. To maintain approximation ratio r, we must have (II) $p'\phi + (1-p')2 \geq \frac{2}{r}$. To maintain strategyproofness, we must have (III) $p'\phi + (1-p') \geq p\phi$. Now, (I) gives $p \geq \frac{\frac{\phi}{r}-1}{\phi-1}$, (II) gives $p' \leq \frac{2-\frac{2}{r}}{2-\phi}$. (III) can be rewritten as $p'(\phi-1) + 1 - p\phi \geq 0$, and so this implies $\frac{2-\frac{2}{r}}{2-\phi}(\phi-1) + 1 - \phi\frac{\frac{\phi}{r}-1}{\phi-1} \geq 0$. Isolating r, this gives $r \geq \frac{5\sqrt{5}-9}{2}$, contradiction. □

2.4.1 NP-Hardness

We note that EQUAL-UTILITY requires solving NP-hard problems. First, computing $OPT(X_1)$ and $OPT(X_2)$ means solving the knapsack problem, which is known to be NP-hard. Solving PROGRAM is NP-hard as well:

Theorem 2.8. *Solving PROGRAM is NP-hard.*

Proof. We prove this by reduction from knapsack. Say we have an instance of the knapsack problem with set of items I. Assume without loss of generality that the knapsack's capacity in this instance is $\frac{1}{2}$ and that the sizes of the items are at most $\frac{1}{2}$ each. Let an optimal solution be OPT^*; we want to know whether or not $v(OPT^*) \geq k$ for some $k > 0$. Set $X_1 = I$, and $X_2 = \{a\}$ where $v(a) = k$ and $s(k) = \frac{1}{2}$, and solve PROGRAM on this instance. Note that agent 2's expected utility can never surpass k, as she only has one item and its value is k. Thus, the optimal value of $PROGRAM$ is at most $2k$, since the utilities must be equal. We claim that the optimal value of $PROGRAM$ is exactly $2k$ iff $v(OPT^*) \geq k$.

1. If $v(OPT^*) \geq k$, then there is a unique solution A to PROGRAM where $A \in \{OPT^* \cup \{a\}, \{a\}\}$ surely (as $v(X_1 \cap (OPT^* \cup \{a\})) = v(OPT^*) \geq k = v(X_2 \cap (OPT^* \cup \{a\})))$. Since a is chosen with probability 1, agent 2 gets an expected utility of exactly k here, and hence so does agent 1. Thus, $\mathbb{E}[v(A)] = 2k$, and therefore the optimal value of PROGRAM is at least $2k$, and thus is exactly $2k$.

2. If $v(OPT^*) < k$, note that whenever $S \subseteq X_1 \cup X_2$ where $v(S) \leq 1$, if $a \in S$, then $v(S \cap X_1) \leq v(OPT^*) < k$ (since if $a \in S$, there is only capacity $\frac{1}{2}$ left for agent 1's items). Thus, it follows that for every solution A of PROGRAM, there is a nonzero probability that $a \notin A$ (otherwise agent 1's expected utility must be strictly less than k, and agent 2's expected utility is exactly k). Thus agent 2's expected utility is strictly less than k, and therefore the optimal value of PROGRAM is strictly less than $2k$.

□

We refer the reader to Appendix A for an informal discussion regarding managing this running-time issue.

2.5 The PACIFY-THE-LIAR Mechanism

We continue exploring UM. We now allow for a general number of agents n, however we restrict ourselves to an environment where there is only one bad apple—specifically, $n - 1$ agents are assumed to be honest. We assume without loss of generality that agent 1 is the manipulative agent (note that our results hold for free even if the honesty of an agent—whether or not that agent is manipulative—is private information of that agent, since we can simply say that if all agents report to be honest, we include nothing in the knapsack).[11] For this environment, we will provide a ϕ-approximate deterministic strategyproof mechanism ($\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$ is the golden ratio), along with a matching lower bound.

Our deterministic mechanism, called PACIFY-THE-LIAR (Algorithm 2.3), is similar in spirit to EQUAL-UTILITY. It begins with a preliminary test, which checks if agent 1 can guarantee an α-approximation on her own (option 1), or if agents 2 through n can guarantee an α-approximation together, without agent 1 (option 2). In the former case, we return $OPT(X_1)$, and in the latter case we return $OPT(\cup_{i\in N\setminus\{1\}} X_i)$. We note a small difference between the preliminary tests in EQUAL-UTILITY and in our new mechanism: in option 2, the benchmark used is the optimal solution $v(OPT(\cup_{i\in N} X_i))$ rather than the upper bound $v(Z_1) + v(Z_2)$. This substitution is crucial in order to maintain a ϕ approximation ratio, and does not violate strategyproofness due to the honesty of all agents other than 1. If the preliminary test fails, we move to option 3, where we look at a collection of solutions that guarantee α-approximation, and attempt to "pacify" agent 1 by choosing her favorite solution within that collection.

Algorithm 2.3 PACIFY-THE-LIAR

Require: Sets of items $X_1, \ldots, X_n$, where $X_i \in \widehat{G_i}$; parameter $\alpha \geq 1$

1: $Z_1 \leftarrow OPT(X_1)$, $Z_2 \leftarrow OPT(\cup_{i\in N\setminus\{1\}} X_i)$
2: **if** $v(Z_1) \geq \frac{1}{\alpha}(v(Z_1) + v(Z_2))$ **then**
3:     **return** Z_1 ▷ option 1
4: **else if** $v(Z_2) \geq \frac{1}{\alpha} v(OPT(\cup_{i\in N} X_i))$ **then**
5:     **return** Z_2 ▷ option 2
6: **else**
7:     $S \leftarrow \{A \subseteq \cup_{i\in N} X_i : v(A) > \alpha v(Z_2)\}$
8:     **return** $\operatorname{argmax}_{A\in S} v(A \cap X_1)$ ▷ option 3
9: **end if**

Note that if we reach option 3, S is nonempty since we did not stop at option 2 (thus S includes

[11] If we naturally extend our definition of mechanism to allow reporting of all private data, including honesty. This observation relies, of course, on the honest agents reporting their honesty correctly.

the optimal solution). By hiding items, agent 1 can only make S smaller; however, since the mechanism returns agent 1's favorite solution in S, she has no incentive to make S smaller, and thus no incentive to misreport.

Theorem 2.9. *In UM, PACIFY-THE-LIAR is strategyproof and α-approximate for $\alpha \geq \phi$.*

Proof. We begin by proving strategyproofness. If the algorithm ends at options 1 or 2, the proof is similar to case (1) in the proof of Theorem 2.5. If the algorithm ends at option 3, the fact that agent 1 cannot benefit from making the mechanism stop at an earlier option follows from a similar argument to the one in case (2) in the proof of Theorem 2.5. Thus, all we need to show is that if the mechanism stops at option 3 under agent 1's misreport, agent 1 does not benefit. Assume $X_1' \subset X_1$ and let $S' = \{A \subseteq X_1' \cup (\cup_{i \in N \setminus \{1\}} X_i) : v(A) > \alpha v(Z_2)\}$. Then, note that $S' \subseteq S$, and for every $A \in S'$, $v(A \cap X_1') = v(A \cap X_1)$; therefore, $\max_{A \in S'} v(A \cap X_1) \leq \max_{A \in S} v(A \cap X_1)$, and so agent 1 does not benefit.

Now that we have established strategyproofness, let us analyze the approximation ratio. Clearly α-approximation is guaranteed when the mechanism ends at options 1 or 2. So let us consider the case where the mechanism ends at option 3. Let A be the output. Since $A \in S$, $v(A) > \alpha v(Z_2)$. Since we did not stop at option 1,

$$v(Z_1) < \frac{1}{\alpha - 1} v(Z_2)$$
$$v(Z_1) + v(Z_2) < \frac{\alpha}{\alpha - 1} v(Z_2).$$

Therefore,

$$\begin{aligned} v(A) &> \alpha v(Z_2) \\ &> (\alpha - 1)(v(Z_1) + v(Z_2)) \\ &\geq (\alpha - 1) v(OPT(\cup_{i \in N} X_i)). \end{aligned}$$

Thus, we are guaranteed an α-approximate mechanism if $\frac{1}{\alpha} \leq (\alpha - 1)$, and this is easily seen to hold true for $\alpha \geq \phi$. □

Finally, we show that no deterministic strategyproof mechanism can do better:

Theorem 2.10. *In the understating model, if the ground sets are unrestricted, no deterministic strategyproof mechanism can provide a worst-case approximation ratio better than ϕ.*

Proof. Let f be a deterministic strategyproof mechanism with approximation ratio $r < \phi$. Consider the profile where $X_1 = \{a\}$ with $s(a) = 1$ and $v(a) = \phi$, $X_2 = \{b\}$ with $s(b) = \frac{1}{2}$ and $v(b) = 1$, and $X_i = \varnothing$ for all $i \geq 3$. To maintain approximation ratio r, we must have $a \in f(\mathbf{X})$. Consider the profile $\mathbf{X}'$ that differs from $\mathbf{X}$ only in $X_1' = \{a, c\}$, where $s(c) = \frac{1}{2}$ and $v(c) = \phi - \epsilon$ for some small $\epsilon > 0$. To maintain strategyproofness, we must have $a \in f(\mathbf{X}')$. Thus, the approximation ratio on that profile is $\frac{\phi-\epsilon+1}{\phi}$, which can be made arbitrarily close to $\frac{\phi+1}{\phi} = \phi$. Therefore, the worst-case approximation ratio of f cannot be better than ϕ. □

2.6 The Known-Quality-Unknown-Quantity Model

In this section, we consider a model which we call 'Known Quality Unknown Quantity' (KQUQ). In this model, the true item profile $\mathbf{X}$ is known, and in fact $|X_i| = 1$ for all $i \in N$; hence, we simply call agent i's item a_i and avoid using $\mathbf{X}$. Furthermore, for each a_i, $r_i = \frac{v(a_i)}{s(a_i)}$ is known, however $v(a_i)$ and $s(a_i)$ themselves are private information of agent i. When $r(a_i)$ is given, $s(a_i)$ determines $v(a_i)$, so we simply ask agent i to report $s(a_i)$. Agent i gets a utility of $v(a_i)$ if her item is chosen and 0 if not; note that her utility from a_i being chosen is $v(a_i)$ even if she misreports. In this model the "quality" of an item is known, but its indivisible value and size are not.

Formally, a deterministic mechanism in this model is a function $f : \mathbb{R}_+^{2n} \to 2^{\{a_1,\ldots,a_n\}}$, which maps $(\mathbf{r}, \mathbf{s})$ to a subset of the items to be included in the knapsack (we define $(\mathbf{r}, \mathbf{s})$ to equal $(r_1, \ldots, r_n, s_1, \ldots, s_n)$, where $s_i = s(a_i)$). We will require feasibility ($s(f(\mathbf{r}, \mathbf{s})) \leq 1$) and strategyproofness ($v(f(\mathbf{r}, \mathbf{s}) \cap \{a_i\}) \geq v(f(\mathbf{r}, (\mathbf{s}_{-i}, s_i')) \cap \{a_i\})$ for all $i \in N$, $s_i' \in (0, 1]$). We will also look for randomized strategyproof mechanisms. The adaptation is similar to before: f maps to a random variable over $2^{\{a_1,\ldots,a_n\}}$, feasibility is $s(f(\mathbf{r}, \mathbf{s})) \leq 1$ surely, and strategyproofness is $\mathbb{E}[v(f(\mathbf{r}, \mathbf{s}) \cap \{a_i\})] \geq \mathbb{E}[v(f(\mathbf{r}, (\mathbf{s}_{-i}, s_i')) \cap \{a_i\})]$. For convenience of presentation, we will allow items with zero value (all of our proofs can easily be adjusted to get rid such items).

In this model, while it is easily seen that HALF-GREEDY is not strategyproof (specifically, MAXIMUM-VALUE is not strategyproof), it can be easily modified to become strategyproof and remain 2-approximate—in fact, the modified mechanism is also well known to be a 2-approximation in non-strategic environments [Burke and Kendall, 2005]. $\succeq'$ is defined as before.

Definition 2.5. *The NEXT mechanism is defined as follows: if* $s(\cup_{i\in N}\{a_i\}) \leq 1$*, return* $\varnothing$*,* *otherwise return* $o_{\cup_{i \in N}\{a_i\}}$

Definition 2.6. *MODIFIED-HALF-GREEDY runs GREEDY with probability* $\frac{1}{2}$ *and NEXT with probability* $\frac{1}{2}$.

MODIFIED-HALF-GREEDY still runs GREEDY with probability $\frac{1}{2}$, but otherwise it doesn't choose the item with the maximal value, but rather the first item to not make it into the knapsack in GREEDY.

Theorem 2.11. *In KQUQ, MODIFIED-HALF-GREEDY is strategyproof and* 2*-approximate.*

Proof. Fix agent $j \in N$. In MODIFIED-HALF-GREEDY, item a_j is chosen with probability either $\frac{1}{2}$ or 0. Specifically, an item is chosen with probability $\frac{1}{2}$ if and only if it is in $L_{\cup_{i \in N}\{a_i\}}(o_{\cup_{i \in N}\{a_i\}}) \cup \{o_{\cup_{i \in N}\{a_i\}}\}$ (if $s(\{a_1, \ldots, a_n\}) \leq 1$, then all items are chosen with probability $\frac{1}{2}$) . If item a_j is chosen with probability $\frac{1}{2}$, agent j has no incentive to manipulate. If item a_j is chosen with probability 0, then $o_{\cup_{i \in N}\{a_i\}} \succ' a_j$, so $s(a_j)$ has no impact on what $L_{\cup_{i \in N}\{a_i\}}(o_{\cup_{i \in N}\{a_i\}}) \cup \{o_{\cup_{i \in N}\{a_i\}}\}$ is, and thus agent j cannot make the item get chosen. So strategyproofness is proven. 2-approximation of MODIFIED-HALF-GREEDY is, as we mentioned, known. □

As in GREEDY, we made some careful choices here to preserve strategyproofness. Specifically, the choice for NEXT to return $\varnothing$ when $s(\cup_{i \in N}\{a_i\}) \leq 1$, despite us being able to include all items in the knapsack in that case, is crucial; if we indeed included all items in this case, strategyproofness would have been violated.

Next, a matching lower bound:

Theorem 2.12. *No randomized strategyproof mechanism can provide a worst-case approximation ratio strictly better than* 2 *in KQUQ.*

Proof. Let f be a randomized strategyproof mechanism with approximation ratio $t < 2$. Consider the case where $r_1 = M$, $r_2 = 1$, and $r_i = 0$ for $i \geq 3$, where M is a large integer. Assume $s(a_1) = 1$ and $s(a_2) = 1$. Let p_1 be the probability with which item a_1 is chosen under f. Then $p_1 M + (1 - p_1) \geq \frac{1}{t} M$ to maintain approximation ratio t, and therefore $p_1 \geq \frac{\frac{M}{t} - 1}{M - 1}$. Now, consider the case where $s(a_1) = \frac{1}{M^2}$ (the rest of the data remains the same), and let p_1' be the probability that item a_1 is chosen under f in this case. To maintain strategyproofness, we must have $p_1 = p_1'$.

Therefore, to maintain approximation ratio t, we must have $p_1 \frac{1}{M} + (1 - p_1) \geq \frac{1}{t}$, which yields $p_1 \leq \frac{1-\frac{1}{t}}{1-\frac{1}{M}}$. Therefore, we must have $\frac{1-\frac{1}{t}}{1-\frac{1}{M}} \geq \frac{\frac{M}{t}-1}{M-1}$, namely $1 + \frac{1}{M} \geq \frac{2}{t}$. However, as $M \to \infty$, the left hand side goes to 1 and the right hand side remains $\frac{2}{t} > 1$. Contradiction. □

Finally, we show that randomization is necessary for good approximation:

Theorem 2.13. *No deterministic strategyproof mechanism can provide a constant worst-case approximation ratio in KQUQ.*

Proof. Let f be a deterministic strategyproof mechanism with approximation ratio t. Consider the case where $r_1 = M$, $r_2 = 1$, and $r_i = 0$ for $i \geq 3$, where $M > t$. Assume $s(a_1) = 1$ and $s(a_2) = 1$. On this instance, to maintain approximation ratio t, f must choose $\{a_1\}$. Now, consider the case where $s(a_1) = \frac{1}{M^2}$ (the rest of the data is the same). On this instance, to maintain approximation ratio t, f must choose $\{a_2\}$. Thus, when agent 1's item is of size $\frac{1}{M^2}$, she will have an incentive to report its size to be 1, violating strategyproofness. □

Chapter 3

Desirable Facility Location

3.1 Introduction

We consider the problem of locating a single facility on the real line. This facility serves a set of n agents, each of whom is located somewhere on the line as well. Each agent cares about his distance to the facility, and incurs a disutility (equivalently, cost) that is equal to his distance to access the facility. An agent's location is assumed to be private information that is known only to him. Agents report their locations to a central planner who decides where to locate the facility based on the reports of the agents. The planner's objective is to minimize a "social" cost function that depends on the vector of distances that the agents need to travel to access the facility. It is natural for the planner to consider locating the facility at a point that minimizes her objective function, but in that case the agents may not have an incentive to report their locations truthfully. As an example, consider the case of 2 agents located at x_1 and x_2 respectively, and suppose the location that optimizes the planner's objective is the mid-point $(x_1+x_2)/2$. Then, assuming $x_1 < x_2$, agent 1 has an incentive to report a location $x_1' < x_1$ so that the planner's decision results in the facility being located closer to his true location. The planner can address this issue by restricting herself to a *strategyproof* mechanism: by this we mean that it should be a (weakly) dominant strategy for each agent to report his location truthfully to the central planner. This, of course, is an attractive property, but it comes at a cost: based on the earlier example, it is clear that the planner cannot hope to optimize her objective. One way to avoid this difficulty is to assume an environment in which agents (and the planner) can make or receive payments; in such a case, the planner selects the location of the facility, and also a payment scheme, which specifies the amount of money an agent pays (or receives) as a function of the reported locations of the agents as well as the location of the facility. This option gives the planner the ability to support the "optimal" solution as the outcome of a strategyproof mechanism by constructing a carefully designed payment scheme in which any potential benefit for a misreporting agent from a change in the location of the facility is offset by an increase in his payment.

There are many settings, however, in which such monetary compensations are either not possible or are undesirable. This motivated Procaccia and Tennenholtz [Procaccia and Tennenholtz, 2013b] to formulate the notion of *Approximate Mechanism Design without Money.* In this model the planner restricts herself to strategyproof mechanisms, but is willing to settle for one that does not necessarily optimize her objective. Instead, the planner's goal is to find a mechanism that effectively *approximates* her objective function. This is captured by the standard notion of approximation that is widely used in the CS literature: for a minimization problem, an algorithm is an α-approximation if the solution it finds is guaranteed to have cost at most α times that of the optimal cost ($\alpha \geq 1$).

Procaccia and Tennenholtz [Procaccia and Tennenholtz, 2013b] apply the notion of approximate mechanism design without money to the facility location problem considered here for two different objectives: (i) *minisum*, where the goal is to minimize the sum of the costs of the agents; and (ii) *minimax*, where the goal is to minimize the maximum agent cost. They show that for the minimax objective choosing any k-th median—picking the kth largest reported location—is a strategyproof, 2-approximate mechanism. They design a randomized mechanism called LRM (Left-Right-Middle) and show that it is a strategyproof, 3/2-approximate mechanism; furthermore, they show that those mechanisms provide the optimal worst-case approximation ratio possible (among all deterministic and randomized strategyproof mechanisms, respectively). For the *minisum* objective, it is known that choosing the median reported location is optimal and strategyproof [Moulin, 1980]. Feldman and Wilf [Feldman and Wilf, 2013b] consider the same facility location problem on a line but with the social cost function being the L_2 norm of the agents' costs (Feldman and Wilf actually used the sum of squares of the agents' costs, however most of their results can be easily converted to the L_2 norm. Of course, the approximation ratios they report need to be adjusted as well). They show that the median is a $\sqrt{2}$-approximate strategyproof mechanism for this objective function, and provide a randomized $(1+\sqrt{2})/2$-approximate strategyproof mechanism. Feldman and Wilf also generalize the median mechanism to maintain strategyproofness and a $\sqrt{2}$ approximation ratio on trees; furthermore, they provide a family of randomized strategyproof mechanisms for trees, and in particular show that a member of this family reduces the approximation ratio to strictly below $\sqrt{2}$. In addition, some consideration has been given in literature to the circle topology, by Alon et al. [Alon *et al.*, 2010b; Alon *et al.*, 2010c]. A general survey of approximate mechanism design without money for facility location problems has been written by Cheng et al. [Cheng and Zhou, 2015].

Aside from the recent literature on approximate mechanism design, our work is loosely related to other strands in the literature with a much longer history. First is the classical work on social choice, which deals with aggregating the preferences of a set of voters over a set of alternatives [Moulin, 2015]. The location problem we consider is a special case in which the alternatives are all possible points on the real line (the location of the facility), and agents have single-peaked preferences. A significant difference, however, is the following: a typical social choice problem is to find an aggregation rule satisfying a desired set of properties, whereas in our case the planner wishes to optimize or approximate a given social objective function. Nevertheless, various techniques and results from this literature are useful in our setting as well. An important result is Moulin's characterization of strategyproof mechanisms on the line [Moulin, 1980]. A parallel characterization result was developed by Schummer and Vohra [Schummer and Vohra, 2002] for general graphs. In both these papers, much like in our work, generalized medians play an important role; also, despite not having a specific objective function, these characterizations assume less specific efficiency related properties, such as Pareto efficiency and onto range. Additional papers along these lines are [Barberà *et al.*, 1998; Danilov, 1994]. We note that impossibility results abound in social choice models; our focus on the simple special case enables us to avoid impossibility results such as the Gibbard-Satterthwaite Theorem [Gibbard, 1973; Satterthwaite, 1975], which implies the non-existence of a *reasonable* social choice function. A second relevant strand of literature is the classical work in operations research on graphical location problems that considers locating the facility at a Condorcet point [Hansen and Thisse, 1981; Labbé, 1985; Bandelt, 1985; Bandelt and Labbé, 1986] (a Condorcet point is one that is preferred by a majority of agents to any other location). This literature seeks to establish bounds on the total cost to all the agents to access the facility divided by the minimum cost, with the understanding that smaller ratios are better. However, this literature does not model individual agent incentives, and moreover does not also explore other mechanisms. Finally, there is a rich literature on facility location problems and variations (such as the k-median and k-center problems) where agent incentives are not taken into account. In such problems, there is typically a single objective function (the planner's), and agent locations are known. In this literature, one resorts to approximation algorithms for a different reason—often, these optimization problems turn out to be computationally intractable, and the focus is on developing computationally efficient heuristics for which a worst-case approximation guarantee can be proved (see [Vygen, 2005], and chapters 25-26 of [Vazirani, 2001]). To our knowledge, most of the algorithms designed in this literature violate our (rather strong) strategyproofness requirements.

In our work, we follow a suggestion of Feldman and Wilf [Feldman and Wilf, 2013b] and study the problem of locating a single facility on a line, but with the objective function being the L_p norm of the vector of agent-costs (for general $p \geq 1$). In the context of real world facility location problems, where the agents must drive to and from the facility, the L_p norm can represent situations where travel time or other cost increases superlinearly with the distance (as suggested in [Brandeau and Chiu, 1988]). For example, when driving over larger distances, there is an increased likelihood (depending on traffic) of the need to stop and refuel, or, in the case of electric cars, stop and recharge–which is even more costly since such recharging can be done at home, without wasting the driver's time. As another example, certain hybrid cars increase their fuel consumption in longer drives— which is relevant if the cost represents fuel consumption rather than travel time. For such problems, our results provide strong lower bounds, robust to the topology of the road network (since they only require a line) and the value of p. We also hope that our results regarding the median will guide the construction of good mechanisms for more general topologies, similarly to the case of $p = 2$ in [Feldman and Wilf, 2013b], where the optimality of the median on the line inspires the construction of a mechanism for tree networks using the appropriate adaptation of the median. Another use of the L_p norm is to strike a balance between efficiency and fairness. The cases of $p = 1$ and $p = \infty$, which were both studied in [Procaccia and Tennenholtz, 2013b], can be viewed as representing the two extremes on the spectrum between maximizing efficiency (minimizing the total social cost) and maximizing fairness (minimizing the cost of the agent who is worst off). Thus, our definition of social cost allows for a controlled tradeoff between efficiency and fairness by varying the value of p. On the line, this interpretation of the L_p norm becomes particularly interesting in the context of voting. Public opinion on many issues is considered to be on a spectrum between political left and right, lending itself naturally to a one dimensional description. One of the common problems in democratic societies is to balance between majority rule and respecting minority rights; thus, the L_p measure allows for a quantitative exploration of this balance. Of course, this interpretation of the L_p norm can be relevant to physical facility location problems as well.

We define the problem formally in Section 3.2. In Section 3.3, we show that the median mechanism (which is strategyproof) provides a $2^{1-\frac{1}{p}}$ approximation ratio, and that this is the optimal approximation ratio among all deterministic strategyproof mechanisms. We move onto randomized mechanisms in Section 3.4. First, we present a negative result: we show that for integer $\infty > p > 2$, *no* mechanism—from a rather large class of randomized mechanisms— has an approximation ratio better than that of the median mechanism, as the number of agents goes to infinity. It is worth noting that all the mechanisms proposed in literature so far— for minimax, minisum, and the L_2 social cost functions— belong to this class of mechanisms. Next, we consider the case of 2 agents, and show that the LRM mechanism provides the optimal approximation ratio among all randomized strategyproof mechanisms (that satisfy certain mild assumptions) for this special case, for every $p \geq 1$. Our result for the special case of 2 agents also gives a lower bound on the approximation ratio for all randomized mechanisms satisfying the assumptions. In the appendix we discuss some omitted technical details, as well as an additional negative result for an alternative definition of the agents' cost.

3.2 Model

Let $N = \{1, 2, \ldots, n\}$, $n \geq 2$, be the set of agents. Each agent $i \in N$ reports a location $x_i \in \mathbb{R}$. A *deterministic* mechanism is a collection of functions $f = \{f_n | \ n \in \mathbb{N}, n \geq 2\}$ such that each $f_n : \mathbb{R}^n \to \mathbb{R}$ maps each location profile $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ to the location of a facility. We will abuse notation and let $f(\mathbf{x})$ denote $f_n(\mathbf{x})$. Under a similar notational abuse, a *randomized* mechanism is a collection of functions f that maps each location profile to a probability distribution over $\mathbb{R}$: if $f(x_1, x_2, \ldots, x_n)$ is the distribution π, then the facility is located by drawing a single sample from π.

Our focus will be on deterministic and randomized mechanisms for the problem of locating a single facility when the location of any agent is *private* information to that agent and cannot be observed or otherwise verified. It is therefore critical that the mechanism be *strategyproof*—it should be optimal for each agent i to report his *true* location x_i rather than something else. To that end we assume that if the facility is located at y, an agent's disutility, equivalently cost, is simply his distance to y. Thus, an agent whose true location is x_i incurs a cost $C(x_i, y) = |x_i - y|$. If the location of the facility is random and according to a distribution π, then the cost of agent i is simply $C(x_i, \pi) = \mathbb{E}_{y\sim\pi}|x_i - y|$, where y is a random variable with distribution π. The formal definition of strategyproofness is now:[1]

Definition 3.1. *A mechanism f is strategyproof if for each $i \in N$, each $x_i, x_i' \in \mathbb{R}$, and for each $\mathbf{x_{-i}} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots x_n) \in \mathbb{R}^{n-1}$,*

$$C(x_i, f(x_i, \mathbf{x_{-i}})) \leq C(x_i, f(x_i', \mathbf{x_{-i}})),$$

where $(\alpha, \mathbf{x_{-i}})$ denotes a vector with the i-th component being α and the j-th component being x_j for all $j \neq i$.

The class of strategyproof mechanisms is quite large: for example, locating the facility at agent 1's reported location is strategyproof, but is not particularly appealing because it fails almost every reasonable notion of fairness and could also be highly "inefficient". To address these issues, and to winnow down the class of acceptable mechanisms, we impose additional requirements that stem from efficiency or fairness considerations. In this chapter we assume that locating a facility at y when the location profile is $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ incurs the *social cost*[2]

$$sc(\mathbf{x}, y) \;=\; \left(\sum_{i\in N} |x_i - y|^p\right)^{1/p}, \quad p \geq 1.$$

For a randomized mechanism f that maps $\mathbf{x}$ to a distribution π, we define the social cost to be

$$sc(\mathbf{x}, \pi) \;=\; \mathbb{E}_{y\sim\pi}\left[\left(\sum_{i\in N} |x_i - y|^p\right)^{1/p}\right].$$

[1]Note that for randomized mechanisms, we require strategyproofness in expectation, rather than ex-post.

[2]For this definition of social cost, an alternative option is to let the agents' costs increase non-linearly with their distance from the facility, in particular $C(x_i, y) = |x_i - y|^p$. In Appendix B.1 we provide an interesting result for this case.

For this definition of social cost, our goal now is to find a strategyproof mechanism that does well with respect to minimizing the social cost. A natural mechanism (and this is the approach taken in the classical literature on facility location) is the "optimal" mechanism: each location profile $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is mapped to $OPT(\mathbf{x})$, defined as $OPT(\mathbf{x}) \in \arg\min_{y\in\mathbb{R}} sc(\mathbf{x}, y)$.[3] This optimal mechanism is not strategyproof as shown in the following example.

Example. Suppose there are two agents located at the points 0 and 1 respectively on the real line. If they report their locations truthfully, the optimal mechanism will locate the facility at $y = 0.5$, for any $p > 1$. Assuming agent 2 reports $x_2 = 1$, if agent 1 reports $x_1' = -1$ instead, the facility will be located at 0, which is best for agent 1.

Given that strategyproofness and optimality cannot be achieved simultaneously, it is necessary to find a tradeoff. In this work we shall focus on strategyproof mechanisms that approximate the optimal social cost as best as possible. The notion of approximation that we use is standard in computer science: an α-approximation algorithm is one that is guaranteed to have cost no more than α times the optimal social cost. Formally, the worst-case approximation ratio of a mechanism f is $\sup_{\mathbf{x}}\{sc(\mathbf{x}, f(\mathbf{x}))/sc(\mathbf{x}, OPT(\mathbf{x}))\}$, where the supremum is taken over all possible instances $\mathbf{x} \in \cup_{k\geq 2}\mathbb{R}^k$ of the problem.[4] Our goal then is to design strategyproof (deterministic or randomized) mechanisms whose worst-case approximation ratio is as close to 1 as possible.

3.3 The Median Mechanism

For the location profile $\mathbf{x} = (x_1, x_2, \dots, x_n)$, the median mechanism is a deterministic mechanism that locates the facility at the "median" of the reported locations. The median is unique if n is odd, but not when n is even, so we need to be more specific in describing the mechanism. For odd n, say $n = 2k - 1$ for some $k \geq 1$, the facility is located at $x_{[k]}$, where $x_{[k]}$ is the kth largest component of the location profile. For even n, say $n = 2k$, the "median" can be any point in the

[3]Strictly speaking, the mechanism is not well defined in cases where the social cost at $\mathbf{x}$ is minimized by multiple locations, but we could pick an exogenous tie-braking rule to deal with such cases.

[4]For the case of randomized mechanisms, it should be noted that this approximation ratio is in expectation rather than with high probability.

interval $[x_{[k]}, x_{[k+1]}]$; to ensure strategyproofness, we need to pick either $x_{[k]}$ or $x_{[k+1]}$, and as a matter of convention we take the median to be $x_{[k]}$. It is well known that the median mechanism is strategyproof.[5] Furthermore, the median mechanism is *anonymous*.[6] Thus we may assume, without loss of generality, that each agent reports her location truthfully.

Our main result in this section is that, for any $p \geq 1$, the median mechanism uniformly achieves the best possible approximation ratio among all deterministic strategyproof mechanisms. We start with two simple observations, which will be used in the proof of this main result.

Lemma 3.1. *For any real numbers a, b, c with $a \leq b \leq c$, and any $p \geq 1$,*

$$(c-a)^p \;\leq\; 2^{p-1}[(c-b)^p + (b-a)^p].$$

Proof. For any $p \geq 1$, $f(x) = x^p$ is a convex function on $[0, \infty)$, and so for any $\lambda \in [0, 1]$ and $x, y \geq 0$,

$$f(\lambda x + (1-\lambda)y)) \;\leq\; \lambda f(x) + (1-\lambda) f(y). \tag{3.1}$$

Setting $\lambda = 1/2$, $x = c - b$, and $y = b - a$, we get:

$$\frac{1}{2^p}(c-a)^p \;\leq\; \frac{1}{2}[(c-b)^p + (b-a)^p]. \tag{3.2}$$

Multiplying both sides of the inequality by 2^p gives the result. □

Lemma 3.2. *For any non-negative real numbers a and b, and any $p \geq 1$,*

$$(a+b)^p \;\geq\; a^p + b^p.$$

Proof. For integer p, the result is a direct consequence of the binomial theorem; the same argument covers the case of rational p as well. Continuity implies the result for all p. □

---

[5] A classical paper of Moulin [Moulin, 1980] for a closely related model shows that all deterministic strategyproof mechanisms are essentially generalized median mechanisms.

[6] In an anonymous mechanism, the facility location is the same for two location profiles that are permutations of each other.

Theorem 3.1. *Suppose there are n agents with the location profile $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Define the social cost of locating a facility at y as $(\sum_{i=1}^{n} |y - x_i|^p)^{\frac{1}{p}}$ for $p \geq 1$. The social cost incurred by the median mechanism is at most $2^{1-\frac{1}{p}}$ times the optimal social cost.*[7]

Proof. We may assume that $x_1 \leq ... \leq x_n$. Let OPT be a facility location that minimizes the social cost, and let m be the median. The inequality we need to prove is

$$\sum_{i=1}^{n} |m - x_i|^p \;\leq\; 2^{p-1} \sum_{i=1}^{n} |OPT - x_i|^p.$$

We do this by pairing each location x_i with its "symmetric" location x_{n+1-i} and arguing that the total cost of these two locations in the median mechanism is within the required bound of their total cost in an optimal solution. For even n, this completes the argument; for odd n the only location without such a pair is the median itself, which incurs *zero* cost in the median mechanism, and so the argument is complete. Formally, the result follows if we can show

$$|m - x_i|^p + |x_{n+1-i} - m|^p \leq 2^{p-1}(|OPT - x_i|^p + |OPT - x_{n+1-i}|^p), \quad \forall\; i \leq \lfloor n/2 \rfloor.$$

We consider two cases, depending on whether OPT is in the interval $[x_i, x_{n+1-i}]$ or not. In each of these cases, OPT may be above the median or below, but the proof remains identical in each subcase, so we give only one.

1. $x_i \leq m \leq OPT \leq x_{n+1-i}$ or $x_i \leq OPT \leq m \leq x_{n+1-i}$. We will prove the first of these subcases; the proof of the second is identical. Applying Lemma 3.1 by setting $a = m$, $b = OPT$, and $c = x_{n+1-i}$, we get

$$|x_{n+1-i} - m|^p \leq 2^{p-1}(|x_{n+1-i} - OPT|^p + |OPT - m|^p).$$

Thus,

$$\begin{aligned} |m - x_i|^p + |x_{n+1-i} - m|^p &\leq |m - x_i|^p + 2^{p-1}(|x_{n+1-i} - OPT|^p + |OPT - m|^p) \\ &\leq 2^{p-1}(|m - x_i|^p + |x_{n+1-i} - OPT|^p + |OPT - m|^p) \\ &\leq 2^{p-1}(|x_{n+1-i} - OPT|^p + |OPT - x_i|^p), \end{aligned}$$

where the last inequality is obtained by applying Lemma 3.2 to the terms $|m - x_i|^p$ and $|OPT - m|^p$.

[7] This is a generalization of the results for $p = 2$ [Feldman and Wilf, 2013b], $p = 1$ and $p = \infty$ [Procaccia and Tennenholtz, 2013b] (when $p = \infty$, the median mechanism provides a 2-approximation).

2. $OPT \leq x_i \leq m \leq x_{n+1-i}$ or $x_i \leq m \leq x_{n+1-i} \leq OPT$. Again, we prove only the first subcase. Note that

$$\begin{aligned} |x_{n+1-i} - m|^p + |m - x_i|^p &\leq |x_{n+1-i} - x_i|^p \\ &\leq |OPT - x_{n+1-i}|^p \\ &\leq 2^{p-1}(|OPT - x_i|^p + |OPT - x_{n+1-i}|^p) \end{aligned}$$

where the first inequality follows from Lemma 3.2. (Note that Lemma 3.1 is not used in the proof of this case.)

□

We end this section by showing that *no* deterministic and strategyproof mechanism can give a better approximation to the social cost.

Lemma 3.3. *Consider the case of two agents and suppose the location profile is* (x_1, x_2) *with* $x_1 < x_2$*. For* $p \geq 1$*, suppose the social cost of locating a facility at* y *is* $(|x_1 - y|^p + |x_2 - y|^p)^{1/p}$*. Any deterministic mechanism that has worst-case approximation ratio better than* $2^{1-\frac{1}{p}}$ *for* $p > 1$ *must locate the facility at* y *for some* $y \in (x_1, x_2)$*.*[8]

Proof. The function $sc(\mathbf{x}, y)$ is strictly convex in y, and its unique minimizer is $y^* = (x_1+x_2)/2$, with the corresponding value $sc(\mathbf{x}, y^*) = |x_2 - x_1|/2^{1-\frac{1}{p}}$. Moreover $sc(\mathbf{x}, x_1) = sc(\mathbf{x}, x_2) = |x_2 - x_1| = 2^{1-\frac{1}{p}} sc(\mathbf{x}, y^*)$. It follows that for the deterministic mechanism to do strictly better than the stated ratio, the facility cannot be located at the reported locations; locating the facility to the left of x_1 or to the right of x_2 only increases the cost of the mechanism, so the only option left for a mechanism to do better is to locate the facility in the interior, i.e., in (x_1, x_2). □

Theorem 3.2. *Any strategyproof deterministic mechanism has an approximation ratio of at least* $2^{1-\frac{1}{p}}$ *for the* L_p *social cost function for any* $p \geq 1$*.*[9]

[8] Ex-post Pareto efficiency (as defined in Section 3.4.2) requires the facility to be located in $[x_1, x_2]$; thus, this property is stronger.

[9] The lower bound of 2 on the approximation ratio holds when $p = \infty$, see Procaccia and Tennenholtz [Procaccia and Tennenholtz, 2013b].

Proof. Using Lemma 3.3, we can now argue similarly to the case of $p = \infty$ (theorem 3.2 in [Procaccia and Tennenholtz, 2013b]).[10] Suppose $p > 1$ (the bound holds trivially for $p = 1$), and suppose a deterministic strategyproof mechanism yields an approximation ratio strictly better than $2^{1-\frac{1}{p}}$ for the L_p social cost. For the two-agent location profile $x_1 = 0, x_2 = 1$, Lemma 3.3 implies the facility is located at some $y \in (0,1)$. Now consider the location profile $x_1 = 0, x_2 = y$. Again, by Lemma 3.3, the mechanism must locate the facility at $y' \in (0, y)$ to guarantee the improved approximation. But if agent 2 is located at $y < 1$, he can misreport his location as 1, forcing the mechanism to locate the facility at y, his true location; this violates strategyproofness. □

We note that our result continues to hold for arbitrary single peaked cost functions, as long as the social cost remains an L_p measure of the distances.

3.4 Randomized Mechanisms

Recall that when the social cost is measured by the L_2 norm or the L_∞ norm, randomization provably improves the approximation ratio. In the former case, Feldman and Wilf [Feldman and Wilf, 2013b] describe an algorithm with an approximation ratio of $(\sqrt{2}+1)/2$; for the latter, Procaccia and Tennenholtz [Procaccia and Tennenholtz, 2013b] design an algorithm with an approximation ratio of 3/2. The mechanisms in both cases are simple and somewhat similar, placing non-negative probabilities *only* on the optimal location and generalized medians (defined shortly), where these probabilities are independent of the reported location profile. In this section we show that this is not enough in general; namely, randomizing over generalized medians and the optimal location does not improve the approximation ratio of the median mechanism for *any* integer $p \in (2, \infty)$. For the case of 2 agents we show that the best approximation ratio is given by the LRM mechanism among all strategyproof mechanisms. Extending this analysis even to the case of 3 agents appears to be non-trivial.

3.4.1 Mixing Dictatorships and Generalized Medians with the Optimal Location

We begin with a definition of generalized medians.

[10] Another argument along this line can be found in the proof of theorem 4.4 in [Feldman and Wilf, 2013b].

Definition 3.2. *Let* $\mathbf{x} \in \mathbb{R}^n$*,* $S \subseteq N$*, and* $m \in \{1, \ldots, |S|\}$*. Let* $S = \{s_1, \ldots, s_{|S|}\}$*, where* $x_{s_i} \leq x_{s_{i+1}}$*. Then, the* m*th generalized median of subset* S *in location profile* $\mathbf{x}$ *is* $x_{[m,S]} = x_{s_m}$*.*[11] *If* $S = N$*, we allow for the shorthand* $x_{[m]} = x_{[m,N]}$*.*

Next, we define the class of mechanisms currently used in literature:

Definition 3.3. *Let* f *be a mechanism which satisfies the following. For every* $n \in \mathbb{N}$*,* $S \subseteq N$*,* $m \in \{1, \ldots, |S|\}$*, there exist non-negative numbers* $v_m^{n,S}$*, and* v_{OPT}^n *with* $v_{OPT}^n + \sum_{S \subseteq N, m \in \{1,\ldots,|S|\}} v_m^{n,S} = 1$*, such that for every profile* $(x_1, x_2, \ldots, x_n)$*,* f *locates the facility at* OPT *with probability* v_{OPT}^n *and at* $x_{[m,S]}$ *with probability* $v_m^{n,S}$ *(where* OPT *is the optimal location for the profile* $(x_1, x_2, \ldots, x_n)$*).*[12] *If* f *satisfies these properties, we say that* f *is a Mixed Generalized Medians Optimal (MGMO) mechanism.*

We now show that for integer $p > 2$, MGMO mechanisms cannot beat the median.

Theorem 3.3. *Let* f *be a strategyproof MGMO mechanism. Then, for any finite integer* $p > 2$*, the approximation ratio of* f *is at least* $2^{1-\frac{1}{p}}$*.*

Proof. Fix $n = 2k$, with $k \in \mathbb{N}$. In all profiles in our proof, the relative order of agents locations remains the same: specifically, $i < j$ implies $x_i \leq x_j$ for all of our profiles $\mathbf{x}$. For every $S \subseteq N$, and every $j \in S$ let $S(j)$ be the number of agents with index weakly smaller than j in S (for example, if $S = \{2, 4, 9\}$, then $S(2) = 1$, $S(4) = 2$, and $S(9) = 3$). On our profiles, the probability that the location of agent $j \in N$ is chosen as a generalized median therefore is $v_j^n = \sum_{S \subseteq N : j \in S} v_{S(j)}^{n,S}$.

For $j = 1, \ldots, k$, define the profile $\mathbf{x}^j$ as follows (where a_j is a parameter to be defined shortly): agents 1 through j are located at $-a_j$; agents $j+1$ through k are located at 0; agents $k+1$ through $2k - j + 1$ are located at 1; and agents $2k - j + 2$ through $2k$ are located at $1 + a_j$ (note the *slight* asymmetry in the location of the agents: while k agents are at or below zero, and k agents are at or above 1, there is an additional agent at 1 compared to zero and so one less agent at $1 + a_j$ compared to $-a_j$). Now, a_j is chosen to be the smallest positive root of the function $g_j(\alpha) = j\alpha^{p-1} - (k - j + 1) - (j - 1)(1 + \alpha)^{p-1}$; such an a_j must exist by the intermediate value theorem, as $g_j(0) < 0$ and $g_j(\alpha)$ is a continuous function of α with $g_j(\alpha) \to \infty$ as $\alpha \to \infty$.

[11] That is, $x_{[m,S]}$ is the mth largest location among the locations of the agents in S, allowing for repetition.

[12] When a location appears more than once in OPT and $x_{[m,S]}$ for $S \subseteq N$ and $m \in \{1, \ldots, |S|\}$, the probabilities add up.

We show that the optimal mechanism locates the facility at zero for the profile $\mathbf{x}^j$, i.e., $OPT = 0$. Note that the social cost for this profile, when locating the facility at $z \in [0,1]$, is $j(z+a_j)^p + (k-j)z^p + (k-j+1)(1-z)^p + (j-1)(1+a_j-z)^p$, and when $z \in (-a_j, 0)$ the social cost becomes $j(z+a_j)^p + (k-j)(-z)^p + (k-j+1)(1-z)^p + (j-1)(1+a_j-z)^p$. Note that the social cost function is differentiable for $z \in (0,1)$ and for $z \in (-a_j, 0)$. The left and right derivatives at 0 are both $pja_j^{p-1} - p(k-j+1) - p(j-1)(1+a_j)^{p-1}$, and thus the social cost function is differentiable on $(-a_j, 1)$ with its derivative at $z = 0$ equal to zero (by our choice of a_j). The fact that this is a global minimum now follows from strict convexity of the social cost function $||\mathbf{x}^j - z(1,\ldots,1)||_p$ (for all $z \in \mathbb{R}$). Thus, indeed, $OPT = 0$.

We now attempt to bound v_{OPT}. For each profile $\mathbf{x}^j$, consider the profile $\mathbf{x}'^j$ that differs only in the location of agent j: namely, $x'^j_j = 0$ instead of $-a_j$. Note that on this profile, $OPT = 0.5$ by symmetry. Strategyproofness implies that a deviation from profile $\mathbf{x}'^j$ to profile $\mathbf{x}^j$ should not be beneficial for agent j, namely $a_j v_j^n - \frac{1}{2} v_{OPT}^n \geq 0$ (where a_j is the increase in agent j's cost caused by that deviation when the facility is built in his reported location, and $\frac{1}{2}$ is the decrease in his cost caused by that deviation when the facility is located at OPT), which implies $v_j^n \geq \frac{v_{OPT}^n}{2a_j}$. Defining a_j for $j = k+1, \ldots, 2k$ in a symmetric fashion, it follows that the same inequality holds for j in that range, and that $a_j = a_{2k-j+1}$. Summing those inequalities up, we get:

$$1 - v_{OPT}^n = \sum_{j=1}^{2k} v_j^n \geq \sum_{j=1}^{2k} \frac{v_{OPT}^n}{2a_j} = 2\sum_{j=1}^{k} \frac{v_{OPT}^n}{2a_j} = \sum_{j=1}^{k} \frac{v_{OPT}^n}{a_j}$$

$$v_{OPT}^n \leq \frac{1}{1 + \sum_{j=1}^{k} \frac{1}{a_j}}$$

Now, we claim it is enough to show that as $n \to \infty$ (or equivalently, as $k \to \infty$), $\sum_{j=1}^{k} \frac{1}{a_j} \to \infty$. The inequality then implies that $v_{OPT}^n \to 0$. Consider the profile which locates k agents at 0 and k agents at 1. The social cost of locating the facility at OPT on this profile is $\sqrt[p]{n}/2$, while the social cost of locating the facility at an agent's location is $\sqrt[p]{n}2^{-\frac{1}{p}}$; thus, the approximation ratio of f on this profile is $\frac{v_{OPT}^n \sqrt[p]{n}/2 + (1-v_{OPT}^n)\sqrt[p]{n}2^{-\frac{1}{p}}}{\sqrt[p]{n}/2} = 2^{1-\frac{1}{p}} - (2^{1-\frac{1}{p}} - 1)v_{OPT}^n$. Thus, as $n \to \infty$, the approximation ratio on these profiles approaches $2^{1-\frac{1}{p}}$, completing the proof.

We are left with the task of showing that $\lim_{k\to\infty} \sum_{j=1}^{k} \frac{1}{a_j} = \infty$. To do so, we first show that for

$j \geq k^{\frac{1}{p-1}}+1$, $2^{p-1}(j-1) > a_j$. Recall that a_j was defined as the smallest positive root of $g_j(\alpha)$, and that $g_j(0) < 0$. Thus, it is enough to show that for j in the appropriate range, $g_j(2^{p-1}(j-1)) > 0$. For notational convenience, we denote $Q = 2^{p-1}$.

$$\begin{aligned}
g_j(Q(j-1)) &= jQ^{p-1}(j-1)^{p-1} - (k-j+1) - (j-1)(1+Q(j-1))^{p-1} \\
&= Q^{p-1}(j-1)^{p-1} - k - (j-1)\sum_{i=1}^{p-2}\binom{p-1}{i}(Q(j-1))^{p-1-i} \\
&\geq Q^{p-1}(j-1)^{p-1} - (j-1)^{p-1} - (j-1)\sum_{i=1}^{p-2}\binom{p-1}{i}(Q(j-1))^{p-1-i} \\
&\geq Q^{p-1}(j-1)^{p-1} - (j-1)^{p-1} - (j-1)\sum_{i=1}^{p-2}\binom{p-1}{i}(Q(j-1))^{p-2} \\
&> Q^{p-1}(j-1)^{p-1} - (j-1)\sum_{i=1}^{p-1}\binom{p-1}{i}(Q(j-1))^{p-2} \\
&= Q^{p-1}(j-1)^{p-1} - (j-1)(Q(j-1))^{p-2}\sum_{i=1}^{p-1}\binom{p-1}{i} \\
&> Q^{p-1}(j-1)^{p-1} - (j-1)(Q(j-1))^{p-2}2^{p-1} = 0.
\end{aligned}$$

Now,

$$\begin{aligned}
\lim_{k\to\infty}\sum_{j=1}^{k}\frac{1}{a_j} &> \lim_{k\to\infty}\sum_{j=\lceil k^{\frac{1}{p-1}}+1\rceil}^{k}\frac{1}{2^{p-1}j} \\
&= \frac{1}{2^{p-1}}\lim_{k\to\infty}\sum_{j=\lceil k^{\frac{1}{p-1}}+1\rceil}^{k}\frac{1}{j} \\
&\geq \frac{1}{2^{p-1}}\lim_{k\to\infty}\int_{k^{\frac{1}{p-1}}+2}^{k}\frac{1}{t}dt \\
&= \frac{1}{2^{p-1}}(\lim_{k\to\infty}\int_{k^{\frac{1}{p-1}}}^{k}\frac{1}{t}dt - \lim_{k\to\infty}\int_{k^{\frac{1}{p-1}}}^{k^{\frac{1}{p-1}}+2}\frac{1}{t}dt) \\
&= \frac{1}{2^{p-1}}((\lim_{k\to\infty}(1-\frac{1}{p-1})\ln k) - 0) = \infty
\end{aligned}$$

which completes our proof. □

3.4.2 Optimality of the LRM Mechanism for 2 Agents

Procaccia and Tennenholtz [Procaccia and Tennenholtz, 2013b] defined the mechanism Left-Right-Middle (LRM) as follows: place the facility with probability $\frac{1}{2}$ at OPT, and with probability $\frac{1}{4}$ at each of $x_{[1]}$ and $x_{[n]}$. They have shown that it is strategyproof, and that it provides a best-possible approximation ratio of $\frac{3}{2}$ when $p = \infty$. Our next result shows that the LRM mechanism provides the best possible approximation ratio among all shift and scale invariant (defined below) strategyproof mechanisms for the case of 2 agents for all L_p social cost functions for $p \geq 1$.

We begin with some definitions: we say that a mechanism f is *shift and scale invariant* if for every location profile $\mathbf{x} = (x_1, x_2)$ and every $c \in \mathbb{R}$, the following two properties are satisfied:[13]

1. Shift Invariance: the random variables $Y' \sim f(x_1 + c, x_2 + c)$ and $Y + c$ s.t. $Y \sim f(\mathbf{x})$ are equal in distribution.[14]

2. Scale Invariance: the random variables $Y' \sim f(cx_1, cx_2)$ and cY s.t. $Y \sim f(\mathbf{x})$ are equal in distribution.

A convenient notation for a given location profile $\mathbf{x}$ is to denote its midpoint as $m_{\mathbf{x}} = \frac{x_1+x_2}{2}$. We say that a mechanism f is *symmetric* if for any location profile $\mathbf{x}$ and for any $y \in \mathbb{R}$, $\mathbb{P}(f(\mathbf{x}) \geq m_{\mathbf{x}} + y) = \mathbb{P}(f(\mathbf{x}) \leq m_{\mathbf{x}} - y)$.

The structure of the proof is as follows. Our goal is to show that within the class of strategyproof, shift invariant and scale invariant mechanisms, we can further limit ourselves to symmetric mechanisms that locate the facility always at the agents' locations or the midpoint; within this further restricted class, it becomes easy to prove that LRM is optimal. We achieve this goal gradually. First we show that we may restrict ourselves to symmetric (and anonymous) mechanisms. We then provide a characterization of strategyproofness for such mechanisms, and use it to show that we can further restrict ourselves to mechanisms which, for each profile $\mathbf{x}$, do not locate the facility

[13]While these two properties are natural and reasonable to expect, it should be noted that they are not implied by strategyproofness- one example is the constant mechanism, which always locates the facility at the same point regardless of the reports. Requiring unanimity in addition to strategyproofness is also not sufficient to guarantee these properties; for example, the mechanism that runs LRM if $x_{[1]} = 0$, and otherwise locates the facility at $x_{[1]}$ and $x_{[2]}$ with probability $1/2$ each, is easily seen to be strategyproof and unanimous but neither shift nor scale invariant.

[14]It is possible to replace shift invariance with symmetry in our assumptions, and preserve our results; see appendix.

both at $(\min\{x_1, x_2\}, \max\{x_1, x_2\})$ and at $(-\infty, \min\{x_1, x_2\}) \cup (\max\{x_1, x_2\}, \infty)$ with positive probability. We then show that we can restrict ourselves to mechanisms that locate the facility always at the agents' locations or the midpoint.

The following lemma allows us to focus on symmetric mechanisms.

Lemma 3.4. *Given any strategyproof, shift and scale invariant mechanism, there exists a symmetric, strategyproof, shift and scale invariant mechanism with the same worst-case approximation ratio.*

Proof. Given a mechanism f, we define the *mirror mechanism* of f, f_{mirror}, to be such that for every profile $\mathbf{x}$, we have that $\mathbb{P}(f_{mirror}(\mathbf{x}) \geq m_{\mathbf{x}} + b) = \mathbb{P}(f(\mathbf{x}) \leq m_{\mathbf{x}} - b)$ for all $b \in \mathbb{R}$.[15]

We will need the following notation: For each profile $\mathbf{x} = (x_1, x_2)$, let $Y_{x_1,x_2} \sim f(\mathbf{x})$, and $Y'_{x_1,x_2} \sim f_{mirror}(\mathbf{x})$. We claim that f_{mirror} is shift invariant, scale invariant and strategyproof (all of the equalities below are in distribution):

1. Shift invariance: let $c \in \mathbb{R}$. Then $Y'_{x_1+c,x_2+c} = 2m_{x_1+c,x_2+c} - Y_{x_1+c,x_2+c} = 2m_{\mathbf{x}} + 2c - Y_{x_1,x_2} - c = Y'_{x_1,x_2} + c$.

2. Scale invariance: let $c \in \mathbb{R}$. Then $Y'_{cx_1,cx_2} = 2cm_{x_1,x_2} - Y_{cx_1,cx_2} = c(2m_{x_1,x_2} - Y_{x_1,x_2}) = cY'_{x_1,x_2}$.

3. Strategyproofness: assume f_{mirror} is not strategyproof, and assume without loss of generality that agent 2 has a profitable misreport: there exist profiles (w_1, w_2) and $(w_1, w_2 + \alpha)$ for some $\alpha \in \mathbb{R}$ such that $\mathbb{E}[|w_2 - Y'_{w_1,w_2}|] > \mathbb{E}[|w_2 - Y'_{w_1,w_2+\alpha}|]$. However, note that $w_2 - Y'_{w_1,w_2+\alpha} = -w_1 - \alpha + Y_{w_1,w_2+\alpha} = Y_{w_1-\alpha,w_2} - w_1$ (the second equality follows from shift invariance), and that $w_2 - Y'_{w_1,w_2} = Y_{w_1,w_2} - w_1$. Thus, it follows that $\mathbb{E}[|w_1 - Y_{w_1,w_2}|] > \mathbb{E}[|Y_{w_1-\alpha,w_2} - w_1|]$, violating strategyproofness for f. Thus f_{mirror} must be strategyproof.

Therefore, the mechanism g that picks f with probability 1/2 and f_{mirror} with probability 1/2 is a strategyproof mechanism that is also symmetric; g trivially satisfies shift and scale invariance. Finally, note that g has the same approximation ratio as f for all location profiles, since f_{mirror} has the same approximation ratio as f. □

[15] Equivalently, the mirror mechanism can be thought of as follows: whenever f locates the facility at $y \in \mathbb{R}$ (that is, the single sampling of $f(\mathbf{x})$ yields y), f_{mirror} "mirrors" that location about $m_{\mathbf{x}}$, meaning it locates the facility at $2m_{\mathbf{x}} - y$.

Mechanisms which satisfy shift and scale invariance as well as symmetry also satisfy anonymity:

Lemma 3.5. *If a mechanism f is shift invariant, scale invariant and symmetric, it is also anonymous.*

Proof. Again, all equalities are in distribution. Let $\mathbf{x}$ be a location profile. We need to prove $Y_{x_1,x_2} = Y_{x_2,x_1}$. Shift and scale invariance gives $Y_{x_2,x_1} = -Y_{x_1,x_2} + x_1 + x_2$; thus, $\mathbb{P}(Y_{x_2,x_1} \leq b) = \mathbb{P}(x_1 + x_2 - b \leq Y_{x_1,x_2})$. But $\mathbb{P}(x_1 + x_2 - b \leq Y_{x_1,x_2}) = \mathbb{P}(Y_{x_1,x_2} \leq b)$ by symmetry about $m_{\mathbf{x}}$, thus $Y_{x_2,x_1} = Y_{x_1,x_2}$. □

The next lemma deals with an equivalent condition for strategyproofness for symmetric, shift and scale invariant mechanisms.

Lemma 3.6. *A symmetric, shift and scale invariant mechanism f is strategyproof if and only if for any profile $\mathbf{x} \in \mathbb{R}^2$ with $x_1 = 0 < x_2$, the following conditions hold:*

1. $-\int_{(-\infty,x_2)} ydF(y) + \int_{(x_2,\infty)} ydF(y) + x_2\mathbb{P}(Y = x_2) \geq 0$
2. $\int_{(-\infty,x_2)} ydF(y) - \int_{(x_2,\infty)} ydF(y) + x_2\mathbb{P}(Y = x_2) \geq 0$

where $Y \sim f(\mathbf{x})$ with c.d.f. F.

Proof. The proof is in the Appendix B.2. □

Given a strategyproof, shift invariant, scale invariant and symmetric mechanism, the upcoming results demonstrate how to find another strategyproof, shift invariant, scale invariant and symmetric mechanism that restricts the probability assignment to x_1, x_2, and $m_{\mathbf{x}}$ for every profile $\mathbf{x}$ and simultaneously gives a weakly better approximation than the original mechanism.

Lemma 3.7. *Let f be a strategyproof, shift invariant, scale invariant and symmetric mechanism. There exists another strategyproof, shift invariant, scale invariant and symmetric mechanism g with a weakly smaller expected social cost on every profile, such that at least one of the following two properties holds:*

(1) For every two-agent profile $\mathbf{x}$, $\mathbb{P}(g(\mathbf{x}) \in (x_1, x_2)) = 0$. (Doesn't utilize interior)[16]

[16]Note that it is possible for such a mechanism to still be ex-post Pareto efficient, if $\mathbb{P}(g(\mathbf{x}) \in \{x_1, x_2\}) = 1$.

(2) For every two-agent profile $\mathbf{x}$, $\mathbb{P}(g(\mathbf{x}) \in (-\infty, x_1) \cup (x_2, \infty)) = 0$. *(Ex-post Pareto efficiency)*

Proof. The proof is in the appendix B.2. □

Lemma 3.8. *Let f be a strategyproof, shift invariant, scale invariant, symmetric mechanism. Assume that f is either ex-post Pareto efficient or doesn't utilize interior. Then there exists another strategyproof mechanism g with a weakly smaller expected social cost on every profile, such that* $\mathbb{P}(g(\mathbf{x}) \in \{x_1, x_2, m_{\mathbf{x}}\}) = 1$ *for every location profile* $\mathbf{x}$. *Furthermore, g satisfies shift invariance, scale invariance and symmetry.*

Proof. We break the proof into two cases.

1. Assume f is ex-post Pareto efficient. Let g be the mechanism that satisfies $\mathbb{P}(g(\mathbf{x}) = x_1) = \mathbb{P}(f(\mathbf{x}) = x_1)$, $\mathbb{P}(g(\mathbf{x}) = x_2) = \mathbb{P}(f(x) = x_2)$, $\mathbb{P}(g(\mathbf{x}) = m_{\mathbf{x}}) = 1 - \mathbb{P}(g(\mathbf{x}) = x_1) - \mathbb{P}(g(\mathbf{x}) = x_2)$. Note that since $m_{\mathbf{x}}$ minimizes the social cost function for the profile $\mathbf{x}$, g certainly provides a weakly better approximation ratio than f. Furthermore, symmetry, shift and scale invariance are preserved.

   Let us prove that condition 1 in Lemma 3.6 holds for g; the proof for condition 2 is similar. Since f is a strategyproof mechanism, the condition implies that for any profile $\mathbf{x} = (x_1, x_2)$ with $x_1 = 0 < x_2$,

$$\begin{aligned} 0 &\leq -\int_{[0,x_2)} y dF(y) + x_2 \mathbb{P}(f(\mathbf{x}) = x_2) \\ &= -\int_{(0,x_2)} y dF(y) + x_2 \mathbb{P}(f(\mathbf{x}) = x_2) \\ &= -\mathbb{E}[f(\mathbf{x}) \mathbb{1}\big(f(\mathbf{x}) \in (x_1, x_2)\big)] + x_2 \mathbb{P}(f(\mathbf{x}) = x_2) \\ &= -m_{\mathbf{x}} \mathbb{P}(f(\mathbf{x}) \in (x_1, x_2)) + x_2 \mathbb{P}(f(\mathbf{x}) = x_2) \\ &= -m_{\mathbf{x}} \mathbb{P}(g(\mathbf{x}) = m_{\mathbf{x}}) + x_2 \mathbb{P}(g(\mathbf{x}) = x_2) \\ &= -\int_{(0,x_2)} y dG(y) + x_2 \mathbb{P}(g(\mathbf{x}) = x_2). \end{aligned}$$

   The third equality holds because the distribution is symmetric around $m_{\mathbf{x}}$. Hence, the condition is satisfied for the mechanism g.

2. Assume f doesn't utilize interior. Let g be the mechanism which, for every profile $\mathbf{x}$, locates $\mathbb{P}(g(\mathbf{x}) = x_1) = \mathbb{P}(g(\mathbf{x}) = x_2) = 0.5$, which is clearly strategyproof, shift invariant, scale

invariant, and symmetric. $sc(\mathbf{x}, x_2)$ minimizes $sc(\mathbf{x}, y)$ among $y \geq x_2$ and $sc(\mathbf{x}, x_1)$ minimizes $sc(\mathbf{x}, y)$ among $y \leq x_1$. Hence, $\mathbb{E}[sc(\mathbf{x}, g(\mathbf{x}))] \leq \mathbb{E}[sc(\mathbf{x}, f(\mathbf{x}))]$.

□

Now we are ready to prove the main theorem.

Theorem 3.4. *The LRM mechanism gives the best approximation ratio among all strategyproof mechanisms that are shift and scale invariant.*

Proof. By the previous lemma, it suffices to search among the class of strategyproof shift invariant, scale invariant and symmetric mechanisms where any element f of the class satisfies the property that $\mathbb{P}(f(\mathbf{x}) \in \{x_1, x_2, m_{\mathbf{x}}\}) = 1$ for every location profile $\mathbf{x}$. Clearly, for such mechanisms, the approximation ratio increases as $\mathbb{P}(f(\mathbf{x}) \in \{x_1, x_2\})$ increases. Assume $\mathbb{P}(f(\mathbf{x}) \in \{x_1, x_2\}) < 0.5$. Then $\mathbb{P}(f(\mathbf{x}) = m_{\mathbf{x}}) > 0.5$, and by symmetry, $\mathbb{P}(f(\mathbf{x}) = x_2) < 0.25$. But this gives, when $x_1 = 0$ and $x_2 > 0$, that $-m_{\mathbf{x}}\mathbb{P}(f(\mathbf{x}) = m_{\mathbf{x}}) + x_2\mathbb{P}(f(\mathbf{x}) = x_2) = -\frac{x_2}{2}\mathbb{P}(f(\mathbf{x}) = m_{\mathbf{x}}) + x_2\mathbb{P}(f(\mathbf{x}) = x_2) < 0$, violating strategyproofness by Lemma 3.6. Thus we must have that $\mathbb{P}(f(\mathbf{x}) \in \{x_1, x_2\}) \geq 0.5$, which implies that among all such mechanisms, LRM provides the best approximation ratio of $0.5(2^{1-\frac{1}{p}} + 1)$. □

An immediate consequence of Theorem 3.4 is the following corollary.

Corollary 3.1. *Any strategyproof shift and scale invariant mechanism has an approximation of at least* $0.5(2^{1-\frac{1}{p}} + 1)$ *in the worst case.*

Chapter 4

Obnoxious Facility Location

4.1 Introduction

In this chapter we consider a problem closely related to the one in Chapter 3. Unlike in Chapter 3, we consider an *obnoxious* facility, such as a landfill: agents prefer the facility to be as far away from them as possible, and hence gain a utility equal to their distance from the facility rather than pay a cost. To emphasize this change, we replace the notation for cost C with u: $u(x_i, y) = |x_i - y|$. This change warrants an additional one: replacing $\mathbb{R}$ with a closed interval, which without loss of generality we take to be $I = [0, 2]$. That is, the agents' locations are in I, and a mechanism f has domain I^n and range I (or distributions over I for randomized mechanisms). This is because any reasonable aggregate measure of the utilities increases as the facility moves farther away from all of the agents; thus, if the interval is not bounded, there is no optimal solution to the problem. The definition of strategyproofness (Definition 3.1) is now modified as follows:

Definition 4.1. *A mechanism f is strategyproof if for each $i \in N$, each $x_i, x_i' \in I$, and for each $\mathbf{x_{-i}} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots x_n) \in I^{n-1}$, $u(x_i, f(x_i, \mathbf{x_{-i}})) \geq u(x_i, f(x_i', \mathbf{x_{-i}}))$.*

The social planner in this case tries to *maximize* some aggregate measure of the utilities, which we call the social benefit function sb. Thus, the worst-case approximation ratio of a mechanism f is $\sup_{\mathbf{x}}\{sb(\mathbf{x}, OPT(\mathbf{x}))/sb(\mathbf{x}, f(\mathbf{x}))\}$. We consider two possible objective functions for the planner in this setting: $sb(\mathbf{x}, y) = \sum_{i \in N} u(x_i, y)$ (maxisum), or $sb(\mathbf{x}, y) = \min_{i \in N} u(x_i, y)$ (egalitarian). For a probability distribution π over I, we define $sb(\mathbf{x}, \pi) = \mathbb{E}_{y \sim \pi}[sb(\mathbf{x}, y)]$; for notational convenience, when f is randomized and there is no confusion, we use $f(\mathbf{x})$ to refer both to the probability distribution and to the associated random variable (which has the same distribution). Barring the changes described here, the rest of the notation remains identical to Section 3.2. The variant we discuss was originally introduced (with the maxisum objective) in [Cheng *et al.*, 2013b], and further explored in [Ibara and Nagamochi, 2012].

The rest of the chapter is as follows: in Section 4.2, we characterize all deterministic strategyproof mechanisms. In Section 4.3 we use the characterization to provide a lower bound of 3 and ∞ on the approximation ratio of deterministic strategyproof mechanisms for the maxisum and egalitarian objectives, respecitvely, matching the best attainable approximation ratio by a deterministic strategyproof mechanism in both cases. In Section 4.4 we provide lower bounds of $\frac{2}{\sqrt{3}}$ and $\frac{3}{2}$ on the approximation ratio of randomized strategyproof mechanisms for the maxisum and egalitarian objectives, respectively. In Section 4.5 we consider a generalized model that allows an agent to control more than one location. In this model, we provide a 3- and $\frac{3}{2}$-approximate strategyproof mechanisms for the maxisum objective in the deterministic and randomized settings respectively (in the randomized setting, we actually provide a *family* of such mechanisms).

4.2 Characterization of Deterministic Mechanisms

In this section, we characterize all deterministic strategyproof mechanisms for the obnoxious facility model. Similar results have been independently obtained by others [Han and Du, 2012; Barber *et al.*, 2012; Manjunath, 2014]. We begin with a temporary, somewhat weak characterization of deterministic mechanisms, in terms of single agent deviations:

Theorem 4.1. *[Reflection Theorem] For any deterministic mechanism f, agent $i \in N$, and partial location profile $\mathbf{x_{-i}}$, define $f_{\mathbf{x_{-i}}}(a) = f(a, \mathbf{x_{-i}})$.*[1] *Then, the mechanism f is strategyproof iff each $f_{\mathbf{x_{-i}}}$ is of the following form: there exists (not necessarily distinct) $\alpha_{\mathbf{x_{-i}}}, \beta_{\mathbf{x_{-i}}} \in I$, such that $\beta_{\mathbf{x_{-i}}} \geq \alpha_{\mathbf{x_{-i}}}$ and:*

1. $f_{\mathbf{x_{-i}}}(a) = \beta_{\mathbf{x_{-i}}}$ *for* $0 \leq a < \frac{\alpha_{\mathbf{x_{-i}}}+\beta_{\mathbf{x_{-i}}}}{2}$
2. $f_{\mathbf{x_{-i}}}(a) = \alpha_{\mathbf{x_{-i}}}$ *for* $\frac{\alpha_{\mathbf{x_{-i}}}+\beta_{\mathbf{x_{-i}}}}{2} < a \leq 2$

[1] Note that when $i \neq j$, $\mathbf{x_{-i}}$ and $\mathbf{x_{-j}}$ are distinct objects, regardless of the values of their coordinates.

3. $f_{\mathbf{x_{-i}}}(\frac{\alpha_{\mathbf{x_{-i}}}+\beta_{\mathbf{x_{-i}}}}{2}) \in \{\alpha_{\mathbf{x_{-i}}}, \beta_{\mathbf{x_{-i}}}\}$

If $\alpha_{\mathbf{x_{-i}}} \neq \beta_{\mathbf{x_{-i}}}$*, we call* $\frac{\alpha_{\mathbf{x_{-i}}}+\beta_{\mathbf{x_{-i}}}}{2}$ *the reflection point of* i *for the partial profile* $\mathbf{x_{-i}}$*.*

Proof. First, assume that f is of the form described above. On a partial location profile $\mathbf{x_{-i}}$, agent i can only get the mechanism to choose one of (up to) two locations: $\alpha_{\mathbf{x_{-i}}}$ or $\beta_{\mathbf{x_{-i}}}$. Let $q = \frac{\alpha_{\mathbf{x_{-i}}}+\beta_{\mathbf{x_{-i}}}}{2}$. If $x_i = q$, his distance from the two locations is equal, and so he is indifferent between them. If $x_i \in [0, q)$, $\beta_{\mathbf{x_{-i}}}$ is weakly farther from him than $\alpha_{\mathbf{x_{-i}}}$, and so he weakly prefers $\beta_{\mathbf{x_{-i}}}$, which is what the mechanism chooses, so he has no incentive to deviate.[2] The case of $x_i \in (q, 2]$ is similar. Thus, f is strategyproof.

On the other hand, assume that f is strategyproof. Fix a location profile $\mathbf{x}$ and an agent i. Let $g = f_{\mathbf{x_{-i}}}$ and let $\beta = g(0)$. Let $S = \{a \in I : g(a) \neq \beta\}$. If S is empty, then g is constant, and we're done. So assume S is nonempty. Consider $m = \inf S$. Note that if $\beta < m$, then an agent located at β can benefit from a deviation to any point in S. Thus, $\beta \geq m$. Let $\alpha = 2m - \beta$ (note that with our knowledge at this point, it might be the case that α is negative and hence not in I; our proof is careful not to assume otherwise). We begin by claiming that either $g(m) = \alpha$, or that m is a limit point of the set $K = \{a \in I : g(a) = \alpha\}$. There are two cases to consider:

1. $m \in S$. Note that in this case $m > 0$. We claim that in this case $g(m) = \alpha$. Assume otherwise, namely $g(m) = \alpha' \neq \alpha$. Note also that $\alpha' \neq \beta$ (since $m \in S$), and thus $m - \alpha \neq |m - \alpha'|$. There are two subcases:

   (a) $m - \alpha < |m - \alpha'|$. In that case, note that as long as the agent is to the left of m, the facility is located at β. Thus, if the agent is located at $m - \epsilon$ for some $\epsilon > 0$, his distance from the facility is $\beta - m + \epsilon$. His distance from α' is at least $|m - \alpha'| - \epsilon$. However, as $\beta - m = m - \alpha < |m - \alpha'|$, we may choose ϵ small enough so that $|m - \alpha'| - \epsilon > \beta - m + \epsilon$. In this case, the agent's deviation from $m - \epsilon$ to m is beneficial to that agent.

   (b) $m - \alpha > |m - \alpha'|$. In this case, it is still true that as long as the agent is to the left of m, the facility is located at β. As the distance of $\beta - m = m - \alpha > |m - \alpha'|$, it follows that an agent located at m will benefit from deviating to the left.

   So indeed, $g(m) = \alpha$.

[2]Weakly because it is possible that $\alpha_{\mathbf{x_{-i}}} = \beta_{\mathbf{x_{-i}}}$.

2. $m \notin S$. Then m is a limit point of S (by definition). In this case we claim that m is a limit point of K. Furthermore, note that $\beta > m$, since if $\beta = m$, then as $m \notin S$, $g(\beta) = \beta$, and the agent can benefit by deviating from β to any point in S. We note that since when the agent is located at m, the facility is located at β, strategyproofness dictates that the facility is always located in $[\alpha, \beta]$, no matter where the agent reports his location to be. Now, assume m is not a limit point of K. Thus, it follows it must be a limit point of either $K_1 = \{a \in I : \alpha < g(a) \leq m\}$ or $K_2 = \{a \in I : m \leq g(a) < \beta\}$ [3]. So, there are two cases:

   (a) m is a limit point of K_1. In particular, there exists some $\epsilon > 0$ such that $m + \epsilon \in K_1$. We consider the following subcases:

      i. There exists $0 < \epsilon' < \epsilon$ s.t. $m + \epsilon' \in K_1$ and $g(m + \epsilon') < g(m + \epsilon)$; in this case, a deviation from $m + \epsilon$ to $m + \epsilon'$ is beneficial.

      ii. There exists $0 < \epsilon' < \epsilon$ s.t. $m + \epsilon' \in K_1$ and $g(m + \epsilon') > g(m + \epsilon)$; in this case, a deviation from $m + \epsilon'$ to $m + \epsilon$ is beneficial.

      iii. $g(m + \epsilon') = g(m + \epsilon)$ for all $0 < \epsilon' < \epsilon$ s.t. $m + \epsilon' \in K_1$. As m is a limit point of K_1 and all points in K_1 are to the right of m, it follows that we may choose ϵ' as small as we want. For ϵ' small enough, this would imply that the deviation from $m + \epsilon'$ to m is beneficial: the distance of $m + \epsilon'$ from $g(m + \epsilon)$ is $m - g(m + \epsilon) + \epsilon'$ and the distance of $m + \epsilon'$ from β is $\beta - m - \epsilon'$. As $\beta - m > m - g(m + \epsilon)$, we may choose ϵ' small enough to make the deviation in question beneficial.

   (b) m is a limit point of K_2. So, there exists $0 < \epsilon < \frac{\beta - m}{2}$ such that $m + \epsilon \in K_2$. The agent can benefit by deviating from $m + \epsilon$ to m (since the facility will be sent from $g(m + \epsilon)$ to β, and since $m \leq g(m + \epsilon) < \beta$, $g(m + \epsilon)$ is closer to $m + \epsilon$ than β).

   Hence, by strategyproofness, we have reached a contradiction, and so m must be a limit point of K.

We have shown that if $m \in S$ then $g(m) = \alpha$, and otherwise by definition of S $g(m) = \beta$. To complete the proof, we must show that $g(a) = \alpha$ for all $a > m$. Assume otherwise for some $a' > m$.

[3] Note that since $g(a) \in [\alpha, \beta]$ for all $a \in I$, and m is a limit point of S, it follows that m is a limit point of $\{a \in I : a \in [\alpha, \beta)\} = K \cup K_1 \cup K_2$.

First, note that since $g(m)$ is either α or β, $g(a) \in [\alpha, \beta]$ for all $a \in I$ by strategyproofness. Note that within this range, α is the point farthest from a'. Thus, the agent has an incentive to deviate from a' to any point a'' for which $g(a'') = \alpha$, where the existance of such a point is guaranteed by the above discussion. This is a contradiction, and so we've completed our proof.[4] □

As a corollary of the above theorem, we can deduce:

Corollary 4.1. *For any deterministic strategyproof mechanism f, and any $n \in \mathbb{N}$, $R_n^f = \{f_n(\mathbf{x}) : \mathbf{x} \in I^n\}$ is finite.*

Proof. Let $\mathbf{x}$ be an arbitrary profile, and set $\mathbf{x}^0 = \mathbf{x}$. For a given profile $\mathbf{x}^{i-1}$, consider the profiles $\mathbf{z} = (0, \mathbf{x}_{-\mathbf{i}}^{i-1})$ and $\mathbf{z}' = (2, \mathbf{x}_{-\mathbf{i}}^{i-1})$. By the reflection theorem, at least one of $f(\mathbf{z}) = f(\mathbf{x}^{i-1})$ or $f(\mathbf{z}') = f(\mathbf{x}^{i-1})$ is true. This is trivial if agent i has no reflection point at $\mathbf{x}_{-\mathbf{i}}^{i-1}$. Otherwise, if he has such a reflection point m, if $x_i^{i-1} > m$ or $x_i^{i-1} < m$, he may deviate to 2 or 0 respectively without changing the facility's location; if $x_i^{i-1} = m$, then still the reflection theorem gives that $f(m, \mathbf{x}_{-\mathbf{i}}^{i-1})$ equals one of $f(\mathbf{z})$ or $f(\mathbf{z}')$, as required [5]. Set $\mathbf{x}^i$ equal to a profile among $\mathbf{z}$ and $\mathbf{z}'$ satisfying the equality. Thus, $\mathbf{x}^n$ is a profile in which all agents are located at the endpoints and $f(\mathbf{x}^n) = f(\mathbf{x})$. Since $\mathbf{x}$ was arbitrary, we have that all elements of R_n^f can be obtained by applying the mechanism to profiles locating all agents at the endpoints. Since there are only finitely many such profiles, R_n^f is finite. □

Now it is time for our strong characterization result. Consider the following definition:

Definition 4.2. *Let f be a deterministic mechanism s.t. $|R_n^f| \leq 2$ for all $n \in \mathbb{N}$. For each $n \in \mathbb{N}$, let $R_n^f = \{\alpha_n, \beta_n\}$ s.t. $\beta_n \geq \alpha_n$, and let $m_n = \frac{\alpha_n + \beta_n}{2}$.*[6] *For any $n \in \mathbb{N}$, for every profile $\mathbf{x} \in I^n$, consider the partition of the agents $L^{\mathbf{x}} = \{i \in N : x_i < m_n\}$, $M^{\mathbf{x}} = \{i \in N : x_i = m_n\}$, and $E^{\mathbf{x}} = \{i \in N : x_i > m_n\}$. We say that f is a midpoint mechanism if it satisfies the following property: for any $n \in \mathbb{N}$, let $\mathbf{x}, \mathbf{y} \in I^n$ be any profiles s.t. $f(\mathbf{x}) = \beta_n$ and $f(\mathbf{y}) = \alpha_n$. If $\beta_n > \alpha_n$, then there exists an agent i which satisfies one of the following:*

---

[4] Note that we didn't actually need that m is a limit point of K in the second case, but merely that K is nonempty. However, this doesn't seem to lead to a much simpler proof, so we stick with the more general argument.

[5] $f(\mathbf{z})$ if $f(m, \mathbf{x}_{-\mathbf{i}}^{i-1}) = \beta_{\mathbf{x}_{-\mathbf{i}}}$, and $f(\mathbf{z}')$ if $f(m, \mathbf{x}_{-\mathbf{i}}^{i-1}) = \alpha_{\mathbf{x}_{-\mathbf{i}}}$.

[6] $\alpha_n = \beta_n$ is possible.

(D-1) $i \in L^{\mathbf{x}}$ *and* $i \in M^{\mathbf{y}}$

(D-2) $i \in L^{\mathbf{x}}$ *and* $i \in E^{\mathbf{y}}$

(D-3) $i \in M^{\mathbf{x}}$ *and* $i \in E^{\mathbf{y}}$

This definition is simple to interpret: the mechanism can switch the facility location from right to left or from left to right only when an agent crosses the midpoint in the opposite direction.

In [Ibara and Nagamochi, 2012], the authors show that for a strategyproof mechanism f, $|R_n^f| \leq 2$ whenever R_n^f is a finite set.[7] Using that, we can now show:

Theorem 4.2. *A deterministic mechanism f is strategyproof iff it is a midpoint mechanism.*

Proof. First, consider a given midpoint mechanism f, and fix $n \in \mathbb{N}$. If f_n is constant, then clearly it is strategyproof. Otherwise, $|R_n^f| = 2$. Consider a profile $\mathbf{x} \in I^n$ and an agent $i \in N$. The facility can only be located at α_n or β_n. If $i \in M_n^{\mathbf{x}}$, he is indifferent between the two points, and thus has no incentive to deviate. If $i \in E_n^{\mathbf{x}}$, he prefers the facility to be located at α_n; however, if $f(\mathbf{x}) \neq \alpha_n$, note that agent i cannot move the facility to α_n by deviating- the rest of the agents remain still, and he himself cannot be the agent required in the definition of the midpoint property (that agent cannot be in $E_n^{\mathbf{x}}$). The proof is similar for $i \in L_n^{\mathbf{x}}$.

Now, assume instead that f is a strategyproof mechanism. Fix $n \in \mathbb{N}$. By Corollary 4.1, R_n^f is finite, and thus by Ibara's and Nagamochi's result, $|R_n^f| \leq 2$. If R_n^f is a singleton there is nothing to prove; thus, assume $|R_n^f| = 2$, and let $\alpha_n, \beta_n \in R_n^f$ s.t. $\beta_n > \alpha_n$. Let $\mathbf{x}, \mathbf{y} \in I^n$ s.t. $f(\mathbf{x}) = \beta_n$ and $f(\mathbf{y}) = \alpha_n$. Consider the sequence of profiles $\mathbf{z}^i$, defined for $i = 0, \ldots, n$ via $\mathbf{z}_j^i = \mathbf{x}_j$ if $j > i$ and $\mathbf{z}_j^i = \mathbf{y}_j$ otherwise. Assume no agent satisfies at least one of (D-1), (D-2) and (D-3). Then, when agent i deviates in $\mathbf{z}^{i-1}$ to create profile $\mathbf{z}^i$, he does not cross m_n from left to right (i.e. moving from $\mathbf{z}_i^{i-1} < m_n$ to $\mathbf{z}_i^i \geq m_n$ or $\mathbf{z}_i^{i-1} \leq m_n$ to $\mathbf{z}_i^i > m_n$). As the possible facility locations are α_n and β_n, m_n is his only candidate for reflection point in $\mathbf{z}_{-\mathbf{i}}^{i-1}$. Thus, the reflection theorem implies that he cannot change the facility location to α_n by deviating. Hence, $f(\mathbf{y}) = f(\mathbf{z}^n) = f(\mathbf{z}^0) = f(\mathbf{x})$, contradiction. □

[7] While they assume anonymity, the proof of this fact does not rely on that assumption. Also, note that this is not an immediate implication of the Gibbard-Satterthwaite theorem, since the agents, by reporting a location, cannot arbitrarily "rank" the locations in R_n^f; only some rankings are feasible.

We note that Ibara and Nagmochi have characterized all anonymous mechanisms under the assumption that R_n^f is finite for all $n \in \mathbb{N}$, using what they called "valid threshold mechanisms". Our proofs easily translate to the anonymous case, and under anonymity, our midpoint mechanisms become equivalent to valid threshold mechanisms. Thus, our work allows the removal of the finite R_n^f assumption for the anonymous case as well.

4.3 Lower Bounds on Deterministic Mechanisms

We can use our characterization to obtain lower bounds on the possible approximation ratios for the maxisum and egalitarian objectives in the deterministic setting.

Theorem 4.3. *No deterministic strategyproof mechanism f can provide an approximation ratio better than 3 for the maxisum objective.*

Proof. Let f be a deterministic strategyproof mechanism. Let $n \in \mathbb{N}$ be even. If f_n is constant, the approximation ratio is clearly unbounded. If R_n^f is not a singleton, then by Theorem 4.2, $|R_n^f| = 2$. Consider the profile $\mathbf{x} \in I^n$ which locates agents 1 through $\frac{n}{2}$ at α_n, agents $\frac{n}{2}+1$ through n at β_n (where $\alpha_n < \beta_n$ are as in the definition of midpoint mechanism). Assume without loss of generality that $f(\mathbf{x}) \neq \alpha_n$. Consider the profile $\mathbf{y}$ which locates agents 1 through $\frac{n}{2}$ at $m_n - \epsilon$ for some $\epsilon > 0$, and agrees with $\mathbf{x}$ on the rest of the agents. Since no deviating agent reaches m_n, the facility location doesn't change, that is $f(\mathbf{y}) = f(\mathbf{x}) = \beta_n$. Locating the facility at β_n (on profile $\mathbf{y}$) leads to a benefit of $\frac{n}{2} \cdot (\frac{\beta_n - \alpha_n}{2} + \epsilon)$, while locating the facility at α_n leads to a benefit of $\frac{n}{2} \cdot (\frac{\beta_n - \alpha_n}{2} - \epsilon) + \frac{n}{2}(\beta_n - \alpha_n)$, and sending $\epsilon \to 0$ gives us the required result.[8] □

By Theorem 4.7, our lower bound is best-possible. Our characterization can also be used to get a lower bound for the egalitarian objective:

Theorem 4.4. *No deterministic strategyproof mechanism f can provide a bounded approximation ratio for the egalitarian objective.*

Proof. For any $n \geq 2$, $|R_n^f| \leq 2$ by Theorem 4.2. Consider any profile which locates at least one agent at each point in R_n^f; any such profile leads to a social benefit of 0 for the mechanism, whereas the optimal benefit is positive. □

[8] If n is odd, we could still make this proof work by locating the additional agent at m_n and send $n \to \infty$.

4.4 Lower Bounds on Randomized Mechanisms

We begin with the maxisum objective. We provide a lower bound of $\frac{2}{\sqrt{3}}$ on the approximation ratio of randomized strategyproof mechanisms.

Theorem 4.5. *No randomized strategyproof mechanism can provide an approximation ratio better than $\frac{2}{\sqrt{3}}$ for the maxisum objective.*

Proof. Let f be a randomized strategyproof mechanism which provides an approximation ratio $c < \frac{2}{\sqrt{3}}$ for the maxisum objective. Consider the case where $N = \{1, 2\}$, and let $a = 2\sqrt{3} - 3$. Let $\mathbf{x}$ be the location profile in which $x_1 = 1 - a$ and $x_2 = 1 + a$. Assume without loss of generality that $P(f(\mathbf{x}) < x_1) \geq P(f(\mathbf{x}) > x_2)$. The expected distance of the facility from x_1 on this profile is at most $(1 - a)P(f(\mathbf{x}) < x_1) + (1 + a)P(f(\mathbf{x}) > x_2) + 2aP(x_1 \leq f(\mathbf{x}) \leq x_2)$; as $P(f(\mathbf{x}) < x_1) \geq P(f(\mathbf{x}) > x_2)$ and $2a \leq 1$, this implies that the expected distance of the facility from x_1 on profile $\mathbf{x}$ is at most 1.

Let $\mathbf{y}$ be the profile in which $y_1 = 0$ and $y_2 = 1 + a$. Let $b = E[f(\mathbf{y})|f(\mathbf{y}) > y_2] - y_2$, and let $p = P(f(\mathbf{x}) > y_2)$. The mechanism's expected benefit is $1 + a + 2bp$, while the optimal cost is $3 - a$. To maintain approximation ratio of c, we must have $1 + a + 2bp \geq \frac{3-a}{c}$, which implies $bp \geq \frac{3-a}{2c} - \frac{1+a}{2}$. Also, as $b \leq 1 - a$, we have that $p \geq \frac{1}{1-a}(\frac{3-a}{2c} - \frac{1+a}{2})$. Now, the expected distance of the facility from x_1 on $\mathbf{y}$ is $(2a + b)p \geq (\frac{2a}{1-a} + 1)(\frac{3-a}{2c} - \frac{1+a}{2}) > (\frac{2a}{1-a} + 1)(\sqrt{3}\frac{3-a}{4} - \frac{1+a}{2}) = 1$. This violates strategyproofness, as agent 1 has an incentive to misreport his location to be 0 when the location profile is $\mathbf{x}$. □

Next, we show a lower bound of $\frac{3}{2}$ for the egalitarian objective:

Theorem 4.6. *No randomized strategyproof mechanism can provide an approximation ratio better than $\frac{3}{2}$ for the egalitarian objective.*

Proof. Let f be such a mechanism, with approximation ratio $c < \frac{3}{2}$. Let the endpoints be 0 and $M + 2$, where M is some large number. Consider the case with $n = 2\lceil\frac{M+1}{\epsilon}\rceil + 4$ agents, where $1 > \epsilon > 0$. Consider the profile $\mathbf{x}$ which locates one agent at 1 and $M + 1$, locates an agent in each $1 + a\epsilon \in (1, M + 1)$ s.t. $a \in \mathbb{N}$, and splits the rest of the agents evenly among the two endpoints (if there is an odd number of agents remaining, locate one agent at $\frac{M+2}{2}$). Note that an optimal facility location is at $M + 1\frac{1}{2}$, with a benefit of $\frac{1}{2}$. Let p be the probability that the facility is located at $[1, M+1]$. It follows that the resulting expected benefit is upper bounded by $\frac{\epsilon}{2}p + \frac{1}{2}(1-p)$. To get the required approximation ratio, we must have $\frac{\epsilon}{2}p + \frac{1}{2}(1-p) \geq \frac{1}{2c}$, which gives $p \leq \frac{\frac{1}{2}(1-\frac{1}{c})}{\frac{1}{2}-\frac{\epsilon}{2}}$. The facility must be located either in $[0,1]$ or in $[M+1, M+2]$ with probability at least $\frac{1-p}{2}$. Assume without loss of generality that it is located with probability at least $\frac{1-p}{2}$ in $[0,1]$. Consider the profile $\mathbf{x}'$, which is obtained from profile $\mathbf{x}$ by relocating the agents from 0 so that there is an agent in every point $a\epsilon \in [0,1)$ s.t. $a \in \mathbb{N}$. Let p' be the probability that on this profile, the facility is located at $[0, M+1]$. Note that the optimal facility location remains $M + 1\frac{1}{2}$ with benefit $\frac{1}{2}$, and on the other hand the expected benefit on this profile is bounded by $\frac{\epsilon}{2}p' + \frac{1}{2}(1-p')$, yielding the bound $p' \leq \frac{\frac{1}{2}(1-\frac{1}{c})}{\frac{1}{2}-\frac{\epsilon}{2}}$. Let us analyze the expected distance of the facility from 0 in the two profiles. For $\mathbf{x}$, the expected distance from 0 is no more than $\frac{1-p}{2} + p(M+1) + \frac{1-p}{2}(M+2)$. On the other hand, for $\mathbf{x}'$, the expected distance from 0 is no less than $(1-p')(M+1)$. Since we have obtained $\mathbf{x}'$ from $\mathbf{x}$ using deviations of agents from 0, strategyproofness dictates $\frac{1-p}{2} + p(M+1) + \frac{1-p}{2}(M+2) \geq (1-p')(M+1)$.[9] Reorganizing this, we get: $\frac{3-p}{2} + \frac{1+p}{2}M \geq (1-p')M + 1 - p'$. Using our bounds for p and p', this implies the inequality $\frac{1}{2} + (\frac{1}{2} + \frac{\frac{1}{2}(1-\frac{1}{c})}{1-\epsilon})M \geq (1 - \frac{\frac{1}{2}(1-\frac{1}{c})}{\frac{1}{2}-\frac{\epsilon}{2}})M - \frac{\frac{1}{2}(1-\frac{1}{c})}{\frac{1}{2}-\frac{\epsilon}{2}}$. Let us reorganize this inequality to $\frac{1}{2} + \frac{\frac{1}{2}(1-\frac{1}{c})}{\frac{1}{2}-\frac{\epsilon}{2}} \geq (\frac{1}{2} - \frac{\frac{3}{2}(1-\frac{1}{c})}{1-\epsilon})M$. As $c < \frac{3}{2}$, we can choose $\epsilon > 0$ small enough so that $\frac{1}{2} - \frac{\frac{3}{2}(1-\frac{1}{c})}{1-\epsilon} > 0$. Sending M to ∞ then causes the r.h.s of the inequality to go to ∞, violating the inequality. □

4.5 Mutiple Locations Per Agent in the Obnoxious Model

In this section we follow the spirit of a suggestion in [Procaccia and Tennenholtz, 2013a] and study a generalized model, in which a single agent may be associated with more than one location. As this multiple location model is a generalization of our previous model, the lower bounds carry over; in particular, for the maxisum objective, we have lower bounds of 3 and $\frac{2}{\sqrt{3}}$ on deterministic and randomized mechanisms respectively. In the deterministic case, we can find a strategyproof

[9] Consider the sequence of profiles $\mathbf{x^0}$ through $\mathbf{x^n}$, such that profile $\mathbf{x^i}$ agrees with $\mathbf{x}'$ on the location of agents 1 through i and with $\mathbf{x}$ on the location of the rest of the agents. Let i^* be the index that maximizes $E_{Y\sim f(\mathbf{x^i})}[Y]$; if there is more than one such index, choose the minimal one. If $i^* > 0$, then it is beneficial for agent i^* to deviate so that the profile changes from $\mathbf{x^{i^*-1}}$ to $\mathbf{x^{i^*}}$, violating strategyproofness. Thus $i^* = 0$, implying the expected distance of the facility from 0 in those profiles satisfies $E_{Y\sim f(\mathbf{x^0})}[Y] \geq E_{Y\sim f(\mathbf{x^n})}[Y]$. But, since $\mathbf{x^0} = \mathbf{x}$ and $\mathbf{x^n} = \mathbf{x}'$, we know that $\frac{1-p}{2} + p(M+1) + \frac{1-p}{2}(M+2) \geq E_{Y\sim f(\mathbf{x})}[Y]$ and $E_{Y\sim f(\mathbf{x}')}[Y] \geq (1-p')(M+1)$.

mechanism to match the lower bound, despite the additional power given to the agents. In addition, we provide a family of $\frac{3}{2}$-approximate randomized strategyproof mechanisms.

Our generalized model can be obtained from our previous model via the following changes. First, let $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{N}^n$. A location profile is now $\mathbf{z} = (\mathbf{z^1}, \mathbf{z^2}, \ldots, \mathbf{z^n})$, where for each $i = 1, ..., n$, $\mathbf{z^i} = (z^i_1, z^i_2, \ldots, z^i_{k_i}) \in I^{k_i}$. A deterministic mechanism is a collection of functions $f = \{f^{\mathbf{k}}_n : n \in \mathbb{N}, \mathbf{k} \in \mathbb{N}^n\}$, such that $f^{\mathbf{k}}_n : I^{k_1} \times \ldots \times I^{k_n} \to I$ is a function that maps each location profile to a facility location. The utility of agent i from facility location y is now defined as $u(\mathbf{z^i}, y) = \sum_{j=1}^{k_i} u(z^i_j, y)$, where $u(x, y)$ is $|x - y|$. The maxisum objective is $\sum_{i=1}^{n} u(\mathbf{z^i}, y)$ as usual. The rest of the notation carries over, and the adjustment to the randomized model is easy and left to the reader. For the approximation ratio, we note that the possible instances of the problem include all possible options for both n and $\mathbf{k}$.

First, we provide a 3-approximate strategyproof deterministic mechanism.

Theorem 4.7. *Let* $R^* = \{i : \frac{\sum_{j=1}^{k_i} z^i_j}{k_i} \leq 1\}$, $L^* = \{i : \frac{\sum_{j=1}^{k_i} z^i_j}{k_i} > 1\}$. *Let f be the mechanism which locates the facility at 2 if* $\sum_{i \in R^*} k_i \geq \sum_{i \in L^*} k_i$ *and at 0 otherwise. This mechanism is strategyproof and 3-approximate for maxisum.*

Proof. Strategyproofness is easy (note that R^* is exactly the set of agents weakly preferring the facility to be located at 2 over 0). The optimal facility location is clearly in $\{0, 2\}$. Assume without loss of generality that $f(\mathbf{z}) = 2$. All we have to prove is that $\frac{sb(\mathbf{z},0)}{sb(\mathbf{z},2)} \leq 3$. But note that every agent $i \in R^*$ receives a utility of at least k_i when the facility is located at 2 and at most k_i when the facility is located at 0. On the other hand, each agent $i \in L^*$ trivially gets a utility between 0 and $2k_i$. Thus, using the fact that $f(\mathbf{z}) = 2$ implies $\sum_{i \in R^*} k_i \geq \sum_{i \in L^*} k_i$, we get $\frac{sb(\mathbf{z},0)}{sb(\mathbf{z},2)} \leq \frac{2\sum_{i \in L^*} k_i + \sum_{i \in R^*} k_i}{\sum_{i \in R^*} k_i} \leq \frac{2\sum_{i \in R^*} k_i + \sum_{i \in R^*} k_i}{\sum_{i \in R^*} k_i} = 3$. □

Note that when $k_i = 1$ for all i, this mechanism reduces to the mechanism proposed in [Cheng *et al.*, 2013b].

Finally, we define a class of randomized strategyproof mechanisms that provide a $\frac{3}{2}$-approximation and show that it is nonempty.

Theorem 4.8. *Let f be a randomized mechanism that, for a profile* $\mathbf{z}$*, locates the facility at 0 with probability* $p_{\mathbf{z}}$ *and at 2 with probability* $(1 - p_{\mathbf{z}})$*. Then the following conditions on* $p_{\mathbf{z}}$ *are sufficient to make the mechanism strategyproof and* $\frac{3}{2}$*-approximate:*

1. *$p_{\mathbf{z}}$ is increasing in $\sum_{i\in L^*} k_i$ and decreasing in $\sum_{i\in R^*} k_i$.*

2. *$\frac{1}{3}+\frac{1}{6}\cdot\frac{\sum_{i\in L^*} k_i}{\sum_{i\in R^*} k_i} \geq p_{\mathbf{z}} \geq \frac{2}{3}-\frac{1}{6}\cdot\frac{\sum_{i\in R^*} k_i}{\sum_{i\in L^*} k_i}$ (if $\sum_{i\in R^*} k_i = 0$, the leftmost term is ∞; if $\sum_{i\in L^*} k_i = 0$, the rightmost term is $-\infty$).*

Furthermore, the class of mechanisms of this form is nonempty.

Proof. Strategyproofness is clear. Fix $\mathbf{z}$, and set $p = p_{\mathbf{z}}$. For the approximation ratio, there are two cases to consider. First, assume that the optimal facility location for profile $\mathbf{z}$ is 0, with social benefit OPT. If 0 and 2 are both optimal, clearly any choice of p yields approximation ratio 1. Assume 2 is not optimal; then $\sum_{i\in L^*} k_i > 0$. As for every $i \in R^*$ we have that $\frac{\sum_{j=1}^{k_i} z_j}{k_i} \leq 1$, his utility from locating the facility at 2 is at least k_i, and so the social benefit from locating the facility there is at least $\sum_{i\in R^*} k_i$. Thus, it is enough to prove that $pOPT + (1-p)\sum_{i\in R^*} k_i \geq \frac{2}{3}OPT$, or equivalently $(1-p)\sum_{i\in R^*} k_i \geq (\frac{2}{3}-p)OPT$. If the right hand side is negative, then this inequality is satisfied. Assume that the right hand side is nonnegative. Note that $OPT \leq 2\sum_{i\in L^*} k_i + \sum_{i\in R^*} k_i$ (the utility of $i \in R^*$ from locating the facility at 0 is bounded by k_i, while the utility of $i \in L^*$ is trivially bounded by $2k_i$). Thus, it is enough to prove that $(1-p)\sum_{i\in R^*} k_i \geq (\frac{2}{3}-p)(2\sum_{i\in L^*} k_i + \sum_{i\in R^*} k_i)$. Isolating p in this inequality gives $p \geq \frac{2}{3} - \frac{1}{6}\frac{\sum_{i\in R^*} k_i}{\sum_{i\in L^*} k_i}$, which is satisfied.

On the other hand, assume that the optimal facility location for profile $\mathbf{z}$ is 2; note that this implies $\sum_{i\in R^*} k_i > 0$. Similarly to the analysis above, we can get a lower bound of $\sum_{i\in L^*} k_i$ on the benefit of locating the facility at 0 and upper bound of $2\sum_{i\in R^*} k_i + \sum_{i\in L^*} k_i$ on OPT. Thus, we need that $(1-p)OPT + p\sum_{i\in L^*} k_i \geq \frac{2}{3}OPT$, and so it is enough to verify that $p\sum_{i\in L^*} k_i \geq (p-\frac{1}{3})(2\sum_{i\in R^*} k_i + \sum_{i\in L^*} k_i)$. Isolating p yields $p \leq \frac{1}{3} + \frac{1}{6}\frac{\sum_{i\in L^*} k_i}{\sum_{i\in R^*} k_i}$.

Finally, we verify that $p = \max\{\frac{2}{3} - \frac{1}{6}\cdot\frac{\sum_{i\in R^*} k_i}{\sum_{i\in L^*} k_i}, 0\}$ (where if $\sum_{i\in L^*} k_i = 0$, $p = 0$) satisfies the above properties. The only thing that requires proof is $p \leq \frac{1}{3}+\frac{1}{6}\cdot\frac{\sum_{i\in L^*} k_i}{\sum_{i\in R^*} k_i}$ (assuming $\sum_{i\in R^*} k_i > 0$; if $\sum_{i\in R^*} k_i = 0$ then there is nothing to prove). Note that the right hand side is positive, so it is enough to show is that $\frac{1}{3}+\frac{1}{6}\cdot\frac{\sum_{i\in L^*} k_i}{\sum_{i\in R^*} k_i} \geq \frac{2}{3}-\frac{1}{6}\cdot\frac{\sum_{i\in R^*} k_i}{\sum_{i\in L^*} k_i}$ when $\sum_{i\in L^*} k_i > 0$. But this is equivalent to $\frac{(\sum_{i\in L^*} k_i)^2+(\sum_{i\in R^*} k_i)^2}{(\sum_{i\in L^*} k_i)(\sum_{i\in R^*} k_i)} \geq 2$, and note that $\frac{(\sum_{i\in L^*} k_i)^2+(\sum_{i\in R^*} k_i)^2}{(\sum_{i\in L^*} k_i)(\sum_{i\in R^*} k_i)} = 2 + \frac{(\sum_{i\in L^*} k_i - \sum_{i\in R^*} k_i)^2}{(\sum_{i\in L^*} k_i)(\sum_{i\in R^*} k_i)} \geq 2$. □

It is worth noting that the randomized mechanism given in [Cheng *et al.*, 2013b], for the special case of $k_i = 1$ for all i, falls into the category of mechanisms we defined here.

Chapter 5

Hybrid Model of Facility Location

5.1 Introduction

In this short chapter, we consider a hybrid of the models discussed in Chapters 3 and 4. The model we discuss generalizes the one in Section 4.1 as follows. The space of locations is again a bounded interval, assumed to be $I = [0, 2]$ without loss of generality. In the hybrid model, there are two types of agents: type 1 agents, who wish for the for the facility to be located as far away from them as possible, and type 2 agents, who wish for the facility to be located as close to them as possible. We denote the type of agent i as θ_i, and the utility of a type θ agent located at x when the facility is located at y as $u(\theta, x, y)$. Like in Chapter 4, when $\theta = 1$, the utility is $u(1, x, y) = |x - y|$. When $\theta = 2$, one might expect the utility for a type 2 agent to be $-|x - y|$. However, negative utilities would lead us to an objective function (which would be maxisum in this chapter) that can take both positive and negative values; for such objective functions, the notion of approximation ratio is ill-defined. To fix this, we define $u(2, x, y) = 2 - |x - y|$, namely we add the length of I to the negative utility. In terms of incentives, this has no impact: the definition of strategyproofness is equivalent whether the length of I is added or not. However, adding this particular constant is the unique choice that simultaneously makes the utilities of both type 1 and 2 lie in the same range $[0, 2]$; thus, in a way, this utility definition does not discriminate between the types. Finally, for a probability distribution over π, we define $u(\theta, x, \pi) = \mathbb{E}_{y\sim\pi} u(\theta, x, y)$. If y is a random variable with distribution π, we make a slight abuse of notation and define $u(\theta, x, y) = u(\theta, x, \pi)$ when there is no risk of confusion.

We consider the type of each agent i to be private information of that agent, which needs to be reported to the planner (although all of our results continue to hold when it is publicly known). Let $T = \{1, 2\}$. A mechanism f is now a collection of functions $\{f_n : n \in \mathbb{N}\}$, where $f_n : T^n \times I^n \to I$ (in case of a randomized mechanism, the range is the space of all probability distributions over I). The definition of strategyproofness therefore becomes:

Definition 5.1. *A mechanism f is strategyproof if for each $i \in N$, each $x_i, x_i' \in I$, $\theta_i, \theta_i' \in T$, and for each $\mathbf{x_{-i}} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots x_n) \in I^{n-1}$,$\theta_{-\mathbf{i}} = (\theta_1, \theta_, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n) \in T^{n-1}$:*

$$u(\theta_i, x_i, f((\theta_i, \theta_{-\mathbf{i}}), (x_i, \mathbf{x_{-i}}))) \geq u(\theta_i, x_i, f((\theta_i', \theta_{-\mathbf{i}}), (x_i', \mathbf{x_{-i}})))$$

.

The objective function we are interested in throughout this chapter is maximizing the sum of utilities $sb(\theta, \mathbf{x}, y) = \sum_{i=1}^{n} u(\theta_i, x_i, y)$ (maxisum). The worst case approximation ratio of a mechanism f is $\sup_{\theta, \mathbf{x}} \frac{sb(\theta, \mathbf{x}, OPT(\theta, \mathbf{x}))}{sb(\theta, \mathbf{x}, f(\theta, \mathbf{x}))}$.

5.2 Deterministic and Randomized Mechanisms for the Hybrid Model

In this section, we design a best-possible 3-approximate deterministic strategyproof mechanism for the maxisum objective, as well as a $\frac{23}{13}$-approximate randomized strategyproof mechanism for the same objective.

Theorem 5.1. *Let $R = \{i : \theta_i = 1, x_i \leq 1\} \cup \{i : \theta_i = 2, x_i \geq 1\}$ and $L = \{i : \theta_i = 1, x_i > 1\} \cup \{i : \theta_i = 2, x_i < 1\}$. Let f be the mechanism that locates the facility at 2 if $|R| \geq |L|$ and at 0 otherwise. Then f is a 3-approximate strategyproof mechanism for the maxisum objective.*

Proof. Strategyproofness is easy. Note that R is the set of agents who weakly prefer the facility to be located at 2 over 0, and L is the set of remaining agents. Since there are only two possible facility locations in this mechanism, it is enough to rule out manipulations by an agent i who prefers the facility to be located at the endpoint *not* chosen by the mechanism. Assume without loss of generality that the facility is located at 2, yet agent i prefers the facility to be located at 0 (that is, $i \in L$). Then, by misreporting, he cannot decrease $|R|$ and cannot increase $|L|$, and therefore regardless of his report, the facility will be located at 2.

For the approximation ratio, let θ and $\mathbf{x}$ be a type and location profile respectively. We would like to show that $\frac{sb(\theta,\mathbf{x},a)}{sb(\theta,\mathbf{x},f(\theta,\mathbf{x}))} \leq 3$ for every possible facility location $a \in I$.[1] We will prove this for the case of $f(\theta,\mathbf{x}) = 2$; the other case is similar. Let R_j be the set of agents of type j in R, and similarly let L_j be the set of agents of type j in L. Define $q_1 := sb(\theta,\mathbf{x},a) = \sum_{i\in R_1\cup L_1} |x_i - a| + \sum_{i\in R_2\cup L_2}(2 - |x_i - a|)$, and $q_2 := sb(\theta,\mathbf{x},f(\theta,\mathbf{x})) = \sum_{i\in R_1\cup L_1}(2 - x_i) + \sum_{i\in R_2\cup L_2} x_i$. We need to show that $q_1 \leq 3q_2$. This is trivial when $q_1 < q_2$, so assume $q_1 \geq q_2$. Note that for $i \in R_1 \cup L_1$, increasing x_i by δ decreases q_2 by δ, and decreases q_1 by at most δ (might even increase q_1 in some cases). Similarly, for $i \in R_2 \cup L_2$, decreasing x_i has the same effect. Given that $x_i \leq 1$ for $i \in R_1$, $x_i \geq 1$ for $i \in R_2$, and $x_i \in [0,2]$ for all i, it is enough to prove $q_1 \leq 3q_2$ on the profile where $x_i = 1$ for $i \in R$, $x_i = 2$ for $i \in L_1$, and $x_i = 0$ for $i \in L_2$.[2] We shall assume this profile for the rest of the proof. On that profile, we have $q_1 = \sum_{i\in R_1} |a-1| + \sum_{i\in L_1\cup L_2}(2-a) + \sum_{i\in R_2}(2 - |a-1|)$ and $q_2 = |R|$. We break our proof into two cases:

1. $a \in [0,1]$: $\sum_{i\in R_1} |a-1| + \sum_{i\in L_1\cup L_2}(2-a) + \sum_{i\in R_2}(2-|a-1|) = |R_1| + 2|L_1| + 2|L_2| + |R_2| + a(-|R_1| - |L_1| - |L_2| + |R_2|) \leq \max\{|R| + 2|L|, |L| + 2|R_2|\}$, where the inequality follows from the fact that the maximum is obtained when $a \in \{0,1\}$ (If $-|R_1| - |L_1| - |L_2| + |R_2| \leq 0$, then it is obtained at $a = 0$, and otherwise at $a = 1$). Note that $|L| \leq |R|$ since $f(\mathbf{x}) = 2$, and that $|R_2| \leq |R|$ by definition. Thus, $\max\{|R| + 2|L|, |L| + 2|R_2|\} \leq 3|R|$. Therefore, $q_1 \leq 3|R| = 3q_2$

2. $a \in [1,2]$: $\sum_{i\in R_1} |a-1| + \sum_{i\in L_1\cup L_2}(2-a) + \sum_{i\in R_2}(2-|a-1|) = -|R_1| + 2|L_1| + 2|L_2| + 3|R_2| + a(|R_1| - |L_1| - |L_2| - |R_2|) \leq \max\{|L| + 2|R_2|, |R|\}$. Again, both terms we're maximizing over are no more than $3|R|$, and so again $q_1 \leq 3|R| = 3q_2$.

□

[1] Note that the statement "$\frac{sb(\theta,\mathbf{x},a)}{sb(\theta,\mathbf{x},f(\theta,\mathbf{x}))} \leq 3$ for every possible facility location $a \in I$" is equivalent to $\frac{sb(\theta,\mathbf{x},OPT(\theta,\mathbf{x}))}{sb(\theta,\mathbf{x},f(\theta,\mathbf{x}))} \leq 3$, as $OPT(\theta,\mathbf{x})$ maximizes the numerator by definition and hence the ratio. However, we choose to analyze an arbitrary fixed a rather than $OPT(\theta,\mathbf{x})$ to avoid having to consider the impact agents' reports have on the optimal location of the facility.

[2] Since, if $q_1 - \delta \leq 3(q_2 - \delta)$, then $q_1 + 2\delta \leq 3q_2$, hence $q_1 \leq 3q_2$.

Note that by Theorem 4.3 no deterministic strategyproof mechanism can do better than 3-approximation (as the obnoxious model is a special case of the hybrid model). Thus, the approximation ratio achieved by our mechanism is best-possible. Moreover, in the obnoxious facility model, the above mechanism reduces to the deterministic mechanism proposed in [Cheng *et al.*, 2013b], who proved that it is a 3-approximation for that special case.

We now use randomization in an attempt to lower the approximation ratio. Getting a 2-approximation is easy: choosing each endpoint with probability $\frac{1}{2}$ is a 2-approximate strategyproof mechanism.[3] However, we can do better:

Theorem 5.2. *Let* $p_1 = \frac{12}{23}$, $p_2 = \frac{8}{23}$, *and* $p_3 = \frac{3}{23}$. *Consider the following randomized mechanism* f. *If* $|R| \geq |L|$, *then* $P(f(\theta, \mathbf{x}) = 2) = p_1$ *and* $P(f(\theta, \mathbf{x}) = 0) = p_2$; *if* $|R| < |L|$, *then* $P(f(\theta, \mathbf{x}) = 2) = p_2$ *and* $P(f(\theta, \mathbf{x}) = 0) = p_1$; *and either way,* $P(f(\theta, \mathbf{x}) = 1) = p_3$. *The mechanism* f *is strategyproof and* $\frac{23}{13}$*-approximate.*

Proof. Strategyproofness is proved similarly to Theorem 5.1. For the approximation ratio, we would like to show that $\frac{sb(\theta,\mathbf{x},a)}{sb(\theta,\mathbf{x},f(\theta,\mathbf{x}))} \leq \frac{23}{13}$ for every possible facility location $a \in I$. We will prove this for the case of $|R| \geq |L|$; the other case is similar. Define R_j and L_j as in the proof of Theorem 5.1. We begin by noting that the approximation ratio is bounded from above by $\frac{46}{19}$ (it is easy to see that every agent is guaranteed a benefit of at least $\frac{19}{23}$ in our mechanism, while the maximal benefit of any agent is 2). Note that the mechanism's expected benefit is $q_2 = -\frac{7}{23}\sum_{i \in R_1} x_i - \frac{1}{23}\sum_{i \in L_1} x_i + \frac{1}{23}\sum_{i \in R_2} x_i + \frac{7}{23}\sum_{i \in L_2} x_i + \frac{27}{23}|R_1| + \frac{21}{23}|L_1| + \frac{25}{23}|R_2| + \frac{19}{23}|L_2|$. We break into cases:

1. $a \in [0, 1]$: in this case, the benefit from locating the facility at a is $q_1 = \sum_{i \in R_1} |a - x_i| + \sum_{i \in L_1}(x_i - a) + \sum_{i \in R_2}(2 + a - x_i) + \sum_{i \in L_2}(2 - |a - x_i|)$. Note that the ratio $\frac{q_1}{q_2}$ increases with x_i for $i \in L_1$, and decreases with x_i for $i \in R_2$; thus, to maximize it, we set $x_i = 2$ for $i \in L_1$ and $x_i = 1$ for $i \in R_2$. For $i \in L_2$, $x_i = a$ maximizes the ratio (note that changing x_i by α decreases the numerator by α and either increases or decreases the denominator by $\frac{7}{23}\alpha$, a change that decreases the ratio since it is known to be no more than $\frac{46}{19} < \frac{23}{7}$). Depending on a, to maximize the ratio we need to set $x_i = 0$ for all $i \in R_1$ or $x_i = 1$ for all $i \in R_1$. We

[3] In [Cheng *et al.*, 2013b], the authors note that this mechanism is 2-approximate for the obnoxious facility model; this still holds true for the hybrid model.

check both cases:

(a) $x_i = 0$ for all $i \in R_1$. Then the ratio becomes $\frac{a(|R_1|-|L_1|+|R_2|)+2|L_1|+|R_2|+2|L_2|}{\frac{27}{23}|R_1|+\frac{19}{23}|L_1|+\frac{26}{23}|R_2|+\frac{19}{23}|L_2|+\frac{7}{23}a|L_2|}$. As $|R_1|+|R_2|-|L_1| \geq |L_2| \geq 0$, this ratio increases with a and hence maximized at $a = 1$ [4], which leads to the ratio: $\frac{|R_1|+|L_1|+2|R_2|+2|L_2|}{\frac{27}{23}|R_1|+\frac{19}{23}|L_1|+\frac{26}{23}|R_2|+\frac{26}{23}|L_2|} \leq \frac{23}{13}$.

(b) $x_i = 1$ for all $i \in R_1$. The ratio becomes $\frac{a(-|R_1|-|L_1|+|R_2|)+|R_1|+2|L_1|+|R_2|+2|L_2|}{\frac{20}{23}|R_1|+\frac{19}{23}|L_1|+\frac{26}{23}|R_2|+\frac{19}{23}|L_2|+\frac{7}{23}a|L_2|}$. Maximization occurs either at $a = 1$ or at $a = 0$. We check both cases:

  i. $a = 1$: the ratio becomes $\frac{|L_1|+2|R_2|+2|L_2|}{\frac{20}{23}|R_1|+\frac{19}{23}|L_1|+\frac{26}{23}|R_2|+\frac{26}{23}|L_2|} \leq \frac{23}{13}$.

  ii. $a = 0$: the ratio becomes $\frac{|R_1|+2|L_1|+|R_2|+2|L_2|}{\frac{20}{23}|R_1|+\frac{19}{23}|L_1|+\frac{26}{23}|R_2|+\frac{19}{23}|L_2|}$. As $|L| \leq |R|$, it follows that the ratio is bounded from above by $\frac{1+2}{\frac{20}{23}+\frac{19}{23}} = \frac{23}{13}$.

2. $a \in [1, 2]$: in this case, the benefit from locating the facility at a is $q_1 = \sum_{i\in R_1}(a - x_i) + \sum_{i\in L_1}|a - x_i| + \sum_{i\in R_2}(2 - |a - x_i|) + \sum_{i\in L_2}(2 + x_i - a)$. Similarly to the previous case, the ratio $\frac{q_1}{q_2}$ is maximized when $x_i = 0$ for $i \in R_1$, $x_i = 1$ for $i \in L_2$, $x_i = a$ for $i \in R_2$, and $x_i = 1$ for all $i \in L_1$ or $x_i = 2$ for all $i \in L_1$. We break into cases:

(a) $x_i = 1$ for all $i \in L_1$: the ratio becomes $\frac{a(|R_1|-|L_2|+|L_1|)-|L_1|+2|R_2|+3|L_2|}{\frac{1}{23}a|R_2|+\frac{27}{23}|R_1|+\frac{20}{23}|L_1|+\frac{25}{23}|R_2|+\frac{26}{23}|L_2|}$. Maximum is obtained at either $a = 1$ or $a = 2$:

  i. $a = 1$: the ratio becomes $\frac{|R_1|+2|R_2|+2|L_2|}{\frac{27}{23}|R_1|+\frac{20}{23}|L_1|+\frac{26}{23}|R_2|+\frac{26}{23}|L_2|} \leq \frac{23}{13}$.

  ii. $a = 2$: the ratio becomes $\frac{2|R_1|+|L_1|+2|R_2|+|L_2|}{\frac{27}{23}|R_1|+\frac{20}{23}|L_1|+\frac{27}{23}|R_2|+\frac{26}{23}|L_2|} \leq \frac{46}{27} < \frac{23}{13}$.

(b) $x_i = 2$ for all $i \in L_1$: the ratio becomes $\frac{a(|R_1|-|L_2|-|L_1|)+2|L_1|+2|R_2|+3|L_2|}{\frac{1}{23}a|R_2|+\frac{27}{23}|R_1|+\frac{19}{23}|L_1|+\frac{25}{23}|R_2|+\frac{26}{23}|L_2|}$. Maximum is obtained at either $a = 1$ or $a = 2$:

  i. $a = 1$: the ratio becomes $\frac{|R_1|+|L_1|+2|R_2|+2|L_2|}{\frac{27}{23}|R_1|+\frac{19}{23}|L_1|+\frac{26}{23}|R_2|+\frac{26}{23}|L_2|} \leq \frac{23}{13}$.

  ii. $a = 2$: the ratio becomes $\frac{2|R_1|+2|R_2|+|L_2|}{\frac{27}{23}|R_1|+\frac{19}{23}|L_1|+\frac{27}{23}|R_2|+\frac{26}{23}|L_2|} \leq \frac{46}{27} < \frac{23}{13}$.

□

The approximation ratio of the mechanism above is tight: when there are two agents of different types, with the type 1 agent at 1 and the type 2 agent at 0, the optimal benefit is 3, whereas the mechanism's expected benefit is $\frac{39}{23}$, and the ratio is exactly $\frac{23}{13}$.

[4] Of course, for agents $i \in L_2$ we technically cannot have $x_i = 1$, but the bound holds nonetheless.

Part II

School Choice

Chapter 6

Reassignment In School Choice

6.1 Introduction

During the past fifteen years, insights from matching theory have informed the design of school choice programs in several cities around the world. The formal study of this mechanism design approach to school choice originated in the paper of Abdulkadiroglu and Sonmez [Abdulkadiroglu and Sönmez, 2003]. They formulated a model in which students have strict preferences over a finite set of schools, each with a given capacity; and each school partitions the set of students into priority classes. There is now a vast and growing literature that explores many aspects of school choice systems. A common feature of most models used in this literature is that they are essentially static. An important aspect of the problem—that has not received as much attention—is the issue of incorporating dynamic considerations, such as changes in student preferences, into the design of assignment mechanisms. This is the main motivation for our work.

6.1.1 Our Approach and Summary of Results

In many public school systems throughout the United States, students and families are required to submit preferences over the schools they are eligible for. For administrative and other reasons, this is done fairly early in the academic year. In particular, at the time these preferences are submitted, students typically do not know their options outside of the public school system. As a consequence, a significant fraction of the students end up not using their allotted seat in a public school. Yet, there is no known systematic way of reallocating these seats to the students, though it is clear that a good reallocation can lead to significant gains in overall welfare.[1] Our goal is to design an explicit reallocation mechanism run at a late stage of the matching process which can better utilize the seats freed up by applicants who choose not to utilize their initial assignments.

We consider a *two*-stage model of school assignment with finitely many schools. Students initially submit their ordinal preferences over schools, and receive a first-round assignment based on these preferences via the Deferred Acceptance (DA) algorithm; school preferences are given by weak priorities, and ties are broken via a single lottery ordering across all schools. Afterwards, some students may be presented with better outside options (such as admission to a private school), and may no longer be interested in the seat allotted to them. In the second round, students are invited to re-submit their (new) ordinal preferences over schools. The goal is to reallocate the seats so that the resulting assignment is fair, efficient, and so that the overall (two-stage) mechanism is strategyproof and does not penalize students for participating in the second-round.

A natural starting point for reallocating seats is to simply re-run DA on the new preferences, using the same school preferences (priorities and tie-breaking) as in the first round. However, this approach may result in a cascade of transfers of students from one school to another, which would be difficult for schools to handle. Alternatively, the second-round mechanism could try to allocate the vacant seats first to those students who were unassigned in the first round so as to reduce transfer costs. Unfortunately, students can readily manipulate such a reassignment mechanism by submitting truncated preference lists initially: such students are either assigned one of their top choices in the first round or they receive high priority in the second round. The challenge is to design a reallocation mechanism that retains the good properties of DA while also avoiding its potentially high transfer costs. The key idea we introduce is to suitably permute the first round lottery numbers and apply a version of DA in which the initial assignment serves as a guarantee and in which the permuted lottery numbers are used to break ties in school priorities. In particular, we show that a mechanism that upgrades students' allocations based on a simple reversal of the first-round lottery order keeps the number of reassignments small while maintaining allocative

[1]Many school assignment systems in the United States and elsewhere use a *supplementary* round in which students are allowed to "appeal" their assignment and resubmit their preferences. However, the supplementary round matching is done fairly soon after the main round and is not tied to when the outside options of the students are realized, so significant allocative inefficiencies remain.

efficiency. In a model with no school priorities, we show that this "reverse lottery" mechanism is optimal in terms of minimizing transfers and maximizing welfare within a large and natural class of mechanisms . Empirical investigations based on data from NYC's high school admissions suggest that this mechanism performs well even in the presence of school priorities.

Our work formulates (Section 6.2) two closely related models for this problem: the discrete model in which there is a finite number of students, and on which the computational and empirical findings are presented; and a continuum model, in which we assume a mass of students, and on which the theoretical results are proved. To keep things simple, the continuum model focuses on the case where schools have a single priority class.

We consider a class of reassignment mechanisms called *Permuted Lottery Deferred Acceptance* (PLDA) mechanisms. Such mechanisms find a reassignment by running DA on the reported second-round preferences of the students, and differ in how ties in school preferences are broken. Each school first prioritizes students who were assigned to it in the first round over those who weren't; within each of the resulting two classes, students are prioritized according to their initial priorities at the school; further ties across all schools are broken via a *permutation* of the lottery numbers. PLDAs differ from each other only in the choice of the permutation. The class of PLDA mechanisms is a natural class to consider if the reassignment mechanism must respect school priorities and cannot reassign students against their will (reassign them to a school they prefer less to their original assignment).

Our main theoretical result (Section 6.3) identifies a technical condition, in the continuum model, under which *all* PLDAs produce 'type-equivalent' assignments, implying that they are equivalent in terms of welfare of the final assignment. Furthermore, when this condition is satisfied, we show that *reverse lottery DA* (RLDA)—the PLDA which permutes the first round lottery numbers by "reversing" them[2]— minimizes the total number of transfers among all PLDAs. We then assess the performance of PLDAs on data from the New York City school system (section 6.4), in the presence of priorities. We observe that different PLDAs including RLDA perform fairly similarly in terms of allocative efficiency, whereas RLDA is able to reduce the number of transfer students significantly. For instance, based on the 2004-05 data set from the NYC public school system we find that the re-running DA using the same lottery numbers results in about 7235 transfer students,

[2]The student with the worst first-round lottery number is given the highest priority in the second round.

whereas reversing the lottery numbers results in only 3079 transfer students.

6.1.2 Related Work

In this subsection we survey related literature. The mechanism design approach to school choice was first formulated by Balinski & Sonmez [Balinski and Sonmez, 1999] and Abdulkadiroglu & Sonmez [Abdulkadiroglu and Sönmez, 2003]. Following the publication of [Abdulkadiroglu and Sönmez, 2003], many economists have worked closely with school authorities to re-design school choice systems, starting with New York City [Abdulkadirolu *et al.*, 2005a] and Boston [Abdulkadirolu *et al.*, 2005b] in 2003 and 2005 respectively, followed by New Orleans (2012), Denver (2012), and Washington DC (2013), among others. These centralized mechanisms appear to outperform the uncoordinated and ad-hoc assignment systems that they replaced [Abdulkadiroğlu *et al.*, 2015]. Furthermore, a significant portion of the theoretical literature has focused on the relative merits of the two canonical mechanisms—Deferred Acceptance and Top Trading Cycles (TTC)—and their variations. There is a rich and growing (theoretical and empirical) literature on school choice problems that explores several aspects including indifferences in preferences and the trade-offs between efficiency and incentives in such models. We do not survey this literature here as it is not directly relevant to our model, but we refer the reader to recent surveys by Pathak [Pathak, 2011] and Abdulkadiroglu and Sonmez [Abdulkadirolu and Snmez, 2011] for an overview.

In our model, students submit preferences, receive an assignment, resubmit preferences and receive their final allocation. If we decouple the two rounds and think of them as independent problems, the first round resembles the models used in the standard literature on school choice problems, whereas the second round resembles the literature on allocation problems where some agents may have an endowment. We next briefly survey the papers that are most relevant to our development.

A common assumption made in school choice literature is that schools are not strategic players and often do not have preferences; instead, students are assumed to have objectively verifiable priorities at the various schools they are eligible for, based on criteria such as whether they live close to the school. The coarseness of these priorities gives the mechanism designer some flexibility in how the mechanisms are actually implemented. An important theme in this literature is the issue of single versus multiple tie breaking. Under the single tie-breaking rule, a strict ordering of the students is drawn by the schools collectively, and all ties within a priority class are broken according to this tie-breaking rule; under multiple tie-breaking, however, each school draws its own random ordering of the students for breaking ties. Abdulkadiroglu et al. [Abdulkadiroglu *et al.*, 2009] empirically compared single tie-breaking verus multiple tie-breaking for the Deferred Acceptance mechanism using data from New York City public schools and observed that single tie-breaking results in more students receiving their top choices, but also in more students receiving a "poor" outcome or being unassigned. Recent papers of Arnosti [Arnosti, 2015], Ashlagi & Nikzad [Ashlagi and Nikzad, 2015] and Ashlagi, Nikzad, and Romm [Ashlagi *et al.*, 2015] study various aspects of this issue carefully and find that single tie-breaking is better in "over-demanded" markets, whereas there is a real trade-off between these rules in under-demanded markets. A concrete design recommendation of Ashlagi and Nikzad [Ashlagi and Nikzad, 2015] is that "popular" schools use single tie-breaking to break ties, which is the tie-breaking rule used in our work. We note that the coarseness in school priorities is an important element in our model: RLDA exploits these indifferences to reduce the number of reassignments, while satisfying natural efficiency and fairness properties.[3]

A second strand of literature that is somewhat relevant to our model is the work of Abdulkadiroglu and Sonmez on house allocation models with existing tenants [Abdulkadiroglu and Sonmez, 1999]. This model is a hybrid version of the housing market problem [Shapley and Scarf, 1974] in which each agent owns a house and has strict preferences over all houses, and the house allocation problem in which a set of agents collectively own houses (but no house is endowed to any individual agent) that they wish to allocate among themselves. In the hybrid model, there are agents of both types (some who are endowed with houses; and some who are not) and houses of both types (some that are endowed to an agent; and some not). The key contribution of Abdulkadiroglu and Sonmez [Abdulkadiroglu and Sonmez, 1999] is a mechanism that simultaneously generalizes TTC and Random Serial Dictatorship in the following sense: when each agent is endowed with a house, their mechanism coincides with the TTC mechanism; and when no agent is endowed with a house, their mechanism coincides with the random serial dictatorship mechanism. Furthermore,

[3] Prior work has exploited indifferences to achieve other ends. For example, Erdil and Ergin [Erdil and Ergin, 2008] show how to improve allocative efficiency; and Ashlagi and Shi [Ashlagi and Shi, 2014] show how to increase community cohesion.

they show that this new mechanism is strategy-proof, Pareto efficient, and individually rational (if an agent's endowed house is allocated to someone else, that agent strictly prefers her assignment to her original endowment), and give other mechanisms for this hybrid model. Given our extensive understanding of the hybrid model, one could proceed by treating the first-round assignment to each student as her endowment, and apply the generalized TTC mechanism. While this is a natural idea, it suffers from some drawbacks. First, the endowments in the hybrid model are exogenous, whereas in our model the endowments are computed based on the preferences; so a strategy-proof mechanism in the hybrid model is not necessarily strategy-proof in ours, as a student could manipulate the mechanism by manipulating the endowments received in the first round. Secondly, while the TTC mechanism is natural in a model with endowments, it may not be a good mechanism for assigning public-school seats to students, where it is perfectly reasonable for a student to have a right to *attend* a school, but not to be able to *trade* these rights with each other. In other words, the first-round assignment in our context should serve only as a guarantee on a student's final allocation, not as a means to improve her eventual allocation.

Our work is among the few papers to consider a dynamic point of view of school admissions. To our knowledge, Compte and Jehiel [Compte and Jehiel, 2008] were the first to consider the Deferred Acceptance mechanism for the problem of reassigning agents to positions in an organization when each agent already holds a position. Agents have preferences over the positions, and each position has a rank-ordering of the agents reflecting how well their skills fit the position. The standard Deferred Acceptance mechanism applied to this problem may make agents worse-off in the reassignment, which may not be acceptable in environments when participation is voluntary. To incentivize participation, they apply the Deferred Acceptance mechanism to a modified instance in which the top-priority agent at any position is the one who currently occupies it. This modified Deferred Acceptance mechanism is strategyproof, stable, and respects individual guarantees (each agent is assigned to a weakly better position). Subsequently, Combe, Tercieux, and Terrier [Combe *et al.*, 2016] study a more general model, motivated by the problem of assigning teachers to positions. In this model the mechanism is faced with two types of teachers: those that already have a position and are looking to improve, and those who are new to the system. As some teachers are already assigned to a school, the mechanism must guarantee an outcome that is at least as good for such teachers (individual rationality). The standard Deferred Acceptance algorithm guarantees stability with respect to the reported preferences, but may fail individual rationality, whereas the modified Deferred Acceptance algorithm guarantees individual rationality but may violate stability with respect to the original school preferences because those are artificially modified to prioritize the teachers who are already at the school. Combe et al. (among many other results) show that the modified Deferred Acceptance mechanism can be improved significantly in terms of overall welfare while also reducing the degree of instability. While the mechanism we use is a many-to-one version of this modified Deferred Acceptance mechanism, a critical distinction between this stream of work and ours is that their models assume the endowment (initial assignment) to be exogenous, whereas in our model this is also obtained by running a Deferred Acceptance mechanism. This opens up the possibility of an agent improving their eventual assignment by manipulating their intermediate endowment, a feature that is notably absent from this literature. A second important distinction is that we are interested in exploiting indifferences in school priorities to reduce the number of reassignments, which is not addressed in this literature.

Finally, Narita [Narita, 2016] analyzes the preference data from NYC school choice system and observes that preferences change after the initial match as students learn more about the schools. His work is focused on developing an empirical model of evolving preferences. Using this model, Narita concludes that the welfare cost of ignoring changes in demand is large, and proposes a centralized reallocation mechanism that can best accommodate these changing preferences. In this context, Narita also considers the modified Deferred Acceptance mechanism and establishes its usefulness. However, Narita does not explore how the mechanism can minimize reassignment costs, which is a central component of our work.

6.2 Model

We start by describing a setup with a finite set of students (and later use it for our empirical analysis). In Section 6.2.2, we translate to a setting with a continuum of students which facilitates our theoretical analysis.

6.2.1 Discrete Model

We consider the problem of allocating seats at a finite set S of schools to a finite set A of students. Students initially submit their preferences over the schools. Schools have weak priorities over students, and ties are broken according to a single lottery order across all schools. Seats are initially allocated according to the student optimal Deferred Acceptance (DA) algorithm [Abdulkadiroglu and Sönmez, 2003]. However, the students who are allotted seats may subsequently be presented with better outside options, such as admission to a private school that is not in S, and may no longer be interested in the seats allotted to them, effectively vacating them. After these outside options are revealed, students submit their new ordinal preferences over schools, and a reassignment is computed. Since the reassignment occurs at a relatively late stage, moving students from one school to another is costly, potentially for both schools and students. Our goal is to design a procedure to find a new assignment that minimizes the amount of student movement with respect to the original assignment, while satisfying appropriate notions of efficiency, fairness, and incentives. We call the initial stage of seat allocation the *first round*, and call the reallocation stage after vacancies are created the *second round*. We emphasize that the first-round mechanism is fixed to be DA with a single uniformly random lottery for tie-breaking, and that the only freedom afforded the planner is the design of the second-round mechanism.

Formally, let $S = \{s_1, \ldots, s_n\}$ be a finite set of schools, and let $s_{n+1} \notin S$ denote the outside option. Let $A = \{1, \ldots, m\}$ be the set of students. For every school $s_i \in S$, let $q_i \in \mathbb{N}_+$ be the *capacity* of school s_i. We assume that the outside option has infinite capacity $q_{n+1} = \infty$. Each school s_i has weak priorities $\succeq_i^S$ over A, which partition the students into priority classes.

Each student $a \in A$ has strict first round preferences $\succ_a$ and second round preferences $\hat{\succ}_a$ over $S \cup \{s_{n+1}\}$. We assume that $\hat{\succ}$ is *consistent* with $\succ$, meaning that the second round preferences are obtained from the first round preferences via truncation: for every $a \in A$: (1) for every $s_i, s_j \in S$, $s_i \succ_a s_j$ iff $s_i \hat{\succ}_a s_j$, and (2) for every $s_i \in S$, $s_{n+1} \succ_a s_i$ implies $s_{n+1} \hat{\succ}_a s_i$. Note that this simply corresponds to the preferences over S being fixed, while the ranking of s_{n+1} weakly improves, corresponding to an outside option (possibly) being realized between the two rounds.

An *assignment* is a function α from A to $S \cup \{s_{n+1}\}$, where $|\alpha(s_i)| \leq q_i$ for all s_i (in a slight abuse of notation, we use $\alpha(s_i)$ to denote $\alpha^{-1}(s_i)$). A *lottery* is a bijection $L : A \to \{1, \cdots, m\}$, and we will call $L(a)$ the *lottery number* of student a. In adherence with current practice [Ashlagi and Nikzad, 2015], we will assume that the first round assignment μ is obtained via DA, with student preferences $\succ$, and school preferences given by taking the priorities $\succeq^S$ and breaking ties according to a (uniformly random) lottery L in favor of the student a with the larger $L(a)$. The goal is to compute a second round assignment function $\hat{\mu}$, which we call a reassignment, based on this first round assignment and students' reported second round preferences (as well as the first round preferences and the lottery).

Let us describe some of the desired properties of such a reassignment. Naturally, any reassignment which requires taking away a student's initial assignment against their will is impractical. Thus, we require our reassignment to respect first-round guarantees:

Definition 6.1. *A reassignment $\hat{\mu}$ **respects guarantees** if every student prefers their second round allocation to their first round allocation, that is, for every $a \in A$, $\hat{\mu}(a) \hat{\succeq}_a \mu(a)$.*

One of the main reasons for using DA in the real world is the way it respects priorities: if a student is not assigned to a school she wants, it is only because that school is full of students of higher priority. A desired property from a reassignment mechanism is to continue respecting priorities, even in the final assignment:

Definition 6.2. *A reassignment $\hat{\mu}$ **respects priorities** if for every school $s_i \in S$ and student $a \in A$ such that $s_i \hat{\succ}_a \hat{\mu}(a)$, we have $|\hat{\mu}(s_i)| = q_i$ and for every a' such that $\hat{\mu}(a') = s_i$, we have $a' \succeq_i^S a$.*

We define what we mean by a transfer student. All else being equal, we want to minimize the number of such students.

Definition 6.3. *A student $a \in A$ is a **transfer student** if she leaves a school in S for another school in S. That is, a is a transfer student under reassignment $\hat{\mu}$ if $\mu(a) \neq \hat{\mu}(a)$ and $\mu(a) \hat{\succ}_a s_{n+1}$.*[4]

The timeline is as follows. First, students submit first round preference reports $\succ^r$ and the mechanism designer obtains the first round assignment μ by running DA with (single) tiebreaking via a uniformly random lottery L. Then, students observe their outside option, and update their

[4]Several alternative definitions of transfer students, such as counting students who are initially in s_{n+1} and end up at a school in S, and/or counting students who no longer find their initial assignment acceptable, could also be considered. We note that our results continue to hold for all of these alternative definitions.

preferences. Finally, students submit their updated second round preference reports $\hat{\succ}^r$ and the mechanism designer obtains the second round assignment $\hat{\mu}$ by running a *reassignment mechanism.* A reassignment mechanism is a function that maps L, μ and the agents' reports $\succ^r$ and $\hat{\succ}^r$ into a reassignment $\hat{\mu}$.

We focus on the following class of reassignment mechanisms, which is a very natural one to consider if respecting guarantees and priorities is required:

Definition 6.4 (Permuted Lottery Deferred Acceptance Mechanisms)**.** *Let P be a permutation of $\{1,\ldots,m\}$. The permuted lottery deferred acceptance mechanism (PLDA) associated with P computes a reassignment by running DA on A with student preferences $\hat{\succ}^r$ and S with school preferences $\hat{\succ}^S$ determined as follows. For each s_i:*

- *A student $a \in A$ for which $\mu(a) = s_i$ is denoted as guaranteed at s_i, and all other students are non-guaranteed at s_i.*
- *Schools prefer all guaranteed students to all non-guaranteed students, that is, for every student $a \in A$ guaranteed at s_i and student $a' \in A$ non-guaranteed at s_i, $a \hat{\succ}_i^S a'$.*
- *Ties within each of the two classes (guaranteed and non-guaranteed) are broken first according to $\succeq_i^S$, and then according to the permuted lottery $P \circ L$ (in favor of the student with the larger permuted lottery number).*

In this paper, we provide some evidence to support the use of the *reverse lottery DA* (RLDA) mechanism, which is a PLDA associated with the permutation $P(a) = m - L(a) + 1$ (namely, the permutation that reverses the order induced by L). We note that PLDA mechanisms respect guarantees and priorities, assuming truthful reports of preferences. Now, these mechanisms are not generally strategyproof, meaning that it is not always optimal for a student to report her preferences truthfully. Note that we refer to the student potentially manipulating both her first and second round preferences in order to improve her final assignment.

Example 6.1. *Consider a setting with $n = 2$ schools and $m = 4$ students. Each school has capacity 1 and a single priority class. For readability, we let $\emptyset$ denote the outside option, $\emptyset = s_{n+1} = s_3$. The students have the following preferences:*

1. $s_1 \succ_1 \emptyset \succ_1 s_2$ *and* $\emptyset \hat{\succ}_1 s_1 \hat{\succ}_1 s_2$,

2. $s_1 \succ_2 s_2 \succ_2 \emptyset$, *second round preferences identical,*

3. $s_2 \succ_3 s_1 \succ_3 \emptyset$, *second round preferences identical,*

4. $s_2 \succ_4 \emptyset \succ_4 s_1$, *second round preferences identical.*

Assume the reassignment mechanism is RLDA. Assume student 2*'s utility is* M *for* s_1, ϵ *for* s_2, *and* 0 *for* s_3 *in both rounds, where* $M >> \epsilon > 0$. *Consider the lottery* $L(i) = i$ *(for* $i = 1, 2, 3, 4$*). If the students report truthfully, the first round assignment is*

$$\mu(A) = (\mu(1), \mu(2), \mu(3), \mu(4)) = (s_1, s_2, \emptyset, \emptyset),$$

and the reassignment is

$$\hat{\mu}(A) = (\hat{\mu}(1), \hat{\mu}(2), \hat{\mu}(3), \hat{\mu}(4)) = (\emptyset, s_2, s_1, \emptyset).$$

However, consider what happens if student 2 *reports* $s_1 \succ_2^r \emptyset \succ_2^r s_2$ *(and the same in round 2). Then, the first round assignment becomes* $\mu(A) = (s_1, \emptyset, s_2, \emptyset)$, *and the reassignment becomes*

$$\hat{\mu}(A) = (\emptyset, s_1, s_2, \emptyset),$$

which is a strictly beneficial change for student 2 *(and in fact, weakly beneficial for all students).*

We remark that this reassignment was not stable in the second round when students reported truthfully, since in that case school s_2 *had second round preferences* $2 \hat{\succ}_2^S 4 \hat{\succ}_2^S 3 \hat{\succ}_2^S 1$, *and so school* s_2 *and student* 4 *form a blocking pair.*

Consider now the expected utility of student 2 *from reporting truthfully and from misreporting, when all other students report truthfully and the expectation is over the first round lottery order. With probability* $\frac{1}{4!}$, *the lottery order is* $1 \succ^B 2 \succ^B 3 \succ^B 4$, *in which case student* 2 *can change her assignment from* s_2 *to* s_1 *by reporting* s_2 *as unacceptable. Moreover, one can verify that for any lottery order, if student* 2 *receives* s_1 *in the first or second round under truthful reporting, then she also receives* s_1 *in the same round by misreporting.*[5] *Hence, by misreporting in this particular fashion, student* 2 *increases her probability of receiving* s_1 *by at least* $\frac{1}{4!}$. *Thus, for* M *sufficiently large relatively to* ϵ, *this violates strategyproofness.*

[5]This is because any stable matching in which student 2 is assigned s_1 remains stable after student 2 truncates. Indeed, student 2 is not part of any blocking pair, as she got her first choice; and any blocking pair not involving student 2 remains blocking under the true preferences, as only student 2 changes her preferences.

Nevertheless, there is reason to expect that students will not strategize when PLDA mechanisms are used. The manipulation seen in Example 6.1 is rather complex, as the misreporting student must manipulate *other* students' first round guarantees and eventual assignments in a delicate fashion. Such "scheming" requires a very high level of sophistication, as well as in-depth knowledge of other students' preferences (both first and second round), and hence is quite unrealistic. On the other hand, in the continuum model presented in Section 6.2.2, where a single student cannot affect other students' guarantees and assignments, it is easily seen that it is dominant strategy for each student to report both her first and second round preferences truthfully.[6] We therefore avoid further distinction between preferences and their reports unless otherwise stated. It is also worth noting that (in the discrete model) PLDA mechanisms satisfy "strategyproofness" when the second round is taken in isolation: if the first round reports are fixed, truthful reporting of second-round preferences is a dominant strategy for each agent. While we do not believe this sort of "strategyproofness" is sufficient on its own for incentivizing truthfulness, it is certainly an attractive property to have in a reassignment mechanism.

6.2.2 Continuum Model

While our empirical analysis is done with respect to the discrete model, our theoretical results are obtained in a model with a continuum of students. One could intuitively think of this model as the case where the number of students is very large. Azevedo and Leshno [Azevedo and Leshno, 2014] have provided an appropriate interpretation of DA in such a continuum model. In what follows we utilize their model for our needs. All of our theoretical results are obtained in the setting where all students initially belong to a single priority class, at all schools. We therefore present our definitions in that restricted setting.

In the continuum model, the set of schools remains $S = \{s_1, \ldots, s_n\}$ with the additional outside option $s_{n+1} \notin S$ as before. The capacities of the schools are $q_1, \ldots, q_n \in \mathbb{R}_+$, and $q_{n+1} = \infty$. A student type λ is described by the triplet $\lambda = (\succ^\lambda, \hat{\succ}^\lambda, l^\lambda)$, where $\succ^\lambda$ and $\hat{\succ}^\lambda$ are strict preferences over $S \cup \{s_{n+1}\}$, which are, respectively, the student's first and second round preferences, and $l^\lambda \in (0, 1)$ is the student's first round lottery number. Let Λ be the set of all student types λ. Let

[6] While we define the continuum model for the case without priorities, this still holds for any reasonable extension to the case with priorities.

η be a probability measure over Λ, which expresses the distribution of lottery numbers across the student types. For convenience of notation, let Θ be the space of all student preferences, meaning the set of all first round and second round pairs of strict preferences over $S \cup \{s_{n+1}\}$, and for each $\theta \in \Theta$ define $\zeta(\theta) = \eta(\{\lambda \in \Lambda : (\succ^\lambda, \hat{\succ}^\lambda) = \theta\})$ to be the measure of all students with preferences θ.

We assume that the first round lottery numbers are uniform and do not discriminate based on preferences (to capture a uniformly random first round lottery order), meaning that for all $\theta \in \Theta$ and intervals (a, b) with $0 \leq a \leq b \leq 1$, the proportion of students with preferences θ who have lottery number in (a, b) is equal to the length of the interval, $\eta(\{\lambda \in \Lambda : (\succ^\lambda, \hat{\succ}^\lambda) = \theta, l^\lambda \in (a, b)\}) = (b - a)\zeta(\theta)$.[7] We also assume that inconsistent student types have total ζ-measure 0, that is, any θ violating consistency satisfies $\zeta(\{\theta\}) = 0$.

An *assignment* α is a function from Λ to $S \cup \{s_{n+1}\}$, where for all $s_i \in S \cup \{s_{n+1}\}$, $\alpha(s_i)$ is η-measurable, and $\eta(\alpha(s_i)) \leq q_i$. We will again let μ denote the first round assignment, and let $\hat{\mu}$ denote the second round assignment.

Next, we define transfers in the continuum; the definition is virtually identical to the discrete case.

Definition 6.5. *We say that $\lambda \in \Lambda$ is a **transfer student** if she leaves a school in S for another school in S. That is, λ is a transfer student under reassignment $\hat{\mu}$ if $\hat{\mu}(\lambda) \neq \mu(\lambda)$ and $\mu(\lambda) \hat{\succ}^\lambda s_{n+1}$.*

We briefly define (static) DA in the continuum in general. We will then use it to define the first round and second round assignments. Let each type λ have strict preferences r^λ over the schools. Define $t_i^\lambda \in \mathbb{R}_+$ to be the score of type λ in school s_i, and assume schools prefer students with higher scores. When the preferences are so defined, Azevedo and Leshno [Azevedo and Leshno, 2014] have shown that, under a few technical conditions (which are trivially satisfied for both of our rounds), there exists an assignment α which is reasonably interpreted as the result of student proposing DA. The assignment α is defined via a vector of cutoffs $\mathbf{T} \in \mathbb{R}_+^{n+1}$, where a student of type λ is assigned to her favorite school among those where her score weakly exceeds the cutoff: $\alpha(\lambda) = \max_{r^\lambda} \{s_i \in S \cup \{s_{n+1}\} : t_i^\lambda \geq T_i\}$. $\mathbf{T}$ is market clearing, namely $\eta(\alpha(s_i)) \leq q_i$ for all $s_i \in S \cup \{s_{n+1}\}$, with equality if $T_i > 0$. Azevedo and Leshno [Azevedo and Leshno, 2014] provide a market clearing cutoff vector that corresponds to student proposing DA. We remark that this set

[7] This can be justified via an axiomatization of the kind obtained in [Al-Najjar, 2004].

of cutoffs minimizes all cutoffs at all schools simultaneously (among all market-clearing cutoffs), although we do not use that fact in our analysis.

We return now to describing our mechanism's operation in the continuum. The first round assignment μ is defined via the market-clearing cutoffs $\mathbf{C} \in \mathbb{R}_+^{n+1}$. $\mathbf{C}$ corresponds to the case where each student type λ has preferences $\succ^\lambda$ and score l^λ at each school. Since we are in the case where there are no first round priorities, the first round is simply random serial dictatorship with the order given by a first round lottery, and so we will use the terms 'first round score' and 'first round lottery' interchangeably to refer to l^λ.

Next, let a permutation P to be a (Lebesgue) measure preserving bijection from $(0,1)$ to $(0,1)$. Under permutation P, a PLDA outputs the second round assignment $\hat{\mu} = \hat{\mu}_P$, defined via the market clearing cutoffs $\hat{\mathbf{C}}^P \in \mathbb{R}_+^{n+1}$. $\hat{\mathbf{C}}^P$ corresponds to the case where each student type λ has preferences $\hat{\succ}^\lambda$ and score $\hat{l}_i^\lambda = P(l^\lambda) + \mathbb{1}(\mu(\lambda) = s_i)$ for each school $s_i \in S \cup \{s_{n+1}\}$. The second round scores are obtained by permuting the lottery number and then adding 1 for the first round assignment. Note that since $l_i^\lambda \in (0,1)$, this makes sure that for every school, students assigned to it in the first round have a higher score than students not assigned to it in the first round.

The permutation that defines the RLDA mechanism is R, where $R(x) = 1 - x$. We will also be interested in the forward lottery deferred acceptance mechanism (FLDA), which preserves the original lottery order and is defined by the identity permutation $F(x) = x$.

6.3 Type Equivalence and Transfer Minimization

In this section, we show that in the continuum model, under a certain technical condition, called the order condition, all PLDAs produce 'type-equivalent' assignments, and hence are equivalent in terms of welfare.[8] Furthermore, when this condition is satisfied, RLDA minimizes transfers among all PLDAs. We also give an example of a concrete class of scenarios where the order condition holds: the case of uniform dropouts.

6.3.1 The Order condition

We begin by defining the order condition, which we will need to state our main results.

[8] A reasonable utility model in the continuum would yield that type equivalence implies welfare equivalence.

Definition 6.6. *We say that the* ***order condition*** *holds on a set of primitives* (S, q, Λ, η) *if:*

1. *For every permutation* P*, for all* $s_i, s_j \in S \cup \{s_{n+1}\}$*, if* $C_i > C_j$*,then* $\hat{C}_i^P \geq \hat{C}_j^P$*.*
2. *For all pairs of permutations* P, P' *and schools* $s_i, s_j \in S \cup \{s_{n+1}\}$*,* $\hat{C}_i^P > \hat{C}_j^P \Rightarrow \hat{C}_i^{P'} \geq \hat{C}_j^{P'}$*.*

The first part of the condition says that the cutoff ordering is preserved from the first to the second round, and second part of the condition simply means that tie breaking is consistent across permutations. We may interpret the order condition as an indication of consistency of relative demand for schools. Informally speaking, it means that the revelation of the outside options does not change the relative popularity of the schools. The requirement that it holds for all permutations is one of robustness.

6.3.2 Type equivalence and Main Results

Next, we define type equivalence, which simply means that, for PLDAs, the measure of each type with preferences $\theta \in \Theta$ assigned to each school is independent of the permutation:

Definition 6.7. *PLDAs are said to produce* ***type-equivalent*** *allocations under permutations* P *and* P' *(or alternatively, permutations* P *and* P' *are said to be type-equivalent) if for all* $\theta \in \Theta$ *and* s_i*,*

$$\eta(\{\lambda \in \Lambda : (\succ^\lambda, \hat{\succ}^\lambda) = \theta, \hat{\mu}_P(\lambda) = s_i\}) = \eta(\{\lambda \in \Lambda : (\succ^\lambda, \hat{\succ}^\lambda) = \theta, \hat{\mu}_{P'}(\lambda) = s_i\}).$$

We are now ready to state the main results of this section.

Theorem 6.1 (Order condition implies type equivalence)**.** *Assume* $\zeta(\{\theta\}) > 0$ *for all consistent* $\theta \in \Theta$*. If the order condition holds, PLDAs produce type equivalent allocations under all permutations.*

Proof. Assume the order condition holds. For convenience, we slightly change the second round scoring function of a PLDA with permutation P to be $\hat{l}_i^\lambda = P(l^\lambda) + \mathbb{1}(l^\lambda \geq C_i)$, meaning that we give each student a guarantee at any school for which they met the cutoff in the first round. By consistency of preferences, it is easily seen that this has no effect on the resulting assignment or cutoffs. We re-index the schools in $S \cup \{s_{n+1}\}$ so that $C_i \geq C_{i+1}$. Moreover, we assume that this indexing is done such that $\hat{C}_i^P \geq \hat{C}_{i+1}^P$ (as defined by PLDA) holds simultaneously for all permutations P (this is possible by the order condition). Define $X_i = \{s_i, \ldots, s_{n+1}\}$. We say a student can *afford* a school in a round if her score in that round is at least as large as the school's cutoff in that round. Let

$$\beta_{i,j} = \eta(\{\lambda \in \Lambda : s_i \text{ is the most desirable school in } X_j \text{ with respect to } \hat{\succ}^\lambda\})$$

be the measure of the students who, when their set of affordable schools is X_j, will choose s_i (when following their second round preferences). Let $E^\lambda(\mathbf{C})$ and $\hat{E}^\lambda_P(\hat{\mathbf{C}}^P)$ be the sets of schools affordable for type λ in the first and second round, respectively, when running PLDA with lottery P.

Let P be a permutation. Note that $\beta_{i,j} > 0$ for all $j \leq i$ since $\zeta(\theta) > 0$ for all $\theta \in \Theta$ (and $\beta_{i,j} = 0$ for all $j > i$). Also note that, for each student $\lambda \in \Lambda$, there exists some i such that $E^\lambda(\mathbf{C}) = X_i$, and since the order condition is satisfied, there exists some $j \leq i$ such that $\hat{E}^\lambda_P(\hat{\mathbf{C}}^P) = X_j$. The fact that $\hat{E}^\lambda_P(\hat{\mathbf{C}}^P) = X_j$ for some j is a result of the order condition; our modified scoring function guarantees that $E^\lambda(\mathbf{C}) \subseteq \hat{E}^\lambda_P(\hat{\mathbf{C}}^P)$ (every school affordable in the first round is guaranteed in the second) and hence that $j \leq i$. Let $\gamma^P_i = \eta(\{\lambda \in \Lambda : \hat{E}^\lambda_P(\hat{\mathbf{C}}^P) = X_i\})$ is the measure of students whose affordable set in the second round of PLDA with permutation P is X_i.

Let P' be another permutation (and define $\gamma^{P'}_i$ similarly). We will prove by induction that $\gamma^{P'}_i = \gamma^P_i$ for all $s_i \in S \cup \{s_{n+1}\}$. Note that this implies type equivalence, since by uniformity of the lottery, we have that for all $\theta \in \Theta$, $\eta(\{\lambda \in \Lambda : (\succ^\lambda, \hat{\succ}^\lambda) = \theta, \hat{E}^\lambda_P(\hat{\mathbf{C}}^P)\}) = \gamma^P_i \zeta(\theta)$ and similarly for P'.[9] Assume $\gamma^{P'}_j = \gamma^P_j$ for $j = 1, \ldots, i-1$. Then we have that $\sum_{j \leq i-1} \beta_{i,j}\gamma^P_j = \sum_{j \leq i-1} \beta_{i,j}\gamma^{P'}_j$. Assume that $\gamma^P_i \neq \gamma^{P'}_i$, and assume w.l.o.g. $\gamma^P_i > \gamma^{P'}_i$. It follows that $q_i \geq \sum_{j \leq i} \beta_{i,j}\gamma^P_j > \sum_{j \leq i} \beta_{i,j}\gamma^{P'}_j$, where the first inequality follows since s_i cannot be filled beyond capacity. Thus, under P', s_i is not full, and therefore $\hat{C}^{P'}_i = 0$. However, this means that all students can afford s_i under P', and therefore $\gamma^{P'}_i = 1 - \sum_{j<i} \gamma^{P'}_j = 1 - \sum_{j<i} \gamma^P_j \geq \gamma^P_i$. This provides the required contradiction, completing the proof. □

Theorem 6.2 (Reverse lottery minimizes transfer)**.** *If PLDAs produce type equivalent allocations under all permutations, RLDA minimizes the measure of transfer students among PLDA mechanisms.*

Proof. Fix $\theta = (\succ^\theta, \hat{\succ}^\theta) \in \Theta$ and school $s_i \in S$. We will show that R minimizes transfers of students with preferences θ out of school s_i. Let P be a permutation. The measure of students

[9] Note that by the change we made to the second round scoring function, the choice a student makes in the first round does not impact her affordable schools in the second.

with preferences θ leaving school s_i in the second round under PLDA with permutation P is $h_P = \eta(\{\lambda \in \Lambda : (\succ^\lambda, \hat{\succ}^\lambda) = \theta, \mu(\lambda) = s_i, \hat{\mu}_P(\lambda) \neq s_i\})$. The measure of students with preferences θ entering school s_i under permutation P is $e_P = \eta(\{\lambda \in \Lambda : (\succ^\lambda, \hat{\succ}^\lambda) = \theta, \mu(\lambda) \neq s_i, \hat{\mu}_P(\lambda) = s_i\})$. Due to type equivalence, there is a constant c, independent of P, such that $h_P = e_P - c$. If $s_{n+1} \hat{\succ}^\theta s_i$, then there are no transfers of students with preferences θ out of school s_i. So assume $s_i \hat{\succ}^\theta s_{n+1}$.

We will show that $h_R \leq h_P$ for all permutations P. If $e_R = 0$, then $h_R = e_R - c \leq e_P - c = h_P$ for all permutations P. So assume $e_R > 0$. We claim that in this case $h_R = 0$, which will complete our proof. Since $e_R > 0$, there exists some student $\lambda \in \Lambda$ with $(\succ^\lambda, \hat{\succ}^\lambda) = \theta$ for which $s_i = \hat{\mu}_R(\lambda) \hat{\succ}^\theta \mu(\lambda)$. By consistency, we have $s_i \succ^\theta \mu(\lambda)$, and therefore λ could not afford (meet the cutoff for) s_i in the first round. Assume $h_R > 0$ (we aim to obtain a contradiction). By similar reasoning, there exists some student $\lambda' \in \Lambda$ with $(\succ^{\lambda'}, \hat{\succ}^{\lambda'}) = \theta$ for which $s_j = \hat{\mu}_R(\lambda') \hat{\succ}^\theta \mu(\lambda') = s_i$. By definition, λ' could afford s_i in the first round, and hence $l^{\lambda'} > l^\lambda$. Note that since $s_i \hat{\succ}^\theta s_{n+1}$, then $s_j \hat{\succ}^\theta s_{n+1}$, and thus by consistency $s_j \succ^\theta s_i$. Thus, λ' could not afford s_j in the first round, and so, since also $l^{\lambda'} > l^\lambda$, the definition of RLDA yields $\hat{l}_j^\lambda > \hat{l}_j^{\lambda'}$ under RLDA. Thus, since λ' can afford s_j in the second round, so can λ, but— as we saw— $s_j \hat{\succ}^\theta s_i = \hat{\mu}_R(\lambda)$. This is a contradiction, and therefore $h_R = 0$, completing our proof. □

Thus, when the order condition holds, Theorem 6.1 essentially means that all PLDAs will generate identical welfare; therefore, it is reasonable to focus on the objective of transfer minimization, and Theorem 6.2 implies that RLDA performs best relatively to that objective.

One important case covered by our results is the case of uniform dropouts. In this case, students either remain with the same preferences in the second round, or drop out of the system entirely. If dropouts occur in an i.i.d.-like fashion, the order condition is satisfied.[10] One real world scenario that can be represented this way is when dropouts occur not due to a better private option, but rather exogenously—say, due to moving to a different city—and thus should not be expected to be

[10]Formally, this is the case where there exists $p \in [0,1]$ s.t. for every strict preferences over schools $\succ$, $\zeta(\{\theta = (\succ^\theta, \hat{\succ}^\theta) \in \Theta : \succ^\theta = \succ, \hat{\succ}^\theta = s_{n+1} \succ \ldots\}) = p\zeta(\{\theta = (\succ^\theta, \hat{\succ}^\theta) \in \Theta : \succ^\theta = \succ\})$, and $\zeta(\{\theta = (\succ^\theta, \hat{\succ}^\theta) \in \Theta : \succ^\theta = \succ, \hat{\succ}^\theta = \succ\}) = (1-p)\zeta(\{\theta = (\succ^\theta, \hat{\succ}^\theta) \in \Theta : \succ^\theta = \succ\})$. We note that this is essentially the case where dropouts are i.i.d. with probability p. We do not phrase it in that language because of the well-known technical measurability issue w.r.t. a continuum of random variables, but it should be noted that this issue can be handled— see, for example, [Al-Najjar, 2004].

correlated with preferences or first round assignments.

6.4 Empirical Analysis

In this section, we use actual data from New York City's (NYC) school choice program to simulate and evaluate the performance of PLDA mechanisms under different permutations P. The simulations indicate that our theoretical results are real-world relevant. Different choices of P are found to yield similar allocative efficiency: the number of students assigned to their k-th choice for each rank k, and number of students remaining unassigned, are very similar for different permutations P. At the same time, the difference in the number of transfer students is significant, and is minimized under RLDA.

6.4.1 The Data and Simulations

We use data from the high school admissions process in NYC for the academic years 2004-05, 2005-06 and 2006-07, as follows:[11]

1. First round preferences: In our simulation, we take the first round preferences $\succ$ of every student to be the preferences they submitted in the main round of admissions. Except for the fact that students may rank no more than 12 schools in reality, the algorithm used in practice is strategyproof [Abdulkadiroğlu *et al.*, 2005], so it may be reasonable to assume these preferences to be true (but see [Hassidim *et al.*, 2015]).

2. Second round preferences: In our simulation, students either drop out from the system entirely in the second round or maintain the same preferences. Students are considered to drop out if the data does not record them as attending any public high school in NYC the following year. For a minority of the students ($9.2\% - 10.45\%$), their attendance in the following year could not be determined by our data, and hence we make them drop out randomly at a rate equal to the rate of dropouts for the rest of the students ($8.9\% - 9.2\%$).

[11] We performed an initial cleanup of the data, such as removing preference entries which did not correspond to an existing school code.

3. School capacities: Each school's capacity is set equal to the number of students assigned to it in the data.[12] This is a lower bound on the actual capacity; however, it should be noted that schools for which this lower bound is strict are schools which remained non-full in reality. Informally speaking, such schools could be considered "unpopular", and one could intuitively expect the additional capacity in such schools to have little impact.

4. School priorities over students are obtained directly from the data.[13]

We consider the following family of permutations, parameterized by a single parameter α, that smoothly interpolates between RLDA and FLDA. Each student a receives a uniform i.i.d. first round lottery number l_a (a normal variable with mean 0 and variance 1), generating a uniformly random lottery order as needed. The second round lottery number of a is specified as $\alpha l_a + l'_a$, where l'_a is a new i.i.d. normal variable with mean 0 and variance 1, and α is identical for all the students. Note that $\alpha = -\infty$ corresponds to RLDA and $\alpha = \infty$ corresponds to FLDA. For a fixed real α, every realization of second round lottery numbers corresponds to some permutation of first round lottery numbers, with α roughly capturing the correlation of the second round order with that of the first round. We remark that two different iterations using the same α do not generally give the same PLDA, since in a PLDA a permutation is deterministic, while l'_a is random. We quote averages across simulations.

6.4.2 Results

The results of our computational experiments based on 2004-05 NYC high school admissions data appear in Figure 6.1 and Table 6.1 (The results for 2005-06 and 2006-07 were similar). Allocative efficiency appears to not vary much across values of α: The number of students receiving their k-th choice for each $1 \leq k \leq 12$, as well as the number of unassigned students, vary by less than 1% of the total number of students (larger values of α give more students their first choice, but only slightly). This is consistent with what we would expect based on our theoretical finding of type equivalence (Theorem 6.1) of the final allocation under different PLDA mechanisms.

[12] As per the final assignment produced by centralized allocation.

[13] Unlike in the theoretical analysis, where we assumed no priorities, we take them into consideration here. We obtained similar results to the ones described below in simulations with no school priorities.

Table 6.1: Simulation Results, 2004-2005 NYC High School admissions

α	Transfers (number)	Transfers (%)	$k=1$ (%)	$k=2$ (%)	$k=3$ (%)	Unassigned (%)
FLDA: ∞	7235	8.84	50.27	13.21	7.53	4.79
7.39	7035	8.59	50.26	13.22	7.53	4.79
2.72	6551	8.00	50.21	13.24	7.55	4.79
1.00	5830	7.12	50.10	13.31	7.57	4.78
0.37	5240	6.40	49.98	13.38	7.61	4.76
0.00	4792	5.85	49.86	13.45	7.64	4.76
-0.37	4336	5.30	49.75	13.49	7.68	4.74
-1.00	3751	4.58	49.59	13.55	7.72	4.73
-2.72	3253	3.97	49.47	13.57	7.76	4.71
-7.39	3106	3.79	49.44	13.57	7.76	4.70
RLDA: $-\infty$	3079	3.76	49.43	13.56	7.76	4.70

The table above describes the mean number/percentage of students transferring, and the mean percentage getting their k-th choice or remaining unassigned, in terms of their percentage out of the total number of students, including those that drop out. The data contained 81,884 students and 653 schools. The percentage of students who dropped out was 9.22%. We averaged across 50 realizations for each value of α. Variation across realizations in number of transfers was ~ 100.

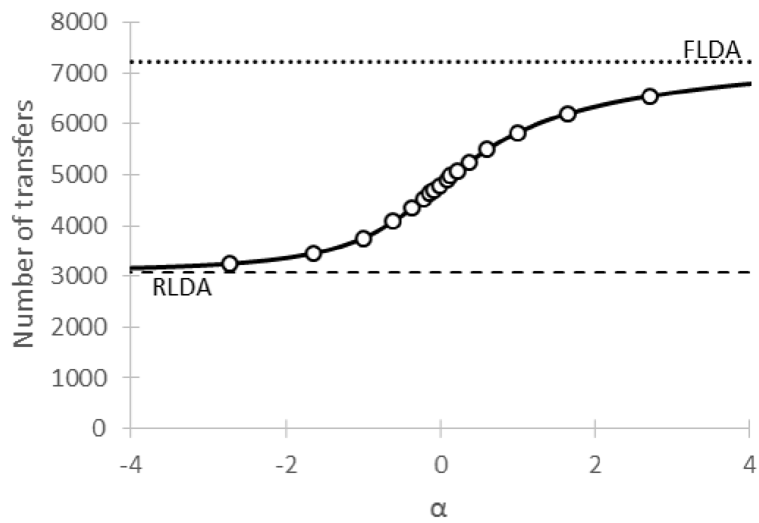

Figure 6.1: Number of transfers of students versus α. The number of transfers under the extreme values of α, namely, $\alpha = \infty$ (FLDA) and $\alpha = -\infty$ (RLDA) are shown via dotted lines.

We find that the mean number of transfers is minimized at $\alpha = -\infty$ (RLDA), and increases with α, see Figure 6.1, again consistent with our theoretical results (Theorem 6.2). The mean number of transfers is as large as about $7,200$ under FLDA compared to just $3,100$ under RLDA. Overall our findings suggest that RLDA would be the best choice of mechanism in the family of PLDA mechanisms considered.

Part III

Bibliography

Bibliography

[Abdulkadiroglu and Sonmez, 1999] Atila Abdulkadiroglu and Tayfun Sonmez. House Allocation with Existing Tenants. *Journal of Economic Theory*, 88(2):233–260, October 1999.

[Abdulkadiroglu and Sönmez, 2003] Atila Abdulkadiroglu and Tayfun Sönmez. School choice: A mechanism design approach. *The American Economic Review*, 93(3):729–747, 2003.

[Abdulkadiroğlu *et al.*, 2005] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. The new york city high school match. *American Economic Review*, pages 364–367, 2005.

[Abdulkadiroglu *et al.*, 2009] Atila Abdulkadiroglu, Parag A. Pathak, and Alvin E. Roth. Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *American Economic Review*, 99(5):1954–78, December 2009.

[Abdulkadiroğlu *et al.*, 2015] Atila Abdulkadiroğlu, Nikhil Agarwal, and Parag A Pathak. The welfare effects of coordinated assignment: Evidence from the nyc hs match. Technical report, National Bureau of Economic Research, 2015.

[Abdulkadirolu and Snmez, 2011] Atila Abdulkadirolu and Tayfun Snmez. *Matching markets: Theory and practice*, pages 1–47. Cambridge University Press, 1 2011.

[Abdulkadirolu *et al.*, 2005a] Atila Abdulkadirolu, Parag A. Pathak, and Alvin E. Roth. The New York City High School Match. *American Economic Review*, 95(2):364–367, May 2005.

[Abdulkadirolu *et al.*, 2005b] Atila Abdulkadirolu, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez. The Boston Public School Match. *American Economic Review*, 95(2):368–371, May 2005.

[Al-Najjar, 2004] Nabil I Al-Najjar. Aggregation and the law of large numbers in large economies. *Games and Economic Behavior*, 47(1):1–35, 2004.

[Alon *et al.*, 2010a] Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. Strategyproof approximation of the minimax on networks. *Math. Oper. Res.*, 35(3):513–526, 2010.

[Alon *et al.*, 2010b] Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. Strategyproof approximation of the minimax on networks. *Math. Oper. Res.*, 35(3):513–526, 2010.

[Alon *et al.*, 2010c] Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. Walking in circles. *Discrete Mathematics*, 310(23):3432–3435, 2010.

[Arnosti, 2015] Nick Arnosti. Short lists in centralized clearinghouses. In Roughgarden et al. [2015], page 751.

[Ashlagi and Nikzad, 2015] Itai Ashlagi and Afshin Nikzad. What matters in tie-breaking rules? how competition guides design. 2015.

[Ashlagi and Roth, 2014] Itai Ashlagi and Alvin E Roth. Free riding and participation in large scale, multi-hospital kidney exchange. *Theoretical Economics*, 9(3):817–863, 2014.

[Ashlagi and Shi, 2014] Itai Ashlagi and Peng Shi. Improving community cohesion in school choice via correlated-lottery implementation. *Operations Research*, 62(6):1247–1264, 2014.

[Ashlagi *et al.*, 2013] Itai Ashlagi, Felix Fischer, Ian A Kash, and Ariel D Procaccia. Mix and match: A strategyproof mechanism for multi-hospital kidney exchange. *Games and Economic Behavior*, 2013.

[Ashlagi *et al.*, 2015] Itai Ashlagi, Afshin Nikzad, and Assaf Romm. Assigning more students to their top choices: A tiebreaking rule comparison. In Roughgarden et al. [2015], pages 755–756.

[Atlamaz and Klaus, 2007] Murat Atlamaz and Bettina Klaus. Manipulation via endowments in exchange markets with indivisible goods. *Social Choice and Welfare*, 28(1):1–18, 2007.

[Azevedo and Leshno, 2014] Eduardo M Azevedo and Jacob D Leshno. A supply and demand framework for two-sided matching markets. *Available at SSRN 2260567*, 2014.

[Balinski and Sonmez, 1999] Michel Balinski and Tayfun Sonmez. A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, 84(1):73–94, January 1999.

[Bandelt and Labbé, 1986] Hans-J Bandelt and Martine Labbé. How bad can a voting location be. *Social Choice and Welfare*, 3(2):125–145, 1986.

[Bandelt, 1985] Hans-Jrgen Bandelt. Networks with condorcet solutions. *European journal of operational research : EJOR*, 20(3):314–326, 1985.

[Bar-Noy *et al.*, 2012] Amotz Bar-Noy, Yi Gai, Matthew P Johnson, Bhaskar Krishnamachari, and George Rabanca. Funding games: the truth but not the whole truth. In *Internet and Network Economics*, pages 128–141. Springer, 2012.

[Barberà *et al.*, 1998] Salvador Barberà, Jordi Massó, and Shigehiro Serizawa. Strategy-proof voting on compact ranges. *Games and Economic Behavior*, 25(2):272–291, 1998.

[Barber *et al.*, 2012] Salvador Barber, Dolors Berga, and Bernardo Moreno. Domains, ranges and strategy-proofness: the case of single-dipped preferences. *Social Choice and Welfare*, 39(2-3):335–352, 2012.

[Bhawalkar and Roughgarden, 2011] Kshipra Bhawalkar and Tim Roughgarden. Welfare guarantees for combinatorial auctions with item bidding. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 700–709. SIAM, 2011.

[Borgers *et al.*, 2015] Tilman Borgers, Daniel Krahmer, and Roland Strausz. *An introduction to the theory of mechanism design.* Oxford University Press, USA, 2015.

[Brandeau and Chiu, 1988] Margaret L Brandeau and Samuel S Chiu. Parametric facility location on a tree network with an l p-norm cost function. *Transportation Science*, 22(1):59–69, 1988.

[Burke and Kendall, 2005] Edmund K Burke and Graham Kendall. *Search methodologies*. Springer, 2005.

[Caragiannis *et al.*, 2015] Ioannis Caragiannis, Christos Kaklamanis, Panagiotis Kanellopoulos, Maria Kyropoulou, Brendan Lucier, Renato Paes Leme, and Éva Tardos. Bounding the inefficiency of outcomes in generalized second price auctions. *Journal of Economic Theory*, 156:343–388, 2015.

[Chen *et al.*, 2011] Ning Chen, Nick Gravin, and Pinyan Lu. Mechanism design without money via stable matching. *CoRR*, abs/1104.2872, 2011.

[Cheng and Zhou, 2015] Yukun Cheng and Sanming Zhou. A survey on approximation mechanism design without money for facility games. In *Advances in Global Optimization*, pages 117–128. Springer, 2015.

[Cheng *et al.*, 2013a] Yukun Cheng, Wei Yu, and Guochuan Zhang. Strategy-proof approximation mechanisms for an obnoxious facility game on networks. *Theoretical Computer Science*, 497:154–163, 2013.

[Cheng *et al.*, 2013b] Yukun Cheng, Wei Yu, and Guochuan Zhang. Strategy-proof approximation mechanisms for an obnoxious facility game on networks. *Theor. Comput. Sci.*, 497:154–163, 2013.

[Combe *et al.*, 2016] Julien Combe, Olivier Tercieux, and Camille Terrier. The design of teacher assignment: Theory and evidence. 2016.

[Compte and Jehiel, 2008] Olivier Compte and Philippe Jehiel. Voluntary participation and re-assignment in two-sided matching. Technical report, working paper, 2008.

[Cramton, 2013] Peter Cramton. Spectrum auction design. *Review of Industrial Organization*, 42(2):161–190, 2013.

[Danilov, 1994] Vladimir I Danilov. The structure of non-manipulable social choice rules on a tree. *Mathematical Social Sciences*, 27(2):123–131, 1994.

[Dughmi and Ghosh, 2010] Shaddin Dughmi and Arpita Ghosh. Truthful assignment without money. In *11th ACM conference on Electronic Commerce*, pages 325–334. ACM, 2010.

[Edelman *et al.*, 2007] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American economic review*, 97(1):242–259, 2007.

[Erdil and Ergin, 2008] Aytek Erdil and Haluk Ergin. What's the matter with tie-breaking? improving efficiency in school choice. *The American Economic Review*, 98(3):669–689, 2008.

[Feigenbaum and Johnson, 2015] Itai Feigenbaum and Matthew P Johnson. Selfish knapsack. *arXiv preprint arXiv:1510.07358*, 2015.

[Feigenbaum and Sethuraman, 2014] Itai Feigenbaum and Jay Sethuraman. Strategyproof mechanisms for one-dimensional hybrid and obnoxious facility location. *arXiv preprint arXiv:1412.3414*, 2014.

[Feigenbaum *et al.*, 2013a] Itai Feigenbaum, Jay Sethuraman, and Chun Ye. Approximately optimal mechanisms for strategyproof facility location: Minimizing l_p norm of costs. *arXiv preprint arXiv:1305.2446*, 2013.

[Feigenbaum *et al.*, 2013b] Itai Feigenbaum, Jay Sethuraman, and Chun Ye. Approximately optimal mechanisms for strategyproof facility location: Minimizing l_p norm of costs. *CoRR*, abs/1305.2446, 2013.

[Feldman and Wilf, 2013a] Michal Feldman and Yoav Wilf. Strategyproof facility location and the least squares objective. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 873–890. ACM, 2013.

[Feldman and Wilf, 2013b] Michal Feldman and Yoav Wilf. Strategyproof facility location and the least squares objective. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 873–890, New York, NY, USA, 2013. ACM.

[Gibbard, 1973] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

[Groves, 1973] Theodore Groves. Incentives in teams. *Econometrica: Journal of the Econometric Society*, pages 617–631, 1973.

[Hajaj *et al.*, 2015] Chen Hajaj, John P Dickerson, Avinatan Hassidim, Tuomas Sandholm, and David Sarne. Strategy-proof and efficient kidney exchange using a credit mechanism. In *AAAI*, pages 921–928, 2015.

[Han and Du, 2012] Qiaoming Han and Donglei Du. *Moneyless Strategy-proof Mechanism on Single-sinked Policy Domain: Characterization and Applications*. Faculty of Business Administration, University of New Brunswick, 2012.

[Hansen and Thisse, 1981] Pierre Hansen and Jacques Thisse. Outcomes of voting and planning: Condorcet, weber and rawls locations. *Journal of Public Economics*, 16(1):1–15, 1981.

[Hassidim *et al.*, 2015] Avinatan Hassidim, Deborah Marciano-Romm, Assaf Romm, and Ran I Shorrer. Strategic behavior in a strategy-proof environment. Technical report, working paper, 2015.

[Ibara and Nagamochi, 2012] Ken Ibara and Hiroshi Nagamochi. Characterizing mechanisms in obnoxious facility game. In Guohui Lin, editor, *COCOA*, volume 7402 of *Lecture Notes in Computer Science*, pages 301–311. Springer, 2012.

[Labbé, 1985] M. Labbé. Outcomes of voting and planning in single facility location problems. *European journal of operational research : EJOR*, 20(3):299–313, 1985.

[Manjunath, 2014] Vikram Manjunath. Efficient and strategy-proof social choice when preferences are single-dipped. *International Journal of Game Theory*, 43(3):579–597, 2014.

[Moulin, 1980] Herve Moulin. On strategy-proofness and single-peakedness. *Public Choice*, 35:437–455, 1980.

[Moulin, 2015] Hervé Moulin. One dimensional mechanism design. In *Web and Internet Economics: 11th International Conference, WINE 2015, Amsterdam, The Netherlands, December 9-12, 2015, Proceedings*, volume 9470, page 436. Springer, 2015.

[Myerson, 1981] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[Narita, 2016] Yusuke Narita. Match or mismatch: Learning and inertia in school choice. 2016.

[Nisan and Ronen, 2007] Noam Nisan and Amir Ronen. Computationally feasible vcg mechanisms. *J. Artif. Intell. Res.(JAIR)*, 29:19–47, 2007.

[Pathak, 2011] Parag A. Pathak. The Mechanism Design Approach to Student Assignment. *Annual Review of Economics*, 3(1):513–536, 09 2011.

[Postlewaite, 1979] Andrew Postlewaite. Manipulation via endowments. *The review of economic studies*, 46(2):255–262, 1979.

[Procaccia and Tennenholtz, 2009] Ariel D Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 177–186. ACM, 2009.

[Procaccia and Tennenholtz, 2013a] Ariel D. Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. *ACM Trans. Economics and Comput.*, 1(4):18, 2013.

[Procaccia and Tennenholtz, 2013b] Ariel D Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. *ACM Transactions on Economics and Computation*, 1(4):18, 2013.

[Roth and Sotomayor, 1992] Alvin E Roth and Marilda Sotomayor. Two-sided matching. *Handbook of game theory with economic applications*, 1:485–541, 1992.

[Roth, 2008] Alvin E Roth. Deferred acceptance algorithms: History, theory, practice, and open questions. *international Journal of game Theory*, 36(3-4):537–569, 2008.

[Roughgarden *et al.*, 2015] Tim Roughgarden, Michal Feldman, and Michael Schwarz, editors. *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC '15, Portland, OR, USA, June 15-19, 2015.* ACM, 2015.

[Roughgarden, 2009] Tim Roughgarden. Intrinsic robustness of the price of anarchy. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 513–522. ACM, 2009.

[Roughgarden, 2015] Tim Roughgarden. The price of anarchy in games of incomplete information. *ACM Transactions on Economics and Computation*, 3(1):6, 2015.

[Satterthwaite, 1975] Mark Allen Satterthwaite. 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, 1975.

[Schummer and Vohra, 2002] James Schummer and Rakesh V Vohra. Strategy-proof location on a network. *Journal of Economic Theory*, 104(2):405–428, 2002.

[Schummer and Vohra, 2013] James Schummer and Rakesh V Vohra. Assignment of arrival slots. *American Economic Journal: Microeconomics*, 5(2):164–185, 2013.

[Shapley and Scarf, 1974] L. S. Shapley and H. E. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1(1):23–37, 1974.

[Vazirani, 2001] Vijay V. Vazirani. *Approximation Algorithms*. Springer-Verlag New York, Inc., New York, NY, USA, 2001.

[Vazirani, 2013] Vijay V Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013.

[Vulkan *et al.*, 2013] Nir Vulkan, Alvin E Roth, and Zvika Neeman. *The handbook of market design*. OUP Oxford, 2013.

[Vygen, 2005] Jens Vygen. *Approximation algorithms facility location problems*. Forschungsinstitut für Diskrete Mathematik, Rheinische Friedrich-Wilhelms-Universität, 2005.

Part IV

Appendices

Appendix A

Appendix for Chapter 2

In this appendix, we (highly) informally discuss a possible "fix" to EQUAL-UTILITY's running time issues. As we have mentioned, EQUAL-UTILITY requires solving two NP-Hard problems: the knapsack problem and PROGRAM. However, we do have approximation algorithms for these NP-hard problems: there is a known FPTAS for knapsack [Vazirani, 2013], and from the proof of Theorem 2.5 we can easily deduce an algorithm which produces $\frac{1}{1-\epsilon}\alpha$ approximation to PROGRAM in time polynomial in $\frac{1}{\epsilon}$ whenever EQUAL-UTILITY needs to solve it. The problem is that if we use these approximation algorithms, strategyproofness can technically be violated in a strange fashion.

To understand this, consider trying to compute $OPT(X_i)$. Suppose that the knapsack FPTAS is "non-monotonic", meaning that there exist $X_i, X_i' \in G_i$ where $X_i' \subset X_i$, and yet the FPTAS provides a higher value solution when it is run on X_i' (than when it is run on X_i). In such a situation, in EQUAL-UTILITY, it could be in agent i's best interest to report X_i' when her true set of items is X_i. Therefore, despite the fact that the goals of the agent and the designer are fully aligned with each other in this case (they both want to maximize the value of items chosen from X_i), the agent can potentially benefit from reporting X_i' instead. Nevertheless, if she does so, it is in an attempt to *assist*, rather than mislead, the designer by narrowing down the search space. The violation of strategyproofness in the approximation algorithm to PROGRAM is of the same kind.

These violations are of the type observed in [Nisan and Ronen, 2007]. Informally speaking, these strange violations could be eliminated if we allow the designer to "listen" to the agent's solution proposals for such problems, and choose the best between her own computed solution and the agent's. Thus, in situations where such communication is possible, EQUAL-UTILITY's running time issues can be handled. A formalization of this idea can be made by an adjustment of what is called "Second Chance Mechanisms" in [Nisan and Ronen, 2007].

Appendix B

Appendix for Chapter 3

B.1 An Alternative Definition of Individual Cost

Let g be a strictly increasing and convex $\mathcal{C}_1$ function on $[0, \infty)$ with $g(0) = g'(0) = 0$. Note that $g(x) = x^p$ satisfies this description for all $p > 1$. We consider a scenario where the cost of agent i is $C(x_i, y) = g(|x_i - y|)$ when the mechanism is deterministic and locates the facility at y. Similarly $C(x_i, \pi) = \mathbb{E}_{y \sim \pi}[g(|x_i - y|)]$ when the mechanism is randomized and locates the facility according to distribution π. The social cost function (for $n = 2$) $h(|x_1 - y|, |x_2 - y|)$ is only assumed to be (1) anonymous ($h(d, d') = h(d', d)$) and (2) satisfy that for all $a \in (\min\{x_1, x_2\}, \max\{x_1, x_2\})$ where $x_1 \neq x_2$, $h(|x_1 - a|, |x_2 - a|) < h(|x_2 - x_1|, 0)$. Note that for $p > 1$, the L_p norm of the distances and the L_p norm of the costs (for the general g above) both satisfy these conditions. We show that in this case, no randomized strategyproof mechanism satisfying shift invariance, scale invariance and ex-post Pareto efficiency for $n = 2$ can help us improve the approximation ratio relatively to the median mechanism.

Theorem B.1. *Let f be a randomized mechanism satisfying shift invariance, scale invariance, and ex-post Pareto efficiency for $n = 2$. Assume f is strategyproof with respect to the individual cost function $C(x_i, y) = g(|x_i - y|)$, where g is a strictly increasing and convex $\mathcal{C}_1$ function on $[0, \infty)$ with $g(0) = g'(0) = 0$. If the social cost function satisfies (1) and (2), then the approximation ratio of f is at least as large as the median's.*

Proof. Using a proof similar to that of Lemma 3.4, we may assume without loss of generality that

f is symmetric. Consider a profile where $n = 2$ and $x_1 = 0$, $x_2 = 1$. Let $Y = f(0, 1)$. We would like to show that $\mathbb{P}(Y \in (0, 1)) = 0$. Suppose for the sake of contradiction that there there exists $x \in (0, \frac{1}{2})$ such that $\mathbb{P}(Y \in (x, 1-x)) = q > 0$. Now suppose agent 2 now misreports his location to $1+\epsilon$ for some small $\epsilon > 0$ such that $\frac{1}{1+\epsilon} > 1-x$. By shift and scale invariance, $f(0, 1+\epsilon) = (1+\epsilon)Y$. Then the difference in cost for agent 2 between the two profile of reports is

$$
\begin{aligned}
&\mathbb{E}[g\big(|1-(1+\epsilon)Y|\big)] - \mathbb{E}[g\big(|1-Y|\big)] = \\
&= -\int_0^{\frac{1}{1+\epsilon}} (g(1-y) - g(1-(1+\epsilon)y))dF(y) + \int_{\frac{1}{1+\epsilon}}^1 (g((1+\epsilon)y - 1) - g(1-y))dF(y) \leq \\
&\leq \mathbb{P}(Y \in [\frac{1}{1+\epsilon}, 1])g(\epsilon) - q\Big(g(1-x^*) - g(1-(1+\epsilon)x^*)\Big)
\end{aligned}
$$

where $x^* \in \arg\min_{y \in [x, 1-x]} g(1-y) - g(1-(1+\epsilon)y)$. The inequality follows from the fact that $g((1+\epsilon)y - 1) - g(1-y) \leq g(\epsilon)$ for all $y \in [\frac{1}{1+\epsilon}, 1]$ and that $g(1-y) - g(1-(1+\epsilon)y) \geq g(1-x^*) - g(1-(1+\epsilon)x^*)$ for all $y \in [x, 1-x]$. Note that

$$
\begin{aligned}
&\lim_{\epsilon \to 0^+} \frac{\mathbb{E}[g\big(|1-(1+\epsilon)Y|\big)] - \mathbb{E}[g\big(|1-Y|\big)]}{\epsilon} \leq \\
&\leq \lim_{\epsilon \to 0^+} \mathbb{P}(Y \in [\frac{1}{1+\epsilon}, 1])\frac{g(\epsilon)}{\epsilon} - q\frac{g(1-x^*) - g(1-(1+\epsilon)x^*)}{\epsilon} \leq \\
&\leq \mathbb{P}(Y = 1)g'(0) - qg'(1-x^*)x^* < 0
\end{aligned}
$$

The third inequality follows from $g'(0) = 0$ and $g'(1-x^*) > 0$ (since g is strictly convex). This implies that $\mathbb{E}[g\big(|1-(1+\epsilon)Y|\big)] - \mathbb{E}[g\big(|1-Y|\big)] < 0$ for ϵ sufficiently small, implying that there is a profitable deviation for agent 2. □

B.2 Omitted Proofs from Section 3.4.2

Lemma 3.6. *A symmetric, shift and scale invariant mechanism f is strategyproof if and only if for any profile $\mathbf{x} \in \mathbb{R}^2$ with $x_1 = 0 < x_2$, the following conditions hold:*

1. $-\int_{(-\infty, x_2)} y dF(y) + \int_{(x_2, \infty)} y dF(y) + x_2\mathbb{P}(Y = x_2) \geq 0$
2. $\int_{(-\infty, x_2)} y dF(y) - \int_{(x_2, \infty)} y dF(y) + x_2\mathbb{P}(Y = x_2) \geq 0$

where $Y \sim f(\mathbf{x})$ *with c.d.f.* F*.*

Proof. First, let us prove that the two conditions imply strategyproofness. By shift invariance and anonymity, it suffices to check strategyproofness for profiles where $x_1 = 0$ and $x_2 \geq 0$. Moreover, any scale invariant mechanism is trivially strategyproof with respect to the profile $(0, 0)$ since scale invariance implies $f(0, 0) = 0$, which means that no agent has incentive to misreport his location.[1] Thus, we can assume that $x_2 > 0$. It suffices to show that agent 2 cannot benefit by deviating from his true location if the two aforementioned conditions hold. Since $x_2 > 0$, we can denote agent 2's deviation x'_2 as cx_2 for some $c \in \mathbb{R}$. Moreover, we can assume that $c \geq 0$. This can be justified as follows. Assume $c < 0$. Note that by symmetry, in any fixed profile $\mathbf{z}$, the closer a point is to $m_{\mathbf{z}}$, the smaller the expected distance of the facility is from that point. In particular, this implies that $C(x_2, f(0, -cx_2)) \leq C(-x_2, f(0, -cx_2))$. But also note that by scale invariance, $C(-x_2, f(0, -cx_2)) = C(x_2, f(0, cx_2))$. Thus, $C(x_2, f(0, -cx_2)) \leq C(x_2, f(0, cx_2))$. Consequently, if reporting cx_2 is a profitable deviation for agent 2 for some $c < 0$, then reporting $-cx_2$ is also a profitable deviation for the agent.

When agent 2 reports his location to be cx_2, where $c > 1$, the change in cost incurred by agent 2 is (where C_{orig} is the expected cost of agent 2 under truthful reporting and C_{dev} is the expected

[1] $f(0, 0) = 0$ follows from, say, $f(0, 0) = f(0 \cdot 1, 0 \cdot 1) = 0 \cdot f(1, 1) = 0$, where the second equality is by scale invariance.

cost of agent 2 under misreporting):

$$C_{dev} - C_{orig} =$$
$$= -(c-1)\int_{(-\infty,\frac{x_2}{c})} ydF(y) + \int_{[\frac{x_2}{c},x_2)} ((c+1)y - 2x_2)dF(y) +$$
$$(c-1)\int_{(x_2,\infty)} ydF(y) + (c-1)x_2\mathbb{P}(Y = x_2) =$$
$$= -(c-1)\int_{(-\infty,x_2)} ydF(y) + \int_{[\frac{x_2}{c},x_2)} (2cy - 2x_2)dF(y) +$$
$$(c-1)\int_{(x_2,\infty)} ydF(y) + (c-1)x_2\mathbb{P}(Y = x_2) \geq$$
$$\geq -(c-1)\int_{(-\infty,x_2)} ydF(y) + (c-1)\int_{(x_2,\infty)} ydF(y) + (c-1)x_2\mathbb{P}(Y = x_2).$$

Hence, when condition 1 holds, we have that $-(c-1)\int_{(-\infty,x_2)} ydF(y) + (c-1)\int_{(x_2,\infty)} yF(y) + (c-1)x_2\mathbb{P}(Y = x_2) \geq 0$, which means that $C_{dev} - C_{orig} \geq 0$.

Similarly, when $0 \leq c < 1$, the change in cost incurred by agent 2 is:

$$C_{dev} - C_{orig} =$$
$$= (1-c)\int_{(-\infty,x_2)} ydF(y) + \int_{(x_2,\frac{x_2}{c}]} (2x_2 - (c+1)y)dF(y) -$$
$$(1-c)\int_{(\frac{x_2}{c},\infty)} ydF(y) + (1-c)x_2\mathbb{P}(Y = x_2) =$$
$$= (1-c)\int_{(-\infty,x_2)} ydF(y) + \int_{(x_2,\frac{x_2}{c}]} (2x_2 - 2cy)dF(y) -$$
$$(1-c)\int_{(x_2,\infty)} ydF(y) + (1-c)x_2\mathbb{P}(Y = x_2) \geq$$
$$\geq (1-c)\int_{(-\infty,x_2)} ydF(y) - (1-c)\int_{(x_2,\infty)} ydF(y) + (1-c)x_2\mathbb{P}(Y = x_2).$$

Hence, when condition 2 holds, we have that $(1-c)\int_{(-\infty,x_2)} ydF(y) - (1-c)\int_{(x_2,\infty)} ydF(y) + (1-c)x_2\mathbb{P}(Y = x_2) \geq 0$, which means that $C_{dev} - C_{orig} \geq 0$. Hence, the mechanism is strategyproof for any profile $\mathbf{x}$ with $x_1 = 0 < x_2$.

To prove the other direction, suppose condition 1 does not hold for some profile $\mathbf{x}$ with $x_1 = 0 <$

x_2. Then there exists $\epsilon > 0$ small enough such that $-\int_{(-\infty,x_2)} ydF(y) + \int_{(x_2,\infty)} ydF(y) + x_2\mathbb{P}(Y = x_2) \leq -\epsilon$ for some $x_2 > 0$. We choose $c > 1$ s.t. $\mathbb{P}(Y \in [\frac{x_2}{c}, x_2)) < \frac{\epsilon}{4x_2}$, then we have that

$$
\begin{aligned}
&C_{dev} - C_{orig} = \\
&= -(c-1)\int_{(-\infty,x_2)} ydF(y) + \int_{[\frac{x_2}{c},x_2)} (2cy - 2x_2)dF(y) + (c-1)\int_{(x_2,\infty)} ydF(y) + \\
&(c-1)x_2\mathbb{P}(Y = x_2) \leq \\
&\leq (c-1)(-\int_{(-\infty,x_2)} ydF(y) + \int_{[\frac{x_2}{c},x_2)} (2x_2)dF(y) + \int_{(x_2,\infty)} ydF(y) + x_2\mathbb{P}(Y = x_2)) < \\
&< -(c-1)\frac{\epsilon}{2} < 0,
\end{aligned}
$$

which contradicts strategyproofness of the mechanism.

Similarly, suppose condition 2 does not hold for some profile $\mathbf{x}$ with $x_1 = 0 < x_2$. Then there exists $\epsilon > 0$ small enough such that $\int_{(-\infty,x_2)} ydF(y) - \int_{(x_2,\infty)} ydF(y) + x_2\mathbb{P}(Y = x_2) \leq -\epsilon$ for some $x_2 > 0$. We choose $0 < c < 1$ s.t. $\mathbb{P}(Y \in (x_2, \frac{x_2}{c}])) < \frac{\epsilon}{4x_2}$, then we have that

$$
\begin{aligned}
&C_{dev} - C_{orig} = \\
&= (1-c)\int_{(-\infty,x_2)} ydF(y) + \int_{(x_2,\frac{x_2}{c}]} (2x_2 - 2cy)dF(y) - (1-c)\int_{(x_2,\infty)} ydF(y) + \\
&(1-c)x_2\mathbb{P}(Y = x_2) \leq \\
&\leq (1-c)(\int_{(-\infty,x_2)} ydF(y) + \int_{[\frac{x_2}{c},x_2)} (2x_2)dF(y) - \int_{(x_2,\infty)} ydF(y) + x_2\mathbb{P}(Y = x_2)) < \\
&< -(1-c)\frac{\epsilon}{2} < 0,
\end{aligned}
$$

which contradicts strategyproofness of the mechanism. □

Lemma 3.7. *Let f be a strategyproof, shift invariant, scale invariant and symmetric mechanism. There exists another strategyproof, shift invariant, scale invariant and symmetric mechanism g with a weakly smaller expected social cost on every profile, such that at least one of the following two properties holds:*

(1) For every two-agent profile $\mathbf{x}$*,* $\mathbb{P}(g(\mathbf{x}) \in (x_1, x_2)) = 0$*. (Doesn't utilize interior)*[2]

(2) For every two-agent profile $\mathbf{x}$*,* $\mathbb{P}(g(\mathbf{x}) \in (-\infty, x_1) \cup (x_2, \infty)) = 0$*. (Ex-post Pareto efficiency)*

Proof. Let f be as given above. Assume f violates both (1) and (2) on some profile $\mathbf{x}$ (otherwise, there is nothing to prove: we can take $g = f$). By shift invariance we may assume without loss of generality that $x_1 = 0$. We may assume by anonymity and shift invariance that $x_1 = 0 < x_2$. Let $Y \sim f(\mathbf{x})$. Let $p_1 = \mathbb{P}(Y \in (m_{\mathbf{x}}, x_2)) + \frac{\mathbb{P}(Y = m_{\mathbf{x}})}{2} = \mathbb{P}(Y \in (x_1, m_{\mathbf{x}})) + \frac{\mathbb{P}(Y = m_{\mathbf{x}})}{2}$, $p_2 = \mathbb{P}(x_2, \infty) = \mathbb{P}(-\infty, x_1)$, $z_1 = \frac{\mathbb{E}[Y\mathbb{1}(Y \in (x_1, m_{\mathbf{x}}))] + m_{\mathbf{x}}\mathbb{P}(Y = m_{\mathbf{x}})/2}{p_1}$, and $z_1' = \mathbb{E}[Y | Y \in (-\infty, x_1)]$, $z_2 = \frac{\mathbb{E}[Y\mathbb{1}(Y \in (m_{\mathbf{x}}, x_2))] + m_{\mathbf{x}}\mathbb{P}(Y = m_{\mathbf{x}})/2}{p_1}$, $z_2' = \mathbb{E}[Y | Y \in (x_2, \infty)]$.[3]

Consider a random variable Y'' defined as follows: $\mathbb{P}(Y'' \in \{z_1', x_1, z_1, z_2, x_2, z_2'\}) = 1$, $\mathbb{P}(Y'' = z_1') = \mathbb{P}(Y'' = z_2') = p_2$, $\mathbb{P}(Y'' = z_1) = \mathbb{P}(Y'' = z_2) = p_1$, and $\mathbb{P}(Y'' = x_1) = \mathbb{P}(Y'' = x_2) = \mathbb{P}(Y = x_1) = \mathbb{P}(Y = x_2)$. Clearly, Y'' is symmetric about the midpoint $m_{\mathbf{x}}$. Since the social cost function is convex, it follows that $\mathbb{E}[sc(\mathbf{x}, Y'')] \leq \mathbb{E}[sc(\mathbf{x}, Y)]$.

Now, consider a random variable Y' obtained from Y'' as follows. We construct Y' from Y'' by shifting parts of the probability mass at z_1 and z_1' to x_1 as well as by shifting parts of the probability mass at z_2 and z_2' to x_2 while ensuring that $\mathbb{E}[Y'] = \mathbb{E}[Y'']$. Specifically, since $z_1 < x_1 < z_1'$, we can write $x_1 = \lambda z_1 + (1 - \lambda) z_1'$ for some $0 < \lambda < 1$. One way to shift the probability mass is to subtract probability λp and $(1 - \lambda)p$ from z_1 and z_1' respectively and add probability p to x_1 for p sufficiently small (do the same transformation for points z_2, z_2', and x_2). This transformation ensures $\mathbb{E}[Y'] = \mathbb{E}[Y'']$ because

$$(p_1 - \lambda p) z_1 + (p_2 - (1 - \lambda)p) z_2 + (\mathbb{P}(Y'' = x_1) + p) x_1 = p_1 z_1 + p_2 z_2 + \mathbb{P}(Y'' = x_1) x_1.$$

In order to maximize the shift in probability mass, we choose the largest p possible or $p = \min(\frac{p_1}{\lambda}, \frac{p_2}{1-\lambda})$. If $p = \frac{p_1}{\lambda}$, then $\mathbb{P}(Y' \in \{z_1', x_1, x_2, z_2'\}) = 1$, as $\mathbb{P}(Y' = z_1') = \mathbb{P}(Y' = z_2') = p_2 - (1 - \lambda)p$, and $\mathbb{P}(Y' = x_1) = \mathbb{P}(Y' = x_2) = \mathbb{P}(Y'' = x_1) + p$. Else if $p = \frac{p_2}{1-\lambda}$, then

[2] Note that it is possible for such a mechanism to still be ex-post Pareto efficient, if $\mathbb{P}(g(\mathbf{x}) \in \{x_1, x_2\}) = 1$.

[3] Note that if $\mathbb{P}(Y = m_{\mathbf{x}}) = 0$, then z_1 is the conditional expectation of Y given that $Y \in (x_1, m_{\mathbf{x}})$. When $\mathbb{P}(Y = m_{\mathbf{x}}) > 0$, imagine that whenever $Y = m_{\mathbf{x}}$, we flip a fair coin; then z_1 is the conditional expectation of Y given that $Y \in (x_1, m_{\mathbf{x}})$ or $Y = m_{\mathbf{x}}$ and the coin lands on heads. z_2 can be defined in a similar manner (replace $(x_1, m_{\mathbf{x}})$ with $(m_{\mathbf{x}}, x_2)$ and heads with tails). From this description it is clear that $z_1 \in (x_1, m_{\mathbf{x}}]$, $z_2 \in [m_{\mathbf{x}}, x_2)$, and that they are symmetric about $m_{\mathbf{x}}$.

$\mathbb{P}(Y' \in \{x_1, z_1, z_2, x_2\}) = 1$, $\mathbb{P}(Y' = z_1) = \mathbb{P}(Y' = z_2) = p_1 - \lambda p$, and $\mathbb{P}(Y' = x_1) = \mathbb{P}(Y' = x_2) = \mathbb{P}(Y'' = x_1) + p$. It is clear from construction that Y' is symmetric about $m_{\mathbf{x}}$. Convexity implies $\mathbb{E}[sc(\mathbf{x}, Y')] \leq \mathbb{E}[sc(\mathbf{x}, Y'')]$, and so $\mathbb{E}[sc(\mathbf{x}, Y')] \leq \mathbb{E}[sc(\mathbf{x}, Y)]$.

Now, let g be a mechanism that locates the facility according to Y' given profile $\mathbf{x}$. Note that there is a unique way to extend the definition of g to all other two-agent profiles such that g is shift and scale invariant as well as symmetric; let us extend the definition of g that way. Furthermore, this extension is easily seen to imply the following:

1. Since $\mathbb{E}[sc(\mathbf{x}, g(\mathbf{x}))] \leq \mathbb{E}[sc(\mathbf{x}, f(\mathbf{x}))]$ for the profile $\mathbf{x}$, the social cost obtained by mechanism g via the extension is no more than the one obtained by mechanism f for all two-agent profiles.

2. If $\mathbb{P}(g(\mathbf{x}) \in (x_1, x_2)) = 0$, then $\mathbb{P}(g(\mathbf{q}) \in (q_1, q_2)) = 0$ for all two-agent profiles $\mathbf{q}$. Similarly, if $\mathbb{P}(g(\mathbf{x}) \in (-\infty, x_1) \cup (x_2, \infty)) = 0$, then $\mathbb{P}(g(\mathbf{q}) \in (-\infty, q_1) \cup (q_2, \infty)) = 0$ for all two-agent profiles $\mathbf{q}$.

Thus, all that is left for us to do is to show strategyproofness of g. We can do so by verifying the conditions in Lemma 3.6 (the fact that it holds for all the required profiles is then again immediate by shift and scale invariance). When $p = \frac{p_1}{\lambda}$, we claim that it suffices to show that:

$$z_2'(p_2 - (1-\lambda)p) - z_1'(p_2 - (1-\lambda)p) + x_2(\mathbb{P}(Y' = x_2) + p) \geq (z_2' - z_1')p_2 - (z_1 + z_2)p_1 + x_2\mathbb{P}(Y = x_2)$$

and that

$$z_1'(p_2 - (1-\lambda)p) - z_2'(p_2 - (1-\lambda)p) + x_2(\mathbb{P}(Y' = x_2) + p) \geq (z_1' - z_2')p_2 + (z_1 + z_2)p_1 + x_2\mathbb{P}(Y = x_2)$$

To justify this claim, we need to show that the right hand sides are always greater than or equal to 0. But note that z_1, z_2, p_1, z_1', z_2', and p_2 were defined so that the right hand sides amount exactly to the conditions of Lemma 3.6 for f on the profile $\mathbf{x}$, and thus must be greater than or equal to zero. After some algebra, the two inequalities above reduce to:

$$z_1'(1-\lambda)p - z_2'(1-\lambda)p + x_2 p \geq -z_1 p_1 - z_2 p_1, \tag{B.1}$$

and

$$-z_1'(1-\lambda)p + z_2'(1-\lambda)p + x_2 p \geq z_1 p_1 + z_2 p_1. \tag{B.2}$$

To show (B.1), we know that

$$z_1'(1-\lambda)p + z_1p_1 = (z_1'(1-\lambda) + z_1\lambda)p = x_1p = 0, \quad \text{that is} \quad z_1'(1-\lambda)p = -z_1p_1,$$

and that

$$x_2p = (z_2'(1-\lambda) + z_2\lambda)p \geq z_2'(1-\lambda)p - z_2p_1, \quad \text{that is} \quad x_2p - z_2'(1-\lambda)p \geq -z_2p_1.$$

Combining the two above expressions gives us the desired result. Similarly, (B.2) follows from the fact that $z_1'(1-\lambda)p + z_1p_1 = 0$ and that $x_2p = (z_2'(1-\lambda) + z_2\lambda)p \geq -z_2'(1-\lambda)p + z_2p_1$. The proof for the case where $p = \frac{p_2}{1-\lambda}$ is similar and so will be omitted. □

B.3 Alternative Assumptions in Section 3.4.2

Theorem 3.4 holds if we replace the assumption of shift invariance with symmetry. It is clear from the structure of the proof that it is enough to replace Lemma 3.4 with the following lemma:

Lemma B.1. *Given any strategyproof, symmetric and scale invariant mechanism, there exists a strategyproof, symmetric, scale and shift invariant mechanism with a weakly smaller worst-case approximation ratio.*

Proof. Given a mechanism f, define $g(\mathbf{x}) = f(0, x_2 - x_1) + x_1$. Assume f is strategyproof, symmetric and scale invariant. We claim that g is strategyproof, symmetric, scale and shift invariant with a weakly smaller worst-case approximation ratio. The fact that g is shift invariant and has a weakly smaller worst-case approximation ratio than f is immediate. Let $Y_{x_1,x_2} \sim f(\mathbf{x})$ and $Y'_{x_1,x_2} \sim g(\mathbf{x})$; the relevant equalities below are in distribution.

1. g is symmetric: let $\mathbf{x} \in \mathbb{R}^2$, and let $b \in \mathbb{R}$. Then $\mathbb{P}(Y'_{x_1,x_2} \geq m_{\mathbf{x}} + b) = \mathbb{P}(Y_{0,x_2-x_1} \geq m_{\mathbf{x}} + b - x_1) = \mathbb{P}(Y_{0,x_2-x_1} \geq m_{(0,x_2-x_1)} + b) = \mathbb{P}(Y_{0,x_2-x_1} \leq m_{(0,x_2-x_1)} - b) = \mathbb{P}(Y_{0,x_2-x_1} \leq m_{\mathbf{x}} - b - x_1) = \mathbb{P}(Y_{0,x_2-x_1} + x_1 \leq m_{\mathbf{x}} - b) = \mathbb{P}(Y'_{x_1,x_2} \leq m_{\mathbf{x}} - b)$.

2. g is scale invariant: let $\mathbf{x} \in \mathbb{R}^2$ and let $c \in \mathbb{R}$. Then $Y_{cx_1,cx_2} = Y_{0,c(x_2-x_1)} + cx_1 = cY_{0,x_2-x_1} + cx_1 = c(Y_{0,x_2-x_1} + x_1) = cY'_{x_1,x_2}$. The second equality follows from scale invariance of f.

3. g is strategyproof: let $\mathbf{x} \in \mathbb{R}^2$, $b, x_2' \in \mathbb{R}$. There are two cases:

(a) Assume $\mathbb{E}[|x_2 - Y'_{x_1,x_2}|] > \mathbb{E}[|x_2 - Y'_{x_1,x'_2}|]$. Note that $\mathbb{E}[|x_2 - Y'_{x_1,x_2}|] = \mathbb{E}[|(x_2 - x_1) - Y_{0,x_2-x_1}|]$ and $\mathbb{E}[|x_2 - Y'_{x_1,x'_2}|] = \mathbb{E}[|(x_2 - x_1) - Y_{0,x'_2-x_1}|]$. Thus, it follows that when agent 1's location is 0 and agent 2's location is $x_2 - x_1$, agent 2 can benefit under f when reporting $x'_2 - x_1$ instead, violating strategyproofness of f. Contradiction.

(b) Assume $\mathbb{E}[|x_1 - Y'_{x_1,x_2}|] > \mathbb{E}[|x_1 - Y'_{x_1+b,x_2}|]$. Note that $\mathbb{E}[|x_1 - Y'_{x_1,x_2}|] = \mathbb{E}[| - Y_{0,x_2-x_1}|] = \mathbb{E}[|(x_2 - x_1) - Y_{0,x_2-x_1}|]$, where the last equality follows from symmetry of f. Also note that $\mathbb{E}[|x_1 - Y'_{x_1+b,x_2}|] = \mathbb{E}[| - b - Y_{0,x_2-x_1-b}|] = \mathbb{E}[|(x_2 - x_1) - Y_{0,x_2-x_1-b}|]$, where again the last equality follows from symmetry of f. Thus, when agent 1's true location is 0 and agent 2's true location is $x_2 - x_1$, then agent 2 benefits under f by reporting $x_2 - x_1 - b$, violating strategyproofness of f. Contradiction.

□