(Non-)Obvious Manipulability of Rank-Minimizing Mechanisms

Peter Troyan
Department of Economics
University of Virginia*

March 2, 2024

Abstract

In assignment problems, the rank distribution of assigned objects is often used to evaluate match quality. Rank-minimizing (RM) mechanisms directly optimize for average rank. While appealing, a drawback is RM mechanisms are not strategyproof. This paper investigates whether RM satisfies the weaker incentive notion of non-obvious manipulability (NOM, Troyan and Morrill, 2020). I show any RM mechanism with full support—placing positive probability on all rank-minimizing allocations—is NOM. In particular, uniform randomization satisfies this condition. Without full support, whether an RM mechanism is NOM or not depends on the details of the selection rule.

*Address: P.O. Box 400182, Charlottesville, VA, 22904. Email: troyan@virginia.edu. I would like to thank Josue Ortega for comments on this paper, as well the Associate Editor and two anonymous referees, whose comments have greatly improved it.

1 Introduction

Many institutions make assignments by collecting agents' ordinal preferences over the possible alternatives and using them as an input to a rule that outputs an assignment. Common examples include public school choice (Abdulkadiroğlu and Sönmez, 2003), medical residency matching (Roth and Peranson, 1999), teacher assignment (Combe et al., 2022), course allocation (Budish and Cantillon, 2012; Budish and Kessler, 2017), and refugee resettlement (Delacrétaz et al., 2023). A natural metric to measure the success of the outcome is the rank distribution: how many students get their first choice, how many get their second choice, and so on. Indeed, many school districts such as those in New York City and San Francisco publicly release statistics on the ranks as a measure of the goodness of the match.

In the context of school choice specifically, most cities rely on either some version of Gale and Shapley's celebrated deferred acceptance mechanism (Gale and Shapley, 1962) or on the so-called Boston mechanism (also sometimes called the immediate acceptance mechanism) to determine the assignment, and then evaluate the rank distribution produced by these mechanisms.[1] However, given the importance placed on the rank distribution, it is also natural to consider mechanisms that optimize directly for this objective. Indeed, Featherstone (2020) notes that Teach for America does exactly this, and uses the rank distribution when *selecting* its assignment. Featherstone (2020) is also one of the few papers that has undertaken a serious analysis of mechanisms based explicitly on the rank distribution (a few other papers in this relatively small but growing literature are discussed below).

While using mechanisms that select assignments based explicitly on the rank distribution is naturally appealing, an important consideration in any mechanism design problem is the incentives of the agents. Indeed, one of the most appealing properties arguing for the use of DA-based mechanisms is that they generally give agents strong incentives to report their true preferences. On the other hand, Proposition 10 of Featherstone (2020) shows that no

[1] There is another class of mechanisms based on the top trading cycles (TTC) algorithm of Shapley and Scarf (1974) that has been studied extensively in the theoretical literature, but has found few adherents in practice. The only use of it in a real-world school choice setting that I am aware of is New Orleans, where it was used for one year before being abandoned (Abdulkadiroğlu et al., 2020).

ordinal assignment mechanism is both rank-efficient (a refinement of Pareto efficiency that he defines that takes into account the rank distribution) and strategyproof.

However, strategyproofness is a very demanding property, and just because a mechanism *can* be manipulated does not mean that it is *will* be manipulated. As part of a recent strand of literature on "obviousness" in mechanism design, Troyan and Morrill (2020) introduce the concept of *non-obvious manipulability (NOM)* as a way to relax strategyproofness. They use their definition to taxonomize non-strategyproof mechanisms into two classes: those that are obviously manipulable (such as the Boston mechanism and pay-as-bid auctions) and thus are likely to be easily manipulated in practice, and those that are non-obviously manipulable (such as school-proposing DA, Kesten's (2010) efficiency-adjusted DA, and uniform price auctions) which, while formally manipulable, have manipulations that are more difficult for cognitively-limited agents to recognize and enact successfully.[2]

In this paper, I consider a canonical assignment model in which there is a set of agents (such as students) to be assigned to a set of objects (such as schools), each of which has some fixed capacity. Agents participate in a mechanism in which they report their preferences over the objects. I consider the class of *rank-minimizing (RM) mechanisms,* which take the reported agent preferences, and implement an assignment that minimizes the average rank of the objects received. As there can in general be many allocations that minimize the average rank at a given preference profile, there are many possible RM mechanisms, depending on how these ties are broken. Because the aforementioned result of Featherstone (2020) implies that no RM mechanism is strategyproof, I consider instead the weaker notion of NOM from Troyan and Morrill (2020), and ask whether RM mechanisms satisfy this property.

An issue that arises in answering this question is that Troyan and Morrill's definition applies only to deterministic mechanisms that select a single outcome at each preference profile. Because there may be many allocations that minimize the average rank for a given preference profile, it is more natural to model a mechanism as returning a probability distribution over such allocations, and use an extension of NOM for probabilistic mechanisms that was first proposed by Demeulemeester and Pereyra (2022).

[2]Experimentally, Cerrone et al. (2023) find high rates of truth-telling in the non-strategyproof, but also not obviously manipulable, EADA mechanism.

My main result (Theorem 1) shows that any RM mechanism with full support is not obviously manipulable. A full support RM mechanism is one that places strictly positive probability weight on every allocation that is rank-minimizing at the submitted preference profile. For instance, one natural way to run an RM mechanism is to select uniformly at random from the set of all RM allocations, which we call the *uniform rank-minimizing (URM) mechanism*.[3] This is a full support mechanism, and so Theorem 1 implies that the URM mechanism is NOM.

After showing Theorem 1, I investigate whether NOM extends beyond the full support assumption. This may be important, for instance, to a designer who may not view all RM allocations as equal, but rather wants to optimize some further objective within this class. For instance, if a designer wants to advantage agents who belong to certain groups within the class of RM allocations, she might want to place zero probability on RM allocations where these agents do worse, and higher probability on RM allocations where they do better.

Without full support, whether an RM mechanism is NOM will depend on the details of the selection rule. Consider the following mechanism, called the *rank-minimizing serial dictatorship (RM-SD)*: Fix an exogenous ordering of the agents, and ask the agents in this order to choose their favorite (remaining) object that they could be assigned at any RM allocation that is also consistent with the choices of the earlier agents. This is a deterministic mechanism (it always produces a single deterministic allocation), and so does not have full support. I show that if each object has capacity 1 (what I call *unit capacity markets*), the RM-SD mechanism is NOM (Proposition 2). However, I also show that if some object has a capacity greater than 1, then the RM-SD mechanism is not NOM (Proposition 3). Further, even in unity capacity markets, there are RM mechanisms without full support that are not NOM, i.e., that are obviously manipulable (Proposition 4).[4]

In sum, the class of RM mechanisms is an appealing class of mechanisms

[3]This is a commonly used implementation of RM mechanisms in the literature; see, e.g., Nikzad (2022) or Ortega and Klein (2023).

[4]An important additional distinction between these mechanisms and the uniform RM mechanism discussed in the previous paragraph is that the latter treats all of the agents fairly. In particular, URM satisfies the fairness property of *equal treatment of equals (ETE)* while the others do not.

for policymakers who are interested in the average rank as a desirable objective. To the extent that the main shortcoming of the RM mechanism is its lack of strategyproofness, my results suggest that this problem may be not so severe, so long as the mechanism has full support (e.g., the uniform RM mechanism). At the same time, another takeaway from my results is that designers who desire to use different selection rules without full support should be prudent in doing so, as this choice could have consequences for the manipulability of the resulting mechanism. While more empirical work is needed to test the theory, given my results and their other appealing properties, RM mechanisms at the very least seem worthy of further consideration for practical market design applications.

Related literature

Besides Featherstone (2020), who provides a detailed analysis of rank efficiency criteria and related mechanisms and was discussed above, there are only three other papers I am aware of in the economics literature that analyze RM mechanisms.[5] Nikzad (2022) studies large markets, and provides an upper bound on the expected average rank of rank-minimizing assignments. Sethuraman (2022) provides another proof of the same result. Nikzad (2022) also shows that the uniform RM mechanism is Bayesian incentive compatible when agent preference rankings are also drawn uniformly at random.

Lastly, Ortega and Klein (2023) study the average rank of RM, DA, and TTC in large markets, both theoretically and using simulations, and find that RM outperforms DA and TTC on important dimensions such as efficiency and fairness. They also use data from secondary school admissions in Hungary to analyze the three mechanisms in an empirical setting, and find similar support for RM mechanisms. The data is from DA, a strategyproof mechanism, and they conduct their analysis of RM mechanisms assuming the DA reports are truthful and that agents will continue to report these truthful preferences in a counterfactual in which they play RM, even though it is not strategyproof. They argue for this approach as follows:

> In our view, it is unclear whether students would misrepresent

[5] There is a large literature in mathematics and operations research that has studied related problems, though they are generally not concerned with incentive issues, which are the focus of this paper. See Krokhmal and Pardalos (2009) for a survey.

> their preferences in RM. The potential gains from manipulation are tiny...and manipulations are risky and could lead to worse outcomes...Furthermore, there is evidence of high truth-telling rates in not obviously manipulable mechanisms (Cerrone et al., 2023).

My results can be seen as a formalization of this final point of Ortega and Klein (2023).

2 Model

2.1 Preferences and allocations

There is a set of N agents $I = \{i_1, \ldots, i_N\}$ and a set of M objects $O = \{o_1, \ldots, o_M\}$. Each object o_m has a **capacity** q_m which denotes the maximum number of agents who can be assigned to it. I assume that $\sum_{m=1}^{M} q_m \geq N$, which is common in school choice settings where all students must be offered a seat at a school, and is without loss of generality if one of the objects is an "outside option" that has enough capacity for all agents. Each agent has a **strict preference relation** $\succ_i$ defined on the set of objects O, where $o \succ_i o'$ denotes that agent i strictly prefers object o to object o'. I use $o \succsim_i o'$ when either $o \succ_i o'$ or $o = o'$. I will sometimes refer to $\succ_i$ as an agent's **type**. I will also write $\succ_i$: $o, o', \ldots,$ to denote an agent who has preferences such that her favorite object is o, second favorite object is o', and the rest of her preferences can be arbitrary. The set $\mathcal{P}_i$ is agent i's preference domain, which consists of all strict rankings over O, and $\mathcal{P}^I = \mathcal{P}_1 \times \cdots \times \mathcal{P}_N$ is the set of all preference profiles for all agents. I write $\succ_I = (\succ_1, \ldots, \succ_N) \in \mathcal{P}^I$ to denote a profile of preferences, one for each agent $i_1, \ldots, i_N$, and sometimes write $\succ_I = (\succ_i, \succ_{-i})$ to separate i's preferences $\succ_i$ from those of the remaining agents, $\succ_{-i}$.

For an agent i with type $\succ_i$, I define $r_i(o) = |\{o' \in O : o' \succsim_i o\}|$ to be the **rank** of object o according to i's preferences; in other words, i's favorite object has rank 1, i's second-favorite object has rank 2, etc. While the rank of an object also depends on an agent's preferences $\succ_i$, for readability, I suppress this from the notation. Also, notice that I use the convention that lower numbers correspond to more preferred objects.

A (deterministic) **allocation** $\alpha : I \to O$ is a function that assigns each agent to an object. I use α_i to denote the object assigned to agent i in allocation α. Any allocation must satisfy $|\{i \in I : \alpha_i = o\}| \leq q_m$ for all $o_m \in O$, i.e., each object cannot be assigned to more agents than its capacity. Let $\mathcal{A}$ be the set of all possible allocations. A **random allocation** $\mu : \mathcal{A} \to [0,1]$ is a probability distribution over $\mathcal{A}$, where $\sum_{\alpha \in \mathcal{A}} \mu(\alpha) = 1$. It is necessary to introduce random allocations to be able to deal with tie-breaking when there are many possible deterministic allocations that minimize the average rank. Let $\mathcal{M}$ be the set of random allocations.

Given a preference profile $\succ_I$, denote the **average rank** of any (deterministic) allocation α (with respect to $\succ_I$) by[6]

$$\bar{r}(\alpha) = \frac{1}{N} \sum_{i \in I} r_i(\alpha_i),$$

and let

$$\bar{\mathcal{A}}(\succ_I) = \{\alpha \in \mathcal{A} : \bar{r}(\alpha) \leq \bar{r}(\alpha') \text{ for all } \alpha' \in \mathcal{A}\}.$$

We call $\bar{\mathcal{A}}(\succ_I)$ the set of **rank-minimizing deterministic allocations** (with respect to $\succ_I$). Since $\mathcal{A}$ is finite, the set $\bar{\mathcal{A}}(\succ_I)$ is non-empty for all $\succ_I$, though it may contain more than one element. A random allocation μ is a **rank-minimizing random allocation** (with respect to $\succ_I$) if $\alpha \notin \bar{\mathcal{A}}(\succ_I)$ implies $\mu(\alpha) = 0$; in other words, μ places positive probability weight on only those deterministic allocations that are rank-minimizing given the preference profile $\succ_I$. Let $\bar{\mathcal{M}}(\succ_I)$ denote the set of rank-minimizing random allocations. Notice that a rank-minimizing random allocation may have $\mu(\alpha) = 0$ for some rank-minimizing (deterministic) allocations α. When $\mu(\alpha) > 0$ for all $\alpha \in \bar{\mathcal{A}}(\succ_I)$, we call μ a **full-support rank-minimizing random allocation**. This distinction will be important for the results below.

2.2 Mechanisms

A **mechanism** is a function $\psi : \mathcal{P}^I \to \mathcal{M}$. For each preference profile $\succ_I \in \mathcal{P}^I$, mechanism ψ returns the random allocation $\psi(\succ_I) \in \mathcal{M}$. For any deterministic allocation $\alpha \in \mathcal{A}$, we write $\psi(\succ_I)(\alpha)$ to denote the probability that the random allocation $\psi(\succ_I)$ places on α. A mechanism ψ is a **rank-minimizing**

[6]Once again, $\bar{r}(\alpha)$ depends on the preference profile $\succ_I$, but this is suppressed to avoid notational clutter.

(RM) mechanism if $\psi(\succ_I)$ is a rank-minimizing random allocation for all $\succ_I$. In other words, $\psi(\succ_I)$ places strictly positive probability only on allocations that are rank-minimizing; formally, $\psi(\succ_I) \in \bar{\mathcal{M}}(\succ_I)$ for all $\succ_I \in \mathcal{P}^I$. If $\psi(\succ_I)$ is a further full-support rank-minimizing allocation for all $\succ_I$, we say that ψ is a **full-support RM mechanism**. One natural tie-breaking rule is to choose uniformly at random from the entire set $\bar{\mathcal{A}}(\succ_I)$ of rank-minimizing deterministic allocations. This results in a particular full-support RM mechanism that we call the **uniform rank-minimizing (URM) mechanism**. Of course, other tie-breaking rules—both with full support and without—are possible as well, and we discuss some of these below.

2.3 Non-obvious manipulability

Troyan and Morrill (2020) define non-obvious manipulability for deterministic direct revelation mechanisms. Informally, a mechanism is not obviously manipulable if, for every agent and every type $\succ_i$, and every possible misreport $\succ_i' \neq \succ_i$: (i) the worst-case outcome under $\succ_i$ is weakly better than the worst-case outcome under $\succ_i'$ and (ii) the best-case outcome under $\succ_i$ is weakly better than the best-case outcome under $\succ_i'$, where the worst and best cases are taken over all possible reports of the other agents, $\succ_{-i}$. They provide a characterization of non-obvious manipulations as those that cannot be recognized by a cognitively limited agent, and classify mechanisms as either obviously manipulable or not obviously manipulable in a wide variety of settings that is in line with empirical evidence.

Formally, Troyan and Morrill (2020)'s definition applies only to deterministic mechanisms, and so it must be extended to deal with the probabilistic mechanisms used in this paper. The natural extension is to simply treat Nature as another player, and calculate the worst and best possible outcomes over both all possible reports of the other agents $\succ_{-i}$ as well as all possible realizations of random draws by Nature.[7] This extension of Troyan and Morrill

[7]Non-obvious manipulability shares a similar motivation with the seminal paper of Li (2017) on obvious dominance (and indeed, the model of cognitive limitations used by Troyan and Morrill (2020) to characterize obvious manipulations is the same used by Li (2017) to characterize obvious dominance). In his paper, Li writes: "Weak dominance treats chance moves and other players asymmetrically...By contrast, obvious dominance treats chance moves and other players symmetrically". In the original model of Troyan and Morrill (2020), Nature (or, what Li (2017) refers to as "chance") does not play a role,

(2020) to random mechanisms first appears (to my knowledge) in Demeulemeester and Pereyra (2022), and the definition below is equivalent to theirs. When the mechanism itself is deterministic, it reduces to the definition of Troyan and Morrill (2020).

Formally, given a random allocation $\mu \in \mathcal{M}$, let $\text{supp}(\mu) = \{\alpha \in \mathcal{A} : \mu(\alpha) > 0\}$ be the **support** of μ and define:

$$\bar{\rho}_i(\mu) = \max_{\alpha \in \text{supp}(\mu)} r_i(\alpha_i)$$

$$\underline{\rho}_i(\mu) = \min_{\alpha \in \text{supp}(\mu)} r_i(\alpha_i).$$

That is, $\bar{\rho}_i(\mu)$ is the rank of i's least-preferred outcome among those that are selected by μ with strictly positive probability (recall that a higher rank corresponds to a worse school, and so in our context, the worst-case is given by taking the maximum). Similarly, $\underline{\rho}_i(\mu)$ is the rank of agent i's best outcome over all of the allocations in the support of μ.

Definition 1. A mechanism ψ is **not obviously manipulable (NOM)** if, for any agent i of type $\succ_i$ and any $\succ_i' \neq \succ_i$, the following are true:

(i) $\max_{\succ_{-i}} \bar{\rho}_i(\psi(\succ_i, \succ_{-i})) \leq \max_{\succ_{-i}} \bar{\rho}_i(\psi(\succ_i', \succ_{-i}))$

(ii) $\min_{\succ_{-i}} \underline{\rho}_i(\psi(\succ_i, \succ_{-i})) \leq \min_{\succ_{-i}} \underline{\rho}_i(\psi(\succ_i', \succ_{-i}))$

If either of (i) or (ii) does not hold for some agent and type, then $\succ_i'$ is an **obvious manipulation** for agent i of type $\succ_i$, and the mechanism ψ is said to be **obviously manipulable**.

To understand the definition of an obvious manipulation, first consider part (i). On the LHS of the inequality, $\bar{\rho}_i(\psi(\succ_i, \succ_{-i}))$ is the rank of i's worst-case outcome that might arise when the implemented random allocation is $\psi(\succ_i, \succ_{-i})$. We then take the maximum over $\succ_{-i}$. This process returns the worst-case outcome for i over all possible reports of other agents, $\succ_{-i}$, and all possible realizations of Nature. The RHS is the same, just replacing $\succ_i$ with a misreport $\succ_i'$. The inequality in (i) says that the rank of the

but here, I must incorporate it. The definition I use continues in the same spirit of Li (2017), by treating Nature and other players symmetrically.

worst-case outcome under the misreport should be weakly worse (i.e., weakly higher) than that under the truth. Part (ii) of Definition 1 is analogous, except it compares the best-case outcomes instead, and, since lower numbers correspond to more preferred outcomes, we replace max with min.

There are several justifications for why a designer might be concerned with obvious manipulations. As discussed in the introduction, just because a mechanism *can* be manipulated does not mean that it *will* be manipulated, and Definition 1 is a way to separate those manipulations that are "obvious", and are thus likely to be identified by participants, from those that are not. Formally, Theorem 1 of Troyan and Morrill (2020) shows that obvious manipulations are precisely those manipulations that can be recognized by an agent who is cognitively-limited in the sense defined by Li (2017), and is unable to contingently reason about outcomes state-by-state. Mathematically, such agents know the range of the function ψ conditional on their own reports, but not the full function itself, state-by-state.[8] Allowing some manipulations so long as they are non-obvious widens the space of mechanisms available to the designer, which may allow for improvements on other dimensions, such as the average rank. Further, unlike other relaxations of strategyproofness such as Bayesian incentive compatibility or approximate notions such as strategyproofness-in-the-large (SPL, Azevedo and Budish, 2019), NOM requires no assumptions on how preferences are drawn or agent beliefs. Rather, NOM is defined using best and worst case scenarios, which are likely to be particularly salient.[9]

[8] This is particularly relevant in the context of school choice. For instance, Troyan and Morrill (2020) write: "...this could be a neighborhood parent group that does not fully understand (or has not been told) the assignment algorithm but has kept track of what preferences have been submitted and what the resulting assignments were." Such parent groups are indeed quite common; see Pathak and Sönmez (2008).

[9] Following Troyan and Morrill (2020), several other papers have applied non-obvious manipulability to various settings, including Aziz and Lam (2021), Ortega and Segal-Halevi (2022), Archbold et al. (2022), Troyan et al. (2020), and Cerrone et al. (2023). Also similar in spirit, though technically different, is Li and Dworczak (2021) who show that a designer can sometimes be better off using a non-SP mechanism even when agents are unsophisticated, no matter how they resolve their "strategic confusion".

3 Results

In this section, I provide my main results on the non-obvious manipulability of RM mechanisms. An important distinction is whether the RM mechanism under consideration is a full-support mechanism or not. We start by considering full-support mechanisms, and then move to discuss other selection rules.

3.1 Full-support RM mechanisms

Recall that we say a rank-minimizing mechanism has full support if, for all $\succ_I$, $\psi(\succ_I)$ is a full-support RM allocation, or, equivalently, $\psi(\succ_I)(\alpha) > 0$ for all $\alpha \in \bar{\mathcal{A}}(\succ_I)$. In words, for each preference profile $\succ_I$, the mechanism selects every rank-minimizing allocation with strictly positive probability.

Theorem 1. *Let ψ be a rank-minimizing mechanism with full support. Then, ψ is not obviously manipulable.*

The proof of Theorem 1 can be found in the appendix. The full-support assumption requires only that the mechanism place non-zero probability on each rank-minimizing allocation, but otherwise the distribution can be arbitrary. One example is the mechanism that, for each preference profile $\succ_I$, selects an allocation uniformly at random from $\bar{\mathcal{A}}(\succ_I)$, the set of all rank-minimizing allocations. We call this mechanism the *uniform rank-minimizing (URM) mechanism.*

URM is a natural choice from the class of RM mechanisms, and is one that has received attention in the literature. Nikzad (2022) shows that URM is Bayesian incentive compatible in markets in which agent preferences are drawn iid and uniformly at random. Ortega and Klein (2023) also use the URM implementation of RM in both their theoretical results comparing RM to TTC and DA, as well as in their simulations.[10]

[10] An additional issue for implementing URM (or, any full-support RM mechanism) is computational: formally, the mechanism requires finding all allocations that minimize the average rank. While it is known that the problem of finding one RM allocation has complexity $O(n^3)$ (Krokhmal and Pardalos, 2009; Parviainen, 2004), I am not aware of any results on the computational complexity of finding all RM allocations, and am unable to speculate on the impact of this issue in practice. For instance, Ortega and Klein (2023) run simulations and counterfactual analysis using Hungarian school choice data, and use

By Theorem 1, URM is not obviously manipulable. Besides satisfying appealing efficiency (rank-minimizing) and strategic (NOM) properties, a final advantage of URM is that because it randomizes uniformly, it treats all agents fairly. This can be formalized as follows. Let $\psi_i(\succ_I)$ be agent i's lottery over objects she is assigned that is induced by the random allocation $\psi(\succ_I)$. A mechanism ψ satisfies **equal treatment of equals (ETE)** if for all $\succ_I \in \mathcal{P}^I$ and all $i, j \in I$, $\succ_i = \succ_j$ implies $\psi_i(\succ_I) = \psi_j(\succ_I)$.

Proposition 1. *The uniform rank-minimizing mechanism satisfies equal treatment of equals.*

The proof of this proposition is simple to see. Consider two agents i and j that have the same preferences. Because i and j have the same preferences, starting from any allocation and swapping their assignments results in no change in the average rank. In particular, for any rank-minimizing allocation $\alpha \in \text{supp}(\psi(\succ_I))$, there is another allocation $\alpha' \in \text{supp}(\psi(\succ_I))$ such that $\alpha'_i = \alpha_j$, $\alpha'_j = \alpha_i$, and $\alpha'_k = \alpha_k$ for all $k \neq i, j$. Since the URM mechanism selects each element in $\text{supp}(\psi(\succ_I))$ with equal probability, by symmetry, agents i and j will have the same lottery over final objects.

3.2 RM mechanisms without full support

While full support RM mechanisms—in particular, the URM mechanism—seem natural, and are always NOM, it is also possible to consider other RM mechanisms that violate this assumption, and so, it is also necessary to investigate whether the NOM property extends beyond full-support mechanisms. As I will show in this section, the answer is "it depends".

I start by showing that full support is not necessary for NOM implementation by showing that there exist deterministic RM mechanisms that are NOM.[11] In particular, Featherstone (2020) suggests the following tie-breaking

an algorithm that searches for one rank-minimizing allocation. While this is not exactly equivalent to URM, by perturbing the algorithm slightly, they are able to find many different rank-minimizing allocations, and they find that there is little variance across them, suggesting that this may not be a major issue in practice.

[11]By deterministic I mean that for each $\succ_I$, $\psi(\succ_I)(\alpha) = 1$ for one allocation α and $\psi(\succ_I)(\alpha') = 0$ for all $\alpha' \neq \alpha$.

procedure:[12] find the set of deterministic RM allocations $\bar{\mathcal{A}}(\succ_I)$, order the agents in some (exogenous) way, and run a serial dictatorship starting with $\bar{\mathcal{A}}(\succ_I)$. In the serial dictatorship phase, when each agent's turn comes, she finds the remaining allocation(s) that give her her most preferred object, and eliminates all others. We call this mechanism the *rank-minimizing serial dictatorship (RM-SD) mechanism*. It is obvious that the RM-SD mechanism always ends with a single deterministic allocation at the end of the serial dictatorship phase, and so this is a deterministic mechanism (in particular, it does not have full support).

Rank-Minimizing Serial Dictatorship (RM-SD)

For a set of allocations $A' \subseteq \mathcal{A}$ and an agent preference $\succ_i$, let

$$Top_{\succ_i}(A') = \{\alpha \in A' : \alpha_i \succsim_i \hat{\alpha}_i \text{ for all } \hat{\alpha}_i \in A'\}.$$

In words, $Top_{\succ_i}(A')$ consists of all of the allocations in the set A' at which agent i gets her top choice. The RM-SD mechanism is defined as follows.

- Fix a bijection $f : \{1, \ldots, N\} \to I$. This bijection produces an ordering of the agents, where agent $f(1)$ is first, agent $f(2)$ is second, etc.
- Given a reported preference profile $\succ_I$, let $\bar{\mathcal{A}}(\succ_I)$ be the set of rank-minimizing allocations at $\succ_I$. Initialize $A_0 = \bar{\mathcal{A}}(\succ_I)$.
- Consider agent $f(1)$, and calculate the set $A_1 = Top_{\succ_{f(1)}}(A_0)$.
- Consider agent $f(2)$, and calculate the set $A_2 = Top_{\succ_{f(2)}}(A_1)$
- etc.

The mechanism ends with a final set $A_N = Top_{\succ_{f(N)}}(A_{N-1})$, where $A_N = \{\alpha\}$. That is, A_N contains a unique allocation α. This allocation α is the final output of the RM-SD mechanism.

For the next result, I focus on **unit capacity markets** in which $|I| = |O| = N$, and $q_o = 1$ for all $o \in O$. When each object has unit capacity, the RM-SD mechanism is not obviously manipulable.

[12] Featherstone (2020) applies this procedure to his broader class of welfare-maximization mechanisms, which contains the class of RM mechanisms.

Proposition 2. *In unit capacity markets, the RM-SD mechanism is not obviously manipulable.*

The proof of this proposition is in the appendix. Proposition 2 shows that the full-support assumption of Theorem 1 is not necessary for RM to be NOM. However, the unit capacity assumption is needed to get this result, and if we do not require it, RM-SD is no longer NOM.

Proposition 3. *Assume that there are at least 3 objects and at least 4 agents, and at least one object has capacity $q_o > 1$. Then, RM-SD is obviously manipulable.*

The proof considers a market with 4 agents and 3 objects with capacities $q_1 = q_3 = 1$ and $q_2 = 2$, and focuses on an agent i with preferences $o_1 \succ_i o_2 \succ_i o_3$.[13] When agent i is second in the serial dictatorship stage, his favorite object o_1 might be taken by the first agent, and then the rank-minimizing constraints relegate him to o_3, his worst object. If he instead reports $o_2 \succ_1' o_1 \succ_1' o_3$, then, because o_2 has capacity 2 and agent i is second in the serial dictatorship, he is able to guarantee himself o_2, which is an obvious manipulation. The full details are in the appendix.

I close this section by noting that even in unit capacity markets, there exist RM mechanisms that are obviously manipulable.

Proposition 4. *Even in unit capacity markets, there exist rank-minimizing mechanisms (without full support) that are obviously manipulable.*

The proof of this proposition (also in the appendix) is by again example. I include this result because the mechanism works differently than the one used to prove Proposition 3, and I think is instructive. In particular, I consider a market of three agents $I = \{i, j, k\}$ and three objects $O = \{o_1, o_2, o_3\}$.[14] When all agents report the same preferences, say $o_1 \succ o_2 \succ o_3$, all allocations are rank-minimizing. Since we do not require full-support, the mechanism always selects a single allocation such that agent i receives her worst choice:

[13] The same proof can be easily embedded in larger markets. The key feature is that at least one object must have greater than unit capacity. Also, the proof of Proposition 2 above points out where the argument no longer holds when moving from unit capacity to more general markets.

[14] Again, this can be easily embedded in larger markets.

$\alpha_i = o_3$. However, when agent i reports a different preference, say $o_2 \succ_i' o_1 \succ_i' o_3$, the mechanism always selects an allocation in which agent i does *not* receive her worst choice: $\alpha_i' \neq o_3$. (The main work of the proof is to show that no matter the preferences of the other agents, it is always possible to find a rank-minimizing allocation where $\alpha_i' \neq o_3$.) Thus, if agent i reports her true type $o_1 \succ_i o_2 \succ_i o_3$, her worst case is o_3, while if she misreports $o_2 \succ_i' o_1 \succ_i' o_3$, her worst case is strictly better than o_3. Therefore, $\succ_i'$ is an obvious manipulation. for type $\succ_i$.

Notice that in this mechanism, agent i is treated differently than the other agents: effectively, the mechanism "protects" i from ever receiving her worst choice if she reports $\succ_i'$. Even though there may be rank-minimizing allocations at which i receives o_3, the mechanism never selects these, whereas a full-support mechanism sometimes would. Indeed, it seems intuitive that if a mechanism is protecting an agent from her worst choice at some preference profiles, but not at others, it will be "obvious" that this mechanism can be manipulated. Because agent i is treated differently than the others, this mechanism also violates the fairness condition of equal treatment of equals introduced in the previous subsection, and that is satisfied by the URM mechanism.[15]

The above results provide just a sampling of possible selection rules for RM, and there are of course many others. For instance, while RM-SD is obviously manipulable when objects can have capacity greater than 1, there could be other selection rules for which RM is NOM. While such a broad characterization is beyond the scope of this paper, the results here suggest that when going beyond full support RM mechanisms, the details of the selection rule will matter for the NOM properties of RM mechanisms.

4 Conclusion

This paper investigates conditions under which RM mechanisms, while not strategyproof, are at least not obviously manipulable. I show that as long as the mechanism has full support, RM will be NOM. In particular, the uniform

[15]The RM-SD mechanism also violates equal treatment of equals, as the agent who goes first in the serial dictatorship phase has an advantage relative to any later agents who submit the same preferences.

RM mechanism in which the designer selects the final allocation uniformly at random from among all of those allocations that minimize the average rank is an NOM mechanism. While this is quite a natural selection rule, and one that has been used by other papers in the literature on RM, there are instances in which a designer may have some interest in favoring some RM allocations over others, and thus may find it desirable to use a selection rule that does not have full-support. My results suggest that care should be taken when doing so. While such mechanisms may achieve the designer's secondary objectives while retaining the rank-minimizing property, they might also compromise the (non-)obvious manipulability of the resulting mechanism and ultimately undermine these objectives.

In sum, the rank-minimizing mechanism is appealing for assignment problems, because it directly optimizes a natural objective that is desirable to policy-makers. Thus, it is somewhat striking that there has been thus far relatively little written about this mechanism in the economics (and in particular, school choice) literature, while there have been perhaps hundreds of papers written about mechanisms such as DA and TTC. One possible explanation for this gap is that the literature is overly-focused on strategyproofness as an incentive property, which is satisfied by both DA and TTC. While strategyproofness is a very appealing desideratum, it limits the flexibility for a designer to optimize on other important dimensions. Considering weaker properties such as NOM allows access to a broader class of mechanisms, including (some) rank-minimizing ones. Whether such mechanisms will be manipulated in practice is ultimately an empirical question, but given my results and their other desirable properties, I argue that RM mechanisms are a class of mechanisms that is worthy of further investigation for practical market design applications.

References

Abdulkadiroğlu, A., Y.-K. Che, P. A. Pathak, A. E. Roth, and O. Tercieux (2020): "Efficiency, Justified Envy, and Incentives in Priority-Based Matching," *American Economic Review: Insights*, 2, 425–442.

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

Archbold, T., B. de Keijzer, and C. Ventre (2022): "Non-Obvious Manipulability for Single-Parameter Agents and Bilateral Trade," *arXiv preprint arXiv:2202.06660*.

Azevedo, E. M. and E. Budish (2019): "Strategy-proofness in the large," *The Review of Economic Studies*, 86, 81–116.

Aziz, H. and A. Lam (2021): "Obvious Manipulability of Voting Rules," in *International Conference on Algorithmic Decision Theory*, Springer, 179–193.

Budish, E. and E. Cantillon (2012): "The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard," *American Economic Review*, 102, 2237–71.

Budish, E. B. and J. B. Kessler (2017): "Can Agents "Report Their Types"? An Experiment that Changed the Course Allocation Mechanism at Wharton," *Chicago Booth Research Paper*.

Cerrone, C., Y. Hermstrüwer, and O. Kesten (2023): "School choice with consent: an experiment," *Economic Journal*, forthcoming.

Combe, J., O. Tercieux, and C. Terrier (2022): "The design of teacher assignment: Theory and evidence," *Review of Economic Studies, forthcoming*.

Delacrétaz, D., S. D. Kominers, and A. Teytelboym (2023): "Matching mechanisms for refugee resettlement," *American Economic Review*, 113, 2689–2717.

Demeulemeester, T. and J. S. Pereyra (2022): "Rawlsian Assignments," *arXiv preprint arXiv:2207.02930*.

Featherstone, C. R. (2020): "Rank efficiency: Investigating a widespread ordinal welfare criterion," Working paper.

Gale, D. and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69, 9–15.

Kesten, O. (2010): "School Choice with Consent," *Quarterly Journal of Economics*, 125, 1297–1348.

Krokhmal, P. A. and P. M. Pardalos (2009): "Random assignment problems," *European Journal of Operational Research*, 194, 1–17.

Li, J. and P. Dworczak (2021): "Are simple mechanisms optimal when agents are unsophisticated?" in *Proceedings of the 22nd ACM Conference on Economics and Computation*, 685–686.

Li, S. (2017): "Obviously Strategy-Proof Mechanisms," *American Economic Review*, 107, 3257–87.

Nikzad, A. (2022): "Rank-optimal assignments in uniform markets," *Theoretical Economics*, 17, 25–55.

Ortega, J. and T. Klein (2023): "The cost of strategy-proofness in school choice," *Games and Economic Behavior*, 141, 515–528.

Ortega, J. and E. Segal-Halevi (2022): "Obvious manipulations in cake-cutting," *Social Choice and Welfare*.

Parviainen, R. (2004): "Random assignment with integer costs," *Combinatorics, Probability and Computing*, 13, 103–113.

Pathak, P. A. and T. Sönmez (2008): "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," 98, 1636–1652.

Roth, A. E. and E. Peranson (1999): "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," 89, 748–780.

Sethuraman, J. (2022): "A note on the average rank of rank-optimal assignments," Working paper, Columbia University.

Shapley, L. and H. Scarf (1974): "On Cores and Indivisibility," *Journal of Mathematical Economics*, 1, 23–37.

Troyan, P., D. Delacrétaz, and A. Kloosterman (2020): "Essentially stable matchings," *Games and Economic Behavior*, 120, 370–390.

Troyan, P. and T. Morrill (2020): "Obvious manipulations," *Journal of Economic Theory*, 185, 104970.

Appendix: Proofs

Proof of Theorem 1

Let ψ^{RM} be a rank-minimizing mechanism with full support. I start with the following lemma.

Lemma 1. *Consider a preference profile* $\succ'_I$ *such that for* $\succ'_i = \succ'_j$ *for all* $i, j \in I$*, and wlog, let this preference ranking be* $\succ'_j$*:* $o_1, o_2, \ldots, o_M$*. Define* $m^* = \min\{m : \sum_{m'=1}^{m} q_{m'} \geq N\}$ *and* $O' = \{o_1, \ldots, o_{m^*}\}$*, and let* $A^* \subseteq \mathcal{A}$ *be the subset of allocations that satisfy:*

(I) All $o_m \in O' \setminus \{o_{m^*}\}$ *are assigned to exactly* q_m *agents.*

(II) Object o_{m^*} *is assigned to exactly* $N - \sum_{m'=1}^{m^*-1} q_{m'}$ *agents.*

Then, $\bar{\mathcal{A}}(\succ'_I) = A^*$*. Further, for any* $\tilde{o} \in O' = \{o_1, \ldots, o_{m^*}\}$*, there is at least one allocation* $\alpha \in \bar{\mathcal{A}}(\succ'_I)$ *such that* $\alpha_i = \tilde{o}$*.*

In words, object o_{m^*} is the critical object, in the sense that the set $O' = \{o_{m_1}, \ldots, o_{m^*}\}$ has enough total capacity to accommodate all agents, while the set $O' \setminus \{o_{m^*}\}$ does not. The set A^* is then the set of all allocations that consist of all possible ways of assigning the N agents to the objects in O' such that $o_{m_1}, \ldots, o_{m^*-1}$ are filled to capacity, and all remaining agents are assigned to o_{m^*}. The lemma says that when all agents have the same preferences, any rank-minimizing allocation must satisfy (I) and (II).

Proof of Lemma 1. Under the preference profile given, since all agents rank everything the same, the average rank of any allocation α is:

$$\bar{r}(\alpha) = \frac{1}{N}\left(\sum_{m=1}^{M} m|\{i \in I : \alpha_i = o_m\}|\right).$$

This is clearly minimized by any allocation that assigns q_1 students to o_1, q_2 students to o_2, etc., until all students are assigned. This is precisely the set of allocations A^*, and so by construction, any allocation $\alpha \in A^*$ achieves the minimum average rank. For any allocation $\alpha' \notin A^*$, either (I) or (II) must fail. If (I) fails, let $o_{\hat{m}} \in \{o_1, \ldots, o_{m^*-1}\}$ be an object that is not filled to capacity. Then, there must be some agent j assigned to some object $o_{m^*}, \ldots, o_M$. Reassigning agent j to object $o_{\hat{m}}$ and leaving all other assignments the same lowers the average rank, and so $\alpha' \notin \bar{\mathcal{A}}(\succ_I)$. Therefore, (I) must hold. If (II) fails, then once again, there is some agent j assigned to some object $o_{m^*+1}, \ldots, o_M$, and we can move this agent to object o_{m^*} and lower the average rank. Therefore, $\bar{\mathcal{A}}(\succ'_I) = A^*$. Lastly, because all agents have the same preferences, it does not matter precisely which agents are assigned to which objects, and so by symmetry, there exists at least one $\alpha \in \bar{\mathcal{A}}(\succ'_I)$ such that $\alpha_i = \tilde{o}$ for all $\tilde{o} \in O'$.

$\blacksquare$

Now, consider an agent i with type $\succ_i$. I show that for any $\succ_i' \neq \succ_i$, each part of Definition 1 holds for any rank-minimizing mechanism with full support.

Part (i): $\max_{\succ_{-i}} \bar{\rho}_i(\psi^{RM}(\succ_i, \succ_{-i})) \leq \max_{\succ_{-i}} \bar{\rho}_i(\psi^{RM}(\succ_i', \succ_{-i}))$**.**

Without loss of generality, index agent i's true type as $\succ_i$: $o_1, o_2, \ldots, o_M$. I first show that when i reports her true preferences, $\max_{\succ_{-i}'} \bar{\rho}_i(\psi^{RM}(\succ_i, \succ_{-i}'$ $)) = m^*$, where m^* is as defined in Lemma 1. When $\succ_j = \succ_i$ for all j, there is at least one allocation $\alpha \in \bar{\mathcal{A}}(\succ_I)$ such that $\alpha_i = o_{m^*}$, by Lemma 1. Since ψ^{RM} has full support, this implies $\max_{\succ_{-i}'} \bar{\rho}_i(\psi^{RM}(\succ_i, \succ_{-i}')) \geq m^*$. To show equality, assume that $\max_{\succ_{-i}} \bar{\rho}_i(\psi^{RM}(\succ_i, \succ_{-i})) > m^*$. Then, there must be some $\succ_{-i}'$ and some $\alpha' \in \bar{\mathcal{A}}(\succ_i, \succ_{-i}')$ such that $\alpha_i' = o_{m'}$ for some $m' > m^*$. This implies that there is some $m'' \leq m^*$ such that object $o_{m''}$ is not filled to capacity. Thus, consider an alternative allocation α'' where agent i is reassigned to $o_{m''}$ and all other agents have the same assignment as in α'. Then, we have $\bar{r}(\alpha'') < \bar{r}(\alpha')$, i.e., this lowers the average rank, which contradicts that $\alpha' \in \bar{\mathcal{A}}(\succ_i, \succ_{-i}')$. Therefore, $\max_{\succ_{-i}'} \bar{\rho}_i(\psi^{RM}(\succ_i, \succ_{-i}'$ $)) = m^*$.

Thus, I have shown that when i reports her true preferences, her worst-case outcome is $\max_{\succ_{-i}'} \bar{\rho}_{-i}(\psi^{RM}(\succ_i, \succ_{-i}')) = m^*$. What remains to show is that for any misreport $\succ_i' \neq \succ_i$, we have $\max_{\succ_{-i}'} \bar{\rho}_i(\psi^{RM}(\succ_i', \succ_{-i}')) \geq m^*$, where of course the maximum is evaluated with respect to i's true preferences. Consider a misreport $\succ_i' \neq \succ_i$. For notational purposes, index this preference profile as

$$\succ_i'\colon o_{r_1}, o_{r_2}, \ldots, o_{r_M}.$$

Let $m^{**} = \min\{m : \sum_{m'=1}^{m} q_{r_{m'}} \geq N\}$. Similar to Lemma 1, this is the index of the critical object in the sense that the set $O'' = \{o_{r_1}, \ldots, o_{r_{m^{**}}}\}$ has enough total capacity for all agents, but the set $\{o_{r_1}, \ldots, o_{r_{m^{**}-1}}\}$ does not. Consider a preference profile $\succ_I'$ such that $\succ_j' = \succ_i'$ for all $j \in I$. By Lemma 1, we have $\bar{\mathcal{A}}(\succ_I') = A^{**}$, where A^{**} is defined analogously to A^* in Lemma 1, replacing O' with the set O''.

Case 1: $O'' = O'$**.**

Since $o_{m^*} \in O' = O''$, by Lemma 1, there is at least one allocation $\alpha \in \bar{\mathcal{A}}(\succ_I')$ such that $\alpha_i = o_{m^*}$. Since ψ^{RM} has full support, $\max_{\succ_{-i}'} \bar{\rho}_i(\psi^{RM}(\succ_i'$

$, \succ'_{-i}) \geq m^*$, as desired.

Case 2: $O'' \neq O'$.

By definition, we cannot have $O'' \subsetneq O'$,[16] and so, if $O'' \neq O'$, there is some $o_m \in O''$ such that $m > m^*$, and thus $o_{m^*} \succ_i o_m$ according to i's true preferences $\succ_i$. By Lemma 1, there is at least one allocation $\alpha \in \bar{\mathcal{A}}(\succ'_I)$ such that $\alpha_i = o_m$. Because $o_{m^*} \succ_i o_m$ and ψ^{RM} has full support, this implies that $\max_{\succ'_{-i}} \bar{\rho}_i(\psi^{RM}(\succ'_i, \succ'_{-i}) > m^*$, as desired.

This completes the argument for part (i).

Part (ii): $\min_{\succ_{-i}} \underline{\rho}_i(\psi(\succ_i, \succ_{-i})) \leq \min_{\succ_{-i}} \underline{\rho}_i(\psi(\succ'_i, \succ_{-i}))$.

Without loss of generality, consider agent i whose preferences are $\succ_i$: $o_1, \ldots, o_M$. As in part (i), let $m^* = \min\{m : \sum_{m'=1}^{m} q_m \geq N\}$. Consider preference profile $\succ_{-i}$ for the other agents constructed as follows:

- Exactly $q_1 - 1$ agents have preferences such that $\succ_j$: $o_1, \ldots$.
- For all $m' = 2, \ldots, m^* - 1$, exactly $q_{m'}$ agents have preferences such that $\succ_j$: $o_{m'}, \ldots$.
- Exactly $N - \sum_{m'=1}^{m^*-1} q_{m'}$ agents have preferences such that $\succ_j$: $o_{m^*}, \ldots$.

In words, the constructed profile $\succ_I$ is such that each object o_m has exactly q_m agents who have ranked it first. This is possible by the definition of m^* and the assumption that $\sum_m q_m \geq N$. Now, notice that at this profile, there is a unique rank-minimizing allocation, $\psi^{RM}(\succ_I) = \{\alpha^*\}$, where α^* is the allocation such that each agent is assigned to her first-choice object. Thus, $\underline{\rho}_i(\psi(\succ_i, \succ_{-i})) = 1$, and so $\min_{\succ_{-i}} \underline{\rho}_i(\psi(\succ_i, \succ_{-i})) = 1$. Since it is obvious that $\underline{\rho}_i(\psi(\succ'_i, \succ_{-i})) \geq 1$ for any $(\succ'_i, \succ_{-i})$, we have $\min_{\succ_{-i}} \underline{\rho}_i(\psi(\succ_i, \succ_{-i})) \leq \min_{\succ_{-i}} \underline{\rho}_i(\psi(\succ'_i, \succ_{-i}))$, and therefore part (ii) of Definition 1 holds. ■

[16] This follows because $O' = \{o_1, \ldots, o_{m^*}\}$ where m^* is the smallest index such that O' contains enough capacity for all agents. If $O'' \subsetneq O'$, then O'' cannot contain enough capacity for all agents, which contradicts the definition of O''.

Proof of Proposition 2

Consider a market such that $|I| = |O| = N$, and $q_o = 1$ for all $o \in O$. Let ψ denote the RM-SD mechanism with fixed agent ordering $f(1), \ldots, f(N)$ in the SD stage. I start with the following claim.

Claim 1. Consider an agent, wlog labeled agent 1, with preferences $o_1 \succ_1 o_2 \succ_1 \cdots \succ_1 o_N$.[17] For all objects $o \neq o_N$, there exists a preference profile for the remaining agents $\succ_{-1}$ such that there is a unique rank-minimizing allocation α at $\succ_I = (\succ_1, \succ_{-1})$, and at this allocation $\alpha_1 = o$.

In other words, this claim says that, for any preferences the agent submits, under any rank-minimizing mechanism, the agent could receive any object, with the possible exception of the object she ranks last.

Proof of Claim 1. Wlog, let $o = o_m$, where $m \leq N-1$. Consider a profile of preferences defined as follows:

$\succ_1$	$\succ_2$	$\succ_3$	$\cdots$	$\succ_m$	$\succ_{m+1}$	$\cdots$	$\succ_{N-1}$	$\succ_N$
$\vdots$	$\boxed{o_1}$	$\boxed{o_2}$		$\boxed{o_{m-1}}$	$\boxed{o_{m+1}}$		$\boxed{o_{N-1}}$	$\boxed{o_N}$
$\vdots$								
$\boxed{o_m}$	$\vdots$	$\vdots$	$\cdots$			$\cdots$	$\vdots$	$\vdots$
$\vdots$								
o_N	o_m	o_m		o_m	o_m		o_m	o_m

In other words, all agents besides agent 1 rank o_m last, and each of these $N-1$ agents has a distinct favorite object in the set $O \setminus \{o_m\}$ (the dots indicate that the remaining parts of the preference profile can be arbitrary). Let the allocation in boxes be denoted by α.

For any allocation α', let $R(\alpha') = \sum_{i=1}^{N} r_i(\alpha'_i)$ be the total sum of ranks.[18] We claim that α in the boxes is the unique rank-minimizing allocation for

[17] Agent 1 need not be the first agent in the SD ordering, i.e., $f(1) \neq 1$. Further, the indexing of the true preference ordering as $o_1 \succ_1 o_2 \succ_1 \cdots$ is without loss of generality. The same arguments will apply to any agent with any true preference ordering.

[18] Obviously, minimizing average rank is equivalent to minimizing the total sum of ranks. I work with the latter here to avoid having to carry around the $1/N$ notation everywhere.

this preference profile. To see this, note that

$$R(\alpha) = \overbrace{m}^{\text{Agent 1}} + \overbrace{1 + 1 \cdots + 1}^{N-1 \text{ times}} = m + (N-1)$$

It is immediate to see that any alternative allocation α' in which $\alpha'_1 = o_m$ must have $R(\alpha') > R(\alpha)$, because agent 1's rank remains unchanged, and the total sum of ranks of agents $2, \ldots, N$ is clearly minimized by giving them all their first choice. Next, consider an allocation α' such that $\alpha'_1 = o_{m'}$ for some $m' > m$. At this allocation, $r_1(\alpha'_1) > r_1(\alpha_1)$, and $r_j(\alpha'_j) \geq 1 = r_j(\alpha_j)$ for all $j \neq 1$, and thus $R(\alpha') > R(\alpha)$.

Finally, consider an allocation such that $\alpha'_1 = o_{m'}$ for some $m' < m$. We can write:

$$R(\alpha') - R(\alpha) = \sum_{i=1}^{N} (r_i(\alpha'_i) - r_i(\alpha_i)).$$

Notice that $r_1(\alpha'_1) - r_1(\alpha_1) = m' - m$. Since some agent j must be assigned to o_m, and all of the other agents rank it last, we have for the agent j that receives o_m, $r_j(\alpha'_j) - r_j(\alpha_j) = N-1$. For all other agents, $r_k(\alpha'_k) - r_k(\alpha_k) \geq 0$ (since $r_k(\alpha_k) = 1$ and $r_k(\alpha'_k) \geq 1$). Thus, we have

$$R(\alpha') - R(\alpha) = m' - m + (N-1) + \sum_{k \neq i,j} (r_k(\alpha'_k) - r_k(\alpha_k))$$

The last summation is bounded below by 0, so $R(\alpha') - R(\alpha) \geq m' - m + (N-1) \geq 2 - N + (N-1) = 1$,[19] i.e., $R(\alpha') > R(\alpha)$. Thus, α is the unique rank-minimizing allocation, and at this allocation, $\alpha_1 = o_m$.

■

Note that Claim 1 does *not* apply to o_N (or, more generally, the object that agent 1 ranks last). In particular, the preferences used in the proof no longer work, because when all agents rank the same object last, there might be multiple rank-minimizing allocations.

Now, consider an agent 1 with true preferences $o_1 \succ_1 o_2 \succ_1 \cdots \succ_1 o_N$.

[19] The second inequality follows because $m \leq N-1$ and $m' \geq 1$, which implies that $m' - m \geq 2 - N$.

By Claim 1, we have that

$$\max_{\succ_{-1}} \bar{\rho}_1(\psi(\succ_1, \succ_{-1})) \geq N - 1. \tag{1}$$

Consider a false report $\succ_1'$, and index this report as $p_1 \succ_1' p_2 \succ_1' \cdots \succ_1' p_N$ (note that p_m—the m^{th} ranked object for $\succ_1'$—may be different from o_m, the m^{th} ranked object for the true preferences $\succ_1$). By Claim 1 applied to $\succ_1'$, for all $o \neq p_N$ there exist preference profiles $\succ_{-1}'$ such that there is a unique rank-minimizing allocation α at $(\succ_1', \succ_{-1}')$, and at this allocation, $\alpha_1 = o$. In particular, this implies that either o_{N-1} or o_N is a possible outcome for agent 1, and thus $\max_{\succ_{-1}} \bar{\rho}_1(\psi(\succ_1', \succ_{-1})) \geq N - 1$. If equation 1 holds with equality, then the proof is complete, as $\max_{\succ_{-1}} \bar{\rho}_1(\psi(\succ_1', \succ_{-1})) \geq N - 1 = \max_{\succ_{-1}} \bar{\rho}_1(\psi(\succ_1, \succ_{-1}))$.

Thus, we need to last consider the case that equation (1) is a strict inequality, which means that $\max_{\succ_{-1}} \bar{\rho}_1(\psi(\succ_1, \succ_{-1})) = N$. In this case, when agent 1 submits her true preferences $o_1 \succ_1 \cdots \succ_1 o_N$, there is some preference profile $\succ_{-1}$ where she might receive object o_N. We must show that, for any $\succ_1'$ that agent 1 might submit, there exists some $\succ_{-1}'$ where she receives object o_N. If $\succ_1'$ ranks o_N anything other than last, then, as in the previous paragraph, we can apply Claim 1 to the preference $\succ_1'$ and conclude there exists some $\succ_{-1}'$ such that agent 1 receives o_N for sure, and we are done.

Thus, consider the case that $\succ_1'$ ranks o_N last. Let $\succ_{-1}$ be the preference profile of the remaining agents such that, when 1 reports the truth, agent 1 receives o_N. In the RM-SD mechanism at $(\succ_1', \succ_{-1})$, when it is agent 1's turn to choose in the serial dictatorship step, it must be that all remaining allocations she can select from assign her to o_N (as otherwise, she would eliminate these allocations and choose a better object for herself). In particular, this implies that $f(1) > 1$, i.e., agent 1 cannot select first.[20]

Index agent 1's false report $\succ_1'$ as:

$$\succ_1' \colon p_1, \ldots, p_N.$$

[20]Indeed, if $f(1) = 1$, then it must be that agent 1 receives o_N at all allocations in $\bar{\mathcal{A}}(\succ_I')$, and thus some other agent j receives o_1 at some allocation $\alpha \in \bar{\mathcal{A}}(\succ_I')$. Let α' denote the assignment at which agent 1 and agent j swap, and all other agents' assignments remain the same. Then, α' cannot have a worse average rank than α, and so $\alpha' \in \bar{\mathcal{A}}(\succ_I')$ as well, and so agent 1 would eliminate α if she chose first.

$\succ'_1$	$\succ'_2$	$\succ'_3$	$\cdots$	$\succ'_N$
p_1	$\boxed{p_1}$	$\boxed{p_2}$		$\boxed{p_{N-1}}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$\boxed{p_N}$	p_N	p_N		p_N

Table 1: Preference profile used for proving part of Proposition 2. Recall that $p_N = o_N$. The boxed allocation represents the outcome of the RM-SD mechanism at this preference profile.

Notice that p_n need not be equal to o_n; we use p's to convey that this is a report that ranks the objects differently than the true report (which was indexed $\succ_1$: $o_1, \ldots, o_N$) while still being able to make easy reference to the n^{th} ranked object. However, we are in the case that $\succ'_1$ ranks o_N—agent 1's worst object according to her true preferences—last, so $p_N = o_N$.

Let $k = f(1)$ the order in which agent 1 chooses in the SD phase. By footnote 20, $k > 1$ and, without loss of generality, assume that the (exogenous) SD ordering of the agents is $2, 3, \ldots, k, 1, k+1, \ldots, N$ (in other words, the ordering not including agent 1 is just $2, 3, 4, \ldots$, and agent 1 is slotted in the k^{th} position). Consider a preference profile $\succ'_I = (\succ'_1, \succ'_{-1})$ that takes the form shown in Table 1.

Claim 2. For any rank-minimizing allocation $\alpha \in \bar{\mathcal{A}}(\succ'_I)$, object p_N is assigned to either agent 1 or agent 2.

Proof. First, notice that every agent ranks p_N last, and it must be given to someone. Thus, the lowest possible total sum of ranks, $N + (N-1) \times 1 = 2N - 1$, and indeed, this is achievable by, for instance, the allocation in boxes in the table. What we show is that at any allocation α' such that $\alpha_j = p_N$ for some $j \neq 1, 2$, $R(\alpha') > 2N - 1$. Write

$$\begin{aligned}
R(\alpha') &= (r'_1(\alpha'_1) + r'_2(\alpha'_2)) + (r'_j(\alpha'_j)) + \left(\sum_{i \neq 1,2,j} r'_i(\alpha'_i)\right) \\
&\geq (3) + (N) + (N-3) \\
&= 2N \\
&> 2N - 1.
\end{aligned}$$

The first inequality follows because (i) only one of agent 1 or 2 can be getting their first choice,[21] and so $r'_1(\alpha'_1)+r'_2(\alpha'_2) \geq 3$; (ii) by assumption, $r'_j(\alpha'_j) = N$ and (iii) $\sum_{i\neq 1,2,j} r'_i(\alpha'_i) \geq N-3$. ■

Now, agent 2 chooses ahead of agent 1 in the serial dictatorship phase of the mechanism, and so will eliminate all allocations in $\bar{\mathcal{A}}(\succ'_I)$ that give her p_N. Thus, when it is agent 1's turn to choose, all remaining allocations are such that $\alpha_1 = p_N$. Recalling that $p_N = o_N$, we have shown that for any false report $\succ'_1$: $p_1, \ldots, p_N$, there is a preference profile of the other agents $\succ'_{-1}$ such that at $(\succ'_1, \succ'_{-1})$ agent 1 gets o_N.

We have therefore shown that, for all $\succ'_1$, $\max_{\succ_{-1}} \bar{\rho}_1(\psi(\succ'_1, \succ_{-1})) \geq \max_{\succ_{-1}} \bar{\rho}_1(\psi(\succ_1, \succ_{-1}))$, which is part (i) of the definition of NOM. Part (ii) follows trivially: when all agents have a unique first choice, the unique rank-minimizing allocation assigns them each their first choice. So, for any true preferences $\succ_1$: $o, o', o'', \cdots$, $\min_{\succ_{-1}} \underline{\rho}_1(\psi(\succ_1, \succ_{-1})) = 1$, and it is immediate that $\min_{\succ_{-1}} \underline{\rho}_1(\psi(\succ_1, \succ_{-1})) \leq \min \underline{\rho}_1(\psi(\succ'_1, \succ_{-1}))$. ■

Proof of Proposition 3

Let $I = \{i, j, k, \ell\}$, $O = \{o_1, o_2, o_3\}$ and $q_1 = q_3 = 1$, while $q_2 = 2$. Consider a serial dictatorship ordering that is j, i, k, ℓ. Let agent i's true preferences be $\succ_i$: o_1, o_2, o_3, and consider a preference profile given in the following table:

$\succ_i$	$\succ_j$	$\succ_k$	$\succ_\ell$
o_1	$\boxed{o_1}$	$\boxed{o_2}$	$\boxed{o_2}$
o_2	o_2	o_1	o_1
$\boxed{o_3}$	o_3	o_3	o_3

At these preferences, the boxed allocation, denoted α, is easily seen to be rank-minimizing. Since j chooses first in the SD phase of the mechanism, she will eliminate all $\alpha' \in \bar{\mathcal{A}}(\succ_I)$ such that $\alpha'_j \neq o_1$.

Claim 3. Let $\alpha' \in \bar{\mathcal{A}}(\succ_I)$ be such that $\alpha'_j = o_1$. Then, $\alpha'_i = o_3$.

To see this claim, simply note that if $\alpha'_i = o_2$, then one of k or ℓ must be assigned o_3, and so the total sum of ranks at α' is 7. The allocation α denoted

[21] This is the point at which the proof breaks down for the general case (beyond unit capacity). In the general case, it is possible that the first choice of agent 1 and 2 has more than one unit of capacity, and so they both might receive it.

in boxes has total sum of ranks equal to 6, and so α' is not rank-minimizing, which is a contradiction.

The upshot of the above claim is that after j moves in the serial dictatorship phase of the mechanism, the only allocations remaining for i to choose from at the second step have $\alpha'_i = o_3$. Thus, we have $\max_{\succ_{-i}} \bar{\rho}(\psi(\succ_i, \succ_{-i})) = 3$.

Consider a false report $\succ'_i$: o_2, o_1, o_3. We claim that $\max_{\succ_{-i}} \bar{\rho}(\psi(\succ'_i, \succ_{-i})) < 3$. Assume not, i.e., $\max_{\succ_{-i}} \bar{\rho}(\psi(\succ'_i, \succ_{-i})) = 3$, and choose some $\succ'_{-i}$ such that $\bar{\rho}(\psi(\succ'_i, \succ'_{-i})) = 3$. Let the set of available allocations when it is i's turn to choose in the serial dictatorship be A'. Since $\bar{\rho}(\psi(\succ'_i, \succ'_{-i})) = 3$, this implies that for all $\alpha' \in A'$, we have $\alpha_i = o_3$, which implies that both o_1 and o_2 are filled to capacity with other agents. In particular, at all $\alpha' \in A'$, there are two agents assigned to o_2, and, since i chooses 2nd in the serial dictatorship, at least one of these agents must be ordered after i. Assume this agent is k, i.e., $\alpha'_k = o_2$ (the same argument works for agent ℓ). Now, agent k must rank $o_2 \succ_k o_3$, which means that the preference profile must be one of the following:

$\succ'_i$	$\succ'_k$
$*o_2$	o_1
o_1	$\boxed{o_2}$
$\boxed{o_3}$	$*o_3$

$\succ'_i$	$\succ'_k$
$*o_2$	$\boxed{o_2}$
o_1	$*o_3$
$\boxed{o_3}$	o_1

$\succ'_i$	$\succ'_k$
$*o_2$	$\boxed{o_2}$
o_1	o_1
$\boxed{o_3}$	$*o_3$

.

In the tables, the boxes denote the assignment α', while the stars denote an alternative assignment α^* where i and j swap their objects, and everything else remains unchanged (we do not show the assignments of the other agents in the tables). Notice that in the left two panels, swapping the assignments of i and j strictly lowers the average rank, which means that in fact, the preferences must be that in the right panel.

To summarize: the preferences of i and k are in the following table, and the assignment α' in boxes is such that $\alpha' \in A'$:

$\succ'_i$	$\succ'_k$
$*o_2$	$\boxed{o_2}$
o_1	o_1
$\boxed{o_3}$	$*o_3$

Since $\alpha' \in A'$, this means that α' is rank-minimizing at $\succ'_I$. As α^* has the same average rank as α', α^* is also rank-minimizing at $\succ'_I$. Further, since the assignments of all other agents remain the same, α^* is not eliminated by agent j at the first step of the serial dictatorship. Thus, α^* is in i's opportunity set when it is her turn to choose, and thus she would choose it and receive o_2, which contradicts that $\max_{\succ_{-i}} \bar{\rho}(\psi(\succ'_i, \succ_{-i})) = 3$. ■

Proof of Proposition 4

Proof. The proof is by example in a market with three students $I = \{i, j, k\}$ and three objects $O = \{o_1, o_2, o_3\}$. We will build a RM mechanism ψ that is obviously manipulable. To start, consider the following profile of preferences, $\succ^1_I$:

$\succ^1_i$	$\succ^1_j$	$\succ^1_k$
o_1	o_1	$\boxed{o_1}$
o_2	$\boxed{o_2}$	o_2
$\boxed{o_3}$	o_3	o_3

Denote the allocation in the boxes by α. Notice that all agents have the exact same preferences, so any allocation minimizes the average rank, and a rank-minimizing mechanism can select any allocation at this preference profile. In particular, we set $\psi(\succ^1_I)(\alpha) = 1$, and $\psi(\succ^1_I)(\alpha') = 0$ for all $\alpha' \neq \alpha$.

Next, consider the following preferences for agent i, $\succ^2_i$:

$\succ^2_i$
o_2
o_1
o_3

Claim 4. For all $\succ'_{-i}$, there exists an α' such that (i) α' is rank-minimizing with respect to $(\succ^2_i, \succ'_{-i})$ and (ii) $\alpha'_i \neq o_3$.

Proof of Claim 4. Assume not, i.e., there exists a preference profile $\succ'_{-i}$ such that all rank-minimizing allocations with respect to $(\succ^2_i, \succ'_{-i})$ assign i to o_3. Let α' be one such rank-minimizing allocation, and thus $\alpha'_i = o_3$. Under α', either agent j or k must be assigned to o_2; wlog, assume that $\alpha'_j = o_2$. Notice that j must rank $o_2 \succ_j o_3$; if not, then α' is not Pareto efficient (i and j

can engage in a Pareto-improving trade), and so is also not rank-minimizing. Thus, the overall preference profile must be one of the following (where k receives o_1 by construction, but k's actual preferences do not matter for the argument, and so are indicated by dots):

$\succ_i^2$	$\succ_j'$	$\succ_k'$
$*o_2$	o_1	$\vdots$
o_1	$\boxed{o_2}$	
$\boxed{o_3}$	$*o_3$	

$\succ_i^2$	$\succ_j'$	$\succ_k'$
$*o_2$	$\boxed{o_2}$	$\vdots$
o_1	$*o_3$	
$\boxed{o_3}$	o_1	

$\succ_i^2$	$\succ_j'$	$\succ_k'$
$*o_2$	$\boxed{o_2}$	$\vdots$
o_1	o_1	
$\boxed{o_3}$	$*o_3$	

The boxes in each table indicate the allocation α', and the stars indicate the alternative allocation where i and j swap: $\alpha_i^* = o_2$, $\alpha_j^* = o_3$, and $\alpha_k^* = \alpha_k' = o_1$. Notice that in the two leftmost panels, $\bar{r}(\alpha^*) < \bar{r}(\alpha')$, which contradicts that α' was rank-minimizing. In the rightmost panel, $\bar{r}(\alpha^*) = \bar{r}(\alpha')$. Since α' was assumed to be rank-minimizing, α^* is also rank-minimizing. However, this contradicts that all rank-minimizing allocations assign i to o_3, and completes the proof of the claim.

■

The upshot of Claim 4 is that we can construct an RM mechanism such that, for all $\succ_{-i}'$ and all α' such that $\alpha_i' = o_3$, we have $\psi(\succ_i^2, \succ_{-i}')(\alpha') = 0$.[22] So, consider agent i of type $\succ_i^1$. If she reports her true type, she receives o_3 with probability 1, while if she reports $\succ_i^2$, she receives o_3 with probability 0, and thus, receives some strictly preferred object with probability 1. Since the worst case from reporting $\succ_i^2$ is strictly preferred to o_3, while the worst case from reporting $\succ_i^1$ (the truth) is o_3, this is an obvious manipulation, and mechanism ψ is obviously manipulable.

■

[22] This follows because, by the claim, there is always at least one alternative rank-minimizing assignment where i does not receive o_3, and so we can build the mechanism to only put positive probability on such assignments.