The Power of Local Manipulation Strategies in Assignment Mechanisms

Timo Mennle and **Michael Weiss** and **Basil Philipp** and **Sven Seuken**
Department of Informatics
University of Zurich
{mennle,seuken}@ifi.uzh.ch
{michael.weiss2,basil.philipp}@uzh.ch

Abstract

We consider three important, non-strategyproof assignment mechanisms: *Probabilistic Serial* and *two variants* of the *Boston mechanism*. Under each of these mechanisms, we study the agent's *manipulation problem* of determining a best response, i.e., a report that maximizes the agent's expected utility. In particular, we consider *local manipulation strategies*, which are simple heuristics based on local, greedy search. We make three main contributions. First, we present results from a behavioral experiment (conducted on Amazon Mechanical Turk) which demonstrate that human manipulation strategies can largely be explained by local manipulation strategies. Second, we prove that local manipulation strategies may fail to solve the manipulation problem optimally. Third, we show via large-scale simulations that despite this non-optimality, these strategies are very effective *on average*. Our results demonstrate that while the manipulation problem may be hard in general, even cognitively or computationally bounded (human) agents can find near-optimal solutions almost all the time via simple local search strategies.

1 Introduction

We consider the *assignment problem*, i.e., the problem of allocating indivisible objects to self-interested agents with private preferences when monetary transfers are not permitted. The agents first report ordinal preferences, i.e., rank-ordered lists of objects, and then an *assignment mechanism* determines a (probabilistic) allocation of the objects to the agents. In practice, such problems often arise in situations that are of great importance to people's lives. Prominent examples include the assignment of teachers to schools via the *Teach for America* program, the assignment of MBA students to internships [Featherstone, 2011], and the assignment of children to public schools [Abdulkadiroğlu and Sönmez, 2003].

The assignment domain is plagued with impossibility results: *Random Serial Dictatorship* is strategyproof, anonymous, and ex-post efficient, but it is essentially the only mechanism with these properties [Bade, 2014]. Bogomolnaia and Moulin [2001] have proposed the *Probabilistic Serial (PS)* mechanism as an alternative because of its superior ordinal efficiency properties, but it is not strategyproof. In practice, other non-strategyproof mechanisms like the *Naïve* or the *Adaptive Boston Mechanism (NBM & ABM)* are employed, even though researchers have well documented that these mechanisms are not strategyproof and are being manipulated by participants [Calsamiglia and Güell, 2014].

The agent's *manipulation problem* is defined as the problem of determining a beneficial misreport in a full information environment, where the agent has access to a black-box implementation of the mechanism. In the voting domain, Bartholdi *et al.* [1989a] proposed computational complexity as a barrier to manipulation by cognitively or computationally bounded agents. While Xia [2011] gave worst-case complexity results for many voting rules, Mossel and Rácz [2014] recently showed that the manipulation problem is easy *on average* for essentially all non-strategyproof voting rules. For assignment mechanisms, the manipulation problem was recently shown to be NP-hard under PS [Aziz *et al.*, 2015b], but results for other mechanisms remain outstanding.

In this paper, we study the manipulation problem under PS, NBM, and ABM. Our approach is orthogonal to theoretical research on the computational complexity of the manipulation problem. Instead, we focus on *local manipulation strategies*, which are simple, behaviorally motivated heuristics based on local, greedy search algorithms. We conducted a behavioral experiment on Mechanical Turk to identify the way in which humans approach the manipulation problem. We then analyzed how manipulable PS, NBM, and ABM are in general, and how successful local manipulation strategies are at solving the manipulation problem under each mechanism. Our research addresses the following two hypotheses:

- **Hypothesis #1:** Human manipulation strategies can largely be explained by local manipulation strategies.
- **Hypothesis #2:** Local manipulation strategies are not always optimal, but very powerful on average.

Overview of Contributions

1. We present results from an online behavioral experiment to understand the manipulation strategies that human subjects employ when facing the manipulation problem. We find that human manipulation strategies can largely be explained by local, greedy search strategies, in line with hypothesis #1.

2. We prove that local manipulation strategies may fail to solve the manipulation problem optimally, in line with research hypothesis #2.
3. We provide large-scale simulation results which show that the situations where local strategies fail are the exception. On average, local manipulation strategies capture a large share of the possible gain from optimal manipulation, again in line with hypothesis #2.

2 Related Work

In the field of assignment mechanisms, Zhou [1990] showed that strategyproofness, ex-ante efficiency, and symmetry are incompatible. Bogomolnaia and Moulin [2001] showed that PS is ordinally efficient, but not strategyproof and that in fact no ordinally efficient and symmetric mechanism can be strategyproof. Boston Mechanisms are widely used in practice [Abdulkadiroğlu *et al.*, 2005], and multiple studies have shown that they are often manipulated [Abdulkadiroğlu *et al.*, 2006]. Naïve and adaptive variants of the Boston mechanism (NBM & ABM) were found to facilitate a subtle trade-off between strategyproofness and efficiency [Mennle and Seuken, 2014b]. In this paper, we study PS, NBM, and ABM, which are the three most important non-strategyproof mechanisms.

Lab experiments have repeatedly shown that models assuming bounded rationality are better at predicting behavior than the perfectly-rational agent model [Gabaix *et al.*, 2006]. For example, [Hugh-Jones *et al.*, 2014] showed that humans do not manipulate PS optimally. Halpern *et al.* [2014] provided a nice survey of work using bounded rationality in decision theory. The observation that (human) agents are computationally and cognitively bounded motivates our definition of *local manipulation strategies*, which are heuristics for the manipulation problem, based on local, greedy search.

Since Bartholdi *et al.* [1989b] proposed computational complexity as an obstacle against manipulation, we have obtained a good understanding of the worst-case difficulty of the manipulation problem under various voting rules [Xia, 2011]. However, the standard worst-case notion of complexity does not prevent manipulation strategies from being effective *on average*. A recent stream of work [Xia and Conitzer, 2008; Dobzinski and Procaccia, 2008; Friedgut *et al.*, 2011; Isaksson *et al.*, 2012; Mossel and Rácz, 2014] established that the average complexity of determining some strictly beneficial manipulation is polynomial under almost all non-strategyproof voting rules. They showed that beneficial misreports are sufficiently frequent, so that in expectation *random search* finds such misreports in polynomial time. For assignment mechanisms, our local manipulation strategies prescribe a *particular heuristic* search strategy.

3 The Model

Let N be a set of n *agents* and M a set of m *objects*, and let there be $q \geq 1$ copies of each object available. Agents each demand one copy of some object and have private *preference orders* $\succ_i$ over the objects, where $j \succ_i j'$ indicates that agent i likes object j better than object j'.

An *allocation* is a (possibly probabilistic) assignment of the objects to the agents. It is represented by an $n \times m$-matrix $x = (x_{i,j})_{i \in N, j \in M}$ with $x_{i,j} \in [0,1]$ that satisfies

$$\begin{array}{ll} \sum_{i \in N} x_{i,j} \leq q, & \text{(capacity constraint)} \\ \sum_{j \in M} x_{i,j} = 1. & \text{(fulfillment constraint)} \end{array}$$

The entry $x_{i,j}$ is interpreted as the *probability that i gets j*.

Agents' preferences over objects are extended to preferences over probabilistic allocations via vNM utilities: each agent is endowed with a *utility function* $u_i : M \to \mathbb{R}^m$ *consistent* with the preference order $\succ_i$, i.e., $u_i(j) > u_i(j') \Leftrightarrow j \succ_i j'$. For an allocation x, agent i's *expected utility* (or just *utility*) is given by

$$\mathbb{E}_x [u_i] = \textstyle\sum_{j \in M} u_i(j) \cdot x_{i,j}. \quad (1)$$

P denotes the set of all preference orders, and X denotes the set of all possible allocations. A *mechanism* is a function

$$f : P^n \to X \quad (2)$$

that receives a *preference profile* $\succ = (\succ_1, \ldots, \succ_n)$ as input and selects an allocation $f(\succ) \in X$. We use $\succ_{-i}$ to denote the preference reports from all agents except i. A mechanism that makes truthful reporting a dominant strategy is called *strategyproof*, otherwise it is called *manipulable*.

3.1 Popular Mechanisms

We consider three well-known assignment mechanisms, which we briefly describe in the following. Formal definitions can be found in the referenced literature.

Probabilistic Serial Mechanism: The *Probabilistic Serial (PS)* mechanism was introduced by Bogomolnaia and Moulin [2001]. It collects the agents' preference reports and uses the *simultaneous eating algorithm* to determine an allocation: first, all agents begin collecting probability shares of their reported first choice at equal speeds. Once all shares of an object are exhausted the agents at this object move to their reported second choices and continue collecting probability shares there. This continues with third, fourth, etc. choices until all agents have collected a total of 1.0 probability.

PS is ordinally efficient, which is a true refinement of ex-post efficiency, but it is not strategyproof.

Naïve Boston Mechanism: The *Naïve Boston mechanism (NBM)* collects the agents' preference reports and determines a single *tie-breaker*, i.e., a linear ordering of the agents, uniformly at random. In the first round, all agents "apply" to their reported first choices. Objects are assigned to the applicants, where preference is given to agents with higher rank in the tie-breaker whenever there are more applicants than available capacity. Then the process repeats, i.e., agents who were not assigned an object in the kth round enter the $k+1$st round, where they apply to their $k+1$st choice. The resulting allocation is probabilistic, because the tie-breaker is random and unknown when the agents submit their preferences.

NBM is frequently used in school choice and has been heavily criticized for its manipulability [Ergin and Sönmez, 2006; Kojima and Ünver, 2014]. However, it also has some appealing welfare properties [Abdulkadiroğlu *et al.*, 2015]. The mechanism is "naïve" in the sense that agents might "waste" rounds by applying to exhausted objects.

Adaptive Boston Mechanism: The *Adaptive Boston mechanism (ABM)* works similarly to NBM, except that agents apply to their best *available* object in each round [Mennle and Seuken, 2014b]. Suppose an agent reports $a \succ b \succ c$, and that a and b are exhausted in the first round, but the agent did not get a. Then its application in the second round will be to c, because b is no longer available. This modification eliminates some obvious opportunities for manipulation that exist under NBM. While ABM is not fully strategyproof, it satisfies the relaxed requirement of partial strategyproofness which NBM does not satisfy [Mennle and Seuken, 2014a].

3.2 The Agent's Manipulation Problem

Given agents N, objects M, and capacity q, the *situation* that an agent i faces is described by the tuple $(u_i, \succ_{-i})$, i.e., the utility of agent i and the reports from the other agents. We consider the problem of deciding on a report, given this information and a black-box implementation of the mechanism.

Definition 1. (Manipulation Problem) The agent's *manipulation problem* is the problem of finding a report $\succ_i^*$ such that i's utility in situation $(u_i, \succ_{-i})$ is as high as possible.

4 Manipulation Strategies

Manipulation strategies are algorithms to solve the manipulation problem. We focus on manipulation strategies that rely on sequential evaluation of reports, i.e., the strategy devises a way to (fully or partially) search the space of possible reports.

Definition 2. (OPT) An agent follows an *optimal manipulation strategy* (OPT) if it considers all possible reports, computes the allocation for each, and selects a report that yields the highest utility, i.e., in a given situation $(u_i, \succ_{-i})$,

$$\text{OPT}(u_i, \succ_{-i}) = \arg\max_{\succ_i' \in P} \mathbb{E}_{f(\succ_i', \succ_{-i})}[u_i]. \quad (3)$$

By definition OPT yields maximal gain from misreporting, but this optimality comes at a cost: an agent who follows the OPT strategy must evaluate all $m!$ possible reports. For agents with limited cognitive or computational abilities this is obviously impossible even for small m. For this reason, we focus on manipulation strategies that only explore part of the search space. *Local manipulation strategies* are particularly simple manipulation strategies that rely on local, greedy search.

4.1 Construction of Local Strategies

First, a notion of locality on the space of reports is needed.

Definition 3. (Neighborhood) For a preference order $\succ$, the *neighborhood* $N_\succ$ *of* $\succ$ is the set of preference orders that differ by at most a swap of two adjacent objects, e.g.,

$$a \succ \mathbf{b} \succ \mathbf{c} \succ d \succ e \quad \text{and} \quad a \succ' \mathbf{c} \succ' \mathbf{b} \succ' d \succ' e \quad (4)$$

differ only in the ordering of b and c, so $\succ'$ is in $N_\succ$.

Suppose an agent can only evaluate reports in the neighborhood of its truthful report. This yields the following basic strategy.

Definition 4. (LOC) An agent follows a *local manipulation strategy* (LOC) if it evaluates all reports that differ from its truthul report by at most a swap and selects the one that yields the highest utility, i.e., in a given situation $(u_i, \succ_{-i})$,

$$\text{LOC}(u_i, \succ_{-i}) = \arg\max_{\succ_i' \in N_{\succ_i}} \mathbb{E}_{f(\succ_i', \succ_{-i})}[u_i]. \quad (5)$$

Carroll [2012] showed that if a mechanism is not fully strategyproof, then there exists at least *one* situation $(u_i, \succ_{-i})$ in which the LOC strategy finds *some* beneficial misreport. While LOC may not be optimal, it is computationally easy in the sense that it only requires the evaluation of f at m reports, namely the truthful report and its $m-1$ neighbors.

Applying LOC iteratively yields the next strategy.

Definition 5. (ITERLOC) An agent follows an *iterated local manipulation strategy* (ITERLOC) if it evaluates all reports that can be reached from its truthful report by a sequence of weakly beneficial swaps and selects the one that yields the highest utility, i.e., in a given situation $(u_i, \succ_{-i})$,

$$\text{ITERLOC}(u_i, \succ_{-i}) = \arg\max_{\succ_i' \in N_{u_i}^+} \mathbb{E}_{f(\succ_i', \succ_{-i})}[u_i], \quad (6)$$

where $N_{u_i}^+$ denotes the *extended neighborhood* of $\succ_i$ with respect to u_i, i.e., the set of reports that can be reached from $\succ_i$ via consecutive swaps, each of which is weakly beneficial.

ITERLOC is at least as successful as LOC as $N_{\succ_i} \subset N_{u_i}^+$.

4.2 Four Different Situations

A situation may not admit any beneficial misreport for an agent, in which case we call it non-manipulable.

Definition 6. (NM) A situation $(u_i, \succ_{-i})$ is *non-manipulable (NM)* if truthful reporting is optimal, i.e., $\succ_i \in \text{OPT}(u_i, \succ_{-i})$.

If a situation *is* manipulable, then LOC and ITERLOC are appealing because they are simple. But an important question is *how well* they solve the manipulation problem. To study this question formally, we define 3 kinds of situations that exhibit varying degrees of hardness for the strategies.

Definition 7. (LOM) A situation $(u_i, \succ_{-i})$ is *locally optimally manipulable (LOM)* if a local misreport is optimal, i.e., $\succ_i \notin \text{OPT}(u_i, \succ_{-i})$ and

$$\text{OPT}(u_i, \succ_{-i}) \cap \text{LOC}(u_i, \succ_{-i}) \neq \emptyset. \quad (7)$$

Definition 8. (LNOM) A situation $(u_i, \succ_{-i})$ is *locally non-optimally manipulable (LNOM)* if (i) there exists a local misreport that yields weakly higher utility than truthful reporting, and (ii) there exists a non-local misreport with even higher utility, i.e., for some $\succ_i' \in N_{\succ_i} \setminus \{\succ_i\}$, $\succ_i^* \in P \setminus N_{\succ_i}$ we have

$$\mathbb{E}_{f(\succ_i^*, \succ_{-i})}[u_i] > \mathbb{E}_{f(\succ_i', \succ_{-i})}[u_i] \geq \mathbb{E}_{f(\succ_i, \succ_{-i})}[u_i]. \quad (8)$$

Definition 9. (EGM) A situation $(u_i, \succ_{-i})$ is *exclusively globally manipulable (EGM)* if (i) all local misreports yield *strictly* lower utility than truthful reporting, and (ii) there exists a non-local, strictly beneficial misreport, i.e., for all $\succ_i' \in N_{\succ_i} \setminus \{\succ_i\}$ and some $\succ_i^* \notin N_{\succ_i}$ we have

$$\mathbb{E}_{f(\succ_i^*, \succ_{-i})}[u_i] > \mathbb{E}_{f(\succ_i, \succ_{-i})}[u_i] > \mathbb{E}_{f(\succ_i', \succ_{-i})}[u_i]. \quad (9)$$

In LOM situations, local manipulation strategies will always be optimal. In LNOM situations, they always find some weakly beneficial misreports and may even find optimal misreports through iterated search. In EGM situations, LOC and ITERLOC will select $\succ_i$, i.e., the truthful report, while OPT will select some $\succ_i^* \notin N_{\succ_i}$. Intuitively, in EGM situations, local manipulation strategies *get stuck* in the truthful report, despite the existence of some strictly beneficial misreport.

Remark 1. The existence of EGM situations under PS, NBM, and ABM was an open research question. In Section 6, we will prove that these situations actually exist.

5 Human Manipulation Strategies

In the previous section, we have defined local manipulation strategies. We now turn our attention to *human manipulation strategies* and ask how well local manipulation strategies explain how humans approach the manipulation problem. To this end, we conducted a behavioral experiment.

5.1 Experiment Design

For our experiment we recruited 489 human subjects from Amazon Mechanical Turk [Mason and Suri, 2012].

Set-up: Subjects were instructed about a single mechanism (either PS, NBM, or ABM) in a tutorial video. Next, their understanding of the respective mechanism and the manipulation problem was tested via control questions. Subjects who passed the questions first played 4 practice instances of the manipulation problem. Then they had 10 minutes to complete 8 instances.[1]

Compensation: All subjects received a base payment of \$0.50 if they passed the control questions. They could earn an additional bonus equal to the utility gain they obtained by manipulating in each of the 8 instances. These instances were randomly generated and contained NM, LOM, LNOM, and EGM situations (two of each, in random order). In each manipulable instance, the maximal achievable bonus varied uniformly between 10 and 100 cents, while of course no positive bonus could be attained in NM instances. Values were normalized such that reporting $\boxed{A}\boxed{B}\boxed{C}\boxed{D}\boxed{E}$ truthfully resulted in a bonus of 0. The average total bonus attainable from playing optimally in all 8 instances was \$3.30.

User Interface: Figure 1 shows a screen-shot of the user interface (UI) of the experiment. The UI showed the subject's utility function ("values") and the reports from the other agents. Subjects could alter the report $\left(\boxed{A}\boxed{B}\boxed{C}\boxed{D}\boxed{E}\right)$ by "dragging-and-dropping" the objects into new positions. At any point in time, the subjects could click on the "Preview" button to calculate the bonus from submitting the currently visible report. Each *preview*-action corresponded to evaluating the mechanism for a particular report. Once subjects were satisfied with an ordering, they could click on the "Submit & next round" button to advance to the next instance.[2]

Data Cleaning and Analysis: 489 subjects watched the tutorial video and attempted the control questions, and 387 of them passed the control questions. 36 were removed from the sample because they aborted early or encountered an error, and 12 were removed because their *achieved share of maximal gain* (or *gain* for short) deviated by more than 2sd from the mean. This left 339 subjects in the final sample (PS: 117, NBM: 114, ABM: 108).

[1] On average subjects spent 46 seconds on each instance, and 80% had over 1 minute left after the last instance. Thus, the time constraint was not binding.

[2] The UI had no "memory" feature as we wanted to keep the UI simple for the subjects from Mechanical Turk. Since subjects returned to the best report they found in 97.5% of the instances, it appears that the lack of such a device did not impair their performance.

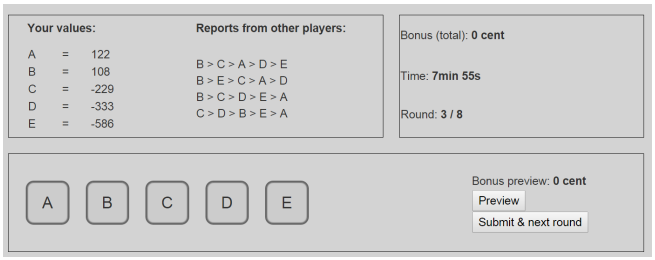

Figure 1: Screen shot of the UI of the experiment.

Our initial analysis revealed that *effort* (in terms of *time spent* and *number of previews*) and *success* (in term of *gain*) varied significantly across subjects. Both effort indicators had significant positive effects on success, but explained less than 20% of the variability. To account for the fact that some subjects simply "tried harder" in our analysis, we use mixed-effects models, controlling for subject random effects. Thus, our conclusions are valid independent of the subjects' individual propensities to exert effort. Depending on whether the response variable is continuous or binary, we use linear or logistic regressions.

5.2 Reliance on Swaps

The basic interaction of subjects with the UI was to move an object to a new positions.

Definition 10. (Elementary Operations) An *elementary operation* is a change in the position of one object. This is a

- *swap*, if one object moves by one position, e.g.,

$$\boxed{A}\boxed{\mathbf{B}}\boxed{\mathbf{C}}\boxed{D}\boxed{E} \text{ to } \boxed{A}\boxed{\mathbf{C}}\boxed{\mathbf{B}}\boxed{D}\boxed{E}.$$

- *slide*, if an object moves by more than one position, e.g.,

$$\boxed{A}\boxed{B}\boxed{C}\boxed{\mathbf{D}}\boxed{E} \text{ to } \boxed{A}\boxed{\mathbf{D}}\boxed{B}\boxed{C}\boxed{E}.$$

In line with intuition, the vast majority of elementary operations were swaps. The odds ratios of choosing swap over slide are given in the first line of Table 1.

Between preview-actions subjects performed one or several elementary operations.

Definition 11. (Report Change) The sequence of elementary operations between two preview-actions is called a *report change*. The report change is a *swap* if it is equivalent to a single swap operation, otherwise it is *complex*.

Result 1. *Subjects primarily relied on swaps for their report changes.*

The second row in Table 1 shows the odds ratios for choosing *swap* over *complex* in report changes. The majority of these report changes were swaps, despite the fact that subjects could perform any number of elementary operations between preview-actions. Thus, swaps were the main step in their search, which is consistent with the notion of locality in the definition of local manipulation strategies.

Remark 2. Note that we do not claim that swaps yield the *only* conceivable notion of locality. Furthermore, we did not analyze the effect that different UI designs may have on the

Odds ratio of *swap* as	PS	NBM	ABM
Elementary operation	2.46***	2.30***	1.87***
Report change	1.46***	1.34***	1.14°

Table 1: Odds ratios of choosing *swaps* over *slide*/*complex* from logistic regression; significance levels indicated for difference from 1.00: ***0.1%, **1%, *5%, °10%.

behavior of human subjects. Nonetheless, the prevalence of swaps as report changes in our data shows that a notion of locality based on swaps is sufficiently interesting to warrant further analysis.

Note also that our next results on *greedy search* and *subjects' success* are largely independent of the particular UI design, because the decision to follow a greedy approach and the relative success of subjects do not depend on the way in which the subjects enter their report or the cost of performing preview actions.

5.3 Reliance on Greedy Search

We distinguish two different events that can occur after a preview-action, and we consider the subjects' possible reactions to these events.

Definition 12. (Hit & Flop) A *hit event* occurs if the previewed bonus strictly increases from one preview-action to the next. A *flop event* occurs if the value strictly decreases.

Definition 13. (Continue & Backtrack) Consider the sequence of actions after a hit or flop event until the next preview-action. Subjects *continue* if this sequence only passes new reports. Otherwise, if some known report is passed, they *backtrack*.

Result 2. *Subjects primarily employed greedy search.*

Table 2 shows the odds ratios of choosing to *continue* (instead of *backtrack*) when a top or flop event occurred. We first observe that subjects frequently chose to continue, even after flop events, which is consistent with the natural desire of humans to "explore." However, we also see that encountering a hit event radically increased the odds of choosing to continue by a factor of almost 3, and this increase is statistically significant. This is consistent with the greedy search assumption in the construction of the ITERLOC strategy.

Kind of event	PS	NBM	ABM
Flop (constant)	1.26**	1.80***	2.02***
Hit	2.74***	2.90***	2.90***

Table 2: Odds ratios of choosing *continue* over *backtrack* from logistic regression with *event* as factor; significance levels indicated for difference from 1.00: ***0.1%, **1%.

5.4 Effect of Difficulty on Performance

Results 1 and 2 show that the micro-structure of our subjects' behavior is largely consistent with local manipulation strategies. In addition, we now verify that the success predicted by local manipulation strategies is also consistent with the success of the subjects in our experiment. We exposed subjects to all four kinds of situations from Section 3.2, i.e., non-manipulable (NM), locally optimally manipulable (LOM), locally non-optimally manipulable (LNOM), and exclusively globally manipulable (EGM). Obviously, in NM situations, subjects cannot achieve any gain by misreporting. For all other (manipulable) situations, let $g_{LOM}, g_{LNOM}, g_{EGM}$ denote the gain in the respective situations, i.e., the bonus attained in a particular instance divided by the bonus attainable by manipulating optimally in that instance.

Kind of situation	PS	NBM	ABM
LNOM (const.)	46.5%***	45.2%***	66.7%***
EGM	-20.4%***	-34.7%***	-46.6%***
LOM	+6.7%°	+21.7%***	+8.9%*

Table 3: *Achieved share of maximal gain* from linear regression with *kind of situation* as factor; significance levels indicated for difference from 0.00: ***0.1%, **1%, *5%, °10%.

Result 3. *Subjects achieved more gain in situations that were predicted to be easier for local manipulation strategies:*

$$g_{LOM} > g_{LNOM} > g_{EGM}. \tag{10}$$

Table 3 shows the effect of *kind of situation* on *gain*, relative to LNOM (the constant in the regression). EGM and LOM situations had negative and positive effects on gain, respectively, and these effects were significant.[3] This is consistent with the performance predicted for agents using local manipulation strategies.

To conclude, Results 1, 2, and 3 support hypothesis #1 that human manipulation strategies can largely be explained by local manipulation strategies: our subjects relied on swaps for their report changes, they performed greedy search, and they gained more in situations that were predicted to be more easily manipulable by local manipulation strategies.

6 Non-optimality of Local Strategies

We now study the ability of local manipulation strategies to solve the manipulation problem. Since the discussion of local manipulation strategies by Carroll [2012], the existence of EGM situations has been an open question for PS, NBM, and ABM. Even though we have already used EGM situations in the experiment, we now establish their existence formally. This in turn implies non-optimality of local manipulation strategies for the manipulation problem.

Proposition 1. *There exist EGM situations for PS.*

Proof. Let $N = \{1, 2, 3, 4\}$, $M = \{a, b, c, d\}$, and $q = 1$. Consider the preference profile

$$\succ_1, \succ_2\colon a \succ b \succ c \succ d; \qquad \succ_3\colon c \succ d \succ b \succ a;$$
$$\succ_4\colon b \succ c \succ d \succ a,$$

[3]Note that the negative values for EGM situations in Table 3 do not imply that subjects received a negative gain in those situations. Instead, these values must be *added* to the values for LNOM situations (the constant in the regression).

and utility function $u_1 = (1.21, 1.1, 1.0, 0.0)$. We can calculate agent 1's allocations for truthfully reporting $\succ_1$ and for any misreport from the neighborhood $N_{\succ_1}$, as well as the utilities. Truthful reporting yields 0.87 and the local misreports yield 0.85, 0.86, and 0.79, which are all strictly lower. However, the optimal misreport is $\succ_1^*$: $b \succ c \succ a \succ d$, which yields 0.92. Thus, $(u_i, \succ_{-i})$ is an EGM situation under PS. □

Proposition 2. *There exist EGM situations for NBM.*

Proof. The proof proceeds analogously to Proposition 1. Let $N = \{1, \ldots, 5\}$, $M = \{a, \ldots, e\}$, $q = 1$ and consider

$$\succ_1: a \succ b \succ c \succ d \succ e, \qquad \succ_2, \succ_3: a \succ b \succ c \succ e \succ d,$$
$$\succ_4: a \succ b \succ d \succ e \succ c, \qquad \succ_5: a \succ c \succ d \succ b \succ e,$$

and utility function $u_1 = (43.0, 3.3, 2.1, 1.0, 0.0)$. The optimal misreport is $\succ_1^*$: $a \succ c \succ d \succ e \succ b$. □

Proposition 3. *There exist EGM situations for ABM.*

Proof. Analogous to Propositions 1 and 2, consider the same setting as in Proposition 2 with

$$\succ_1: a \succ b \succ c \succ d \succ e, \qquad \succ_2: a \succ b \succ d \succ e \succ c,$$
$$\succ_3: a \succ b \succ d \succ c \succ e, \qquad \succ_4: a \succ e \succ d \succ c \succ b,$$
$$\succ_5: a \succ c \succ d \succ b \succ e,$$

and utility function $u_1 = (20.0, 1.15, 1.05, 1.0, 0.0)$. The optimal misreport is $\succ_1^*$: $a \succ d \succ c \succ e \succ b$. □

Remark 3. Finding these EGM situations was challenging, because they are extremely rare. We designed a search algorithm that constructs "border-line" utility functions because random sampling proved unfruitful.

7 Average Performance of Local Strategies

The non-optimality of local manipulation strategies in the worst case (EGM situations) tells us little about their average performance. In this section, we first study the manipulability of PS, NBM, and ABM in general, and then we analyze how LOC and ITERLOC perform *on average* under these mechanisms. We ask the following questions:

1. How manipulable (by OPT) are PS, NBM, and ABM, and how high is the gain from manipulation?
2. How effective are LOC and ITERLOC for the manipulation problem?

7.1 Simulation Set-up

We answer these questions via large-scale simulations. In choosing a stochastic model for the agents' preference orders and utility functions we follow prior work in the matching community that employed simulations [Kominers *et al.*, 2010; Budish and Cantillon, 2012; Abdulkadiroğlu *et al.*, 2015; Erdil and Ergin, 2008]. Our model comprises all models used in these papers.

Treatment Variations: To create a general model, we include the following variations: a *treatment* consists of

- a *mechanism* $f \in \{\text{PS, NBM, ABM}\}$,
- a *number of objects* $m \in \{3, 4, 5\}$
- a *capacity* $q \in \{1, 2, 3\}$ (and fixing the number of agents $n = q \cdot m$, such that supply exactly satisfies demand),
- a *correlation* $\alpha \in \{0, \frac{1}{3}, \frac{2}{3}\}$.

Sample preference profiles were obtained by drawing utility profiles with normally distributed utility values and then correlating them by α. We sampled 10,000 preference profiles for each treatment and then used Gibbs sampling to sample 1,000 utility functions for agent 1.

Remark 4. The treatment parameters were chosen to keep the computational effort manageable. Using the best known algorithms for ABM and NBM, determining OPT for a single preference profile requires $n!m!$ non-trivial computations. However, the evaluations for our largest treatments ($n = 15, m = 5$) already took a full day on a powerful compute cluster. Evaluations for one additional object would have taken $30,000$ times as long. Thus, exact simulations for more than $n = 15$ agents or more than $m = 5$ objects are computationally infeasible.

Manipulability and Effectiveness: Without loss of generality, we evaluated each situation from the perspective of agent 1. A situation is called OPT-*manipulable* if OPT yields a strictly beneficial misreport. To measure OPT-*manipulability* in a particular treatment, we consider the likelihood that a randomly sampled situation is OPT-manipulable. LOC- and ITERLOC-manipulability are defined analogously. These measures follow prior work on manipulability, e.g., [Laslier, 2010; Aziz *et al.*, 2015a].

To measure OPT-*effectiveness*, we normalize agent 1's utility function such that $\min\{u_1(j) | j \in M\} = 0$ and consider the percentage gain from using OPT instead of reporting truthfully. To measure LOC- and ITERLOC-*effectiveness*, we consider what share of the maximum possible gain each strategy captures if the situation is OPT-manipulable.

7.2 Simulation Results

Result 4. *NBM is most vulnerable to manipulation, ABM has intermediate manipulability, and PS is the least manipulable mechanism. The effectiveness of optimal manipulation is low under PS and higher under NBM and ABM.*

Result 4 is apparent from rows (A) and (B) in Figure 2. The values in (A) tell us *how likely* it is that a situation sampled from that particular treatment is susceptible to manipulation by agent 1, and the values in (B) tell us *by how much* agent 1 expects to profit by using OPT. These findings are consistent with the insights about the manipulability of the three mechanisms from prior work: PS is generally considered to have the best incentive properties of the three non-strategyproof mechanisms we studied. It is weakly strategyproof [Bogomolnaia and Moulin, 2001] and incentives improve in larger markets [Kojima and Manea, 2010]. The finding that NBM is more manipulable than ABM is consistent with the theoretical result that ABM is partially strategyproof while NBM is not [Mennle and Seuken, 2014b].

		mechanism = PS									NBM									ABM								
		m = 3			4			5			3			4			5			3			4			5		
	Cor.	n = 3	6	9	4	8	12	5	10	15	3	6	9	4	8	12	5	10	15	3	6	9	4	8	12	5	10	15
(A) OPT-manipulability, in %	α = 0	3	2	1	8	4	3	11	6	4	3	3	3	12	13	13	20	22	23	3	3	3	8	7	6	12	10	8
	1/3	3	2	2	9	5	3	13	7	4	5	7	7	20	23	24	30	34	37	6	7	7	13	13	13	19	18	18
	2/3	1	1	1	5	2	2	7	4	2	19	24	24	53	56	57	66	68	69	20	24	23	43	44	45	56	58	57
(B) OPT-effectiveness, conditional on OPT-manipulability, in %	0	6	5	4	5	3	2	4	2	2	17	14	13	15	10	8	13	8	6	17	14	13	14	12	10	12	10	9
	1/3	6	5	3	5	3	2	3	2	1	18	17	17	18	13	13	17	13	11	19	17	17	16	15	16	15	15	14
	2/3	6	4	2	4	2	1	2	1	1	22	23	25	25	26	27	31	31	31	22	23	25	24	27	28	27	30	31
(C) LOC: captured share of max. gain, conditional on OPT-manipulability, in %	0	100	100	100	97	99	100	96	99	99	100	100	100	95	97	97	89	88	88	100	100	100	97	99	99	97	98	98
	1/3	100	100	100	97	99	100	97	99	99	100	100	100	95	97	97	88	89	89	100	100	100	97	99	99	95	97	97
	2/3	100	100	100	99	100	100	98	99	99	100	100	100	97	97	98	93	94	95	100	100	100	97	98	99	93	94	95
(D) ITERLOC: captured share of max. gain, conditional on OPT-manipulability, in %	0	100	100	100	99	100	100	98	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	1/3	100	100	100	99	100	100	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	2/3	100	100	100	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Figure 2: Heat maps: **(A)** estimated likelihood (in %) of OPT-manipulability, **(B)** expected gain (in %) from OPT in OPT-manipulable situations, **(C) & (D)** expected captured share of maximum gain (in %) from LOC and ITERLOC, respectively, conditional on OPT-manipulability.

Result 5. *The* LOC *strategy often provides a very good solution to the manipulation problem.*

We have evaluated LOC-effectiveness, conditional on OPT-manipulability. The results are shown in row (C) of Figure 2. For most treatments, LOC captures more than 95% of the maximum possible gain, with the lowest value of 88% under NBM with $m = 5$, $n = 15$, and correlation $\alpha = 0$. We also evaluated LOC-manipulability, conditional on OPT-manipulability (not shown) and found that LOC can be expected to find *some* beneficial misreport with probabilities of at least 88%, but usually essentially 100% across all treatments. The success of LOC is particularly interesting because of its extreme simplicity: instead of $m!$ it only requires the evaluation of m reports. Thus, computational effort is drastically reduced relative to OPT. Result 5 implies that this reduction in effort leads to only a minor loss in utility for the manipulating agent.

Result 6. *The* ITERLOC *strategy almost always provides a near-optimal solution to the manipulation problem.*

Analogous to the LOC strategy, we have evaluated ITERLOC-effectiveness, conditional on OPT-manipulability. Row (D) of Figure 2 shows the results: ITERLOC captures more than 98% of the maximal gain across all treatments. ITERLOC-manipulability, conditional on OPT-manipulability (not shown) was at least 99%, but usually essentially 100% across all treatments. Thus, if a situation is manipulable then ITERLOC can be expected to almost always find a near-optimal manipulation. This makes ITERLOC an extremely effective strategy for solving the manipulation problem *approximately*.

Remark 5. We also ran the same simulations with *uniformly* (instead of normally) distributed utility profiles, and for utility profiles with *tiers*, where a set of objects is universally preferred by all agents (reflecting the "ghetto effect" [Urquiola, 2005]). All results were qualitatively the same and are omitted due to space constraints.

8 Conclusion

In this paper, we have analyzed local manipulation strategies for the three important assignment mechanisms PS, NBM, and ABM. These simple strategies arise when agents follow local, greedy search to solve the manipulation problem. Under the LOC strategy agents search for a misreport only in the neighborhood of their truthful report, and under ITERLOC they follow paths with weakly increasing utility. We have studied (1) how well the search behavior of humans can be explained by local manipulation strategies and (2) how well local manipulation strategies solve the manipulation problem.

Towards the first question, we have found that the human subjects in our experiment relied on swaps for their report changes, they searched in a greedy manner, and their performance was better in situations that are predicted to be easier for agents with local manipulation strategies. This evidence suggests that local manipulation strategies largely explain the way in which humans approach the manipulation problem.

Towards the second question, we have proven that both LOC and ITERLOC can fail to solve the manipulation problem optimally for PS, NBM, and ABM. However, using large-scale simulations we have shown that, on average, local manipulation strategies are very powerful heuristics for the manipulation problem which usually find a beneficial manipulation (in almost 100% of the cases) and capture a large share of the possible gain (more than 95% for most treatments using LOC, and more than 98% using ITERLOC).

In general, determining optimal manipulations in assignment problems may be computationally hard. But our findings demonstrate that even cognitively or computationally bounded (human) agents can capture a large part of the utility gain with very low effort by using simple, local strategies. In addition, our results motivate two new research agendas: first, simple, e.g., local best response strategies could be used to define new approximate equilibrium concepts. Second, a more sophisticated behavioral model (beyond local, greedy search) could be derived via experimental studies to capture human manipulation strategies in a more complete way.

Acknowledgements

We would like to thank Katharina Huesmann for insightful discussions and the anonymous reviewers for their helpful comments. Part of this research was supported by the Hasler Foundation and the SNSF (Swiss National Science Foundation).

References

[Abdulkadiroğlu and Sönmez, 2003] Atila Abdulkadiroğlu and Tayfun Sönmez. School Choice: A Mechanism Design Approach. *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*, 95(2):364–367, 2005.

[Abdulkadiroğlu *et al.*, 2006] Atila Abdulkadiroğlu, Alvin E. Roth, Parag A Pathak, and Tayfun Sönmez. Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access. Working Paper, 2006.

[Abdulkadiroğlu *et al.*, 2015] Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda. Expanding "Choice" in School Choice. *American Economic Journal: Microeconomics*, 7(1):1–42, 2015.

[Aziz *et al.*, 2015a] Haris Aziz, Serge Gaspers, Nicholas Mattei, Simon Mackenzie, Nina Narodytska, and Toby Walsh. Equilibria Under the Probabilistic Serial Rule. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2015.

[Aziz *et al.*, 2015b] Haris Aziz, Serge Gaspers, Nicholas Mattei, Simon Mackenzie, Nina Narodytska, and Toby Walsh. Manipulating the Probabilistic Serial Rule. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems*, 2015.

[Bade, 2014] Sophie Bade. Random Serial Dictatorship: The One and Only. Mimeo, 2014.

[Bartholdi *et al.*, 1989a] J. J. Bartholdi, C. A. Tovey, and M. A. Trick. The Computational Difficulty of Manipulating an Election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[Bartholdi *et al.*, 1989b] J. J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting Schemes for Which it can be Difficult to Tell who Won the Election. *Social Choice and Welfare*, 6(2):157–165, 1989.

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

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

[Calsamiglia and Güell, 2014] Caterina Calsamiglia and Maia Güell. The Illusion of School Choice: Empirical Evidence from Barcelona. Working Paper, 2014.

[Carroll, 2012] Gabriel Carroll. When Are Local Incentive Constraints Sufficient? *Econometrica*, 80(2):661–686, 2012.

[Dobzinski and Procaccia, 2008] Shahar Dobzinski and Ariel D. Procaccia. Frequent Manipulability of Elections: the Case of Two Voters. In *Proceedings of the 4th International Workshop on Internet and Network Economics*, 2008.

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

[Ergin and Sönmez, 2006] Haluk Ergin and Tayfun Sönmez. Games of School Choice Under the Boston Mechanism. *Journal of Public Economics*, 90(1-2):215–237, 2006.

[Featherstone, 2011] Clayton R Featherstone. A Rank-based Refinement of Ordinal Efficiency and a new (but Familiar) Class of Ordinal Assignment Mechanisms. Working Paper, 2011.

[Friedgut *et al.*, 2011] Ehud Friedgut, Gil Kalai, Nathan Keller, and Noam Nisan. A Quantitative Version of the Gibbard-Satterthwaite Theorem for Three Alternatives. *SIAM Journal on Computing*, 40(3):934–952, 2011.

[Gabaix *et al.*, 2006] Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg. Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review*, 96 (4):1043–1068, 2006.

[Halpern *et al.*, 2014] Joseph Y. Halpern, Rafael Pass, and Lior Seeman. Decision Theory with Resource-Bounded Agents. *Topics in Cognitive Science*, 6(2):245–257, 2014.

[Hugh-Jones *et al.*, 2014] David Hugh-Jones, Morimitsu Kurino, and Christoph Vanberg. An Experimental Study on the Incentives of the Probabilistic Serial Mechanism. *Games and Economic Behavior*, 87:367–380, 2014.

[Isaksson *et al.*, 2012] Marcus Isaksson, Guy Kindler, and Elchanan Mossel. The Geometry of Manipulation: A Quantitative Proof of the Gibbard-Satterthwaite Theorem. *Combinatorica*, 32(2):221–250, 2012.

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

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

[Kominers *et al.*, 2010] Scott Duke Kominers, Mike Ruberry, and Jonathan Ullman. Course Allocation by Proxy Auction. In *Proceedings of the 6th International Workshop on Internet and Network Economics*, 2010.

[Laslier, 2010] Jean-Francois Laslier. In Silico Voting Experiments. In *Handbook on Approval Voting*, Studies in Choice and Welfare, pages 311–335. Springer, 2010.

[Mason and Suri, 2012] Winter Mason and Siddharth Suri. Conducting Behavioral Research on Amazon's Mechanical Turk. *Behavioral Research Methods*, 44(1):1–23, 2012.

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

[Mennle and Seuken, 2014b] Timo Mennle and Sven Seuken. The Naïve versus the Adaptive Boston Mechanism. Working Paper, 2014.

[Mossel and Rácz, 2014] Elchanan Mossel and Miklós Z Rácz. A Quantitative Gibbard-Satterthwaite Theorem Without Neutrality. *Combinatorica*, pages 1–71, 2014.

[Urquiola, 2005] Miguel Urquiola. Does School Choice Lead to Sorting? Evidence from Tiebout Variation. *American Economic Review*, 95(4):1310–1326, 2005.

[Xia and Conitzer, 2008] Lirong Xia and Vincent Conitzer. A Sufficient Condition for Voting Rules to Be Frequently Manipulable. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, 2008.

[Xia, 2011] Lirong Xia. *Computational Voting Theory: Game-Theoretic and Combinatorial Aspects*. PhD thesis, Computer Science Department, Duke University, Durham, NC, 2011.

[Zhou, 1990] Lin Zhou. On a Conjecture by Gale about One-sided Matching Problems. *Journal of Economic Theory*, 52(1):123–135, 1990.