THE PRICE OF DECEPTION IN SOCIAL CHOICE

A Dissertation
Presented to
The Academic Faculty

By

James P. Bailey

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Algorithms, Combinatorics, and Optimization

Georgia Institute of Technology

December 2017

THE PRICE OF DECEPTION IN SOCIAL CHOICE

Approved by:

Dr. Craig Tovey, Advisor
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Paul Griffin
School of Industrial Engineering
Purdue University

Dr. Vijay Vazirani
College of Computing
Georgia Institute of Technology

Dr. Santosh Vempala
College of Computing
Georgia Institute of Technology

Dr. Prasad Tetali
School of Mathematics
Georgia Institute of Technology

Date Approved: August 4, 2017

"People behave dishonestly enough to profit but honestly enough to delude themselves of their own integrity."

– Mazar *et al.* [46]

To my friends and family

ACKNOWLEDGEMENTS

First and foremost I thank the staff of ISyE that keep things running. I especially thank Pam Morrison and Marc Reese. In addition to the excellent operational work they did for everyone, Pam always helped me remember forgotten things and Mark made sure I could get into my office every time I locked my keys inside. I hope both of you are enjoying your retirement.

I would also like to thank the Division of Civil, Mechanical, and Manufacturing Innovation of the National Science Foundation. The funds provided (grant #1335301) made it far easier to efficiently work.

I thank my adviser, Craig Tovey. Craig and I often spent hours discussing politics, religion, education and many other things. These conversations were always fascinating and I treasure them far more than the occasional conversations about research.

I thank each of my committee members. Santosh, Vijay, and Prasad were all my instructors at one point; each of your classes were fascinating and insightful and I am thankful that I had them with you. Although I never had a class with Paul, I did speak to him on occasion and he is by far the nicest person I ever met. I hope you are enjoy your new position at Purdue and know that you are missed at Georgia Tech.

I would also like to thank Jörg Rothe for agreeing to be the reader for my dissertation. I've enjoyed our past conversations about the research area. I appreciate all the feedback you've provided.

Finally I would like to thank my significant other, Bahar Çavdar. In addition to being wonderful, you motivated me to work a little bit harder. I confess that my time in Turkey was the most productive time in my life. I appreciate the focus you give me and I hope I can do the same for you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Most social choice algorithms rely on private data from individuals to maximize a social utility. However, many algorithms are manipulable – an individual can manipulate her reported data to obtain an outcome she prefers often at the cost of social good. Literature addresses this by declaring an algorithm as "manipulable" or "strategy-proof". However, for many decisions we are forced to either use a manipulable algorithm or an algorithm with negative properties; for instance, a dictatorship is the only strategy-proof way to decide an election. Thus, we view it as unwise to take an all-or-nothing approach to manipulation since we so often have to settle for nothing.

In this dissertation, we focus on algorithmic design with strategic behavior in mind. Specifically, we develop the framework to examine the effect of manipulation on social choice using a game theoretic model. We show that the standard Nash equlibria solution concept is insufficient to capture human behavior as it relates to deception. The combinatorial nature of many problems results in Nash equilibria where individuals are lying to their own detriment. To remove these absurd equilibria and better capture human behavior, we introduce the Minimal Dishonesty refinement. In our model each individual tells the smallest possible lie to obtain the best possible result and therefore if the individual tries to be more honest then they will get a worse outcome. Our model for human behavior is supported by a vast amount of experimental evidence in psychology and economics and allows us to better understand the likely outcomes of strategic behavior.

We also introduce a measure of manipulation – the Price of Deception – that *quantifies* the impact of strategic behavior. Specifically, the Price of Deception is the worst-case ratio between the social utility when each individual is sincere, and the social utility at a Nash equilibrium. With Minimal Dishonesty and the Price of Deception we are able to identify algorithms that are negligibly impacted by manipulation, algorithms where strategic behavior leads to arbitrarily poor outcomes, and anything in-between. We demonstrate

the power of Minimal Dishonesty and the Price of Deception by answering open problems in assignments, facility location, elections, and stable marriages including a 28-year open problem by Gusfield and Irving. Our results demonstrate that the Price of Deception, like computational complexity, provides finer distinctions of manipulability than between "yes" and "no".

Assignments: We calculate the Price of Deception of various assignment procedures. We show that manipulation has significantly different impacts on different procedures. We find the Price of Deception tends to be larger when minimizing social costs than when maximizing social welfare – manipulation has a larger impact on minimization problems. We also find that the Price of Deception is significantly lower when ties are broken randomly. This indicates that even a small change to a decision mechanism, such as a tie-breaking rule, can have a large effect on the impact of manipulation.

Facility Location: We calculate the Price of Deception for standard facility location algorithms, including 1-Median and 1-Mean, and find significant differences among them. We also find that the Price of Deception can increase significantly as the structure of allowed facility locations deviates from a hyper-rectangle. Although some algorithms have high Prices of Deception, we do give a family of fair tie-breaking rules for the 1-Median Problem for which the Price of Deception is as low as possible – the algorithm remains manipulable, but manipulation has no impact on social utility.

Elections: We calculate the Price of Deception for several standard voting rules, including Borda count, Copeland, veto, plurality, and approval voting, and find significant differences among them. In general, the more information a voting rule requires, the higher its Price of Deception. However, plurality voting has the largest Price of Deception despite using little information. Furthermore, we observe that tie-breaking rules in an election can have a significant impact on the Price of Deception. For instance, breaking ties randomly with Majority Judgment leads to better outcomes than lexicographic tie-breaking whereas lexicographic tie-breaking is better for plurality elections.

Stable Marriage: We show that for all marriage algorithms representable by a monotonic and INS function, every minimally dishonest equilibrium yields a sincerely stable marriage. This result supports the use of algorithms less biased than the (Gale-Shapley) man-optimal, which we prove yields the woman-optimal marriage in every minimally dishonest equilibrium. However, bias cannot be totally eliminated, in the sense that no monotonic INS-representable stable marriage algorithm is certain to yield the egalitarian marriage in a minimally dishonest equilibrium, thus answering a 28-year old open question of Gusfield and Irving's in the negative. Finally, we show that these results extend to the Student Placement Problem, where women are polygamous and honest, but not to the College Admissions Problem, where women are both polygamous and strategic. Specifically, we show no algorithm can guarantee stability in the College Admissions Problem.

1

INTRODUCTION

In this thesis we examine the impact of manipulation in the assignment problem, facility location procedures, election schemes and stable marriage mechanisms. We make significant contributions to each of these areas and highlight them more in Section 1.3. More importantly however, we establish the necessary framework to understand the impact of manipulation for any algorithm that relies on private information. Specifically, we introduce a model of manipulation to represent human behavior that we justify inuitively, theoretically, and experimentally in Section 2.3. We introduce a measure, the **Price of Deception**, that quantifies, not categorizes, the effect of manipulation. Our studies of the four aforementioned areas demonstrate the power of the measure to understand the impact of manipulation for any algorithm. The methods we present in this dissertation can readily be extended to any decision problem that relies on private information including allocating a limited capacity of goods within a network, truck-load sharing, and scheduling based on consumer constraints.

1.1 The Price of Deception: A Finer Measure of Manipulability

The literature on strategic behavior in algorithmic design primarily labels algorithms as "manipulable" or "strategy-proof". Yet we know that strategy-proofness is often unobtainable: Zhou [69] tells us that no symmetric, pareto optimal assignment procedure is strategy-proof; The Gibbard-Satterthwaite [28, 65], Gardenfors [27] and related theorems tell us that every election rule that one would consider to be reasonable is manipulable by a strategic voter; Roth [60] tells us any matching algorithm that guarantees stability is also manipuable. Therefore, we always sacrifice non-manipulability in lieu of other properties. But is it wise to take such an all-or-nothing position with respect to manipulability, especially since we almost always settle for nothing?

We introduce a measure of how much strategic behavior can alter the outcome of an algorithm. We call this measure the **Price of Deception**. For example, let the *sincere Borda count* of a candidate in a strategic Borda election be his Borda count with respect to the sincere voter preferences. Let a *sincere Borda winner* be a candidate with largest sincere Borda count, i.e. a candidate who would win were all voters non-strategic. Suppose that the sincere Borda count of a sincere Borda winner could not be more than twice the sincere Borda count of the winner of a strategic Borda election. Then the Price of Deception of Borda voting would be at most 2. As another example, suppose that approval voting had an infinitely large Price of Deception. Then it might be that a candidate sincerely approved by the fewest voters could win an approval vote election when voters are strategic.

We propose that the Price of Deception should be one of the criteria by which an algorithm is assessed. In general, any algorithm can be represented as a maximizer of some utility function. If the utility function has an intrinsic correspondence to the quality of the outcome, then the Price of Deception has a correspondence to the capability of an algorithm's ability to select a quality outcome. Like the computational complexity of manipulation [9, 48] the Price of Deception offers finer distinctions than simply between "manipulable" and "non-manipulable."

1.2 Minimal Dishonesty

The standard complete-normal form game and the Nash equilibrium solution concept [24] used to analyze games give absurd outcomes for "Games of Deception" for combinatorial problems. For instance, consider a majority election between A and B where all individuals sincerely prefer A to B. Suppose instead that everyone indicates they prefer to B to A. Since the election is decided by majority, no single individual could alter their submitted preferences to change the outcome of the election. Therefore, the Nash equilibrium solution concept labels this as possible outcome despite its obvious absurdity.

In Section 2.3, we introduce our refinement. We assume that individuals are **minimally dishonest** – individuals will be dishonest enough to obtain a better result but will not be more dishonest than necessary. We provide logical reasons for minimal dishonesty; lying causes guilt and it is cognitively easier to be somewhat truthful than it is to lie spuriously. We prove (Theorem 2.3.3) that our minimal dishonesty refinement is consistent with the current literature's assumptions regarding strategy-proof algorithms. Most importantly, we show that our model to explain human behavior is backed by a myriad of evidence in psychology and experimental economics.

Re-examine the majority election where everyone sincerely prefers A to B but indicates that they prefer B to A. As mentioned, the Nash equilibrium solution concept indicates that this is a possible outcome despite its absurdity. Once we apply the minimal dishonesty refinement, the result changes; if any individual tries to alter their preferences and instead honestly report they prefer A to B, then the outcome of the election will not change but the individual will be more honest. Therefore the individual is not minimally dishonest and we do not consider everyone voting B a reasonable outcome. For this particular example, everyone honestly reporting that they prefer A to B in the only minimally dishonest equilibrium which is precisely what we would expect.

Our model of minimal dishonesty allows us to make significant strides in understanding the effects of manipulation in assignment problems, facility location procedures, election schemes and stable marriage mechanisms as highlighted in the next section. Moreover, minimal dishonesty establishes the necessary framework to understand the impact of manipulation in algorithms that rely on private information.

1.3 Contributions of this Thesis

The most important contribution of this thesis is the Price of Deception measure of the impact of manipulation and the minimal dishonesty model to represent strategic behavior. We demonstrate the power of these two concepts by making significant progress in understanding the effect of manipulation in various settings.

In Chapter 3 we examine the Assignment Problem and find that algorithms that minimize social cost are more heavily impacted by manipulation (they have higher Prices of Deception) than algorithms that maximize social welfare. Moreover, we establish that small changes to an algorithm can have significant effects on the impact of manipulation – on the Price of Deception. We find that a randomized tie-breaking rule in this settings has a significantly lower Price of Deception than a deterministic tie-breaking rule.

However, when examining the Facility Location Problem in Chapter 4 we establish that deterministic tie-breaking rules yield good societal outcomes despite manipulation while randomized tie-breaking can lead to poor outcomes. Something as seemingly minor as a tie-breaking rule can have a significant impact on the effect of manipulation and that effect can be drastically different in different problem settings. In addition, we give a deterministic and fair variant of the 1-Median Problem that is not strategy-proof but manipulation has no impact on social utility; strategic behavior may cause the outcome of the algorithm to change but the *quality* of the outcome is unaffected by manipulation.

In Chapter 5 when studying election schemes we again observe that randomized and deterministic tie-breaking rules yield different Prices of Deception. Moreover, for some election schemes randomized tie-breaking is better than lexicographic and in other schemes lexicographic tie-breaking is better. We prove this by determining the Price of Deception for an array of popular voting schemes. Typically, the more information a voter provides to the algorithm the higher the Price the Deception – the more strategic behavior impacts the quality of the outcome. This seems intuitive as more information gives a voter more opportunities to manipulate the system. However, it is not a universal truth; Plurality elections have the largest Price of Deception despite voters only providing a single piece of information to the algorithm – their most preferred candidate.

Chapters 3-5 prove that the Price of Deception is able to discriminate between different algorithms based on the impact of strategic behavior. Our results show that the impact of manipulation varies greatly between algorithms and is sensitive to small changes such as a tie-breaking rule. Moreover, we show that the effect of these sensitivities can vary drastically depending on the setting. The Price of Deception captures this behavior. Therefore we propose that the Price of Deception, like computational complexity of manipulation, should be one of the criteria used to access an algorithm.

Finally in Chapter 6, we study the Stable Marriage Problem. Using the minimal dishonesty solution concept, we answer a 28-year open problem by Gusfield and Irving by showing that every reasonable stable marriage algorithm will guarantee stability even when individuals are strategic. We also extend these results to the Student Placement Problem. This result supports the use of less biased algorithms than the (Gale-Shapley) man-optimal algorithm, the only algorithm currently known to guarantee stability when individuals are strategic. Moreover we show that the Gale-Shapley algorithm is biased because it guarantees the best solution for women and the worst solution for men when individuals are strategic. However, we also show that no reasonable algorithm can guarantee the fairest outcome when individuals are strategic.

2

PRELIMINARIES

2.1 The Game of Deception in Social Choice

In this chapter, we introduce the necessary definitions and solution concepts that will be used throughout this dissertation. Each subsequent chapter addresses *Games of Deception* where a central decision mechanism (algorithm, procedure, or rule) makes a social choice (selects a winning candidate, allocates items, etc.) after receiving information (preferences over candidates, valuations for items, etc.) from each individual (voter, bidder, etc.). Our central decision mechanisms tend to have nice properties; we may guarantee an "equitable" allocation, or a "stable" marriage, or a "facility location" that maximizes social utility. However, to guarantee such properties we often must assume that individuals truthfully reveal their information. We introduce many of the necessary concepts for this dissertation in the context of the Fair Division Problem.

Definition 2.1.1. The Fair Division Problem consists of n individuals and m divisible goods where individual i has an integer value of $\pi_{ij} \in [0, cm]$ for item j. The integer value c is a parameter of the Fair Division Problem. Furthermore, $\sum_{j=1}^{m} \pi_{ij} = cm$ for each individual i. A central decision maker in the Fair Division Problem assigns x_{ij} of item j to individual i where $\sum_{i=1}^{n} x_{ij} = 1$.

The set of feasible allocations in the Fair Division Problem is a polytope. Specifically, an allocation is feasible if it satisfies the following linear constraints:

$$\sum_{i=1}^{n} x_{ij} \leq 1 \qquad \forall j = 1, ..., m \tag{2.1}$$

$$x_{ij} \geq 0 \qquad \forall j = 1, ..., m \; \forall i = 1, ..., n \tag{2.2}$$

Therefore if the central decision maker maximizes a concave utility function to determine the division of goods, then an allocation can be found quickly.

Example 2.1.2. Maximizing the Minimum Utility in Fair Division.

In this example, we have a decision maker who seeks to maximize the minimum utility for each individual. Let $c = 40$, $n = 2$ *and* $m = 3$. *Individual* i *assigns item* j *a value* $\pi_{ij} \in [0, cm] = [0, 120]$. *Furthermore, for each* i, *the sum of* π_{ij} *is* $cm = 120$. *Individuals' values for the items are given by* Π *below.*

$$(\Pi) \qquad \begin{array}{lll} \pi_{11} = 50 & \pi_{12} = 50 & \pi_{13} = 20 \\ \pi_{21} = 10 & \pi_{22} = 10 & \pi_{23} = 100 \end{array} \qquad \begin{array}{r} (2.3) \\ (2.4) \end{array}$$

If we assign $X_i = \{x_{i1}, x_{i2}, ..., x_{im}\}$ *to individual* i, *then individual* i *receives a utility of* $\sum_{j=1}^{m} \pi_{ij} x_{ij}$. *Therefore, the central decision maker seeks to maximize the minimum of* $\sum_{j=1}^{m} \pi_{1j} x_{1j}$ *and* $\sum_{j=1}^{m} \pi_{2j} x_{2j}$. *Given the individuals' valuations, an optimal assignment can be found with the following linear program:*

$$\max z \tag{2.5}$$

$$\sum_{j=1}^{m} \pi_{ij} x_{ij} \geq z \qquad \forall i = 1, ..., n \tag{2.6}$$

$$\sum_{i=1}^{n} x_{ij} \leq 1 \qquad \forall j = 1, ..., m \tag{2.7}$$

$$x_{ij} \geq 0 \qquad \forall j = 1, ..., m \; \forall i = 1, ..., n \tag{2.8}$$

Thus, to maximize the minimum of each individuals' utility, the central decision maker selects the assignment $x_{11} = x_{12} = x_{23} = 1$ *assigning items 1 and 2 to individual 1 and item 3 to individual 2. With this assignment, each individual receives a utility of 100.*

Each chapter addresses specific Games of Deception. In each game each individual player has information, usually consisting of preference data. Ideally, individual i truthfully reveals their information π_i to the centralized decision mechanism. If the mechanism selects an outcome deterministically, the central mechanism selects some outcome $r(\Pi)$ and individual i's utility $u_i(\pi_i, r(\Pi))$ depends on the selected outcome.

In the Fair Division Problem individual i's preferences are given by $\pi_i = (\pi_{i1}, \pi_{i2}, ..., \pi_{im})$. The central decision mechanism assigns x_{ij} of item j in individual i. Therefore the outcome is given by $X = r(\Pi)$ according to (2.5)-(2.8) and individual i's utility is given by $u_i(\pi_i, r(\Pi)) = \pi_i \cdot X_i$.

If individuals truthfully report Π to the central decision mechanism, then we can often guarantee desirable properties. For instance, when maximizing the minimum utility in the Fair Division Problem, the Maximin Problem guarantees a solution that Pareto dominates all *equitable* outcomes. Furthermore if $\pi_{ij} > 0$ for all i and j, then the outcome is equitable.

Definition 2.1.3. An outcome r is *equitable* if $u_i(\pi_i, r) = u_j(\pi_j, r)$ for each i and j.

However, these guarantees rely on the assumption of honesty. Regrettably, honesty is not a realistic assumption.

Example 2.1.4. Individuals can have an incentive to be dishonest.

We continue Example 2.1.2. If each individual truthfully reveals their preferences as given in Π *in Example 2.1.2, then individual 1 will be assigned items 1 and 2 and individual 2 will be assigned item 3. Furthermore, the allocation is equitable as each person has utility* 100*. However, individual 1 can get a strictly better outcome.*

$$(\bar{\Pi}) \qquad \begin{aligned} &\bar{\pi}_{11} = 20 && \bar{\pi}_{12} = 20 && \bar{\pi}_{13} = 80 && (2.9)\\ &\bar{\pi}_{21} = \pi_{21} = 10 && \bar{\pi}_{22} = \pi_{22} = 10 && \bar{\pi}_{23} = \pi_{23} = 100 && (2.10) \end{aligned}$$

If individual 1 alters her submitted preferences according to $\bar{\Pi}$*, then the outcome changes. Using the linear program described by Equations 2.5-2.8, we determine that individual 1 still receives all of item 1 and item 2, but now also receives* $\frac{1}{3}$ *of item 3. With respect to the putative preferences* $\bar{\Pi}$*, each individual receives a utility of* $\frac{200}{3} \approx 67$ *and the central decision maker believes that it has selected an equitable outcome. However, with respect to the sincere preferences* Π*, individual 1 receives a utility of* $50 + 50 + \frac{20}{3} = \frac{320}{3} \approx 107$ *while individual 2 receives a utility of only* $100\frac{2}{3} \approx 67$.

Since individual 1 receives a utility of $107 > 100$*, individual 1 has incentive to misrepresent her preferences.*

Example 2.1.4 not only shows individuals may have incentive to lie, it highlights the disturbing possibility that we cannot guarantee desirable properties when individuals behave strategically. As discussed in Section 1.1, existing literature acknowledges that mechanisms can be manipulable. However, much of the literature focuses on simply labeling a mechanism as "manipulable" or "strategy-proof". We have already described the pitfalls of such labels in Section 1.1. Instead, in this dissertation we will focus on the outcomes obtained when individuals behave strategically by using game theoretic techniques to analyze Games of Deception. We generalize the earlier notation to allow for games that select outcomes randomly.

The Game of Deception $G(N, \mathcal{P}, \Pi, r, u, \Omega, \mu)$

- Each individual i has information $\pi_i \in \mathcal{P}_i$ describing their preferences. The collection of all information is the (sincere) profile $\Pi = \{\pi_i\}_{i=1}^n$ where $|N| = n$.
- To play the game, individual i submits putative preference data $\bar{\pi}_i \in \mathcal{P}_i$. The collection of all submitted data is denoted $\bar{\Pi} = \{\bar{\pi}_i\}_{i=1}^n$.
- It is common knowledge that a central decision mechanism will select outcome $r(\bar{\Pi}, \omega)$ when given input $\bar{\Pi}$ and random event ω.
- The random event $\omega \in \Omega$ is selected according to μ. We denote $r(\bar{\Pi})$ as the distribution of outcomes according to Ω and μ.
- Individual i evaluates $r(\bar{\Pi})$ according to i's sincere preferences π_i. Specifically, individual i's utility of the set of outcomes $r(\bar{\Pi})$ is $u_i(\pi_i, r(\bar{\Pi}))$.

If the decision mechanism always selects the outcome deterministically, then we simply refer to the Game of Deception as $G(N, \mathcal{P}, \Pi, r, u)$. In many settings, u is either equal to Π or can be deduced from Π and the game can be further reduced to $G(N, \mathcal{P}, \Pi, r)$. For instance, in an election voter v's preferences π_v typically are an ordering of the candidates. In this case, v prefers candidate A to candidate B (with respect to π_v) if and only if $u_v(\pi_v, A) > u_v(\pi_v, B)$. In the Fair Division setting, an individual i's information is precisely their values for each object (i.e. i's utility of the allocation X_i is $u_i(\pi_i, X_i) = \pi_i \cdot X_i$).

We remark that u_i is not necessarily unique and may be an uncountably large set of data. Since individual i submits π_i and not u_i to the central decision mechanism, the complexity of u_i is irrelevant when computing the social outcome $r(\Pi)$. The utility u_i is a parameter of a game used to determine the information $\bar{\pi}_i$ that individual i is likely to submit. The complexity of u_i does not impede our ability to predict u_i when we assume some properties of the individual utility functions. For instance, in the Facility Location Problem where individual i submits a preferred location π_i, we assume i has a utility of $u_i(\pi_i, x) = ||\pi_i - x||_{p_i}$ of the facilty location x for some $p_i \in (0, \infty)$.

While there are many ways to analyze games, in this dissertation we focus on the most common approach – the Nash equilibrium solution concept for normal-form games under complete information [24].

Definition 2.1.5. The preference profile $[\bar{\Pi}_{-i}, \bar{\pi}'_i]$ if obtained from $\bar{\Pi}$ by replacing $\bar{\pi}_i$ with $\bar{\pi}'_i$.

Definition 2.1.6. The preference profile $\bar{\Pi} \in \mathcal{P}$ is a Nash equilibrium for $G(n, \mathcal{P}, \Pi, r, u, \Omega, \mu)$ if $u_i(\pi_i, r(\bar{\Pi})) \geq u_i(\pi_i, r([\bar{\Pi}_{-i}, \bar{\pi}'_i]))$ for each individual i and $\bar{\pi}'_i \in \mathcal{P}_i$.

Example 2.1.7. A Nash Equilibrium of the Fair Division Game

We continue Examples 2.1.2 and 2.1.4. The sincere preference profile Π *is given again below.*

$$\pi_1 = (50, 50, 20) \tag{2.11}$$

$$\pi_2 = (10, 10, 100) \tag{2.12}$$

Recall from Example 2.1.2, that if everyone is honest then the assignment $r(\Pi)$ *gives all of items* 1 *and* 2 *to individual* 1 *and gives item 3 to individual* 2*.*

Example 2.1.2 shows that individual 1 has incentive to lie by updating her preference list to $\bar{\pi}_1 = (20, 20, 80)$ *and therefore* Π *is not a Nash equilibrium. Furthermore the preferences given in Example 2.1.4 are not a Nash equilibrium since individual 2 could obtain a strictly better result by submitting* $\bar{\pi}'_2 = (19, 19, 82)$*. With respect to the new preferences, the assignment gives all of items 1 and 2 and* $\frac{42}{162}$ *of item 3 to player 1 and* $\frac{120}{162}$ *of item 3 to individual* 2*.*

We now might ask if $\{\bar{\pi}_1, \bar{\pi}'_2\}$ *forms a Nash equilibrium. The answer depends on how the decision mechanism breaks ties. If individual 2 instead submits* $\bar{\pi}_2 = (20, 20, 80)$ *there are a variety of optimal allocations with respect to the sincere preferences. For this example, we assume the decision mechanism break ties by assigning the lowest indexed individual the most of the lowest indexed good as possible (i.e. decision mechanism prioritizes assigning item 1 to individual 1, then item 2 to individual 1,..., then item 1 to individual 2,..., item m to individual 2,..., item m to individual n). Therefore if both individuals submit* $(20, 20, 80)$ *then individual* 1 *is assigned items 1 and 2 and* $\frac{1}{4}$ *of item 3.*

With this tie breaking rule, $\bar{\Pi} = \{\bar{\pi}_1, \bar{\pi}_2\}$ *is a Nash equilibrium. We begin by showing that individual* 1 *cannot obtain an better outcome. Let* (a, b, c) *denote any allocation where individual* 1 *is assigned* a*,* b*, and* c *of items* 1*,* 2*, and* 3 *respectively. Individual* 2 *then receives* $1 - a$*,* $1 - b$*, and* $1 - c$ *of items* 1*,* 2*, and* 3*. Regardless of the preferences submitted,* $(a, b, c) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ *is an allocation where both individuals receive a utility of* $\frac{cm}{n} = 60$ *implying that individual* 2 *must receive a utility of at least* 60 *with respect to* $\bar{\pi}_2$ *if* (a, b, c) *is the result of the decision mechanism.*

Therefore we know that

$$\bar{\pi}_{21} \cdot (1-a) + \bar{\pi}_{22} \cdot (1-b) + \bar{\pi}_{23} \cdot (1-c) \geq 60 \tag{2.13}$$

$$\Rightarrow 20a + 20b + 80c \leq 60 \tag{2.14}$$

$$\Rightarrow 5a + 5b + 20c \leq 15 \tag{2.15}$$

for any allocation (a, b, c) *assigned when* $\bar{\pi}_2$ *is submitted. Suppose individual* 1 *alters her preferences to get a different outcome* (a, b, c)*. We know that* a *and* b *are at most* 1 *implying* $45a \leq 45$ *and* $45b \leq 45$*. Combining these two inequalities with (2.15) we have*

$$\pi_1 \cdot (a, b, c) = 50a + 50b + 20c \leq 105 \tag{2.16}$$

implying that no matter how individual 1 *alters her preferences she will not receive a utility more than 105. When she submits* $\bar{\pi}_1 = (20, 20, 80)$ *then she receives a utility of* $\pi_1 \cdot (1, 1, \frac{1}{4}) = (50, 50, 20) \cdot (1, 1, \frac{1}{4}) = 105$ *and therefore* $\bar{\pi}_1$ *is a best response. Therefore individual 1 cannot alter her preferences to get a better outcome.*

Similarly, we show that individual 2 *is reporting a best response with the inequalities* $\frac{5}{4}(20a + 20b + 80c) \leq \frac{5}{4}60$*,* $-15a \leq 0$*, and* $-15b \leq 0$*, to get* $\pi_1 \cdot (a, b, c) = 10a + 10b + 100c \leq 75$*. This implies that individual* 2 *can obtain a utility of at most* 75*, which is precisely the value he gets when submitting* $\bar{\pi}_2$*. Therefore* $\bar{\Pi}$ *is a Nash equilibrium.*

2.2 The Price of Deception

A centralized mechanism makes a decision that optimizes a measure of societal benefit based on private information submitted by individuals. However, the individuals have their own valuations of each possible decision. Therefore, there is a Game of Deception in which they provide possibly untruthful information, and experience outcomes in accordance with their own true valuations of the centralized decision that is made based on the information they provide. As a result of this game, social benefit may decrease. The Price of Deception is the worst-case ratio between the optimum possible overall benefit and the expected overall benefit resulting from a pure strategy Nash equilibrium of the game. If the mechanism minimizes a societal cost the Price of Deception is defined as the reciprocal of that ratio, so that its value is always at least 1.

We remark that the revelation principle [49, 50] is irrelevant to the Price of Deception. This is because revelation elicits sincere information only by yielding the same outcome that strategic information would yield. The revelation principle can be a powerful tool for analyzing outcomes. But for our purposes, the elicitation of sincere preference information is not an end in itself.

The Price of Deception is similar to a concept known in the computer science community as the "Price of Anarchy" [62]. However, the two concepts are not equivalent. The Price of Anarchy measures the difference between the social values of an optimal solution selected by a central coordinator and an equilibrium solution when all agents act selfishly. As we discuss in Section 2.4, one can view the Price of Deception as a Price of Anarchy only by assuming that a central coordinator can force individuals to be truthful, thus negating the existence of private information. Also, the term "anarchy" is a singularly inapt descriptor of a democratic process. Although the Price of Deception is not a type of Price of Anarchy, the converse turns out to be true. In Section 2.4 we show that, in a formal sense, one can view any Price of Anarchy as a Price of Deception. We proceed to the formal definition of Price of Deception.

Every Game of Deception, $G(N, \mathcal{P}, \Pi, r, u, \Omega, \mu)$, can be modeled such that r is the maximizer of some social utility function U while respecting some constraints. For instance, in Examples 2.1.2-2.3.1,

$$r(\bar{\Pi}, \omega) \in \operatorname*{argmax}_{X \geq 0} \left\{ \min_{i \in [n]} u_i(\bar{\pi}_i, X) : \sum_{i=1}^{n} x_{ij} \leq 1 \; \forall j \in [m] \right\} \tag{2.17}$$

Therefore a social utility that represents the mechanism is $U(\bar{\Pi}, X) = \min_{i \in [n]} u_i(\bar{\pi}_i, X)$.

In the earlier examples, we established that $r(\Pi)$ is not always the same as $r(\bar{\Pi})$ even when $\bar{\Pi}$ is a Nash equilibrium. Therefore there is a disquieting possibility that $U(\Pi, r(\bar{\Pi})) < U(\Pi, r(\Pi)$ and social utility may decrease when individuals act strategically. The majority of this dissertation focuses on quantifying **how much** manipulation impacts social utility.

Notably, another social utility that represents the mechanism is

$$U'(\bar{\Pi}, X) = \begin{cases} 1 & \text{if } X \in r(\bar{\Pi}) \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

With respect to the definition of U', manipulation impacts social utility if and only if the outcome changes. Any analysis using U' is equivalent to simply labeling a system as manipulable or strategy-proof. Therefore, to describe how much manipulation impacts social utility, it is important we work with a social utility that has a meaningful interpretation. Fortunately, many mechanisms either have a well defined social utility function given by the literature or are naturally defined by their social utilities.

Definition 2.2.1. For the Game of Deception $G(N, \mathcal{P}, \Pi, r, u, \Omega, \mu)$ the set of Nash equilibria for the sincere preference Π is $NE(\Pi) \subseteq \mathcal{P}$.

When analyzing a specific mechanism, ideally $NE(\Pi) = \Pi$ or at least $r(\Pi) = r(\bar{\Pi})$ for all $\bar{\Pi} \in NE(\Pi)$ and manipulation does not impact the social outcome. However, this is often not the case. Next, we would hope that the social utility does not change (i.e. $U(\Pi, r(\Pi)) = U(\Pi, r(\bar{\Pi}))$ for all $\bar{\Pi} \in NE(\Pi)$). But in this dissertation we find very few mechanisms where this is true. Thus we are left hoping that $U(\Pi, r(\bar{\Pi}))$ is at least close to $U(\Pi, r(\Pi))$.

Definition 2.2.2. Given an instance of preferences Π, the Price of Deception of the game

$G(N, \mathcal{P}, \Pi, r, u, \Omega, \mu)$ is

$$\sup_{\bar{\Pi} \in NE(\Pi)} \frac{E\Big(U\big(\Pi, r(\Pi)\big)\Big)}{E\Big(U\big(\Pi, r(\bar{\Pi})\big)\Big)} \tag{2.19}$$

where $E\Big(U\big(\Pi, r(\bar{\Pi})\big)\Big)$ is the expected utility of $r(\bar{\Pi})$ given Ω and μ. By definition, $U(\Pi, x) = U(\Pi, y)$ for all $x, y \in r(\Pi)$ and therefore we simply write $U(\Pi, r(\Pi))$ for the numerator.

If the Price of Deception of an instance is 100, then manipulation can cause the social utility to decrease by a factor of 100. But if the Price of Deception is 2, then the social utility of the Nash equilibrium is always within a factor of 2 of the social utility when everyone is honest. If the Price of Deception is 1 then manipulation does not negatively impact social utility! If the mechanism uses costs instead of utilities, then we define the Price of Deception using the reciprocal of (2.19).

Example 2.2.3. The Price of Deception for Two Instances of the Fair Division Problem.

We continue with the Fair Division Problem using $U(\bar{\Pi}, X) = \min_{i \in [n]} u_i(\bar{\pi}_i, X)$ while breaking ties lexicographically as in Example 2.3.1. For the first instance, individuals' utilities given by Pi below.

$$\pi_{11} = 50 \qquad \pi_{12} = 50 \tag{2.20}$$

$$\pi_{21} = 0 \qquad \pi_{22} = 100 \tag{2.21}$$

If both individuals are honest, then individual 1 receives all of item 1 and $\frac{1}{3}$ of item 2 while individual 2 receives all of item 2. Both individuals would receive a utility of $\frac{200}{3} \approx 66.7$.

Computing the Price of Deception for this instance now requires finding a tight lower bound l on the social utility of an equilibrium. This requires (i) a proof that the social utility cannot be below l and (ii) an example (or family of examples) to show the bound is tight (asymptotically tight). We begin with (i).

A feasible allocation for this problem is $x_{ij} = \frac{1}{n}$ *for all* $i = 1, ..., n$ *and* $j = 1, ..., m$. *Furthermore, for each player* i *the utility of this allocation is* $\sum_{j=1}^{m} v_{ij}x_{ij} = \sum_{j=1}^{m} \frac{v_{ij}}{n} = \frac{cm}{n}$. *Since the mechanism maximizes the minimum utility, if an individual is honest then they must receive a utility of at least* $\frac{cm}{n}$. *In this example,* $c = 50$, $m = 2$ *and* $n = 2$ *implying each individual most receive a utility of at least* 50 *in every Nash equilibrium. Therefore the Price of Deception for this instance is at most* $\frac{200/3}{50} = \frac{4}{3}$.

Next, we must produce an example showing this bound is tight (ii). Consider the submitted preferences $\bar{\Pi}$ *below.*

$$\bar{\pi}_{11} = 0 \qquad \bar{\pi}_{12} = 100 \tag{2.22}$$

$$\bar{\pi}_{21} = 0 \qquad \bar{\pi}_{22} = 100 \tag{2.23}$$

These submitted preferences correspond to a Nash equilibrium. With these submitted preferences, individual 1 receives all of item 1 and half of item 2. Therefore with respect to the sincere preferences, individual 1 *receives a utility of* 75 *while individual 2 only has a utility of* 50. *Furthermore if individual 1 increases their submitted preference for item 1, then they still receive all of item 1 but less of item 2. Similarly, if individual 2 increases their submitted preference for item 1, then they receive all of item 1 and less of item 2. Therefore, for both individuals, altering their preferences results in a worse solution and therefore the preference profile in Table* $\bar{\Pi}$ *is a Nash equilibrium. The social outcome of this Nash equilibrium is* 50 *and the Price of Deception is at least* $\frac{200}{3} \cdot \frac{1}{50} = \frac{4}{3}$. *Combined with the previous bound, we conclude the Price of Deception of this instance is* $\frac{4}{3}$ *– it is possible that manipulation causes the social utility to decrease by a factor of* $\frac{4}{3} \approx 1.33$.

We also consider a second set of sincere preferences given in the Π' *below. With respect to these preferences, individual 1 is assigned all of item 1 and* $\frac{1}{6}$ *of item 2. Both individuals receive a utility of* $\frac{250}{3} \approx 83.33$. *Similar to the first instance, the preferences given by* $\bar{\Pi}$ *above are a Nash equilibrium with social utility of* 50 *(with respect to the sincere preferences). As before, at every Nash equilibrium every individual must have a utility of at least* 50 *and therefore the Price of*

Deception of this instance if $\frac{250}{3} \cdot \frac{1}{50} = \frac{5}{3} = 1.6$.

$$\pi'_{11} = 80 \qquad \pi'_{12} = 20 \tag{2.24}$$

$$\pi_{21} = 0 \qquad \pi_{22} = 100 \tag{2.25}$$

In Example 2.2.3, we see that the Price of Deception of an instance can vary for different sincere preference profiles. As mechanism designers, we only see individuals submitted preferences. Therefore, if individuals submit the preferences given by $\bar{\Pi}$, we cannot determine if their sincere preferences are given by Π or Π' or something else and we cannot conclude how close we are to the intended result. However, we can bound the Price of Deception of an instance by conducting a worse-case analysis over all sincere preference profiles.

Definition 2.2.4. The Price of Deception of a mechanism with U, N, $\mathcal{P}$, r, u, Ω, and μ is

$$\sup_{\Pi \in \mathcal{P}} \sup_{\bar{\Pi} \in NE(\Pi)} \frac{U\big(\Pi, r(\Pi)\big)}{E\Big(U\big(\Pi, r(\bar{\Pi})\big)\Big)} \tag{2.26}$$

where the expectation is taken with respect to Ω and μ.

If the mechanism uses costs instead of utilities, then we define the Price of Deception using the reciprocal of (2.26).

Theorem 2.2.5. *The Price of Deception for the Fair Division Problem with* n *individuals and* n *objects when breaking ties lexicographically is between* $\frac{cn-1+\frac{1}{1+cn(n-1)}}{c+\frac{cn-c}{cn-1}} \to n$ *as* $c \to \infty$ *and* n.

For small c The Price of Deception may be slightly less than n. However, as c grows large the problem approaches the continuous version of the Fair Division Problem and the Price of Deception grows slightly. Fortunately the Price of Deception is bounded by n instead of going to ∞.

Proof of Theorem 2.2.5. As with Example 2.2.3, we must (i) show that the Price of Deception is at least $\frac{cn-1+\frac{1}{1+cn(n-1)}}{c+\frac{cn-c}{cn-1}}$ (i.e. there is an instance with Price of Deception $\frac{cn-1+\frac{1}{1+cn(n-1)}}{c+\frac{cn-c}{cn-1}}$) and (ii) the Price of Deception is at most n (i.e. for every sincere $\Pi \in \mathcal{P}$ and every Nash equilibrium $\bar{\Pi} \in NE(\Pi)$, the utility of $r(\Pi)$ is at most n times more than the utility of $r(\bar{\Pi})$.

(i) The Price of Deception of the Fair Division Problem is at least $\frac{2}{3}n\frac{c}{1+c}$: Consider the sincere preferences below with $m = n$ objects where person i most prefers object i:

$$\pi_{ii} = cn - 1 \; \forall i = 1, ..., n-1 \tag{2.27}$$

$$\pi_{in} = 1 \; \forall i = 1, ..., n-1 \tag{2.28}$$

$$\pi_{nn} = cn \tag{2.29}$$

$$\pi_{ij} = 0 \; \forall \; other \; i, j \tag{2.30}$$

It may also be easier to consider the preferences in the form

$$\Pi = \begin{pmatrix} cn-1 & 0 & ... & 0 & 1 \\ 0 & cn-1 & ... & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & cn-1 & 1 \\ 0 & 0 & ... & 0 & cn \end{pmatrix} \tag{2.31}$$

When individual i is assigned object i then everyone receives utility of at least $cn - 1$. The optimal choice is to give all of object i and $\frac{1}{1+cn(n-1)}$ of object n to individual i for $i \leq n-1$, and $1 - \frac{n-1}{1+cn(n-1)}$ of object n to individual n for a social utility of $U(\Pi, r(\Pi)) =$

$cn-1+\frac{1}{1+cn(n-1)}$. The following putative preferences are a Nash equilibrium for Π:

$$\bar{\pi}_{ii} = 1 \; \forall i = 1, ..., n-1 \tag{2.32}$$

$$\bar{\pi}_{in} = cn-1 \; \forall i = 1, ..., n-1 \tag{2.33}$$

$$\bar{\pi}_{n1} = 1 \tag{2.34}$$

$$\bar{\pi}_{nn} = cn-1 \tag{2.35}$$

$$\bar{\pi}_{ij} = 0 \; \forall \; other \; i, j \tag{2.36}$$

Alternatively,

$$\bar{\Pi} = \begin{pmatrix} 1 & 0 & ... & 0 & cn-1 \\ 0 & 1 & ... & 0 & cn-1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & cn-1 \\ 1 & 0 & ... & 0 & cn-1 \end{pmatrix} \tag{2.37}$$

With respect to these preferences, since ties are broken lexicographically individual i receives all of object i and $\frac{1}{n}\left(1-\frac{1}{cn-1}\right)$ of object n for $i \leq n-1$ and individual n is assigned $\frac{1}{n}+\frac{n-1}{n}\left(\frac{1}{cn-1}\right)$ of object n. With respect to the sincere preferences Π, individual i's utility for this allocation is $u_i(\pi_i, r(\bar{\Pi})) = cn-1+\frac{1}{n}\left(1-\frac{1}{cn-1}\right)$ for $i \leq n-1$ – slightly more than the $cn-1+\frac{1}{1+cn(n-1)}$ received when everyone was honest – and individual n has a sincere utility of only $u_n(\pi_n, r(\bar{\Pi})) = c+\frac{cn-c}{cn-1}$ – significantly less than the $cn-1+\frac{1}{1+cn(n-1)}$ received when everyone was honest. The social utility of this outcome with respect to the sincere preferences is $U(\Pi, r(\bar{\Pi})) = c+\frac{cn-c}{cn-1}$. If $\bar{\Pi}$ is a Nash equilibrium, then the Price of Deception of this instance is:

$$\frac{U(\Pi, r(\Pi))}{U(\Pi, r(\bar{\Pi}))} = \frac{cn-1+\frac{1}{1+cn(n-1)}}{c+\frac{cn-c}{cn-1}} \tag{2.38}$$

It remains to show that $\bar{\Pi}$ is a Nash equilibrium. We first examine individual $i \neq n$. Let $\bar{\pi}'_i \in \mathcal{P}_i$ a set of weights that i can submit. To determine that Π is a Nash equilibrium we must show that $u_i(\pi_i, r(\bar{\Pi})) \geq u_i(\pi_i, r([\bar{\Pi}_{-i}, \bar{\pi}'_i]))$. Let $\bar{\pi}'_i$ be such that individual i gets a utility of at least $u_i(\pi_i, r(\bar{\Pi}))$. We conclude $\bar{\pi}'_{ii} \geq 1$ since otherwise individual 1 is assigned item i (if $i = 1$ then individual n is assigned item 1) and this corresponds to a strictly worse outcome for player i. Next we claim $\bar{\pi}'_{in} \geq cn - 1$. If this is not the case, then either $\bar{\pi}_{ii} > 1$ which results in i receiving all of i and less of item n, or $\bar{\pi}_{ij} > 1$ for some $j \notin \{i, n\}$ and individual i receives item j and less of item n. Both outcomes are worse than i and therefore $\bar{\pi}_i = \bar{\pi}'_i$. The argument to show that individual n is giving a best response is similar implying $\bar{\Pi}$ is a Nash equilibrium.

(ii) The Price of Deception of the Fair Division Problem is at most n: As in Example 2.2.3, if an individual is honest then they receive an allocation with value at least $\sum_{j=1}^{m} \frac{v_{ij}}{n} = \frac{cm}{n}$. Therefore for every sincere profile Π and equilibrium $\bar{\Pi} \in NE(\Pi)$, $U(\Pi, r(\bar{\Pi})) = \min_{i \in N} u_i(\pi_i, r(\Pi)) \geq \frac{cm}{n}$. Furthermore, since an individual can receive a utility of at most $\sum_{j=1}^{m} v_{ij} = cm$, the social utility of $r(\bar{\Pi})$ is

$$U(\Pi, r(\bar{\Pi})) = \min_{i \in N} u_i(\pi_i, r(\bar{\Pi})) \leq cm. \tag{2.39}$$

Therefore the Price of Deception is at most

$$\sup_{\bar{\Pi} \in NE(\Pi)} \frac{U\big(\Pi, r(\Pi)\big)}{E\Big(U\big(\Pi, r(\bar{\Pi})\big)\Big)} \leq \frac{cm}{\frac{cm}{n}} = n \tag{2.40}$$

□

2.3 Minimal Dishonesty

Anaylzing of "Games of Deception" is by no means novel [28, 65, 27, 64, 31, 53, 43, 13]. Yet the solution concept is neglected in many areas as researchers still simply label mechanisms as "manipulable" or "strategy-proof". While we make no attempt to describe the beliefs or intentions of other researchers, we believe there is a good reason to neglect the Nash equilibrium solution concept as we have described it; the outcomes of Games of Deception often make little sense.

Example 2.3.1. Games of Deception May Lead to Absurd Outcomes.

We continue with the Fair Division Problem. Once again we will maximize the minimum utility with respect to the submitted preferences while breaking ties by assigning the lowest indexed individual the most of the lowest indexed good as possible. Consider the sincere preference profile Π *given by* Π *below.*

$$\pi_{11} = 100 \qquad \pi_{12} = 0 \tag{2.41}$$

$$\pi_{21} = 0 \qquad \pi_{22} = 100 \tag{2.42}$$

As a decision maker, these are the best preferences we can possibly hope for. We can give each individual everything they want! The optimal allocation gives item i *to individual* i *with a utility of 100. The sincere preference profile* Π *is even a Nash equilibrium. Regrettably, it is not the only Nash equilibrium.*

$$\bar{\pi}_{11} = 0 \qquad \bar{\pi}_{12} = 100 \tag{2.43}$$

$$\bar{\pi}_{21} = 0 \qquad \bar{\pi}_{22} = 100 \tag{2.44}$$

Given the submitted preference profile $\bar{\Pi}$*, the central decision mechanism assigns item 1 to individual 1 and evenly splits item 2 between individuals 1 and 2. With respect to the submitted preferences, the central decision mechanism believes that each individual obtains a utility of 50. However, with respect to the sincere preferences, individual 1 has a utility of 100 and individual 2 has a utility of 50. Moreover,* $\bar{\Pi}$ *is a Nash equilibrium since neither individual can alter their preferences to get a* ***strictly*** *better outcome.*

Nash equilibria are meant to describe or predict the outcome of events [24]. However, the outcome in Example 2.3.1 is absurd; individual 1 is lying spuriously. The absurdity shown in Example 2.3.1 is fairly tame compared to the others we mention in this dissertation. If individual 1's sincere preferences were slightly different, then the lie observed in 2.3.1 makes sense; if $\pi_1 = (99, 1)$ then individual 1 lied to receive half of item 2 and all of item 1 which would be a strict improvement for individual 1. Moreover, individual 1's lie is not self-detrimental even though it negatively impacts others. This is not true in all settings.

Consider a majority election between two candidates A and B and suppose every individual prefers A to B. Then obviously A should win the election. However, the Nash equilibrium solution concept tells us something different. If everyone lies and indicates they prefer B to A, then B wins the election. Furthermore, since the election is determined by majority rule, if any single individual alters their submitted preferences then B still wins the election. Therefore there is a Nash equilibrium where B wins the election.

We believe these absurd equilibria are the biggest obstacle to understanding the effects of strategic behavior.

To overcome this obstacle, we introduce the minimal dishonesty refinement to the Game of Deception.

Definition 2.3.2. Let the putative preferences $\bar{\Pi}$ be a Nash equilibrium for the sincere preferences Π in the Game of Deception $G(N, \mathcal{P}, \Pi, r, u, \Omega, \mu)$. Player i is minimally dishonest if $u_i(\pi_i, r([\bar{\Pi}_{-i}, \bar{\pi}_i'])) < u_i(\pi_i, r(\bar{\Pi}))$ for each $\bar{\pi}_i' \in \mathcal{P}_i$ that is more consistent with π_i than $\bar{\pi}_i$. In other words, if player i is minimally dishonest, then being more honest should yield a worse outcome for i. A Nash equilibrium $\bar{\Pi}$ is a minimally dishonest Nash equilibrium if each player is minimally dishonest.

We provide an intuitive definition of minimial dishonesty, show it is consistent with the current literature and provide a large amount of experimental evidence backing our refinement. If individual i can be more honest and get at least as good a result, then the individual would do so because lying causes guilt and because being somewhat truthful is cognitively easier than lying spuriously. Thus, there is some "utility" associated with being more honest. Therefore an individual is not minimally dishonest and is not acting in their best interest. Hence, we view minimally dishonest Nash equilibria as the set of reasonable outcomes in a Game of Deception.

The assumption of minimal dishonesty is also consistent with the current literature's assumptions of "strategy-proof" or "non-manipulable" mechanisms. Current literature assumes that if a mechanism is strategy-proof then every individual will be honest. But to be strategy-proof mechanism only requires that the sincere profile Π be at least one of the Nash equilibria. It does not require Π to be the only Nash equilibrium. The set of Nash equilibria predicts the outcome of events, yet the literature ignores all other equilibria when the mechanism is strategy-proof. This is reasonable because it makes little sense for people to lie when there is not an incentive to do so. Minimal dishonesty captures this behavior. Furthermore, if individuals are minimally dishonest, then being honest is the only Nash equilibrium for a strategy-proof mechanism.

Theorem 2.3.3. *A mechanism is strategy proof if and only if the sincere profile* Π *is the only minimally dishonest equilibrium for all* $\Pi \in \mathcal{P}$.

Proof. The second direction follows by definition; if Π is a minimally dishonest Nash equilibrium then it is also a Nash equilibrium and therefore the mechanism is strategy-proof.

For the first direction, since the mechanism is strategy-proof, Π is a Nash equilibrium for Π. Certainly individual v is minimally dishonest since v is completely honest and Π is a minimally dishonest Nash equilibrium. Now we need to show there can be no other minimally dishonest equilibria. Suppose $\bar{\Pi} \neq \Pi$ is a minimally dishonest Nash equilibrium and let v be such that $\bar{\pi}_v \neq \pi_v$. Since the mechanism is strategy proof, individual gets at least as good an outcome if they submit the more honest π_v. Therefore $\bar{\Pi}$ is not a minimally dishonest Nash equilibrium and the theorem statement holds. $\square$

We've argued that minimal dishonesty is logically intuitive and that it explains the assumptions researchers make when using strategy-proof mechanisms. Most importantly, our hypothesis is supported by a substantial body of empirical evidence from the experimental economics and psychology literatures that people are averse to lying. Gneezy [29] experimentally finds that people do not lie unless there is a benefit. Hurkens and Kartik [32] perform additional experiments that confirm an aversion to lying, and show their and Gneezy's data to be consistent with the assumption that some people never lie and others always lie if it is to their benefit. Charness and Dufwenberg [14] experimentally find an aversion to lying and show that it is consistent with guilt avoidance. Battigalli [10] experimentally find some contexts in which guilt is insufficient to explain aversion to deception. Several papers report evidence of a "pure" i.e., context-independent, aversion to lying [42, 15, 29, 32] that is significant but not sufficient to fully explain experimental data.

The set of research results we have cited here is by no means exhaustive. Two additional ones are of particular relevance to our concept of minimal dishonesty. Mazar *et al.* [46] find that "people behave dishonestly enough to profit but honestly enough to delude themselves of their own integrity." Lundquist *et al.* [44] find that people have an aversion to lying which increases with the size of the lie. Both of these studies support our hypothesis that people will not lie more than is necessary to achieve a desirable outcome.

Some experimental evidence is less confirmatory of our hypothesis. Several studies, beginning with [29], have found an aversion to lying if doing so would disbenefit someone else substantially more than the benefit one would accrue.

2.4 The Price of Anarchy and The Price of Deception

Let us now elaborate on the differences between the Price of Deception and the Price of Anarchy. The latter measure was initially employed in [19] and then [8], later aptly named by [40, 52], and now is widely employed in computer science. It is defined as the worst-case ratio between the maximum overall benefit achievable by a theoretical centralized authority and the overall benefit that would be achieved if all agents behaved autonomously to maximize their individual utilities. For example, each packet traversing the internet might "selfishly" route itself to arrive as quickly as possible, slowed by other packets utilizing the same communication links, but without regard for the overall performance of the internet [63]. The overall internet performance with selfish autonomous packets would be compared with the overall performance if a centralized authority routed each packet. Thus, the Price of Anarchy measures the decrease in benefit motivated by differences among individual utility functions that is enabled by individual control of *actions*. The Price of Deception is the decrease in benefit motivated by differences among individual utility functions that is enabled by individual control of *information*. The submission of preference information can be viewed as an action, but the Price of Deception cannot be viewed as a Price of Anarchy unless we deny the existence of private information, because the centralized authority would have to know each voter's sincere preferences. However, one can view a Price of Anarchy as a Price of Deception through a valid transformation.

2.4.1 Games of Anarchy

A Game of Anarchy, $G_A(N, \mathcal{S}, \mu)$, consists of

$N :=$ a set of n players.

$\mathcal{S}_i :=$ the set of actions player i can commit.

$S = (s_1, ..., s_n) \in \mathcal{S}_1 \times ... \times \mathcal{S}_n = \mathcal{S} :=$ set of actions committed by all players.

$\mu_i(S) :=$ individual i's value of S.

$W(S) :=$ society's value for S.

$Eq :=$ set of $S' \in \mathcal{S}$ that form a Nash equilibrium.

In a Game of Anarchy, individual i decides to commit an action $s_i \in \mathcal{S}_i$. Individual i then receives a utility of $\mu_i(S)$ and society receives a utility of $W(S)$. If instead a centralized coordinator controlled the actions of others, society would receive a utility of $\max_{S' \in \mathcal{S}} W(S')$. The set of equilibria Eq are the set of $S' \in \mathcal{S}$ where no individual i can alter s'_i in order to increase the value of $\mu_i(S')$. The Price of Anarchy is the worst case ratio between the societal value obtained with a centralized coordinator and the societal value at an equilibrium without a coordinator. Formally, the Price of Anarchy is

$$\sup_{S \in \mathcal{S}} \sup_{\overline{S} \in Eq} \frac{W(S)}{W(\overline{S})}. \tag{2.45}$$

The Price of Anarchy naturally occurs when information is public but each individual has control of his actions.

2.4.2 The Obstacle to Reducing Deception to Anarchy

The Price of Anarchy arises naturally in settings where information is public but control of actions is determined in private whereas the Price of Deception arises naturally in settings where information is private but control of actions is determined by the public. Conceptually, Price of Anarchy and Price of Deception are very different despite their similarity of mathematical form. We now attempt to reduce the latter to the former. Given a deterministic Game of Deception , $G_D(N, \mathcal{P}, \Pi, r, u)$ one can define a Game of Anarchy $G_{A,\Pi}(N, \mathcal{S}, \mu)$ that has the same set of equilibria. Let $S \in \mathcal{P}$ correspond to the action of submitting information S to the decision mechanism. The set of all possible combinations of information that the players can submit is $\mathcal{S} := \mathcal{P}$. Individual i's utility is $\mu_i(S) = u_i(\Pi, r(S))$ and society receives a utility of $W(S) = U(\Pi, r(S))$. If there were a centralized coordinator that could control the information that each player submits, the coordinator would force each player to be honest and society would receive a utility of $\max_{S \in \mathcal{S}} W(S) = W(\Pi)$. The set of equilibria Eq would then equal $E(\Pi)$. We have the following relationship:

$$PriceofDeception\Big(G_D(N, \mathcal{P}, \Pi, r, u)\Big) = PriceofAnarchy\Big(G_{A,\Pi}(N, \mathcal{P}, u(\Pi, \cdot))\Big) \tag{2.46}$$

However, this reduction only yields the Price of Deception of a decision mechanism for a specific set of sincere preferences Π. To obtain the Price of Deception of the decision mechanism, we would have to take the supremum of the Price of Anarchy of a family of games of anarchy. This may be problematic, but it is not the primary obstacle blocking the reduction of Price of Deception to Price of Anarchy. The primary obstacle is that the transformation we have sketched from Price of Deception to Price of Anarchy violates the premise of Price of Deception. The central coordinator required in the Game of Anarchy to obtain an optimal outcome for society would have to force individuals to be honest, which is impossible since a central coordinator would need access to private information to verify that players are honest. Furthermore, there is still a central decision mechanism embedded in the individual's utility functions in the game with anarchy – the Game of Anarchy lacks anarchy.

2.4.3 Reducing Anarchy to Deception

In contrast, Price of Anarchy can reduce to Price of Deception. Given a Game of Anarchy, $G_A(N, \mathcal{S}, \mu)$, we define a Game of Deception, $G_D(N, \mathcal{P}, \Pi, r, u)$, that has the same set of equilibria. Let $S^* \in \text{argmax}_{S \in \mathcal{S}} W(S)$. Individual i's private information π_i corresponds to the action that individual i should take in order to maximize society's utility (i.e. $\pi_i = s_i^*$). However, individual i may act strategically and may tell the decision mechanism that $\bar{\pi}_i \in \mathcal{S}_i$ is the action that i should take in order to maximize society's utility (i.e. $\mathcal{P} = \mathcal{S}$). The central decision mechanism then selects the outcome $r(\bar{\Pi})$ where each individual must complete the action described by $\bar{\Pi}$. Individual i receives utility $u_i(S_i^*, r(\bar{\Pi})) = \mu(\bar{\Pi})$ and society receives a utility of $U(S^*, r(\bar{\Pi})) = W(\bar{\Pi})$. If each individual were honest, then the central mechanism would assign individual i the action described by S_i^*. The set of equilibria $E(S^*)$ is then precisely equal to Eq. We then have the following relationship:

$$PriceofAnarchy\Big(G_A(N, \mathcal{S}, \mu)\Big) = PriceofDeception\Big(G_D(N, \mathcal{S}, S^*, r, u)\Big) \quad (2.47)$$

Whereas our attempted reduction from Price of Deception to Price of Anarchy violates the existence of private information that is a premise of the Price of Deception, reducing Price of Anarchy to Price of Deception can be easily explained by a traveling consultant and a lazy or incompetent central coordinator.

Suppose we have a lazy or incompetent central coordinator who simply lets everyone act on their own. Further suppose that one day a traveling consultant visits, analyzes the system, informs each individual what task they should perform to optimize society's utility and leaves forever. Upon hearing of the consultant's visit, the central coordinator wishes to make everyone act in the optimal way. However, the task an individual should complete is private information known only by the individual. Thus, the coordinator asks every individual to report their private information and instructs each individual to do their reported task. Individuals interested in their own self-gain may strategically alter their information to be assigned any task that they desire.

We view the reduction just given as artificial, despite its logical validity. Hence we prefer to view Price of Anarchy and Price of Deception as distinct concepts.

3

PRICE OF DECEPTION IN THE ASSIGNMENT PROBLEM

3.1 Introduction

In this chapter, we study algorithms that assign indivisible jobs to individuals based on the individuals' valuations of tasks. Some authors have observed that when there are many individuals, each individual's contribution to the problem is small and hence there is no incentive to be dishonest [33]. These assumptions cannot be valid since it is well known that no assignment algorithm satisfies symmetry, pareto optimality, and strategy-proofness [69]. Therefore the use (and analysis) of manipulable assignment algorithms is necessary.

We consider two common models for submitting preference individuals. In the first, individuals submit valuations for each task. The values come from the continuous domain $[0, 1]$. Second, we consider ordinal based information where individuals rank each task. This corresponds to each individual assigning a unique integer value in $\{1, ..., n\}$ to each job. In both variants we find that minimization problems have higher Prices of Deception than maximization problems. We also consider both lexicographic and randomized tie-breaking rules. For this setting, we find that the tie-breaking rule can lead to significant changes in the Price of Deception. Specifically for the Max Benefit Assignment Problem with continuous valuations, the Price of Deception with lexicographic tie-breaking is $\Theta(n^2)$ while random tie-breaking has a Price of Deception of only $O(n^{\frac{2}{3}})$. These analyses prove that manipulable mechanisms may suffer from manipulation in different ways. Moreover, the Price of Deception captures this behavior and is an essential tool for evaluating the quality of a decision mechanism.

All mechanisms we analyze yield Pareto optimal solutions when individuals are sincere. Moreover, when ties are broken randomly, the mechanisms are also symmetric and therefore not strategy-proof. We also consider manipulable non-symmetric lexicographic tie-breaking rules. The results of our analyses are in Table 3.1 below.

Table 3.1: Price of Deceptions for Various Assignment Procedures.

	Lexicographic	Random
$[0,1]$ Max Assignment	$n^2 - n + 1$	$\Omega\left(\sqrt{n}\right), O\left(n^{\frac{2}{3}}\right)$
$[0,1]$ Min Assignment	∞	$\geq n$
Integral Max Assignment	$2\frac{n-1}{n}$	< 4
Integral Min Assignment	$\frac{n}{2}$	$< \frac{3n+1}{4}$

3.1.1 Lexicographic Tie-Breaking

The two Prices of Deception that stand out most in Table 3.1 are $[0,1]$ Min Assignments with lexicographic tie-breaking and Integral Max Assignments. With a Price of Deception of ∞, the $[0,1]$ Min Assignment Problem is heavily impacted by manipulation – strategic behavior can lead to results that are arbitrarily poor relative to the sincere solution. On the other end of the spectrum, the Integral Max Assignment Problem, while still impacted by manipulation, has a constant sized Price of Deception – strategic behavior will never cause the social utility to change by more than a factor of two.

The minimization problems have worse Prices of Deceptions than maximization problems. When individuals submit ordinal data (Integral Min/Mas Assignments) there is a simple reason for this: Both variants can have a solution with value anywhere in the interval of $[n, n^2]$. Moreover, in the worst-case equilibrium for both variants, the value of the solution will have a value of approximately $\frac{n^2}{2}$. While $\frac{n^2}{2}$ is approximately the same distance from n and n^2, it is of a different magnitude of n (the best-case value for the Min Assignment Problem) whereas it is the same magnitude as n^2 (the best-case value for the Max Assignment Problem). Hence why we see that one Price of Deception is approximately $\frac{n}{2}$ while the other is 2. While this difference is significant for measuring how close we remain to the desired outcome, it may be seen as somewhat artificial.

However in the $[0,1]$ variant, the differences between the Prices of Deception are caused by something more significant than an affine scaling of utilities. We show in Theorem 3.3.3 that in the $[0,1]$ Max Assignment Problem with lexicographic tie-breaking that the kth individual will receive one of her k favorite jobs which heavily restricts what an equilibrium solution can look like. But when we analyze the Min Assignment Problem with lexicographic tie-breaking, we can only guarantee that the kth individual can avoid one of her k least favorite jobs. This places almost no restrictions on the set of equilibria and manipulation can lead to very poor results.

3.1.2 Random Tie-Breaking

In the $[0,1]$ Max Assignment Problem where lexicographic tie-breaking yields a Price of Deception of $n^2 - n + 1$ where as random tie-breaking yields a Price of Deception of only at most $n^{\frac{2}{3}}$ – an improvement of $n^{\frac{4}{3}}$. Something as small as a tie-breaking rule can significantly decrease the impact of manipulation. This seems intuitive; lexicographic tie-breaking is by definition unfair and gives lower indexed individuals additional power to manipulate the system. However, we will see in later chapters that random tie-breaking can actually lead to worse outcomes than lexicographic tie-breaking. This means we must be careful when selecting a tie-breaking rule – it seems unlikely that there is a single tie-breaking rule that is always best.

In the remaining algorithms we examine, it is unclear whether random tie-breaking is better or worse than lexicographic tie-breaking. We conjecture that the Prices of Deception of the three remaining mechanisms when breaking ties randomly are lower than when ties are broken lexicographically. We have, however, proved that manipulation still impacts the $[0,1]$ Min Assignment Problem more than the $[0,1]$ Max Assignment Problem and that manipulation has a relatively small impact on the Integral Max Assignment Problem – once again it has a constant sized Price of Deception.

3.1.3 Price of Deception's Response to Indifference

The Price of Deception of ∞ of the $[0, 1]$ Min Assignment Problem with lexicographic tie-breaking does not tell the full story of the effect of manipulation. The proof of Theorem 3.3.7 provides a family of instances where the value of the sincere optimal solutions converge to zero while the value of the strategic solution is always at least some constant c. The value of the constant c is important.

Suppose we have two mechanisms with Prices of Deceptions of ∞. For the first mechanism suppose that $c = 1$ meaning that as the cost of the sincere solution approaches zero, the cost of the strategic solution is 1. For the second mechanism, suppose $c = n$. The second mechanism is n times as bad even though they have the same Prices of Deception.

To capture the difference, we artificially bound each valuation of each job to be at least $\frac{\delta}{n}$ for some $\delta < 1$. This implies that the cost of the sincere solution is bounded below by δ. Therefore for the first mechanism mentioned above, the Price of Deception is $\frac{1}{\delta}$ while the second mechanism has a Price of Deception of $\frac{n}{\delta}$. Thus we are able to distinguish between the two. When we apply this concept to the the $[0, 1]$ Min Assignment Problem, we determine that the Price of Deception is $\Theta(\frac{n}{\delta})$ (Theorem 3.3.8) which is the worst Price of Deception any of our assignment procedures can have.

The process of bounding valuations has another important property; it measures how much individual preferences vary between jobs. Since every valuation sums to 1 (i.e. $\sum_{j=1}^{n} \pi_{ij} = 1$), we know that if all valuations are at least $\frac{\delta}{n}$, then all valuations are also at most $1 - \frac{n-1}{n}\delta$. This implies that the difference in valuations of an individuals most preferred and least preferred tasks is at most $1 - \delta$. Thus, we can think of δ as a quantitative measure of indifference; δ approaching 1 is equivalent to individuals being almost indifferent between all jobs.

In the context of the $[0, 1]$ Min Assignment Problem, the Price of Deception can still be large ($\frac{n}{\delta}$ even when individuals are mostly indifferent between jobs). However, the Price of Deception of the $[0, 1]$ Max Assignment Problem is highly impacted by indifference.

Specifically, if each individual has a value of at least $\frac{\delta}{n}$ on each job, then the Price of Deception is at most $\frac{n^2-n-1+(n-1)^2\delta}{1+(n-1)\delta}$ implying that as individuals become more indifferent between jobs (as $1-\delta$ becomes small), the Price of Deception converges to 1 at a relatively quick rate; that is to say that if people are almost indifferent, then manipulation has little impact on the quality of the outcome.

3.1.4 Serial Dictatorship and Aversion to Lying

In Theorem 3.3.3 and Corollary 3.3.12 we show that if individuals are strategic in the $[0,1]$ Max, Integral Max, and Integral Min Assignment Problems with lexicographic tie-breaking then the kth individual is guaranteed one of their top k ranked jobs; specifically we show that if each individual is strategic then each algorithm is equivalent to individuals iteratively picking their most preferred job. This process is referred to as Serial Dictatorship. Notably in [69], Zhou gives the Serial Dictatorship algorithm as an example of a pareto optimal, strategy-proof algorithm.

Since the outcomes of all three manipulable algorithms are equivalent to the outcome of Serial Dictatorship and since Serial Dictatorship is strategy-proof it may seem an improvement to just use Serial Dictatorship. However, we claim that running Serial Dictatorship does not guarantee the same results as $[0,1]$ Max, Integral Max, and Integral Min algorithms; recall from Section 2.3 that some individuals act strategically but minimally dishonesty while other individuals have a pure aversion to lying. When some individuals are truthful, Theorem 3.3.3 and Corollary 3.3.12 only extend to strategic individuals; i.e. if the kth individual is strategic, she is guaranteed one of her top k ranked jobs. Therefore if our objective is the Integral Max Assignment, then it will still obtain better outcomes than Serial Dictatorship as long as we have some truthful individuals.

We also remark that opting to use Serial Dictatorship, like the revelation principle (see Section 2.2), does not truly remove the impact of manipulation even when all individuals are strategic; while we manage to induce individuals to be honest, if our goal is the outcome of the $[0, 1]$ Max Assignment Problem, then, by using Serial Dictatorship, we may still end up with a solution that is worse by a factor of $n^2 - n + 1$.

3.2 The Model

We begin by reinforcing many of the definitions from Chapter 2 by presenting them in context of the Assignment Problem.

Definition 3.2.1. An instance of the Assignment Problem consists of

- Sets $V = \{v_1, ..., v_n\}$ and $J = \{j_1, ..., j_n\}$ of people and jobs, respectively.
- For each $i \in V$, a set of values or costs π_i where π_{ij} is where j is the ith person's value (cost respectively) of job j for each $j \in J$. Individual i's utility of j, $u_i(j)$ is not necessarily equal to π_{ij} but $u_i(j) > u_i(j')$ if and only if $\pi_{ij} > \pi_{ij'}$.
- Denote by $\mathcal{P}_i$ the set of possible preferences player i can submit.
- Denote by Π the collection of π_i over all $i \in V$. $\Pi \in \mathcal{P}$ is called the *preference profile*.
- The profile Π is submitted to a publicly known matching procedure r. The outcome $r(\Pi)$ corresponds to a perfect matching between people and jobs or a distribution over the set of perfect matchings (for randomized matching procedures).

A mechanism r may impose restrictions on how an individual may report their preferences. For instance, in the $[0, 1]$ Max Assignment Problem, π_{ij} corresponds to the value person i places on job j. Utilities are normalized such that $\sum_{j \in J} \pi_{ij} = 1$ for each individual i. Therefore person i defines π_{ij} as $\frac{u_i(j)}{\sum_{j' \in J} u_i(j')}$.

Let M be a perfect matching between V and J. A matching procedure r selects assignment(s) $r(\Pi) \subseteq \operatorname{argmax}_M U(\Pi, M)$ where $U(\Pi, M) = \sum_{\{i,j\} \in M} \pi_{ij}$ is society's value of the assignment M. If Π instead refers to costs, then replace argmax with argmin. The set

$\sum_{\{i,j\}\in M} \pi_{ij}$ is not necessarily a singleton and therefore we must determine how to break ties to find an assignment of jobs to individuals. We consider two tie-breaking rules: lexicographic tie-breaking – assigning the lowest indexed individual their most valued task – and random tie-breaking – randomly re-indexing individuals and then using the lexicographic tie-breaking rule. When a random tie-breaking rule is used, we assume individuals and society are risk neutral when evaluating the quality of the outcome $r(\Pi)$.

Unfortunately, there is no a priori guarantee that individuals will be honest; individual i may falsely report a strategic $\bar{\pi}_i \neq \pi_i$ to obtain a better outcome. Therefore the submitted preferences $\bar{\Pi}$ may be significantly different than the sincere preferences Π. Not only does this mean the outcome may change when individuals act strategically (i.e. $r(\Pi) \neq r(\bar{\Pi})$), but the social of utility of the outcome may drastically decrease (i.e. $U(\Pi, r(\bar{\Pi})) < U(\Pi, r(\Pi))$). To understand the effect of manipulation of various assignment procedures, we analyze the Strategic Assignment Game.

Strategic Assignment Game

- Each individual i has information $\pi_i \in \mathcal{P}_i$ describing their preferences. The collection of all information is the (sincere) profile $\Pi = \{\pi_i\}_{i=1}^n$.
- To play the game, individual i submits putative preference data $\bar{\pi}_i \in \mathcal{P}_i$. The collection of all submitted data is denoted $\bar{\Pi} = \{\bar{\pi}_i\}_{i=1}^n$.
- It is common knowledge that a central decision mechanism will select assignment $r(\bar{\Pi}, \omega)$ when given input $\bar{\Pi}$ and random event ω.
- The random event $\omega \in \Omega$ is selected according to μ. We denote $r(\bar{\Pi})$ as the distribution of outcomes according to Ω and μ.
- Individual i evaluates $r(\bar{\Pi})$ according to i's sincere preferences π_i. Specifically, individual i's utility of the set of outcomes $r(\bar{\Pi})$ is $u_i(\pi_i, r(\bar{\Pi}))$.

By the definition of the Strategic Assignment Game, a set of submitted preferences $\bar{\Pi}$ forms a pure strategy Nash equilibrium if no individual i would obtain an outcome they sincerely prefer to $r(\bar{\Pi})$ (with respect to π_i) by altering $\bar{\pi}_i$.

Example 3.2.2. A Nash equilibrium of the Strategic Assignment Game.

Consider the $[0,1]$ *Max Assignment Problem where* $\sum_{j \in J} \pi_{ij} = 1$ *and we seek a matching that maximizes* $\sum_{\{i,j\} \in M} \pi_{ij}$*. In the event of more than one such matching, we break ties lexicographically by assigning the lowest index individual their most preferred task. Consider the following sincere preferences:*

$$\pi_1 = (.30, .20, .40, .10) \tag{3.1}$$

$$\pi_2 = (.24, .24, .25, .23) \tag{3.2}$$

$$\pi_3 = (.50, .20, .10, .20) \tag{3.3}$$

$$\pi_4 = (.20, .10, .60, .10) \tag{3.4}$$

When individuals are honest, we obtain the assignment $r(\Pi) = \{\{1,2\},\{2,4\},\{3,1\},\{4,3\}\}$ *indicating that individual* 1 *is assigned job* 2*, individual* 2 *is assigned job* 4*, etc. With respect to the sincere preferences, the social utility of this outcome is* $U(\Pi, r(\Pi)) = \pi_{12}+\pi_{24}+\pi_{31}+\pi_{43} = 1.53$*. However, honesty is not the best policy for each of these individuals. For instance, if individual* 1 *instead submits* $\bar{\pi}_1 = (1,0,0,0)$ *then he will be assigned job* 1*, a job she prefers. A Nash equilibrium is given below:*

$$\bar{\pi}_1 = (0, 0, 1, 0) \tag{3.5}$$

$$\bar{\pi}_2 = (.24, .24, .25, .23) \tag{3.6}$$

$$\bar{\pi}_3 = (1, 0, 0, 0) \tag{3.7}$$

$$\bar{\pi}_4 = (.20, .10, .60, .10) \tag{3.8}$$

When individuals submit the putative $\bar{\Pi}$*, the outcome is* $r(\bar{\Pi}) = \{\{1,3\},\{2,2\},\{3,1\},\{4,4\}\}$*.*

Society believes it has selected an outcome with utility $U(\bar{\Pi}, r(\bar{\Pi})) = \bar{\pi}_{13} + \bar{\pi}_{22} + \bar{\pi}_{31} + \bar{\pi}_{44} = 2.84$. *But society's true utility is defined with respect to* Π *and is* $U(\Pi, r(\bar{\Pi})) = \pi_{13} + \pi_{22} + \pi_{31} + \pi_{44} = 1.24$ *and therefore social utility decreased as a result of strategic behavior.*

To establish that $\bar{\Pi}$ *is a Nash equilibrium, consider each person individually. Neither individual* 1 *nor* 3 *can alter their preferences to get a better result since they already have their most preferred outcome. Individual* 2 *only prefers job* 3 *but since* $\bar{\pi}_1 = (0, 0, 1, 0)$ *individual* 1 *will always be assigned job* 3 *by the lexicographic tie-breaking rule regardless of the other preferences. Similarly, individual* 4 *can not alter her preferences to obtain job* 1 *or* 3*, the only jobs she prefers to her current assignment. Therefore* $\bar{\Pi}$ *is a Nash equilibrium and strategic behavior not only changes the outcome, it decreases social utility from* 1.53 *to* 1.24

Example 3.2.2 shows an instance where the social utility decrease from 1.53 to 1.24; manipulation resulted in an outcome that is $\frac{1.53}{1.24} \approx 1.234$ times as bad. For the remainder of the chapter, we focus on finding a bound, the Price of Deception (see Section 2.2), on how much manipulation impacts social utility of various assignment procedures.

Definition 3.2.3 (Price of Deception for Strategic Assignment Games)**.** Let r be the an assignment procedure. Let U be an associated real-valued function that, given a profile Π of individual preferences over J, outputs for each assignment M a *societal utility (or cost)* $U(\Pi, M)$ of assignment M with respect to Π. The function r must be such that for all Π, $r(\Pi) \subseteq \operatorname{argmax}_M U(\Pi, M)$ (respectively argmin). Let $NE(\Pi)$ denote the set of equilibria of the Strategic Assignment Game. Then the Price of Deception of procedure r is

$$\sup_{Pi \in \mathcal{P}} \sup_{\bar{\Pi} \in NE(\Pi)} \frac{U(\Pi, r(\Pi))}{E(U(\Pi, r(\bar{\Pi})))}. \tag{3.9}$$

If U corresponds to a cost, the Price of Deception is instead the reciprocal of (3.9).

Based on Example 3.2.2, we know the Price of Deception of the $[0, 1]$ Max Assignment Problem is at least 1.234. In this chapter we do not consider minimal dishonesty; unlike facility location, voting and stable marriage algorithms that we analyze later, we show that there is an intuitive characterization for the set of Nash equilibria in the Strategic Assignment Game and that there is no need for minimal dishonesty.

3.3 The Price of Deception with Lexicographic Tie-Breaking

We first find the Price of Deception when ties are broken lexicographically. If there exist multiple optimal solutions, we select one where the lowest indexed individual is assigned their most valued (or least costly) object/task that is lowest indexed iteratively.

3.3.1 $[0,1]$ Max Assignments

We define the $[\delta,\gamma]$ Max Assignment Problem to be such that $\pi_{ij} \geq \frac{\delta}{n}$ and $\sum_j \pi_{ij} = \gamma$ for each individual i. Unlike the Fair Division Problem, the valuations for each item come from a continuous domain. Without loss of generality, it suffices to analyze this model when $\gamma = 1$. We begin with the most common $\delta = 0$.

Lemma 3.3.1. *If $\pi_{11} = 1$, then $r(\Pi)$ assigns individual 1 to job 1 in the $[0,1]$ Max Benefit Assignment Problem.*

Proof. For contradiction, assume that individual 1 is assigned task $k \neq 1$ and that player i gets job 1. Consider switching these two assignments. We have that the change in the objective value is given by $\pi_{11} + \pi_{ik} - \pi_{1k} - \pi_{i1} = 1 + \pi_{ik} - \pi_{i1} \geq 0$. But this implies the objective value is at least as good when assigning k to individual 1. This is a contradiction since ties are broken lexiographically. $\square$

Corollary 3.3.2. *Individual i can alter her preferences to obtain one of her top i ranked jobs.*

The proof of Corollary 3.3.2 follows in the same fashion as Lemma 3.3.1. Individual i simply needs to submit $\bar{\pi}_{ij} = 1$ for a job j where $\bar{\pi}_{i'j} < 1$ for all $i' < i$. Corollary 3.3.2 shows that the $[0,1]$ Max Assignment Problem with lexicographic tie-breaking is so responsive to preferences that strategic behavior results in players iteratively picking the task or object remaining that they most prefer. This process is known as Serial Dictatorship. We describe the set of outcomes of Nash equilibria with the following algorithm.

Algorithm 1 Algorithm for finding outcomes for $[0, 1]$ Max Assignment

1: **procedure** SERIALDICTATORSHIP
2: $\quad Jobs \leftarrow [n]$
3: $\quad x_{ij} \leftarrow 0 \; \forall \{i, j\} \in [n]^2$
4: $\quad$ **for** $i = 1 \; to \; n$ **do**
5: $\quad\quad$ Select $j \in \operatorname{argmax}_{k \in Jobs} \pi_{ik}$
6: $\quad\quad Jobs \leftarrow Jobs \setminus \{j\}$
7: $\quad\quad x_{ij} = 1$
8: $\quad$ **end for**
9: $\quad$ Output r
10: **end procedure**

The output x of Algorithm 1 decribes the matching. Specifically, $x_{ij} = 1$ if and only if individual i is assigned to job j. Similarly, we write $r_{ij}(\Pi) = 1$ if mechanism r assigns i to j when given profile Π.

Theorem 3.3.3. *Let Π be a set of sincere preferences. The assignment x is obtainable at a Nash equilibrium iff x is a possible output from Algorithm 1*

Proof. One direction is simple; if x is not obtainable from Algorithm 1 then there is an individual i that not obtain one of her top i preferred jobs contradicting Corollary 3.3.2.

For the other direction, let x be an output of Algorithm 1 and consider the putative profile $\bar{\Pi} = x$. We have that $\bar{\pi}_{ij} > 0$ iff $\bar{\pi}_{ij} = 1$ iff $x_{ij} = 1$ and therefore $r(\bar{\Pi}) = x$. We now establish that $\bar{\Pi}$ is a Nash equilibrium. Consider individual i and her assigned job j where $\bar{\pi}_{ij} = x_{ij} = 1$. Let job k be such that $\pi_{ij} > \pi_{ik}$. By line 5 of Algorithm 1, $x_{lk} = 1$ for some $l < i$. By definition, $\bar{\pi}_{lk} = 1$ and individual i cannot alter her preferences to obtain job k since $l < i$ and ties are broken lexicographically. Therefore individual i cannot alter her preferences to obtain a better outcome and $\bar{\Pi}$ is a Nash equilibrium. $\square$

Theorem 3.3.4. *The Price of Deception in the $[0, 1]$ Max Benefit Assignment Problem is $n^2 - n + 1$.*

Proof. Let $X_i = u_i(\pi_i, r(\Pi))$ be the value individual i obtains when she is sincere and let $Y_i = u_i(\pi_i, r(\bar{\Pi}))$ be the value at a Nash equilibrium $\bar{\Pi}$. Each X_i is at most one, each Y_i is at least zero, and by Lemma 3.3.1, $Y_1 \geq X_1$. Therefore the Price of Deception is

$$\frac{U(\Pi, r(\Pi))}{U(\Pi, r(\bar{\Pi}))} = \frac{X_1 + X_2 + \cdots + X_n}{Y_1 + Y_2 + \cdots + Y_n} \tag{3.10}$$

$$\leq \frac{X_1 + (n-1)}{Y_1} \tag{3.11}$$

$$\leq \frac{Y_1 + (n-1)}{Y_1}. \tag{3.12}$$

By Lemma 3.3.1, Y_1 corresponds to the value of individual 1's sincerely most preferred task and therefore $Y_1 \geq \sum_{j=1}^{n} \frac{\pi_{1j}}{n} \geq \frac{1}{n}$ and the Price of Deception is at most $n^2 - n + 1$.

We now give an example to show that this is tight:

$$\pi_{11} = \frac{1}{n} + \epsilon \tag{3.13}$$

$$\pi_{1j} = \frac{1}{n} - \frac{\epsilon}{n-1} \ \forall j > 1 \tag{3.14}$$

$$\pi_{ii} = \epsilon \ \forall i > 1 \tag{3.15}$$

$$\pi_{i,i-1} = 1 - \epsilon \ \forall i > 1 \tag{3.16}$$

$$\pi_{ij} = 0 \ \forall i > 1, \ \forall j \notin \{i, i-1\} \tag{3.17}$$

The optimal solution for this problem is $(n-1)(1-\epsilon) + \frac{1}{n} - \frac{\epsilon}{n-1}$ and is obtained by giving everyone but person 1 her favorite job. By Theorem 3.3.3, there exists a Nash equilibrium $\bar{\Pi}$ where $r(\bar{\Pi})_{ii} = 1 \ \forall i \in [n]$ with a value of $U(\Pi, r(\bar{\Pi})) = \frac{1}{n} + \epsilon + (n-1)\epsilon$ yielding a Price of Deception of:

$$\frac{(n-1)(1-\epsilon) + \frac{1}{n} - \frac{\epsilon}{n-1}}{\frac{1}{n} + \epsilon + (n-1)\epsilon} \to n^2 - n + 1 \text{ as } \epsilon \to 0 \tag{3.18}$$

completing the proof of the theorem. □

In the proof of Theorem 3.3.4 we could have easily taken $\epsilon = 0$ for the instance Π to obtain the same result without examining the limiting behavior. However, individual 1 would be indifferent among all tasks. While $\bar{\Pi}$ would still be a Nash equilibrium, it would be an equilibrium where individual 1 is lying spuriously – she had no incentive to lie and would have received the same utility had she been honest. However, by taking $\epsilon > 0$, individual 1 has reason to lie and is minimally dishonest (Section 2.3).

3.3.2 $[\delta, 1]$ Max Assignments

In some settings, we may know a priori that $\pi_{ij} \geq \frac{\delta}{n}$ for some δ. Thus we may impose such a restriction on the individuals for their submitted preferences. As discussed in Section 3.1.3, $1 - \delta$ indicates how much individual preferences vary between tasks.

Both Corollary 3.3.2 and Theorem 3.3.3 extend to this setting (although $\bar{\pi}_{ij} = 1 - \frac{n-1}{n}\delta$ for all i, j where $x_{ij} = 1$ and $\bar{\pi}_{ij} = \frac{\delta}{n}$ in the proof of Theorem 3.3.3).

Theorem 3.3.5. *The Price of Deception in the* $[\delta, 1]$ *Max Benefit Assignment Problem is* $\frac{n^2-n+1-(n-1)^2\delta}{1+(n-1)\delta}$.

Proof. The proof follows in the same fashion as Theorem 3.3.4. Once again, let X_i be the value individual i gets when she is sincere and let Y_i be the value individual i gets in some equilibrium $\bar{\Pi}$. As before, $Y_i \geq \frac{\delta}{n}$ for all i and $\pi_{ij} = 1 - \sum_{k \neq j} \pi_{ik} \leq 1 - \frac{n-1}{n}\delta$ implying

$X_i \leq 1 - \frac{n-1}{n}\delta$. Thus, the Price of Deception is given by

$$\frac{U(\Pi, r(\Pi))}{U(\Pi, r(\bar{\Pi}))} = \frac{X_1 + X_2 + \cdots + X_n}{Y_1 + Y_2 + \cdots + Y_n} \tag{3.19}$$

$$\leq \frac{X_1 + (n-1)(1 - \frac{n-1}{n}\delta)}{Y_1 + (n-1)\frac{\delta}{n}} \tag{3.20}$$

$$\leq \frac{Y_1 + (n-1)(1 - \frac{n-1}{n}\delta)}{Y_1 + (n-1)\frac{\delta}{n}} \tag{3.21}$$

$$\leq \frac{\frac{1}{n} + (n-1)(1 - \frac{n-1}{n}\delta)}{\frac{1}{n} + (n-1)\frac{\delta}{n}}) \tag{3.22}$$

$$= \frac{n^2 - n + 1 - (n-1)^2\delta}{1 + (n-1)\delta} \tag{3.23}$$

The problem instance demonstrating that the bound is tight is similar to Theorem 3.3.4:

$$\pi_{11} = \frac{1}{n} + \epsilon \tag{3.24}$$

$$\pi_{1j} = \frac{1}{n} - \frac{\epsilon}{n-1} \; \forall j > 1 \tag{3.25}$$

$$\pi_{ii} = \frac{\delta}{n} + \epsilon \; \forall i > 1 \tag{3.26}$$

$$\pi_{i,i-1} = 1 - \epsilon - \frac{n-1}{n}\delta \; \forall i > 1 \tag{3.27}$$

$$\pi_{ij} = \frac{\delta}{n} \; \forall i > 1, j \notin \{i, i-1\} \tag{3.28}$$

The optimal value for this problem is $(n-1)(1 - \epsilon - \frac{n-1}{n}\delta) + \frac{1}{n} - \frac{\epsilon}{n-1}$ and is obtained by assigning everyone but person 1 his favorite job. By Theorem 3.3.3, there exists a Nash equilibrium $\bar{\Pi}$ where $r(\bar{\Pi})_{ii} = 1 \; \forall i \in [n]$ with a value of $U(\Pi, r(\bar{\Pi})) = \frac{1}{n} + n\epsilon + \frac{n-1}{n}\delta$ yielding a Price of Deception of:

$$\frac{\frac{1}{n} - \frac{\epsilon}{n-1} + (n-1)(1 - \epsilon - \frac{n-1}{n}\delta)}{\frac{1}{n} + \epsilon + \frac{n-1}{n}\delta} \to \frac{n^2 - n + 1 - (n-1)^2\delta}{1 + (n-1)\delta} \text{ as } \epsilon \to 0 \tag{3.29}$$

completing the proof of the theorem. □

To obtain the guaranteed bound of $\frac{n^2-n+1-(n-1)^2\delta}{1+(n-1)\delta}$, we need not impose the restriction that all submitted preferences have value at least $\frac{\delta}{n}$. It is only necessary that the each sincere value π_{ij} is at least $\frac{\delta}{n}$. This holds because Corollary 3.3.2 holds regardless of the value of $\delta < 1$ we give to the individuals. Thus there may seem to be no purpose in analyzing the $[\delta, 1]$ Max Assignment Problem. However, Theorem 3.3.5 tells us how much manipulation impacts society depending on how much individual preferences vary between jobs. For instance, if everyone is almost indifferent among tasks and $\frac{1}{2n} \leq \pi_{ij} \leq \frac{n+1}{2n}$ ($\delta = \frac{1}{2}$) for each individual then manipulation can only impact society by a factor of $\frac{n^2-n+1-(n-1)^2\frac{1}{2}}{1+(n-1)\frac{1}{2}} = \frac{n^2+1}{n+1} \approx n$.

3.3.3 $[0, 1]$ Min Assignments

In the Min Assignment Problem, $\pi_{ij} \geq 0$ denotes the cost for individual i to complete job j. Once again we require $\sum_j \pi_{ij} = 1$ for each individual i. We seek an assignment of jobs to individuals that minimizes the total cost.

Theorem 3.3.6. *For $n = 2$, the Price of Deception of the $[0, 1]$ Min Cost Assignment Problem is 3.*

Proof. We begin by establishing the upper bound. When there are only two individuals, there is a bijection between instances of the $[0, 1]$ Min Assignment Problem and instances of the $[0, 1]$ Max Assignment Problem; if individual i has a cost (benefit) of π_{ij} for job j in the Min (Max) Assignment Problem, that corresponds to a benefit (cost) of $\pi'_{ij} = 1 - \pi_{ij}$. Moreover $\sum_{j=1}^2 \pi'_{ij} = \sum_{j=1}^2 (1 - \pi_{ij}) = 1$ and Π' is a valid set of preferences. Furthermore, the social choice mechanism selects the same outcome in both instances: an assignment x minimizes (maximizes) $\sum_{i=1}^2 \sum_{j=1}^2 \pi_{ij} x_{ij}$ if and only if it maximizes (minimizes) $2 - \sum_{i=1}^2 \sum_{j=1}^2 \pi_{ij} x_{ij} = \sum_{i=1}^2 \sum_{j=1}^2 \pi'_{ij} x_{ij}$. Therefore Corollary 3.3.2 holds in this setting and individual 1 receives her most preferred job at a Nash equilibrium.

As in Theorem 3.3.4, let $X_i = u_i(\pi_i, r(\Pi))$ be the cost to individual i when everyone is honest and let $Y_i = u_i(\pi_i, r(\bar{\Pi}))$ be the cost at some Nash equilibrium $\bar{\Pi}$. By Corollary 3.3.2, $Y_1 \leq X_1$ and therefore $Y_1 \leq .5$. Since social "utility" is actually a cost we use the reciprocal of (2.19) to find the Price of Deception:

$$\frac{U(\Pi, r(\bar{\Pi}))}{U(\Pi, r(\Pi))} = \frac{Y_1 + Y_2}{X_1 + X_2} \tag{3.30}$$

$$\leq \frac{Y_1 + 1}{X_1} \tag{3.31}$$

$$\leq \frac{Y_1 + 1}{Y_1} \tag{3.32}$$

$$= \frac{.5 + 1}{.5} = 3 \tag{3.33}$$

To establish the lower bound, consider the following preferences:

$$\pi_{11} = .5 - \epsilon \qquad \pi_{12} = .5 + \epsilon \tag{3.34}$$

$$\pi_{21} = 0 \qquad \pi_{22} = 1. \tag{3.35}$$

In the optimal assignment individual 1 completes task 2 and individual 2 completes task 1 for a total cost of $.5 + \epsilon$. By Theorem 3.3.4, there is a Nash equilibrium $\bar{\Pi}$ where individual i is assigned task i for a sincere cost of $1.5 - \epsilon$ yielding a Price of Deception that tends to 3 as $\epsilon \to 0$. □

In the proof of Theorem 3.3.6 we established that the $[0, 1]$ Min Assignment Problem is the same as the $[0, 1]$ Max Assignment Problem when there are only 2 individuals. This is not true when there are more players and Corollary 3.3.2 does not extend. While individuals do not have the power to get whatever job they desire, manipulation has a much larger impact on the $[0, 1]$ Min Cost Assignment Problem.

Theorem 3.3.7. *The Price of Deception of the* $[0, 1]$ *Min Cost Assignment Problem is* ∞

for $n \geq 3$.

Thus far we have only seen mechanisms with finite Prices of Deception. An infinite Price of Deception indicates that manipulation can lead to arbitrarily poor results. The proof of such only requires a family of instances where the Prices of Deception of the instances tend to infinity.

Proof of Theorem 3.3.7. Consider the following sincere profile Π:

$$\pi_{in} = 1 \; \forall i \neq 1 \tag{3.36}$$

$$\pi_{ij} = 0 \; \forall i \neq 1 \; \forall j \neq n \tag{3.37}$$

$$\pi_{1n} = \epsilon \tag{3.38}$$

$$\pi_{11} = 0 \tag{3.39}$$

$$\pi_{1j} = \frac{1-\epsilon}{n-2} \; \forall \, 2 \leq j \leq n-1 \tag{3.40}$$

The sincere outcome $r(\Pi)$ assigns individual i to job $i-1$ for $i \geq 2$ and assigns individual 1 to job n for a total cost of $U(\Pi, r(\Pi)) = \epsilon$. Individual 1 can alter her preferences such that individual i is assigned to job i with the preferences given below:

$$\bar{\pi}_{1n} = 1 \tag{3.41}$$

$$\bar{\pi}_{1j} = 0 \; \forall j \neq n \tag{3.42}$$

$$\bar{\pi}_{ij} = \pi_{ij} \; \forall i \neq 1 \; \forall j \tag{3.43}$$

The selected outcome $r(\bar{\Pi})$ assigns individual i to job i. This corresponds to a sincere cost of $U(\Pi, r(\bar{\Pi})) = 1$. The profile $\bar{\Pi}$ corresponds to a Nash equilibrium. Only individual n would like to deviate from the solution but she cannot as any alterations to $\bar{\pi}_n$ give the appearance that she can complete job n for the cheapest amount. Thus, the Price of Deception of this instance is $\frac{1}{\epsilon} \to \infty$ as $\epsilon \to 0$. □

When there is an infinite Price of Deception, we actually lose some information. We cannot distinguish between a mechanism where the cost when everyone is honest is ϵ and the cost of the equilibrium is 1 and a mechanism where the cost when everyone is honest is again ϵ but the cost of the equilibrium is n. Both mechanisms have Prices of Deception that approach infinity as $\epsilon \to 0$, the effect of manipulation on the latter is n times worse. Although the instance in the proof of Theorem 3.3.7 makes the $[0, 1]$ Min Assignment Problem look like the former, we will show in the next section that it is more similar to the latter.

3.3.4 $[\delta, 1]$ Min Assignments

Similar to the $[\delta, 1]$ Max Assignment Problem, we restrict π_{ij} such that it is at least $\frac{\delta}{n}$. As with the Max Assignment Problem, such a restriction will give us a better idea of how the Price of Deception behaves with indifference between tasks. As mentioned at end of the previous section, this restriction will also give us an idea of just how bad the Price of Deception of ∞ is in the $[0, 1]$ Min Assignment Problem.

Theorem 3.3.8. *The Price of Deception in the* $[\delta, 1]$ *Min Cost Assignment Problem is in* $\Theta(\frac{n}{\delta})$.

Proof. We begin by establishing the upper bound. The worst possible assignment has a cost of at most $n(1 - \frac{(n-1)\delta}{n}) = n - (n-1)\delta$ since a job costs at most $1 - \frac{(n-1)\delta}{n}$ for an individual to complete. Similarly, the best possible assignment has a cost of δ. Therefore the Price of Deception is at most

$$\frac{n - (n-1)\delta}{\delta} = \frac{n}{\delta} - n + 1 \in \Theta\left(\frac{n}{\delta}\right) \tag{3.44}$$

We now establish the lower bound. Consider the following sincere preferences:

$$\pi_{11} = \frac{\delta}{n} \tag{3.45}$$

$$\pi_{1j} = \frac{1 - 2\frac{\delta}{n} - \epsilon}{n-2} \; \forall j \notin \{1, n\} \tag{3.46}$$

$$\pi_{1n} = \frac{\delta}{n} + \epsilon \tag{3.47}$$

$$\pi_{ii} = \frac{1}{2} - \epsilon - \frac{n-2}{n}\delta \; \forall i \notin \{1, n\} \tag{3.48}$$

$$\pi_{in} = \frac{1}{2} + \epsilon - \frac{n-2}{n}\delta \; \forall i \notin \{1, n\} \tag{3.49}$$

$$\pi_{ij} = \frac{\delta}{n} \; \forall i \notin \{1, n\} \; \forall j \notin \{i, n\} \tag{3.50}$$

$$\pi_{nn} = 1 - \frac{n-1}{n}\delta \tag{3.51}$$

$$\pi_{nj} = \frac{\delta}{n} \; \forall j \neq 1 \tag{3.52}$$

According to the lexicographic tie-breaking rules, $r(\Pi)$ assigns individual i to task $i-1$ for $i \neq 1$ and assigns individual 1 to task n for a total cost of $\delta + \epsilon$.

We now examine a set of putative preferences forming a Nash equilibrium with respect to Π.

$$\bar{\pi}_{in} = 1 - \frac{(n-1)\delta}{n} \; \forall i \tag{3.53}$$

$$\bar{\pi}_{ij} = \frac{\delta}{n} \; \forall i \; \forall j \neq n \tag{3.54}$$

Since ties are broken lexicographically, $r(\bar{\Pi})$ assigns individual i to task i. This corresponds to a real cost of

$$U(\Pi, r(\bar{\Pi})) = \frac{\delta}{n} + (n-2)\left(\frac{1}{2} - \epsilon - \frac{n-2}{n}\delta\right) + \left(1 - \frac{n-1}{n}\delta\right) \tag{3.55}$$

$$= \frac{n}{2} - \frac{(n-1)(n-2)}{n}\delta - (n-2)\epsilon \tag{3.56}$$

Similar to Theorem 3.3.7, $\bar{\Pi}$ is a Nash equilibrium since any deviation from $\bar{\pi}_i$ by individual

i causes i to be assigned to task n. Thus the Price of Deception is

$$\frac{\frac{n}{2} - \frac{(n-1)(n-2)}{n}\delta - (n-2)\epsilon}{\delta + \epsilon} \to \frac{\frac{n}{2} - \frac{(n-1)(n-2)}{n}\delta}{\delta} \; as \; \epsilon \to 0 \tag{3.57}$$

$$= \frac{n}{2\delta} - \frac{(n-1)(n-2)}{n} \in \Theta\left(\frac{n}{\delta}\right) \tag{3.58}$$

completing the proof of the theorem. □

3.3.5 Integral Max Assignments

In some cases of the assignment problem, we collect ordinal information from individuals and then treat it as cardinal data. Specifically, we have them rank their jobs such that $\pi_{ij} = n - k + 1$ if and only if individual i's kth favorite task is job j. In this setting, we seek to maximize the sum of individual benefits. Similar to the $[0, 1]$ Max Assignment Problem we show that the decision mechanism is very responsive to players' preferences.

Lemma 3.3.9. *Individual* 1 *can always change her preferences so there is at least one assignment* M' *that maximizes individuals' reported utilities and assigns individual 1 to job* a *for any* a*.*

Lemma 3.3.9 does not quite guarantee that player 1 can alter her preferences to receive her most preferred job. If she updates her preferences in accordance with Lemma 3.3.9 it does not guarantee the lexicographic tie-breaking rule will match her to her most preferred task. We address this later.

Prior to proving Lemma 3.3.9, we estalish a few preliminaries on finding optimal assignments. Every assignment problem can be represented by a complete bipartite graph with one set of nodes representing individuals and the other set representing jobs. The edge $\{i, j\}$ connecting individual i to job j has weight π_{ij}. An optimal assignment now corresponds to a max weight matching in the bipartite graph. We now present a characterization of the set of max weight matchings.

Definition 3.3.10. Let M be a perfect matching on a complete bipartite graph and let C be an even cycle such that $2 \cdot |M \cap C| = \frac{|C|}{2}$. That is to say that every other edge in C is also in M. The cycle C is **augmenting** if $\sum_{\{i,j\} \in C \setminus M} \pi_{ij} > \sum_{\{i,j\} \in M \setminus C} \pi_{ij}$.

If there exists an augmenting cycle C with respect to M then a better matching M' can be obtained by removing edges in $M \cap C$ and adding edges in $C \setminus M$. Formally, $M' = (M \setminus C) \cup (C \setminus M)$. Since $\sum_{\{i,j\} \in C \setminus M} \pi_{ij} > \sum_{\{i,j\} \in M \cap C} \pi_{ij}$, this will increase the value of the assignment and maintain the validity of the matching. Thus M is optimal if and only if M has no augmenting cycles. The 'only if' is an immediate extension of Berge's lemma [11].

Proof of Lemma 3.3.9. Suppose Π is a set of preferences where individual 1 is not assigned her job a in any optimal matching. Let M be an optimal matching and $\{1, a\} \notin M$. Let b and k be such that $\{1, b\} \in M$ and $\{k, a\} \in M$; individual 1 is assigned task b and individual k is assigned task a according to M.

Let $\bar{\Pi} = [\Pi_{-1}, \pi_k]$ be the profile obtained when individual 1 updates her preferences to k's preferences list π_k. Suppose for contradiction that no assignment that maximizes the sum of individual utilities with respect to $\bar{\Pi}$ matches individual 1 to job a. Let $M' = (M \setminus \{\{1, b\}, \{k, a\}\}) \cup \{\{1, a\}, \{k, b\}\}$ be the matching obtained by forcing individuals 1 and k to switch tasks in the assignment M. Since M' assigns individual 1 to job a, M' is not optimal with respect to Π by assumption. Thus, M' contains an augmenting cycle C with respect to Π.

First we claim that if no optimal matching contains $\{1, a\}$ then there is an augmenting cycle C where $\{1, a\} \in C$. This follows in the same fashion as Berge's lemma. Consider $(M \cup M') \setminus (M \cap M')$ – the symmetric difference between M and M'. Since both M and M' are perfect matchings on a bipartite graph the symmetric difference consists of even cycles. Since $\{1, a\} \in M'$, there is a cycle $C \in (M \cup M') \setminus (M \cap M')$ where $\{1, a\} \in C$. If C is not augmenting, then the matching $M'' = (M \setminus C) \cup (C \setminus M)$ has at least as high a value as M implying M'' is also optimal and contains $\{1,a\}$. By assumption this is not the case and therefore there is an augmenting cycle containing $\{1,a\}$.

Let C be an augmenting cycle containing $\{1,a\}$. We now examine two cases:

Case 1: $\{1,a\} \in C$ and $\{k,b\} \notin C$. This implies the following:

$$\sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\in M'}} \bar{\pi}_{ij} < \sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\notin M'}} \bar{\pi}_{ij} \tag{3.59}$$

$$\Rightarrow \bar{\pi}_{1a} + \sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\in M \\ i\neq 1}} \pi_{ij} < \bar{\pi}_{1w} + \sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\in M \\ i\neq 1}} \pi_{ij} \tag{3.60}$$

$$\Rightarrow \pi_{ka} + \sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\in M \\ i\neq 1}} \pi_{ij} < \pi_{kw} + \sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\in M \\ i\neq 1}} \pi_{ij} \tag{3.61}$$

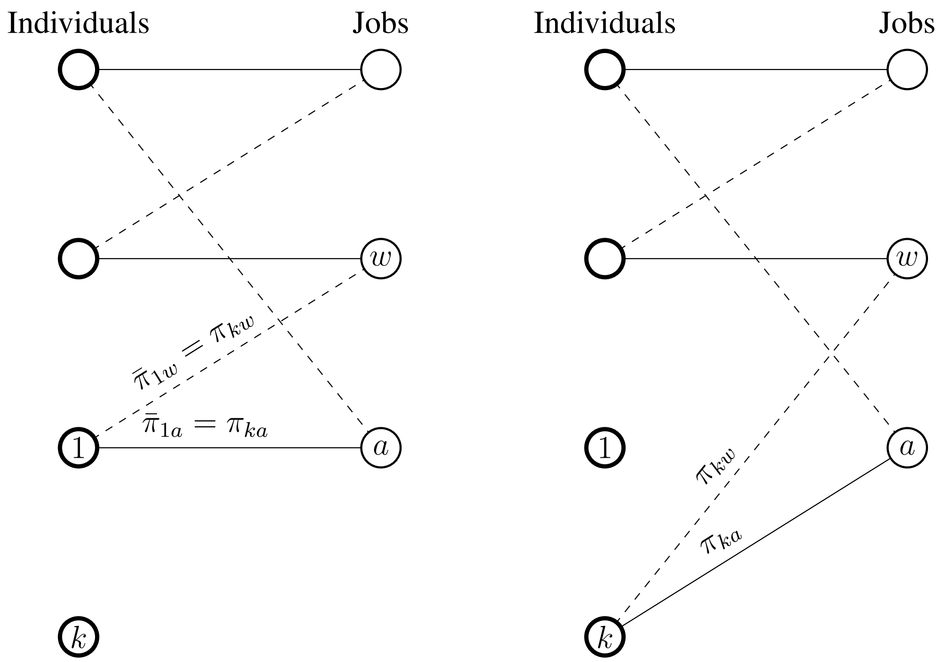
where $w \neq a$ is such that $\{1,w\} \in C \backslash M'$. These inequalities hold because by the definition of $\bar{\Pi}$ and M', $\bar{\pi}_i$ and π_i differ for only $i=1$; if $i \notin \{1,k\}$ then $\{i,j\} \in M'$ if and only if $\{i,j\} \in M$; and $\bar{\pi}_i = \pi_k$. But (3.61) implies that $(C \setminus \{\{1,a\}\{1,w\}\}) \cup \{\{k,a\}\{k,w\}\}$ is a cycle that augments M with respect to Π as shown in Figure 3.1. This is a contradiction since M is optimal with respect to Π. Therefore Case 1 cannot occur.

Case 2: $\{1,a\},\{k,b\} \in C$. Denote P_{1k} as the path that goes from individual 1 to individual k within the cycle C and contains the edge $\{1,a\}$ (see Figure 3.2). Let P_{k1} be the edge-disjoint path going from individual k to individual i in the cycle C. Therefore $P_{k1} \cup P_{1k} = C$. The cycle C is augmenting and therefore

$$\sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\in M'}} \bar{\pi}_{ij} < \sum_{\substack{\{i,j\}\in C: \\ \{i,j\}\notin M'}} \bar{\pi}_{ij} \tag{3.62}$$

$$\Rightarrow \sum_{\substack{\{i,j\}\in P_{1k}: \\ \{i,j\}\in M'}} \bar{\pi}_{ij} + \sum_{\substack{\{i,j\}\in P_{k1}: \\ \{i,j\}\in M'}} \bar{\pi}_{ij} < \sum_{\substack{\{i,j\}\in P_{1k}: \\ \{i,j\}\notin M'}} \bar{\pi}_{ij} + \sum_{\substack{\{i,j\}\in P_{k1}: \\ \{i,j\}\notin M'}} \bar{\pi}_{ij}. \tag{3.63}$$

This implies that either the edges in $P_{1k} \cap M'$ are less valuable than the edges in $P_{1k} \setminus M'$ or the edges in $P_{k1} \cap M'$ are less valuable than the edges in $P_{k1} \setminus M'$ (or both). Either way



(a) Augmenting Cycle for M' and $\bar{\Pi}$.

(b) Augmenting Cycle for M and Π.

Figure 3.1: Augmenting Cycles for Case 1.

we can construct an augmenting cycle as shown in Figure 3.2 with the original matching M and preferences Π. Without loss of generality assume the edges in $P_{1k} \cap M'$ are less valuable than the edges in $P_{1k} \setminus M'$ implying

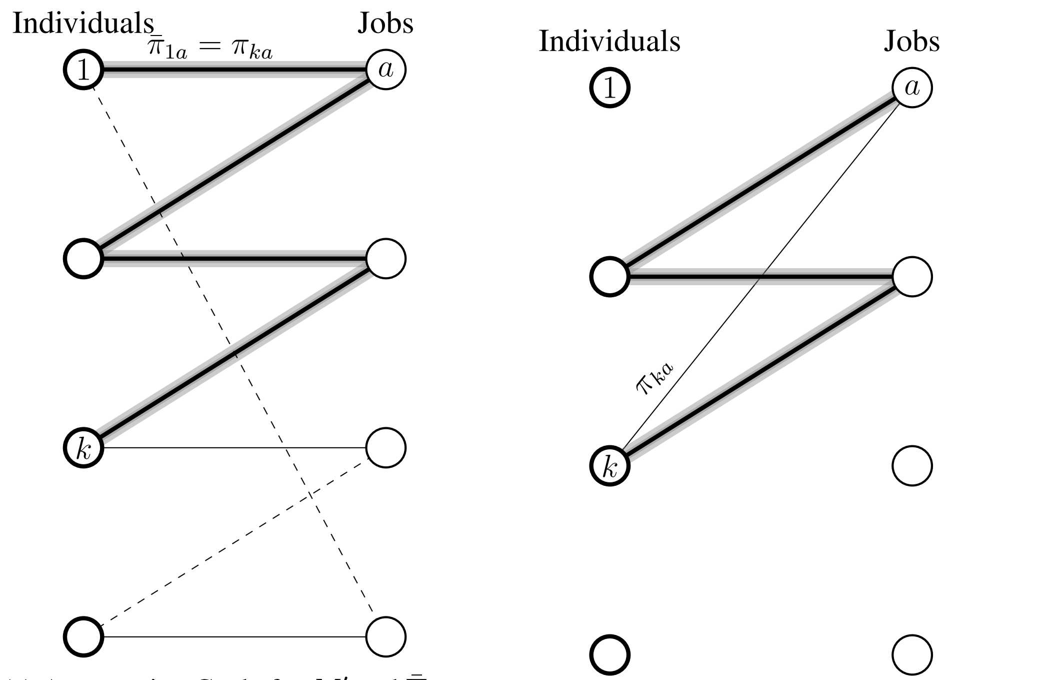
$$\sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\in M'}} \bar{\pi}_{ij} < \sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\notin M'}} \bar{\pi}_{ij} \tag{3.64}$$

$$\Rightarrow \quad \bar{\pi}_{1a} + \sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\in M'\\ i\neq 1}} \pi_{ij} < \sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\notin M'}} \pi_{ij} \tag{3.65}$$

$$\Rightarrow \quad \bar{\pi}_{1a} + \sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\in M\\ i\neq 1}} \pi_{ij} < \sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\notin M}} \pi_{ij} \tag{3.66}$$

$$\Rightarrow \quad \pi_{ka} + \sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\in M\\ i\neq 1}} \pi_{ij} < \sum_{\substack{\{i,j\}\in P_{1k}:\\ \{i,j\}\notin M}} \pi_{ij}. \tag{3.67}$$

Similar to Case 1, all steps hold by the definition of $\bar{\Pi}$ and M'. But (3.67) implies that the cycle $(P_{1k} \setminus \{1, a\}) \cup \{k, a\}$ augments M with respect to Π, a contradiction completing



(a) Augmenting Cycle for M' and $\bar{\Pi}$ with P_{1k} Highlighted and P_{k1} Not Highlighted.

(b) Augmenting Cycle for M and Π

Figure 3.2: Augmenting Cycles for Case 2.

Case 2.

Both cases result in contradictions and thus, individual 1 can change her preferences so that there is an optimal solution where she is assigned job a. □

Corollary 3.3.11. *Individual* 1 *can always change her preferences so she gets her highest ranked job in the Integral Max Assignment Problem when breaking ties lexicographically.*

Proof. By Lemma 3.3.9, individual 1 can adjust her preferences to $\bar{\pi}_1$ so that there is a maximum valued solution where she gets her favorite job. We now show how we can adjust these preferences so that individual 1 gets her favorite job when using the lexicographic ranked order tie-breaking system.

Denote individual 1's favorite job as job a. If $\bar{\pi}_{1a} = n$, then player 1 gets job a by our tie-breaking rule. Otherwise, $\bar{\pi}_{1a} = l < n$ and $\bar{\pi}_{1b} = n$ where $b \neq a$. Let M be an maximum matching assigning individual 1 to job a. If instead individual 1 instead reports $\bar{\pi}_{1a} = n$ and $\bar{\pi}_{1b} = l$ then M remains maximum. Moreover, in every maximum matching individual 1 is assigned to job a after setting $\bar{\pi}_{ia}$ to n, completing the proof of the corollary. $\square$

We note here that this proof only relies on there being a unique maximum weighted job and hence the same proof technique could have been used in the sections corresponding to $[0,1]$ and $[\delta,1]$ Max Benefit Assignments but not the Min Cost Assignment versions.

Corollary 3.3.12. *Individual i can alter her preferences so she is assigned one of her i highest-ranked jobs in the Integral Max Assignment Problem when ties are broken lexicographically.*

The proof of Corollary 3.3.12 follows in the same fashion as Corollary 3.3.11. Similar to Theorem 3.3.3, Algorithm 1 finds all possible outcomes of Nash equilibria. Algorithm 1 iteratively assigns each individual their most valued task of those that remain. Since each individual assigns each task a unique value, Algorithm 1 has a unique output. More importantly, this output corresponds to the only outcome we can expect at a Nash equilibrium.

Theorem 3.3.13. *Let x be the assignment obtained from Algorithm 1 with sincere preferences Π. Then there is a Nash equilibrium $\bar{\Pi}$ where $r(\bar{\Pi}) = x$. Moreover, for every equilibrium $\bar{\Pi}$, $r(\bar{\Pi}) = x$.*

Proof. The second part of the the theorem holds immediately by Corollary 3.3.12.

Since lexicographic tie-breaking only relies on the index of the individual, we may relabel the jobs so that x assigns task i to individual i for all i. This implies that individual i prefers task j to i if and only if $j < i$. We will prove that the following set of preferences form a Nash equilibrium:

$$\bar{\pi}_{ij} = n - i + j \; \forall i < n \; \forall j \leq i \tag{3.68}$$

$$\bar{\pi}_{ij} = n - j + 1 \; \forall i < n \; \forall j > i \tag{3.69}$$

$$\bar{\pi}_{nj} = n - j + 1 \; \forall j \tag{3.70}$$

Equivalently,

$$\bar{\Pi} = \begin{pmatrix} n & n-1 & n-2 & n-3 & \dots & 2 & 1 \\ n-1 & n & n-2 & n-3 & \dots & 2 & 1 \\ n-2 & n-1 & n & n-3 & \dots & 2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & 3 & 4 & 5 & \dots & n & 1 \\ n & n-1 & n-2 & n-3 & \dots & 2 & 1 \end{pmatrix}. \tag{3.71}$$

Since ties are broken lexicographically $r(\bar{\Pi}) = x$ and assigns individual i to task i for all i. We show $\bar{\Pi}$ is a Nash equilibrium via contradiction.

Suppose that $\bar{\Pi}$ is not a Nash equilibrium and that individual k can alter her preferences to $\bar{\pi}'_k$ such that she is assigned task $l < k$. Let M be the assignment obtained after k updates her preferences.

We first claim that for all $i \notin \{k, l\}$ individual i is assigned job i. We proceed by minimal counter example and let $i \notin \{k, l\}$ be the lowest indexed individual that is assigned to $j \neq i$. Suppose instead that individual w is assigned i. By minimality, one of the following holds: $i = n, j = k, w = l$; $i < n, w > i, j > i$; $i < n, w > i, j = k < i$; $i < n, i > w = l, j > i$; or $i < n, i > w = l, j = k < i$. In each case we will either give an augmenting cycle for M that yields a higher valued matching, or a value-neutral cycle for M that yields a matching with same value that is preferred according to the lexicographic tie-breaking rule. Either way M is not selected after k updates her preferences, a contradiction.

Case 1: $i = n, j = k, w = l$. Update M with the cycle shown in Figure 3.3.

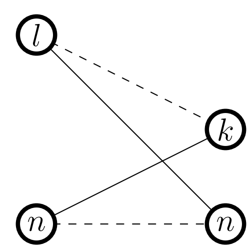


Figure 3.3: Cycle for Case 1.

If we update M according to the cycle in Figure 3.3 by adding edges $\{n, n\}$ and $\{l, k\}$ and removing edges $\{n, k\}$ and $\{l, n\}$ then the value of the matching increases by

$$\bar{\pi}_{ii} + \bar{\pi}_{wj} - \bar{\pi}_{ij} - \bar{\pi}_{wi} = \bar{\pi}_{nn} + \bar{\pi}_{lk} - \bar{\pi}_{nk} - \bar{\pi}_{ln} \tag{3.72}$$

$$= 1 + (n - k + 1) - (n - k + 1) - 1 = 0. \tag{3.73}$$

Hence the value of the matching is the same if we assign l to k and n to n. Since the lower indexed l prefers the new matching, the lexicographic tie-breaking rule does not assign l to n contradicting that M was selected. This completes Case 1.

Case 2: $i < n, w > i, j > i$. Update M with the augmenting cycle shown in Figure 3.4.

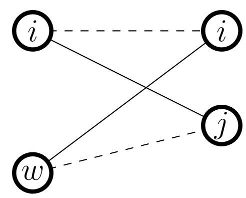


Figure 3.4: Augmenting Cycle for Case 2.

If we update M according to the cycle in Figure 3.4 by adding edges $\{i, i\}$ and $\{w, j\}$ and removing edges $\{i, j\}$ and $\{w, i\}$ then the value of the matching increases by

$$\bar{\pi}_{ii} + \bar{\pi}_{wj} - \bar{\pi}_{ij} - \bar{\pi}_{wi} = n + (n - j + 1) - (n - j + 1) - (n - w + i) \tag{3.74}$$

$$= w - i > 0 \tag{3.75}$$

if $w < j$ and

$$\bar{\pi}_{ii} + \bar{\pi}_{wj} - \bar{\pi}_{ij} - \bar{\pi}_{wi} = n + (n - w + j) - (n - j + 1) - (n - w + i) \tag{3.76}$$

$$= 2j - 1 - i \geq 0 \tag{3.77}$$

if $w \leq j$. Either way the value of the matching does not decrease and the cycle C in Figure 3.4 is such that $(M \setminus C) \cup (C \setminus M)$ is a matching preferred by the decision mechanism. This is a contradiction and completes Case 2.

Case 3: $i < n, w > i, j = k < i$. Update M with the cycle shown in Figure 3.5.

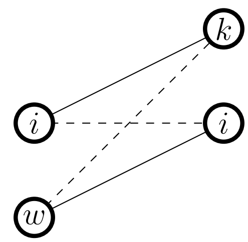


Figure 3.5: Cycle for Case 3.

If we update M according to the cycle in Figure 3.5 by adding edges $\{i, i\}$ and $\{w, k\}$ and removing edges $\{i, k\}$ and $\{w, i\}$ then the value of the matching increases by

$$\bar{\pi}_{ii} + \bar{\pi}_{wj} - \bar{\pi}_{ij} - \bar{\pi}_{wi} = \bar{\pi}_{ii} + \bar{\pi}_{wk} - \bar{\pi}_{ik} - \bar{\pi}_{wi} \tag{3.78}$$

$$= n + (n - w + k) - (n - i + k) - (n - w + i) = 0 \tag{3.79}$$

Thus the value of the matching is the same if we assign i to i and w to k. Similar to Case 1, this is a contradiction and Case 3 is complete.

Case 4: $i < n, i > w = l, j > i$. Update M with the augmenting cycle shown in Figure 3.6.

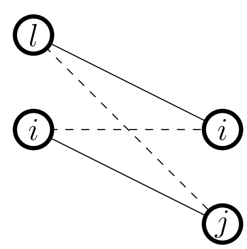


Figure 3.6: Augmenting Cycle for Case 4.

If we update M according to the cycle in Figure 3.6 by adding edges $\{i, i\}$ and $\{l, j\}$ and removing edges $\{i, j\}$ and $\{l, i\}$ then the value of the matching increases by

$$\bar{\pi}_{ii} + \bar{\pi}_{wj} - \bar{\pi}_{ij} - \bar{\pi}_{wi} = \bar{\pi}_{ii} + \bar{\pi}_{lj} - \bar{\pi}_{ij} - \bar{\pi}_{li} \tag{3.80}$$

$$= n + (n - j + 1) - (n - j + 1) - (n - i + 1) = i - 1 > 0 \tag{3.81}$$

since $i > l \geq 1$. Therefore the cycle in Figure 3.6 is augmenting. This contradicts that M

was selected after k updated her preferences and completes Case 4.

Case 5: $i < n, i > w = l, j = k < i$. Update M with the augmenting cycle shown in Figure 3.7.

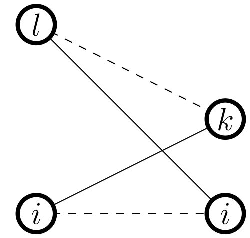


Figure 3.7: Augmenting Cycle for Case 5.

If we update M according to the cycle in Figure 3.6 by adding edges $\{i, i\}$ and $\{l, k\}$ and removing edges $\{i, k\}$ and $\{l, i\}$ then the value of the matching increases by

$$\bar{\pi}_{ii} + \bar{\pi}_{wj} - \bar{\pi}_{ij} - \bar{\pi}_{wi} = \bar{\pi}_{ii} + \bar{\pi}_{lk} - \bar{\pi}_{ik} - \bar{\pi}_{li} \tag{3.82}$$

$$= n + (n - k + 1) - (n - i + k) - (n - i + 1) = 2i - 2k > 0 \tag{3.83}$$

The cycle in Figure 3.7 is augmenting for M. This contradicts that M was selected after k updated her preferences and this completes Case 5.

The claim holds for all five cases and therefore M assigns individual i to task i for all $i \notin \{k, l\}$, individual k to task $l < k$, and individual l to task k. However, this implies there is an augmenting cycle for M after k updates her preferences. Suppose instead that we try to assign individual l to task l, individual n to task k and individual k to task n. This increases the value of the matching by

$$\bar{\pi}_{ll} + \bar{\pi}_{nk} + \bar{\pi}'_{kn} - \bar{\pi}_{lk} - \bar{\pi}_{nn} - \bar{\pi}'_{kl} \tag{3.84}$$

$$= n + (n - k + 1) + \bar{\pi}'_{kn} - (n - k + 1) - 1 - \bar{\pi}'_{kl} \tag{3.85}$$

$$= n - 1 + \bar{\pi}'_{kn} - \bar{\pi}'_{kl} \tag{3.86}$$

$$\geq 0 \tag{3.87}$$

However, by the lexicographic tie-breaking rule, the new matching is preferred to M

since it assigns the lower indexed individual l her more preferred job l. This contradicts that $r([\bar{\Pi}_{-k}, \bar{\pi}'_k]) = M$. Therefore individual k cannot alter her preferences to get a better outcome and therefore $\bar{\Pi}$ is a Nash equilibrium for Π. $\square$

Theorem 3.3.14. *The Price of Deception of the Integral Max Assignment Problem with lexicographic tie-breaking is* $\frac{2(n-1)}{n}$.

Proof. By Theorem 3.3.13, an equilibrium solution haves value at least $\sum_{i=1}^{n}(n-i+1) = \sum_{i=1}^{n} i = \binom{n+1}{2}$. The highest value an assignment can have is $\sum_{i=1}^{n} n = n^2$. However, by Theorem 3.3.13 any set of preferences Π where $U(\Pi, r(\Pi)) = n^2$ also has $U(\Pi, r(\bar{\Pi})) = n^2$ yielding a Price of Deception of 1 for that particular instance. Therefore, if the Price of Deception is more than 1, then $U(\Pi, r(\Pi))$ is at most $n^2 - 1$ yielding an upper bound of $\frac{2(n-1)}{n}$ on the Price of Deception. To obtain the lower bound, consider the following possible preferences:

$$\pi_{11} = n \tag{3.88}$$

$$\pi_{1n} = n - 1 \tag{3.89}$$

$$\pi_{i,i-1} = n \ \forall i > 1 \tag{3.90}$$

$$\pi_{ii} = n - i + 1 \ \forall i > 1 \tag{3.91}$$

$$\pi_{ij} > n - i + 1 \ \forall i > 1 \ \forall j < i - 2 \tag{3.92}$$

$$\pi_{ij} < n - i + 1 \ \forall i > 1 \ \forall j > i + 1 \tag{3.93}$$

where π_{1j} is arbitrary for $j \notin \{1, n\}$. The optimal matching $r(\Pi)$ assigns task $i-1$ to individual i for all $i > 1$ and task n to individual 1 for a total benefit of $U(\Pi, r(\Pi)) = n^2 - 1$. By Theorem 3.3.13, there is a Nash equilibrium $\bar{\Pi}$ that assigns task i to individual i for all i for a sincere cost of $U(\Pi, r(\bar{\Pi})) = \binom{n+1}{2}$ yielding a Price of Deception of $\frac{2(n-1)}{n}$ and completing the proof of the theorem. $\square$

3.3.6 Integral Min Assignments

In Integral Max Assignment Problem, each individual i submits a integer weight $\pi_{ij} \in [n]$ to task j. As in the Max Assignment Problem $\pi_{ij} = \pi_{ik}$ if and only if $j = k$. This time π_{ij} corresponds to the cost. There is an immediate bijection between instances of the Integral Min Assignments and Integral Max Assignments. If an individuals' cost (benefit) in the Min (Max) Assignment Problem is π_{ij}, then this corresponds to a benefit (cost) in the Max (Min) Assignment Problem of $\pi'_{ij} = n - \pi_{ij} + 1$. Therefore Theorem 3.3.13 holds in this setting and we can immediately find the Price of Deception.

Theorem 3.3.15. *The Price of Deception of the Integral Min Assignment Problem with lexicographic tie-breaking is $\frac{n}{2}$.*

Proof. As in the proof of Theorem 3.3.14, by Theorem 3.3.13, an equilibrium has value at most $\sum_{i=1}^{n} i = \binom{n+1}{2}$. The cheapest assignment possible has value $U(\Pi, r(\Pi)) = \sum_{i=1}^{n} 1 = n$ but once again Theorem 3.3.13 implies that $U(\Pi, r(\bar{\Pi})) = n$. Therefore if the Price of Deception is more than one then $U(\Pi, r(\bar{\Pi})) \geq n + 1$ yielding an upper bound of $\frac{n}{2}$ on the Price of Deception. To achieve the matching lower bound, simply consider the preferences $\Pi' = \{\pi'_{ij} = n - \pi_{ij} + 1\}_{i,j=1}^{n}$ where Π is defined in the proof of Theorem 3.3.14. □

3.4 The Price of Deception with Random Tie-Breaking

In this section, we find the Prices of Deception using a random tie-breaking rule. To break ties randomly, we re-index individuals uniformly at random and then use the lexicographic tie-breaking rule with respect to the new indices.

3.4.1 $[0, 1]$ Max Assignment Problem

Theorem 3.4.1. *The Price of Deception in the $[0, 1]$ Max Assignment Problem breaking ties randomly is in $\Omega(\sqrt{n})$.*

Proof. We begin by assuming that $n = k^2$ where $k \in \mathbb{Z}$. We now define our weights in the following way:

$$\pi_{ik+1,ik+1} = 1 \qquad \forall i = 0, 1, \ldots, k-1, \tag{3.94}$$

$$\pi_{ik+1,j} = 0 \qquad \forall i = 0, 1, \ldots, k-1 \; \forall j \neq ik+1, \tag{3.95}$$

$$\pi_{ik+j,ik+1} = \frac{1}{n} + \epsilon \qquad \forall i = 0, 1, \ldots, k-1 \; \forall j = 2, 3, \ldots, k, \tag{3.96}$$

$$\pi_{ik+j,t} = \frac{1}{n} - \frac{\epsilon}{n-1} \qquad \forall i = 0, 1, \ldots, k-1 \; \forall j = 2, 3, \ldots, k \; \forall t \neq ik+1. \tag{3.97}$$

These preferences correspond to k groups of k individuals where each group has a single task they most prefer[1]. One member of the group likes only this job and dislikes all other jobs. The other members of the group are basically indifferent between all jobs. Ignoring ϵ terms, the optimal solution to this has value $U(\Pi, r(\Pi)) = k + (n-k)\frac{1}{n}$ by giving everyone their job if they place a value 1 on it and assigning all other jobs randomly.

We now consider the following set of putative preferences:

$$\bar{\pi}_{ik+j,ik+1} = 1 \qquad \forall i = 0, 1, \ldots, k-1, \forall j = 1, ..., k \tag{3.98}$$

$$\bar{\pi}_{ik+j,t} = 0 \qquad \forall i = 0, 1, \ldots, k-1 \; \forall j = 1, ..., k \; \forall t \neq ik+1. \tag{3.99}$$

With respect to $\bar{\Pi}$, each group selects one member to receive the most preferred job. All other tasks are assigned uniformly at random. Again, ignoring ϵ terms, the optimal solution to this has value $U(\Pi, r(\bar{\Pi})) = k\frac{1}{k} + (n-k)\frac{1}{n}$ since each player with a true value of 1 on their favorite job only receives it if they are the first that gets to decide amongst those which prefer that job (which occurs with probability $\frac{1}{k}$). These values are at equilibrium since any player altering their value ensures she will get her lowest weighted job. This example yields

[1] For n where $\sqrt{n} \notin \mathbb{Z}$ we have various groups of size $\lfloor \sqrt{n} \rfloor$ or $\lceil \sqrt{n} \rceil$. This has a negligible effect on the Price of Deception for large n.

a Price of Deception of

$$\frac{k+1-\frac{k}{n}}{\frac{k^2}{n}+1-\frac{k}{n}} = \frac{\sqrt{n}+1-\frac{1}{\sqrt{n}}}{2-\frac{1}{\sqrt{n}}} \approx \frac{\sqrt{n}}{2} \tag{3.100}$$

completing the proof of the theorem. □

We did not obtain a matching upper bound in the proof of Theorem 3.4.1. We give an upper bound in the following theorem.

Theorem 3.4.2. *The Price of Deception for the* $[0,1]$ *Max Assignment Problem is in* $O(n^{\frac{2}{3}})$.

The proof of Theorem 3.4.2, consists of two parts. First we show that the Price of Deception is in $O(\frac{n^2}{k^2})$ where $k = U(\Pi, r(\Pi))$. Second, we show that the Price of Deception is in $O(k)$. Setting $\frac{n^2}{k^2}$ equal to k, we obtain that the upper bound is at most $n^{\frac{2}{3}}$. Prior to establishing the first part, we present the following lemma.

Lemma 3.4.3. *For* $n \geq k$, $\sum_{i=1}^{k} \frac{1}{x_i} \geq \frac{k^2}{n}$ *when* $\sum_{i=1}^{k} = n$ *and* $x_i \geq 1 \; \forall i = 1, .., k$.

Proof. The value $\sum_{i=1}^{k} \frac{1}{x_i}$ is bounded below by the following mathematical program:

$$min\ z = f(x) := \sum_{i=1}^{k} \frac{1}{x_i} \tag{3.101}$$

$$subject\ to: \tag{3.102}$$

$$\sum_{i=1}^{k} x_i = n \tag{3.103}$$

$$x \geq \mathbf{1} \tag{3.104}$$

The Hessian of f, $\nabla^2 f$, is such that $\nabla^2 f_{ij} = 0$ if $i \neq j$ and $\nabla^2 f_{ii} = \frac{2}{x_i^3}$. This implies that $\nabla^2 f$ is positive definite over the domain $\{x : \sum_{i=1}^{k} x_i = n, x \geq \mathbf{1}\}$ implying that f is strictly convex over its domain. Furthermore, the domain is symmetric and convex and, by strict convexity of f, the optimal solution x^* is symmetric. Therefore, $x_i^* = x_j^* = \frac{n}{k}$ and $f(x^*) = \frac{k^2}{n}$ completing the proof of the lemma. □

Theorem 3.4.4. *The Price of Deception for the* $[0,1]$ *Max Assignment Problem is in* $O(\frac{n^2}{k^2})$ *where* k *is the value of the sincere optimal solution.*

Proof. Since $U(\Pi, r(\Pi)) \geq k$ there are at least $l = \lceil \frac{k}{2} \rceil$ individuals getting a value at least $\frac{k}{2n}$ in an optimal sincere solution. Otherwise the value of the optimal solution is less than $\frac{k}{2} + (n - \frac{k}{2})\frac{k}{2n} = k - \frac{k^2}{4n} < k$. Suppose without loss of generality these individuals are indexed by $\{1, ..., l\}$ and player i gets job i in the sincere optimal solution.

Let $\bar{\Pi}$ be a Nash equililbrium for Π and let α_i be the number of individuals placing a value of 1 on job i (excluding player i) in $\bar{\Pi}$. If player $i \in \{1, ..., l\}$ places 1 on job i, then player i receives job i with probability $\frac{1}{\alpha_i+1}$ indicating player i has a utility in $r(\bar{\Pi})$ of at least $\frac{k}{2n(\alpha_i+1)}$. Thus the value of the equilibrium is at least

$$\sum_{i=1}^{l} \frac{k}{2n(\alpha_i + 1)} = \frac{k}{2n} \sum_{i=1}^{l} \frac{1}{\alpha_i + 1} \geq \frac{k}{2n} \frac{l^2}{n + l} = \frac{kl^2}{2n^2 + 2nl} \tag{3.105}$$

by Lemma 3.4.3 since $\sum_{i=1}^{l} (\alpha_i + 1) \leq n + l$. Thus the Price of Deception is at most

$$\frac{k}{\frac{kl^2}{2n^2+2nl}} = \frac{2n^2 + 2nl}{l^2} \in O\left(\frac{n^2}{k^2}\right) \tag{3.106}$$

□

Theorem 3.4.5. *The Price of Deception for the* $[0,1]$ *Max Assignment Problem is in* $O(k)$ *where* k *is the value of the sincere optimal solution.*

Proof. It suffices to show that for any set of preferences, the value of a solution in an equilibrium is at least some constant c (we will select $c = \frac{1}{2}$).

Let α_i be the number of 1's on job i excluding player 1. As in Theorem 3.4.4, if player i places 1 on job i then she receives job i with probability $\frac{1}{\alpha_i+1}$. Therefore, player 1's value from $r(\bar{\Pi})$ is at least

$$t_1 = \begin{matrix} max \\ 1 \leq j \leq n \end{matrix} \left\{ \frac{\pi_{1j}}{1+\alpha_j} \right\} \tag{3.107}$$

$$\Rightarrow \qquad t_1 \geq \frac{\pi_{1j}}{1+\alpha_j}\ \forall j \tag{3.108}$$

$$\Rightarrow \qquad t_1(1+\alpha_j) \geq \pi_{1j}\ \forall j \tag{3.109}$$

$$\Rightarrow \qquad t_1 \sum_{j=1}^{n} (1+\alpha_j) \geq \sum_{j=1}^{n} \pi_{1j} = 1 \tag{3.110}$$

$$\Rightarrow \qquad t_1(2n-1) \geq t_1 \sum_{j=1}^{n} (1+\alpha_j) \geq 1 \tag{3.111}$$

$$\Rightarrow \qquad t_1 \geq \frac{1}{2n} \tag{3.112}$$

This holds for every player, implying that the value of the solution in equilibrium is at least $\frac{1}{2}$, thus completing the proof. □

Theorems 3.4.4 and 3.4.5 immediately imply Theorem 3.4.2.

3.4.2 $[0, 1]$ Min Assignment Problem

Theorem 3.4.6. *The Price of Deception in the* $[0, 1]$ *Min Assignment Problem breaking ties randomly is at least* n*.*

Proof. Consider the following sincere preferences:

$$\pi_{in} = 1\ \forall i \neq 1 \tag{3.113}$$

$$\pi_{ij} = 0\ \forall i \neq 1\ \forall j \neq n \tag{3.114}$$

$$\pi_{1i} = \frac{1}{n} - \frac{\epsilon}{n-1}\ \forall i \neq n \tag{3.115}$$

$$\pi_{1n} = \frac{1}{n} + \epsilon \tag{3.116}$$

This instance has an optimal solution given by $U(\Pi, r(\Pi)) = \frac{1}{n} + \epsilon$ obtained by assigning job n to individual 1 and all other jobs uniformly at random. We now consider the following set of false preferences:

$$\bar{\pi}_{in} = 1 \; \forall i \tag{3.117}$$

$$\bar{\pi}_{ij} = 0 \; \forall i \; \forall j \neq n. \tag{3.118}$$

The assignment $r(\bar{\Pi})$ gives out a completely random matching since everyone has the same preferences. No players can do better since any change would ensure they get their worst job. This solution has a since expected value of $U(\Pi, r(\bar{\Pi})) = 1$ yielding a Price of Deception of n as $\epsilon \to 0$. □

3.4.3 Integral Max Assignment Problem

Theorem 3.4.7. *The Price of Deception in the Integral Max Assignment Problem breaking ties randomly is at most* $\frac{4(n^2-1)}{n^2+3n} < 4$.

Proof. This proof proceeds in a similar fashion to Theorems 3.4.4 and 3.4.5. Let α_j be the number of individuals that place a value of n on job j excluding player i with respect to equilibrium preferences $\bar{\Pi}$. Let $\alpha = \sum_{j=1}^{n} \alpha_j \leq n$. Following from the same strategy as decribed by Theorem 3.3.13, player 1 can adjust her preferences so that she receives job i with probability at least $\frac{1}{\alpha_i+1}$. If she commits to such a strategy, then her utility of the new outcome will be at least $\frac{\pi_{1j}}{\alpha_j+1} + \frac{\alpha_j}{\alpha_j+1}$. Therefore her value of $r(\bar{\Pi})$ must be at least

$$t_1 = \begin{matrix} max \\ 1 \leq j \leq n \end{matrix} \left\{ \frac{\pi_{1j}}{\alpha_j + 1} + \frac{\alpha_j}{\alpha_j + 1} \right\} \tag{3.119}$$

$$\Rightarrow \qquad t_1(1 + \alpha_j) \geq \pi_{1j} + \alpha_j \; \forall j \tag{3.120}$$

$$\Rightarrow \qquad t_1 \sum_{j=1}^{n} (1 + \alpha_j) \geq \sum_{j=1}^{n} (\pi_{1j} + \alpha_j) = \binom{n+1}{2} + \alpha \tag{3.121}$$

$$\Rightarrow \qquad t_1 \geq \frac{n(n+1) + 2\alpha}{2n + 2\alpha} \tag{3.122}$$

$$\Rightarrow \qquad t_1 \geq \frac{n^2 + 3n}{4n} \tag{3.123}$$

This holds for all individuals and therefore the value of $r(\bar{\Pi})$ is at least

$$U(\Pi, r(\bar{\Pi})) \geq \frac{n^2 + 3n}{4} \tag{3.124}$$

Similar to Theorem 3.3.14, if the Price of Deception is more than one, then $U(\Pi, r(\bar{\Pi})) \leq n^2 - 1$ implying that the Price of Deception of the Max Integral Assignment Problem is at most

$$\frac{4(n^2 - 1)}{n^2 + 3n} < 4. \tag{3.125}$$

□

3.4.4 Integral Min Assignment Problem

Theorem 3.4.8. *The Price of Deception in the Integral Min Assignment Problem breaking ties randomly is at most* $\frac{3n^2+n}{4(n+1)} \leq \frac{3n+1}{4}$.

Proof. As stated earlier, there is a bijection between the Integral Min and Max Assignment Problems. In the proof of Theorem 3.4.7, it was established that each individual receives a utility of at least $n - \frac{n^2+3n}{4} + 1$ in the Integral Max Assignment Problem. As in the proof of Theorem 3.3.15, this means that each individual can select $\bar{\pi}_i$ such that their cost at an equilibrium $\bar{\Pi}$ is at most $n - (n - \frac{n^2+3n}{4} + 1) + 1 = \frac{3n^2+n}{4n}$. As established in Theorem 3.3.15, the Price of Deception is greater than one only if $U(\Pi, r(\Pi)) \geq n + 1$. This implies that the Price of Deception is at most

$$\frac{n\frac{3n^2+n}{4n}}{n+1} = \frac{3n^2 + n}{4(n+1)} \leq \frac{3n+1}{4} \tag{3.126}$$

□

4

PRICE OF DECEPTION IN SPATIAL VOTING AND FACILITY LOCATION

4.1 Introduction

In this chapter, we study the Price of Deception for algorithms that select a single facility location (or political policy) from anywhere in a compact convex region based on a st of individually reported facility locations (political preferences). The standard solution techniques (1-Median and 1-Mean algorithms) give solutions that minimize the sum of distances between the facility location and the locations preferred by each individual. While there has been quite a bit of work focused on developing strategy-proof variants of the 1-Median or p-Median (p facilities) [43, 22, 23], there has not been much working in understanding the type of outcomes obtained when individuals are strategic.

In this chapter we consider three variants of the Facility Location Problem. We first examine the 1-Median Problem where a central coordinator attempts to minimize the sum of individual distances to the facility location where distances are measured with the L_1 norm. We also consider the same problem while measuring distances with the L_2 or euclidean norm. Finally we consider the 1-Mean Problem where the sum of squared euclidean distances is minimized. For the first two mechanisms, we consider both randomized and deterministic tie-breaking rules. For the third, there is always a unique optimal facility location and there is no need to break ties.

When analyzing a player's best response, a common approach is to assume individuals measure the quality of the facility location by measuring the euclidean distance (L_2 norm) between the individual's preferred location and the facility location. We prove our results for individual utilities based on any L_p norm where $p \in (0, \infty)$ (i.e. $u_i(\pi_i, x) = ||\pi_i - x||_{p_i}$).

We begin by analyzing the Facility Location Problem without minimal dishonesty. In

Theorem 4.3.1 we show for the 1-Median Problem when using either the L_1 of L_2 norm, for any feasible facility location, there is a Nash equilibrium that yields that location. That is to say that the Nash equilibrium solution concept considers every facility location an acceptable outcome regardless of the sincere preferences. There is disappointing from a normative perspective; there are infinitely many possible facility locations and the Nash equilibrium concept gives no way to discriminate between them. Morever, it is unrealistic from a predictive perspective; in the proof of Theorem 4.3.1, each individual submits that there preferred outcome is location x. As a result the facility is placed at x and these preferences form a Nash equilibrium because no individual can cause the facility to move at all. This indicates that people are lying without reason and thus the 1-Median Problem needs the minimal dishonesty refinement defined in Section 2.3. We remark that the Price of Deception is the same in the 1-Mean Problem with or without the minimal dishonesty refinement. Unlike the 1-Median Problem, if any individual changes their preference in the 1-Mean Problem then the facility location changes; the responsiveness of the facility location procedure prevents absurd outcomes like the ones we see in Theorem 4.3.1.

After applying the minimal dishonesty refinement we obtain far more satisfying results. We first observe that tie-breaking rules once again have a large effect on the Price of Deception. Unlike the assignment problem however, lexicographic tie-breaking rules yield quality outcomes while random tie-breaking can lead to arbitrarily poor results. Specifically, we give a fair variant of the 1-Median Problem that, while manipulable, has a Price of Deception of one. This represents our best possible hope for a manipulable algorithm; while manipulation can change the outcome, it does not negatively impact the quality of the outcome. We also observe that the set of allowed facility locations can have a significant impact of the Price of Deception; for many of the algorithms we examine the Price of Deception can be arbitrarily large if the set of allowed facility locations is a triangle. However, if the set of allowed facility locations is a rectangle then the Price of Deception is bounded for most mechanisms. These results provide evidence that the Price of Deception helps discriminate among various facility location procedures.

4.1.1 Facility Location in $\mathbb{R}$

We begin by examining the problem in one dimension. Since $||x||_2 = ||x||_1$ in $\mathbb{R}$ we consider only the 1-Median and 1-Mean Problem. The results are given in Table 4.1 below.

Table 4.1: Prices of Deception of the Facility Location Problem in $\mathbb{R}$.

Price of Deception in $\mathbb{R}$		
Number of Individuals	$\|N\|$ is odd	$\|N\|$ is even
Deterministic 1-Median	1	1
Random 1-Median	1	$\sqrt{2}$
1-Mean	$O(n)$	$O(n)$

Perhaps most striking in Table 4.1 is the difference in the Prices of Deception of the 1-Median Problem with deterministic and random tie-breaking rules. In the previous chapter, we observed a significant difference between randomized and lexicographic tie-breaking rules. Specifically, we observed that randomized methods outperformed lexicographic methods due to the biased nature of lexicographic tie-breaking. We examine a stark contrast in this setting; deterministic tie-breaking rules have no impact on manipulation while random tie-breaking can lead to sub-optimal outcomes. This emphasizes the importance of very carefully considering small changes to an algorithm because small differences can have significant impacts on performance and those impacts can vary greatly depending on the setting.

In the 1-Median Problem, the set of optimal facility locations will be given by an interval $[a, b]$ (possibly $a = b$ and always $a = b$ when $|N|$ is odd) for some $a \leq b$. To break ties in the deterministic setting, we select $\lambda \in [0, 1]$ and then place the facility at $(1-\lambda)a + \lambda b$. In Theorem 4.4.6, we establish for $\lambda \in \{0, 1\}$ that the 1-Median Problem is strategy-proof – no individual will have incentive to behave dishonesty – and it immediately follows that the Price of Deception is one. More fascinating however is that for $\lambda \in (0, 1)$ and for $|N|$ even, the 1-Median Problem is manipulable but the Price of Deception is still one; manipulation has an impact on the outcome – the facility location may change due to strategic behavior – but the social utility of the outcome remains unchanged. This is significant; in accordance with current literature's preference for strategy-proof algorithms, we would select $\lambda = 0$ to guarantee that the mechanism is unaltered by manipulation. However, selecting $\lambda = 0$ prioritizes placing the facility to the left which appears unfair to everyone to the right of the facility. As a result of this research, we instead recommend that we set $\lambda = .5$ because it is more fair **and** still guarantees that manipulation has no impact on social utility since the Price of Deception is one.

We also examine the 1-Mean Problem in $\mathbb{R}$. It has a significantly higher Price of Deception $\Theta(n)$ indicating that the impact of manipulation increases linearly with the number of individuals in the game. Denoting Π as the set of locations individuals sincerely prefer, we guarantee that the facility location is placed in $conv.hull(\Pi)$ when individuals are strategic. However, no further guarantees exist and the strategic facility location can be quite far from the facility location obtained when everyone is honest.

4.1.2 Facility Location in $\mathbb{R}^k$ for $k \geq 2$

We also consider the Facility Location Problem in higher dimensions. It is most common to consider the problem in $\mathbb{R}^2$ but we generalize for all higher dimensions. We have only required that the set of allowed facility locations $\mathcal{X}$ be compact and convex. In $\mathbb{R}$, this describes a line segment. However, in higher dimensions, there are an infinite number of possibilities for the shape of the set of allowed locations. We consider both arbitrary shapes as well as hyper-rectangles – the most common set of allowed facility locations. The Prices of Deceptions for our various procedures are given in Table 4.2.

Perhaps what stands out most is that the Price of Deception is ∞ for arbitrary $\mathcal{X}$ regardless of the facility location procedure. Specifically, in Theorems 4.4.26 and 4.4.19 we show that for a triangle with largest angle α that the Price of Deception converges to ∞ as α tends to π. The only exceptions are a few variants of the deterministic version of

Table 4.2: Prices of Deception of the Facility Location Problem in $\mathbb{R}^k$ for $k \geq 2$.

Price of Deception in $\mathbb{R}^k$ for $k \geq 2$				
Allowed Locations $\mathcal{X}$	Arbitrary $\mathcal{X}$		Rectangular $\mathcal{X}$	
Number of Individuals	$\|N\|$ is odd	$\|N\|$ is even	$\|N\|$ is odd	$\|N\|$ is even
Deterministic 1-Median	1	1 or ∞	1	1
Random 1-Median	1	∞	1	$\sqrt{2}$
1-Mean	∞		$O(n)$	
Minimize L_2 Norm	?	∞	?	∞

the 1-Median Problem using the L_1 norm. In k dimensions, the set of optimal solutions is some hyper-rectangle $\{x : a \leq x \leq b\}$. To break ties deterministically, we select some $\lambda \in [0,1]^k$ and select the facility location $(\mathbf{1} - \lambda)a + \lambda b$. In Theorem 4.4.17, we establish that for $\lambda \in \{0,1\}^k$ that the 1-Median Problem is strategy-proof regardless of the set of allowed facility locations and therefore has a Price of Deception of 1.

The results improve significantly when the set of allowed facility locations is a hyper-rectangle. As with the Facility Location Problem in $\mathbb{R}$, we establish that the Price of Deception of the 1-Median Problem using the L_1 norm is one (Theorem 4.4.14). Specifically selecting $\lambda = \{\frac{1}{2}\}^k$, we have a fair procedure that, while not strategy-proof, guarantees that manipulation will have no impact on the quality of the outcome. Moreover, Theorem 4.4.14 holds for all individual utility functions $u_i(\pi_i, x) = ||\pi_i - x||_{p_i}$ for all $p_i \in (0, \infty)$. However, if ties are broken randomly in the 1-Median Problem, then manipulation can lead to sub-optimal outcomes. Even worse, if we use the L_2 norm instead for the social utility, the Price of Deception is unbounded regardless of the tie-breaking rule (Theorem 4.4.26) and we may obtain arbitrarily poor results. For the 1-Mean Problem, in Theorem 4.4.24 we show that the facility location will always be in the smallest bounding box containing all the sincere preferences and therefore the Price of Deception is $\Theta(n)$ and once again the impact of manipulation, while finite, scales linearly with the number of individuals. Once again this result holds for all utility functions.

4.2 The Model

Once again we present the definitions from Chapter 2 as they relate the Facility Location Problem.

Definition 4.2.1. An instance of the Facility Location Problem consists of

- A compact convex set $\mathcal{X} \subseteq \mathbb{R}^k$ of allowed facility locations.
- A Set $V = \{1, 2, ..., n\}$ of individuals.
- For each individual $i \in V$, a location $\pi_i \in \mathcal{P}_i = \mathcal{X}$ representing i's preferred facility location. In voting theory π_i is called individual i's ideal point.
- Denote by Π the collection of π_i over all $i \in V$. $\Pi \in \mathcal{P}$ is called the *preference profile*.
- The profile Π is submitted to a publicly known facility location procedure r. The outcome $r(\Pi)$ corresponds to a facility location or a distribution of facility locations.

In context of this chapter, $r(\Pi)$ is a maximizer of a social utility function. We consider three mechanisms, each having a different social utility function:

- 1-Median Problem: $U(\Pi, x) = \sum_{i=1}^{n} ||\pi_i - x||_1$.
- 1-Mean Problem: $U(\Pi, x) = \sum_{i=1}^{n} ||\pi_i - x||_2^2$.
- Minimize L_2 Norm: $U(\Pi, x) = \sum_{i=1}^{n} ||\pi_i - x||_2$.

Only the 1-Mean Problem is guaranteed a unique optimizer. For the other problems we consider both lexicographic and random tie breaking rules.

Individual i's valuation of the location $x \in \mathcal{X}$ is given by $u_i(\pi_i, x) = ||\pi_i - x||_{p_i}$ for some $p_i \in (0, \infty)$. Typically p_i is taken to be 2 (the Euclidean norm). We derive our results for any p_i. In the event that the procedure r is non-deterministic, we assume that individuals are risk-neutral and that $u_i(\pi_i, X)$ for some distribution X is equal to its expected value.

Strategic Facility Location Game

- Each individual i has information $\pi_i \in \mathcal{X}$ describing their preferred facility location. The collection of all information is the (sincere) profile $\Pi = \{\pi_i\}_{i=1}^n$.
- To play the game, individual i submits putative location $\bar{\pi}_i \in \mathcal{X}$. The collection of all submitted data is denoted $\bar{\Pi} = \{\bar{\pi}_i\}_{i=1}^n$.
- It is common knowledge that a central decision mechanism will select facility location $r(\bar{\Pi}, \omega)$ when given input $\bar{\Pi}$ and random event ω.
- The random event $\omega \in \Omega$ is selected according to μ. We denote $r(\bar{\Pi})$ as the distribution of facility locations according to Ω and μ.
- Individual i evaluates $r(\bar{\Pi})$ according to i's sincere preferences π_i. Specifically, individual i's utility of the distribution of outcomes $r(\bar{\Pi})$ is $u_i(\pi_i, r(\bar{\Pi})) = E(||\pi_i - r(\bar{\Pi})||_{p_i})$ for some $p_i \in (0, \infty)$.

By the definition of the Strategic Facility Location Game, a set of submitted preferences $\bar{\Pi}$ forms a pure strategy Nash equilibrium if no individual i would obtain an outcome they sincerely prefer to $r(\bar{\Pi})$ (with respect to π_i) by altering $\bar{\pi}_i$.

Example 4.2.2. A Nash Equilibrium of the Strategic Location Game.

Consider the 1-Mean Problem where the facility location is $r(\Pi)$ that minimizes $U(\Pi, x) = \sum_{i=1}^n ||\pi_i - x||_2^2$. It is well known that the minimizer of this function is $r(\Pi) = \sum_{i=1}^n \frac{\pi_i}{n}$. Consider the feasible region $\mathcal{X} = \{x \in \mathbb{R}^2 : (0,0) \leq x \leq (1,1)\}$ and the sincere preferences $\pi_1 = (\frac{1}{8}, \frac{1}{8}), \pi_2 = (\frac{1}{8}, \frac{1}{2})$ and $\pi_3 = (\frac{1}{2}, \frac{1}{8})$. With respect to these preferences, the facility location is *located at $r(\Pi) = (\frac{1}{4}, \frac{1}{4})$. This corresponds to a cost of $U(\Pi, r(\Pi)) = \sum_{i=1}^{3} ||\pi_i - r(\Pi)||_2^2 = \frac{\sqrt{2}+2\sqrt{5}}{8} \approx .736$. The region $\mathcal{X}$ and sincere preferences are given in Figure 4.1 below.*

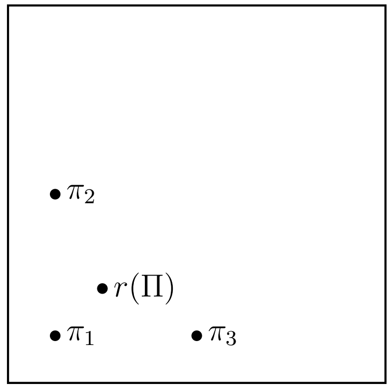


Figure 4.1: Sincere Preferences for Example 4.2.2

Individual 1 would like to move the facility to the lower left, individual 2 to the upper left, and individual 3 to the lower right. A Nash equilibrium where individuals attempt to do just this is given by $\bar{\pi}_1 = (0,0), \bar{\pi}_2 = (0,1)$, and $\bar{\pi}_3 = (1,0)$. With respect to the submitted preferences $\bar{\Pi}$, the facility is located at $r(\bar{\Pi}) = (\frac{1}{3}, \frac{1}{3})$. The central decision mechanism believes it has selected a facility with cost $U(\bar{\Pi}, r(\bar{\Pi})) = \sum_{i=1}^{3} ||\bar{\pi}_i - r(\bar{\Pi})||_2^2 = \frac{\sqrt{2}+2\sqrt{5}}{3} \approx 1.96$. With respect to the true preferences Π, the facility actually costs $U(\Pi, r(\bar{\Pi})) = \sum_{i=1}^{3} ||\pi_i - r(\bar{\Pi})||_2^2 = \frac{5\sqrt{2}+2\sqrt{41}}{24} \approx .828$. The putative preferences are given in Figure 4.2 below.

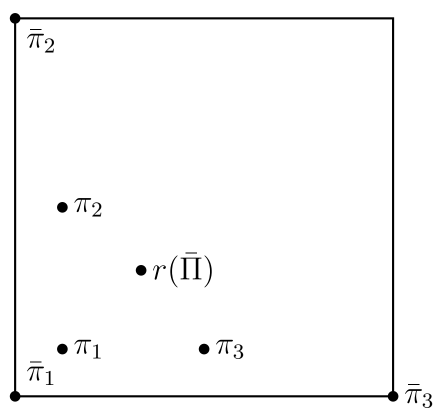


Figure 4.2: Putative Preferences for Example 4.2.2

To see that $\bar{\Pi}$ corresponds to a Nash equilibrium, first consider individual 1. If individual 1 *alters her submitted location $\bar{\pi}_1$, then she must move it up or to the right. Such an action causes the facility location to move up or to the right respectively. Both possibilities cause the facility to move further away from π_1 and therefore individual 1 cannot alter her submitted preferences to get a better result. Individual 2 and Individual 3 are giving best responses by symmetric reasoning and $\bar{\Pi}$ is a pure strategy Nash equilibrium.*

Example 4.2.2 demonstrates that manipulation in the 1-Mean can cause the social cost to increase from $.736$ to $.828$ – a cost that is $\frac{.828}{.736} = 1.25$ times as bad. For the remainder of the chapter, we focus on finding the Price of Deception (see Section 2.2) to measure the impact of manipulation in the Facility Location Problem.

Definition 4.2.3 (Price of Deception for Strategic Facility Location Games)**.** Let r be a facility location procedure. Let U be an associated real-valued cost that, given a profile Π of preferred facility locations, outputs a social cost $U(\Pi, x)$ for each possible facility location $x \in \mathcal{X}$. The decision procedure r is such that $r(\Pi) \subseteq \operatorname{argmax}_{x\in\mathcal{X}} U(\Pi, x)$. Let $NE(\Pi)$ denote the set of equilibria of the Strategic Facility Location Game with procedure r. Then the Price of Deception of r is

$$\sup_{\Pi\in\mathcal{P}} \sup_{\bar{\Pi}\in NE(\Pi)} \frac{E(U(\Pi, r(\bar{\Pi})))}{U(\Pi, r(\Pi))} \tag{4.1}$$

Based on Example 4.2.2, we know that the Price of Deception of the 1-Mean Problem is at least 1.25. Prior to finding the Price of Deception for these algorithms, we first demonstrate the need for minimal dishonesty when we minimize the sum of L_1 or L_2 norms.

4.3 Strategic Facility Location Without Refinement

Theorem 4.3.1. *Let x be an arbitrary point in $\mathcal{X}$ and let $n \geq 3$. Regardless of the sincere preference profile Π, there is a Nash equilibrium $\bar{\Pi}$ where $r(\bar{\Pi}) = x$ when distance is measured with the L_1 or L_2 norm.*

Proof. Suppose each voter v submits $\bar{\pi}_v = x$, then $r(\bar{\Pi}) = x$ and the central decision mechanism believes it has selected a facility that is 0 away from everyone. We now show $\bar{\Pi}$ is a Nash equilibrium by showing that no individual can change the facility location. Suppose voter v submits $\bar{\pi}'_v$ instead. With respect to $[\bar{\Pi}_{-v}, \bar{\pi}'_v]$, the cost to place the facility at $x = r(\bar{\Pi})$ is $||\bar{\pi}'_v - x||$ while the cost to place the facility at $x' = x + d$ is

$$\sum_{v' \neq v} ||\bar{\pi}_{v'} - x'|| + ||\bar{\pi}'_v - x'|| = (n-1)||d|| + ||\bar{\pi}'_v - x'|| \tag{4.2}$$

$$= (n-2)||d|| + ||d|| + ||\bar{\pi}'_v - x'|| \tag{4.3}$$

$$\geq (n-2)||d|| + ||\bar{\pi}'_v - (x' - d)|| \tag{4.4}$$

$$= (n-2)||d|| + ||\bar{\pi}'_v - x|| \tag{4.5}$$

Since $n \geq 3$, then $x' = x + d$ costs more with respect to $[\bar{\Pi}_{-v}, \bar{\pi}'_v]$ than x whenever $||d|| > 0$. Therefore, $r([\bar{\Pi}_{-v}, \bar{\pi}'_v]) = x$ and individual cannot alter their submitted preferences to get a better result and $\bar{\Pi}$ is a Nash equilibrium. □

Corollary 4.3.2. *The Price of Deception of the Facility Location Problem is ∞ when distance is measured with the L_1 or L_2 norm.*

Proof. Let $y \in \mathcal{X}$. Suppose $n \geq 3$ and let $\pi_v = y$ for all $v \in N$. If individuals are honest then the central decision mechanism selects $r(\Pi) = y$ which is 0 away from all individuals. Let $x \in \mathcal{X} \setminus y$. By Theorem 4.3.1 there is a Nash equilibrium $\bar{\Pi}$ where $r(\bar{\Pi}) = x$ which is $||y - x|| > 0$ away from individual v's sincere location π_v. Therefore the Price of Deception is at least $\frac{n||y-x||}{0} = \infty$. □

Theorem 4.3.1 and Corollary 4.3.2 demonstrate that the Nash equilibria concept can lead to illogical outcomes. Therefore we apply the minimal dishonesty solution concept from Section 2.3.

Definition 4.3.3. Let Π be the sincere preferences and let $\bar{\Pi}$ be a Nash Equilibrium in the Strategic Facility Location Game. An individual v is minimally dishonest if $||\pi_v - \bar{\pi}'_v|| <$

$||\pi_v - \bar{\pi}_v||$ implies that v's cost of the more honest outcome, $u_v(\pi_v, r([\bar{\Pi}_{-v}, \bar{\pi}'_v)])$, is more than v's cost of the equilibrium outcome, $u_v(\pi_v, r(\bar{\Pi}))$.

A pure strategy Nash equilibrium is minimally dishonest if each individual is minimally dishonest.

Said in the contrapositive, an individual submitting location $\bar{\pi}_v$ is not minimally dishonest if she can report a location closer $\bar{\pi}'_v$ to her preferred location π_v and obtain at least as good an outcome. We proceed by examining each problem with deterministic and random tie-breaking rules.

4.4 Prices of Deception for Facility Location Procedures

4.4.1 1-Median Problem

We begin by analyzing the 1-Median Problem where the facility location $r(\Pi)$ is a minimizer of $U(\Pi, x) = \sum_{i=1}^{n} ||\pi_i - x||_1 = \sum_{i=1}^{n} \sum_{j=1}^{k} |\pi_{ij} - x_j|$ where π_{ij} is the jth coordinate of π_i. While there is no closed form solution for this problem, the problem is separable and easy to solve. Let a be minimal and b be maximal and such that $|\{\pi_v : \pi_v \leq a\}| \geq \lceil \frac{n}{2} \rceil$ and $|\{\pi_v : \pi_v \geq b\}| \geq \lceil \frac{n}{2} \rceil$. Then the set of optimal facility locations is given by $\{x : a \leq x \leq b\}$. When selecting a facility location we assume $r(\Pi)$'s only dependence on Π is a and b.

1-Median Problem in $\mathbb{R}$.

Breaking Ties Deterministically

We begin by breaking ties deterministically. Consider the λ-1-Median Problem for when voters submit locations in $\mathbb{R}^k$:

Definition 4.4.1. Let $\lambda \in [0, 1]^k$. For the profile Π, let $\{x \in \mathbb{R}^k : a(\Pi) \leq x \leq b(\Pi)\}$ be the set of optimal facility locations. The λ-1-Median Problem selects facility location $r(\Pi) = (\mathbf{1} - \lambda)a(\Pi) + \lambda b(\Pi)$.

Example 4.4.2. Three Nash Equilibria With the Minimal Dishonesty Refinement

Suppose the set of allowed facility locations is $\mathcal{X} = [0,6]$ *and that the sincere preferences are given in Figure 4.3. Furthermore suppose the* $\frac{1}{2}$*-1-Median Problem is used to select a facility location. Given the sincere preferences, any location in* $[1,3]$ *minimizes the sum of* L_1 *norms. If everyone is sincere, then the procedure places the facility at* $r(\Pi) = 2$*. The total cost of this facility is* $\sum_{v=1}^{6} |\pi_v - r(\Pi)| = 12$*.*

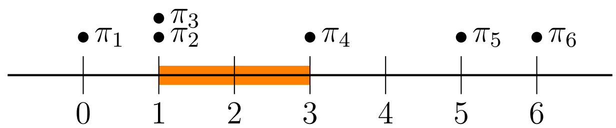


Figure 4.3: Individuals' Sincere Locations for Example 4.4.2.

The sincere preferences Π *do not correspond to a Nash equilibrium since individual* 4 *can submit* $\bar{\pi}_v = 4$ *to obtain a better result. We present 3 minimally dishonest Nash equilibria for* Π *below.*

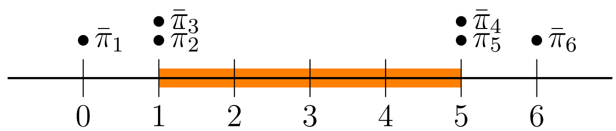


Figure 4.4: First Minimally Dishonest Nash Equilibrium for Example 4.4.2.

In the first equilibrium shown in Figure 4.4, only individual 4 *is dishonest. Any location in* $[1,5]$ *minimizes the sum of the* L_1 *norms with respect to the submitted preferences* $\bar{\Pi}$*. Therefore* $r(\bar{\Pi}) = 3$ *and the actual cost of this facility location (with respect to the sincere locations) is* $\sum_{v=1}^{6} |\pi_v - r(\bar{\Pi})| = 12$*. Although someone has successfully manipulated this system, the total cost for the facility location remains unchanged.*

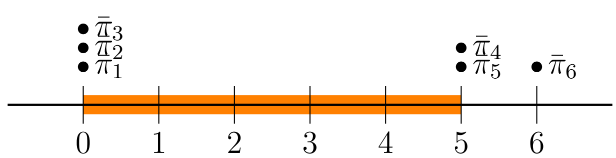


Figure 4.5: Second Minimally Dishonest Nash Equilibrium for Example 4.4.2.

In the second equilibrium, shown in Figure 4.5, individuals 2*,* 3 *and* 4 *are dishonest resulting in the facility location* $r(\bar{\Pi}) = 2.5$*. Once again the actual cost of the facility location is* 12*.*

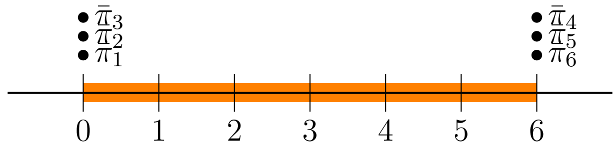


Figure 4.6: Third Minimally Dishonest Nash Equilibrium for Example 4.4.2.

In the third equilibrium, shown in Figure 4.6, only individuals 1 *and* 6 *are honest resulting in the facility location* $r(\bar{\Pi}) = 3$*. The actual cost of the facility location is again* 12*.*

In Example 4.4.2, we observe that we reach the ideal outcome for manipulable systems; The system is manipulable and we have individuals acting strategically by altering their preferences to obtain better results, but **manipulation has no effect on the social utility**. In Theorem 4.4.3, we present a class of mechanisms where manipulation does not impact the social utility when minimizing the sum of the L_1 norms.

Theorem 4.4.3. *When individuals are minimally dishonest, the Price of Deception for the* λ*-1-Median decision mechanism in* $\mathbb{R}$ *is* 1*.*

Let $[a(\bar{\Pi}), b(\bar{\Pi})]$ be the set of facility locations that minimizes the sum of L_1 norms given the the submitted preferences $\bar{\Pi}$. We know $|\{v : \bar{\pi}_v \leq a(\bar{\Pi})\}| \geq \frac{n}{2}$ and that there is a *median* voter such that $\bar{\pi}_v = a(\bar{\Pi})$. An equivalent statement holds for $b(\bar{\Pi})$ implying $a(\bar{\Pi}) = b(\bar{\Pi}) = r(\bar{\Pi})$ when n is odd. Prior to proving Theorem 4.4.3, we begin with the following results describing a relationship between the submitted and the sincere preferences.

Lemma 4.4.4. *Let* Π *be the sincere preferences and* $\bar{\Pi}$ *be a minimally dishonest equilibrium when selecting a facility location by minimizing the* L_1 *norm. Then* $\{v : \pi_v \leq a(\bar{\Pi})\} = \{v : \bar{\pi}_v \leq a(\bar{\Pi})\}$*.*

Proof. First we show $(\subseteq)$. Let v be such that $\pi_v \leq a(\bar{\Pi})$ and suppose for contradiction that $\bar{\pi}_v > a(\bar{\Pi})$. Let $\bar{\Pi}' = [\bar{\Pi}_{-v}, a(\bar{\Pi})]$ be the preference profile obtained after replacing $\bar{\pi}_v$ with $a(\bar{\Pi})$. With respect to $\bar{\Pi}'$, the facility is located at $r(\bar{\Pi}') = a(\bar{\Pi})$ which is at least as close to π_v as $r(\bar{\Pi})$. However, this is a contradiction to minimal dishonesty since v can be more honest and get at least as good a result. Therefore $\{v : \pi_v \leq a(\bar{\Pi})\} \subseteq \{v : \bar{\pi}_v \leq a(\bar{\Pi})\}$.

Showing ($\supseteq$) uses an identical argument. Let v be such that $\bar{\pi}_v \leq a(\bar{\Pi})$ and assume $\pi_v > a(\bar{\Pi})$ for contradiction. However, v can obtain at least as good a result by submitting $\bar{\pi}'_v = \pi_v$ causing the facility to be located somewhere in $[b(\bar{\Pi}), \pi_v]$. This contradicts minimal dishonesty. Therefore $\{v : \pi_v \leq a(\bar{\Pi})\} = \{v : \bar{\pi}_v \leq a(\bar{\Pi})\}$ and the lemma holds. $\square$

An identical equality holds for voters to the right of $b(\bar{\Pi})$. Therefore, Lemma 4.4.4 allows us to give a partial ordering over the voters' sincere preferences when we only know the submitted preferences. Specifically for v where $\bar{\pi}_v \leq a(\bar{\Pi})$ and v' where $\bar{\pi}_{v'} \geq b(\bar{\Pi})$, we know $\pi_v \leq \pi_{v'}$ and the following corollary holds immediately:

Corollary 4.4.5. *Let Π be the sincere preferences and $\bar{\Pi}$ be a minimally dishonest equilibrium when selecting a facility location by minimizing the L_1 norm. Then $\{v : \pi_v \leq a(\Pi)\} = \{v : \bar{\pi}_v \leq a(\bar{\Pi})\}$.*

We are now able to prove Theorem 4.4.3. We show that manipulation has no impact on social utility when we solve the 1-Median Problem and break ties by using the center point.

Proof of Theorem 4.4.3. Let Π and $\bar{\Pi}$ be sincere and submitted preferences where $\bar{\Pi}$ is a minimally dishonest Nash equilibrium. It suffices to show $r(\bar{\Pi}) = (1-\lambda)a(\bar{\Pi}) + \lambda b(\bar{\Pi}) \in [a(\Pi), b(\Pi)]$ since every $x \in [a(\Pi), b(\Pi)]$ is an optimal solution to the 1-Median Problem for preferences Π.

First we show that $a(\bar{\Pi}) \leq a(\Pi)$. For contradiction, suppose this is not the case and that there is a median voter with $\bar{\pi}_v = a(\bar{\Pi}) > a(\Pi) \geq \pi_v$. If v instead is honest and reports π_v then either the facility location will not change or the new facility location will be in the interval $[\pi_v, r(\bar{\Pi}))$. In both cases, v obtains at least as good an outcome contradicting minimal dishonesty and completing the claim. Symmetrically, $b(\bar{\Pi}) \geq b(\Pi)$.

We now show $r(\bar{\Pi}) \in [a(\Pi), b(\Pi)]$. For contradiction, and without loss of generality, assume $r(\bar{\Pi}) < a(\Pi)$ implying $\lambda < 1$. There must exist a voter v such that $\pi_v = a(\Pi)$. By Lemma 4.4.4, $\bar{\pi}_v \leq a(\bar{\Pi}) < \pi_v$. If $b(\bar{\Pi}) \leq \pi_v$, then v can submit $\bar{\pi}'_v = \pi_v$ which will yield an outcome in $[b(\bar{\Pi}), \pi_v]$. If $b(\bar{\Pi}) > \pi_v$, then there is an even number of voters and v can instead submit $\bar{\pi}_v = \frac{\pi_v - \lambda b(\bar{\Pi})}{1-\lambda}$ which will yield the facility location π_v.

In both cases, v can be more honest and get at least as good an outcome contradicting minimal dishonesty. Therefore $r(\bar{\Pi}) \in [a(\Pi), b(\Pi)]$. Thus the social utility of the outcome $r(\bar{\Pi})$, $\sum_{v \in N} |\pi_v - r(\bar{\Pi})|$, is equal to the utility of the intended outcome, $\sum_{v \in N} |\pi_v - r(\Pi)|$. Therefore, the Price of Deception is

$$\frac{\sum_{v \in N} |\pi_v - r(\bar{\Pi})|}{\sum_{v \in N} |\pi_v - r(\Pi)|} = \frac{\sum_{v \in N} |\pi_v - r(\Pi)|}{\sum_{v \in N} |\pi_v - r(\Pi)|} = 1 \tag{4.6}$$

completing the proof of the theorem. □

Example 4.4.2 shows that people can manipulate the system and Theorem 4.4.3 shows that manipulation will not impact social utility. Furthermore, we can select λ such that the mechanism is not manipulable. Specifically, for $\lambda \in \{0, 1\}$, the λ-1-Median Problem is strategy-proof.

Theorem 4.4.6. *Fix $\lambda \in \{0, 1\}$. The λ-1-Median Problem in $\mathbb{R}$ is strategy-proof.*

Proof. By definition of strategy-proof, we need to show that Π is a Nash equilibrium for Π. Let $[a(\Pi), b(\Pi)]$ be the set of optimal facility locations. Without loss of generality assume $\lambda = 1$ and $r(\Pi) = b(\Pi)$. Suppose voter v changes her submitted preferences to a minimally dishonest best response $\bar{\pi}_v$ and the facility is now located at $r([\Pi_{-v}, \bar{\pi}_v]) = b([\Pi_{-v}, \bar{\pi}_v])$. If $\pi_v = b(\Pi)$, then v has no incentive to change her preferences. If $\pi_v > b(\Pi)$ then by Corollary 4.4.5 $\bar{\pi}_v \geq b(\Pi)$. However, there is still a median voter v' where $\pi_{v'} = b(\Pi)$ and therefore $b([\Pi_{-v}, \bar{\pi}_v]) = b(\Pi)$ and v has no incentive to change her preferences. Symmetrically no voter v where $\pi_v \leq a(\Pi)$ can alter her preferences to get a better outcome for herself. Therefore Π is a Nash equilibrium. □

When $|N|$ is odd, $a(\bar{\Pi}) = b(\bar{\Pi})$ and the tie-breaking rule is never invoked. Therefore the 1-Median Problem is strategy-proof with an odd number of voters.

Corollary 4.4.7. *When $|N|$ is odd the 1-Median Problem in $\mathbb{R}$ is strategy-proof.*

Breaking Ties Randomly

Next, we consider breaking ties uniformly at random. If $|N|$ is odd then there are no ties and the 1-Median Problem is strategy-proof. When there are an even number of voters we show that the Price of Deception is $\sqrt{2}$. First we characterize an individual's best response.

Lemma 4.4.8. *Suppose $|N| = 2$ and each voter must submit a location in the interval $\mathcal{P}_i = \mathcal{X} = [l, u]$ for the 1-Median Problem when breaking ties uniformly at random. If $\bar{\pi}_1 \leq \pi_2$, then voter 2's unique minimally dishonest best response is $\bar{\pi}_2 = min\{u, \bar{\pi}_1 + \sqrt{2}(\pi_2 - \bar{\pi}_1)\}$.*

Proof. Trivially if $\bar{\pi}_1 = \pi_2$ then the best response for player 2 is $\bar{\pi}_1 + \sqrt{2}(\pi_2 - \bar{\pi}_1) = \pi_2$. Otherwise $\bar{\pi}_1 < \pi_2$. By scaling and translating we may assume $\bar{\pi}_1 = 0$ and $\pi_2 = 1$ and show that player 2's best response is $min\{u, \sqrt{2}\}$. It is straightforward to show that voter 2 should submit a location $x \geq 1$. The facility location Y is selected uniformly at random from $[0, x]$ and voter 2 has an expected utility of

$$E(|Y-1|) = P(Y \leq 1)E(1-Y|Y \leq 1) + P(Y \geq 1)E(Y-1|Y \geq 1) \tag{4.7}$$

$$= \frac{1}{x} \cdot \frac{1}{2} + \frac{x-1}{x} \cdot \frac{x-1}{2} \tag{4.8}$$

$$= \frac{x^2 - 2x + 2}{2x} \tag{4.9}$$

It is straightforward to verify that this function is uniquely minimized at $x = \sqrt{2}$ on the domain $[1, \infty]$. Therefore if $u \geq \sqrt{2}$ the unique best response for player 2 is $\sqrt{2}$. If $u < \sqrt{2}$ then it is easily verified that $E(|Y-1|)$ is strictly decreasing on the interval $[1, u]$ and therefore player 2's unique best response is $x = u$. Either way, player 2's unique best response is $min\{u, \sqrt{2}\}$. By the uniqueness of the best response, any more honest

$\bar{\pi}_2$ yields a worse outcome for player 2 and therefore $min\{u, \sqrt{2}\}$ is player 2's unique minimally dishonest best response. $\square$

Theorem 4.4.9. *Let* $|N| = 2$. *When both individuals are minimally dishonest and ties are broken uniformly at random, the Price of Deception for the 1-Median decision mechanism in* $\mathbb{R}$ *is* $\sqrt{2}$.

Proof. Without loss of generality $\pi_1 \leq \pi_2$.

First we consider $\pi_1 = \pi_2 = 0$ and show that the Price of Deception is 1. Let $\bar{\Pi}$ be a minimally dishonest Nash equilibrium. If $\bar{\pi}_1 < 0$ then Lemma 4.4.8 implies $\bar{\pi}_2 > 0$ thus without loss of generality we may assume $\bar{\pi}_1 \leq 0 \leq \bar{\pi}_2$. By Lemma 4.4.8,

$$\bar{\pi}_2 = min\{u, \bar{\pi}_1 - \sqrt{2}\bar{\pi}_1\} \tag{4.10}$$

$$\leq \bar{\pi}_1 - \sqrt{2}\bar{\pi}_1 = (1 - \sqrt{2})\bar{\pi}_1 \tag{4.11}$$

This implies player 2 should submit a locations slightly closer to 0 than player 1. Symmetrically, player 1 should submit a location closer to 0 than player 2. Both conditions cannot simultaneously happen and the only $\bar{\Pi}$ satisfying (4.11) and it's symmetric statement is $\bar{\Pi} = \Pi = (0, 0)$. It is straightforward to verify this is also a minimally dishonest Nash equilibrium.

Second we consider $\pi_1 < \pi_2$. By scaling, translating, and reflecting we may assume $\pi_1 = -1$ and $\pi_2 = 1$ and that $u \geq -l$. The sincere outcome $r(\bar{\Pi})$ selects a facility location uniformly at random in $[-1, 1]$ with cost $U(\Pi, r(\Pi)) = 2$. Again let $\bar{\Pi}$ be a Nash equilibrium. By Lemma 4.4.8 $\bar{\pi}_2 = min\{u, \bar{\pi}_1 + \sqrt{2}(1 - \bar{\pi}_1)\}$. Symmetrically, $\bar{\pi}_1 = max\{l, \bar{\pi}_2 + \sqrt{2}(-1 - \bar{\pi}_2)\}$.

Observe that if $\bar{\pi}_2 = u$ then

$$\bar{\pi}_2 + \sqrt{2}(-1 - \bar{\pi}_2) = (1 - \sqrt{2})u - \sqrt{2} \tag{4.12}$$

$$\leq (\sqrt{2} - 1)l - \sqrt{2} < l \tag{4.13}$$

This implies that if $\bar{\pi}_2 = u$ then $\bar{\pi}_1 = l$. Therefore we can break the problem into two cases: $\bar{\pi}_1 > l$, $\bar{\pi}_2 < u$; and $\bar{\pi}_1 = l$.

Case 1: $\bar{\pi}_1 > l$**,** $\bar{\pi}_2 < u$. By Lemma 4.4.8

$$\bar{\pi}_2 = (1-\sqrt{2})\bar{\pi}_1 + \sqrt{2} \tag{4.14}$$

$$\bar{\pi}_1 = (1-\sqrt{2})\bar{\pi}_2 - \sqrt{2} \tag{4.15}$$

which has the unique solution $\bar{\pi}_2 = -\bar{\pi}_1 = 1+\sqrt{2}$ and the facility location is selected uniformly at random from $[-1-\sqrt{2}, 1+\sqrt{2}]$. For brevity, let $U = U(\Pi, r(\bar{\Pi}))$. Let $Y \sim Unif(\bar{\pi}_1, \bar{\pi}_2)$ be the randomly selected facility location. We have the following

$$P(Y \le -1) = \frac{-\bar{\pi}_1 - 1}{\bar{\pi}_2 - \bar{\pi}_1} \qquad E(U|Y \le -1) = 1 - \bar{\pi}_1 \tag{4.16}$$

$$P(Y \ge 1) = \frac{\bar{\pi}_2 - 1}{\bar{\pi}_2 - \bar{\pi}_1} \qquad E(U|Y \ge 1) = 1 + \bar{\pi}_2 \tag{4.17}$$

$$P(-1 \le Y \le 1) = \frac{2}{\bar{\pi}_2 - \bar{\pi}_1} \qquad E(U| -1 \le Y \le 1) = 2. \tag{4.18}$$

Putting all three together,

$$E(U) = \frac{\bar{\pi}_2^2 + \bar{\pi}_1^2 + 2}{\bar{\pi}_2 - \bar{\pi}_1}. \tag{4.19}$$

For Case 1 where $\bar{\pi}_2 = -\bar{\pi}_1 = \sqrt{2}+1$ the expected cost of $\bar{\Pi}$ is $E(U) = 2\sqrt{2}$. Since the cost when everyone is sincere is $U(\Pi, r(\Pi)) = 2$, the Price of Deception in Case 1 is $\sqrt{2}$.

Case 2: $\bar{\pi}_1 = l$. By Lemma 4.4.8, $l \ge (1-\sqrt{2})\bar{\pi}_2 - \sqrt{2}$ and $\bar{\pi}_2 \le (1-\sqrt{2})l + \sqrt{2}$. Combining both inequalities, $l \ge -\sqrt{2} - 1$. Combining the last two inequalities, $\bar{\pi}_2 \le \sqrt{2} + 1$.

The function $(z^2 + y^2 + 2)/(z - y)$ is increasing on the interval $0 \le z \le \sqrt{2} + 1$ when

$y \in [-\sqrt{2}-1, 0]$ and therefore by (4.21),

$$E(U) = \frac{\bar{\pi}_2^2 + \bar{\pi}_1^2 + 2}{\bar{\pi}_2 - \bar{\pi}_1} \leq \frac{(\sqrt{2}+1)^2 + \bar{\pi}_1^2 + 2}{(\sqrt{2}+1) - \bar{\pi}_1}. \tag{4.20}$$

Symmetrically,

$$E(U) \leq \frac{(\sqrt{2}+1)^2 + \bar{\pi}_1^2 + 2}{(\sqrt{2}+1) - \bar{\pi}_1} \leq \frac{(\sqrt{2}+1)^2 + (-\sqrt{2}-1)^2 + 2}{(\sqrt{2}+1) - (-\sqrt{2}-1)} = 2\sqrt{2}. \tag{4.21}$$

As in Case 1, the value of the sincere outcome is always 2 and the Price of Deception is at most $\sqrt{2}$. □

Theorem 4.4.10. *Suppose $|N|$ is even. When individuals are minimally dishonest and ties are broken uniformly at random, the Price of Deception for the 1-Median Problem in $\mathbb{R}$ is $\sqrt{2}$.*

Proof. The lower bound is given by Theorem 4.4.9.

Let Π be a set of sincere preferences and $\bar{\Pi}$ be a corresponding minimally dishonest Nash equilibrium. Given a game with $|N|$ players we show how to reduce the game to 2 players that yields the same $r(\bar{\Pi})$. Finally we show that the Price of Deception of the 2-player game is at least as large as the Price of Deception of the $|N|$-player game. Therefore the Price of Deception is at most $\sqrt{2}$ by Theorem 4.4.9.

Corollary 4.4.5 extends to this setting and $\{\pi_v : \pi_v \leq a(\Pi)\} = \{\bar{\pi}_v : \bar{\pi}_v \leq a(\bar{\Pi})\}$. We begin by showing $\pi_i \leq \pi_j$ implies $\bar{\pi}_i \leq \bar{\pi}_j$. The result immediately holds if $\pi_i \leq a(\Pi)$ but $\pi_j > a(\Pi)$. Therefore by symmetry assume $\pi_j \leq a(\bar{\Pi})$. If $\bar{\pi}_k < a(\bar{\Pi})$ then k obtains the same outcome if they submit $min\{a(\bar{\Pi}), \pi_k\}$. Therefore by minimal dishonesty $\bar{\pi}_k = min\{a(\bar{\Pi}), \pi_k\}$ and $\bar{\pi}_i \leq \bar{\pi}_j$ as desired. We proceed to the main result.

Let $k = |N|/2$ and index players so that $\pi_{-k} \leq \pi_{-k+1} \leq ... \leq \pi_{-1} \leq \pi_1 \leq ... \leq \pi_k$. By the previous claim, if $\bar{\Pi}$ is a minimally dishonest Nash equilibrium then $\bar{\pi}_{-k} \leq \bar{\pi}_{-k+1} \leq ... \leq \bar{\pi}_{-1} \leq \bar{\pi}_1 \leq ... \leq \bar{\pi}_k$ and players 1 and -1 are median voters with respect to Π and $\bar{\Pi}$.

Construct the following 2-player game. Let $u = \bar{\pi}_2$ and $l = \bar{\pi}_{-2}$. Similar to Theorem 4.4.6, $u \geq b(\Pi) = \pi_1$ and $l \leq a(\Pi) = \pi_{-1}$. Players 1 and -1 have preferences $\pi'_i = \pi_i$. This is a valid game since $l \leq \pi'_{-1} \leq \pi'_1 \leq u$. It immediately follows that $\bar{\Pi}' = \{\bar{\pi}_{-1}, \bar{\pi}_1\}$ is a minimally dishonest Nash equilibrium for Π' with the same outcome as $\bar{\Pi}$ since otherwise $\bar{\Pi}$ is not a minimally dishonest Nash equilibrium for Π.

Denote $U_i(\Pi, r)$ as the cost for player i given location π_i and outcome r. Thus $E[U(\Pi, r(\bar{\Pi}))] = \sum_i E[U_i(\Pi, r(\bar{\Pi}))]$. Since $b(\Pi) = \pi_1 \leq \pi_i$ for $i \geq 1$,

$$U_i(\Pi, r(\Pi)) = U_1(\Pi, r(\Pi)) + \pi_i - \pi_1 \; \forall i \geq 1. \tag{4.22}$$

$$= U_1(\Pi', r(\Pi')) + \pi_i - \pi_1 \; \forall i \geq 1. \tag{4.23}$$

By the triangle inequality,

$$E[U_i(\Pi, r(\bar{\Pi}))] \leq E[U_1(\Pi, r(\bar{\Pi}))] + \pi_i - \pi_1 \; \forall i \geq 1. \tag{4.24}$$

$$= E[U_1(\Pi', r(\bar{\Pi}'))] + \pi_i - \pi_1 \; \forall i \geq 1. \tag{4.25}$$

Symmetrically,

$$U_i(\Pi, r(\Pi)) = U_{-1}(\Pi', r(\Pi')) + \pi_{-1} - \pi_i \; \forall i \leq -1. \tag{4.26}$$

$$E[U_i(\Pi, r(\bar{\Pi}))] \leq E[U_{-1}(\Pi', r(\bar{\Pi}'))] + \pi_{-1} - \pi_i \; \forall i \leq -1. \tag{4.27}$$

We proceed by bounding the Price of Deception by grouping players i and $-i$. Observe that for $n_1, n_2, d_1, d_2 \geq 0$ that $\frac{n_1+n_2}{d_1+d_2} \leq max\{\frac{n_1}{d_1}, \frac{n_2}{d_2}\}$ where $\frac{0}{0} = 1$ and $\frac{x}{0} = \infty$ for $x > 0$.

Therefore

$$
\begin{aligned}
&\frac{E[U_i(\Pi, r(\bar{\Pi}))] + E[U_{-i}(\Pi, r(\bar{\Pi}))]}{U_i(\Pi, r(\Pi)) + U_{-i}(\Pi, r(\Pi))} && (4.28)\\
\leq &\frac{E[U_1(\Pi', r(\bar{\Pi}'))] + E[U_{-1}(\Pi', r(\bar{\Pi}'))] + \pi_i - \pi_1 + \pi_{-1} - \pi_{-i}}{U_1(\Pi', r(\Pi'))] + U_{-1}(\Pi', r(\Pi'))] + \pi_i - \pi_1 + \pi_{-1} - \pi_{-i}} && (4.29)\\
\leq &max\left\{\frac{E[U_1(\Pi', r(\bar{\Pi}'))] + E[U_{-1}(\Pi', r(\bar{\Pi}'))]}{U_1(\Pi', r(\Pi'))] + U_{-1}(\Pi', r(\Pi'))]}, \frac{\pi_i - \pi_1 + \pi_{-1} - \pi_{-i}}{\pi_i - \pi_1 + \pi_{-1} - \pi_{-i}}\right\} && (4.30)\\
\leq &max\{\sqrt{2}, 1\} = \sqrt{2} && (4.31)
\end{aligned}
$$

by Theorem 4.4.9 since Π' and $\bar{\Pi}'$ are from a 2-player game.

Thus the Price of Deception for Π and $\bar{\Pi}$ is given by

$$
\begin{aligned}
\frac{E[U(\Pi, r(\bar{\Pi}))]}{U(\Pi, r(\Pi))} &= \frac{\sum_i E[U_i(\Pi, r(\bar{\Pi}))]}{\sum_i U_i(\Pi, r(\Pi))} && (4.32)\\
&= \frac{\sum_{i=1}^k \Big(E[U_i(\Pi, r(\bar{\Pi})) + E[U_{-i}(\Pi, r(\bar{\Pi}))]\Big)}{\sum_{i=1}^k \Big(U_i(\Pi, r(\Pi)) + U_{-i}(\Pi, r(\Pi))\Big)} && (4.33)\\
&\leq max_{i\in[k]} \frac{E[U_i(\Pi, r(\bar{\Pi})) + E[U_{-i}(\Pi, r(\bar{\Pi}))]}{U_i(\Pi, r(\Pi)) + U_{-i}(\Pi, r(\Pi))} && (4.34)\\
&\leq \sqrt{2} && (4.35)
\end{aligned}
$$

completing the proof of the theorem. □

1-Median Problem in $\mathbb{R}^k$

Since the 1-Median Problem is separable we might expect that Theorem 4.4.3 and Theorem 4.4.10 hold in higher dimensions. However, while determining $r(\Pi)$ might be separable, determining a minimally dishonest best response $\bar{\pi}_v$ is not as shown in the following example:

Example 4.4.11. Finding a Minimally Dishonest Best Response is not Separable.

If determining the minimally dishonest best response an individual v cannot separately compute *each component of their best response $\bar{\pi}_v$ for general $\mathcal{X}$. Consider the following preferences where individuals must submit preferences inside $\mathcal{X}$ as given by the triangle in Figure 4.7.*

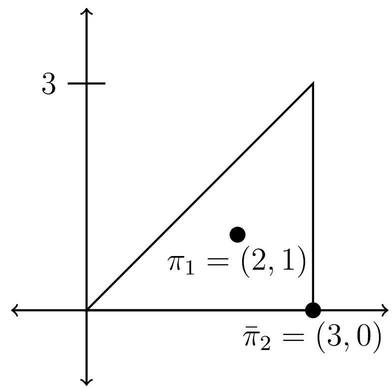


Figure 4.7: Determining a Best Response is not Separable in the 1-Median Problem.

Let $\lambda = (\frac{1}{2}, \frac{1}{2})$. Ideally, voter 1 would like to submit $\bar{\pi}_1$ such that $r(\bar{\Pi}) = (2,1)$. Denote $\bar{\pi}_{ij}$ and $r_j(\bar{\Pi})$ as the j^{th} coordinate of $\bar{\pi}_i$ and $r(\bar{\Pi})$ respectively. Voter 1 can cause $r_1(\bar{\Pi}) = 2$ by submitting $\bar{\pi}_{11} = 1$ and can cause $r_2(\bar{\Pi}) = 1$ by submitting $\bar{\pi}_{12} = 2$. Therefore, voter 1 can only cause $r(\bar{\Pi}) = (2,1)$ by submitting $\bar{\pi}_1 = (1,2)$. This point is not in the triangle and therefore voter 1 cannot submit $(1,2)$. The optimal choice of $\bar{\pi}_{11}$ depends on the choice of $\bar{\pi}_{12}$ and therefore determining a best response is not a separable problem.

Regrettably, the lack of separability in determining a best response can lead to a large Price of Deception in higher dimensions.

Theorem 4.4.12. *Suppose $\lambda_1 \in (0,1)$ and voter v's cost of outcome x is $u_v(\pi_v, x) = ||\pi_v - x||_{p_v}$ for some $p_v \in (0,\infty)$ for each v. Then the Price of Deception of the λ-1-Median Problem in $\mathbb{R}^k$ is ∞ for $k \geq 2$.*

Proof. It suffices to show the result for $k = 2$ since the bad example can always be placed in higher dimensions. Let $|N| = 2k$ for some integer $k \geq 2$. The sincere preferences are $\pi_v = 0$ for all v. The set of feasible facility location placements is $\mathcal{X} = conv.hull(\pi_1, a, b)$ as shown in Figure 4.8.

If everyone is honest, then the facility is placed at $r(\Pi) = (0,0)$ with total cost 0. Suppose half the voters submit $\bar{\pi}_v = a$ and half the voters submit $\bar{\pi}_v = b$. Then $r(\bar{\Pi}) =$

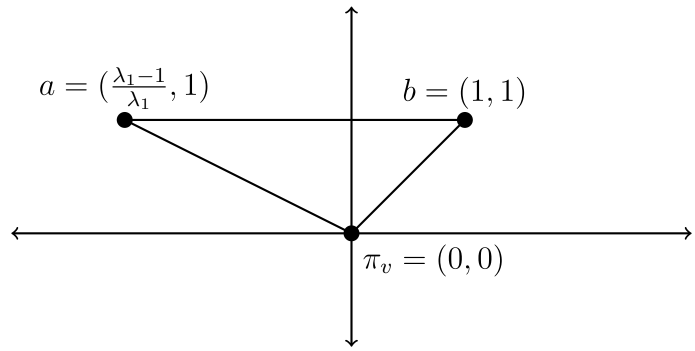


Figure 4.8: Preferences Showing the Price of Deception is ∞ for the λ-1-Median Problem.

$\lambda a + (\mathbf{1} - \lambda)b = (0, 1)$ with a sincere cost of $|N|$. If $\bar{\Pi}$ is a minimally dishonest Nash equilibrium, then the Price of Deception is ∞.

We now show that $\bar{\Pi}$ is a minimally dishonest Nash equilibrium. Consider voter v that submits $\bar{\pi}_v = b$. Regardless of how she alters her submitted information, more than half the voters submit a height of 1 and therefore she cannot change the height of the facility. Moreover, if she alters her preferences at all then the facility moves to the left corresponding to a worse outcome for her. There v is providing a minimally dishonest best response. Symmetrically voter v submitting $\bar{\pi}_v = a$ is providing a minimally dishonest best response and $\bar{\Pi}$ is a minimally dishonest Nash equilibrium. □

Theorem 4.4.13. *Suppose voter v's cost of outcome x is $u_v(\pi_v, x) = ||\pi_v - x||_{p_v}$ for some $p_v \in (0, \infty)$ for each v and that ties are broken uniformly at random. Then the Price of Deception of the 1-Median Problem in $\mathbb{R}^k$ is ∞ for $k \geq 2$.*

Proof. The proof is similar identical to Theorem 4.4.13. Again let $N = [2k]$ for some integer $k \geq 2$. The sincere preferences are $\pi_v = (-1, 0)$ for odd indexed voters and let $\pi_v = (1, 0)$ for even indexed voters. Fix $h > 0$ and let $a = (-1-\sqrt{2}, h)$ and $b = (1+\sqrt{2}, h)$ and let $\mathcal{X} = conv.hull(\pi_1, a, b)$ as shown in Figure 4.9.

If everyone is honest then the facility is again placed at $r(\Pi) = (0, 0)$ with total cost $2|N|$. Suppose the odd indexed voters submit $\bar{\pi}_v = a$ and the even indexed voters submit $\bar{\pi}_v = b$. Then the facility is selected uniformly at random from $[a, b]$ with total cost

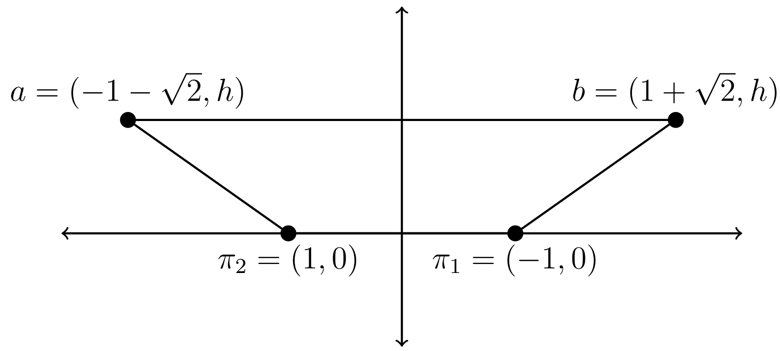


Figure 4.9: Preferences Showing the Price of Deception is ∞ for the 1-Median Problem with Random Tie-Breaking.

$2\sqrt{2}|N| + h \to \infty$ as $h \to \infty$. Therefore if $\bar{\Pi}$ is a minimally dishonest Nash equilibrium then the Price of Deception is ∞.

We now show that $\bar{\Pi}$ is a Nash equilibrium. By symmetry consider voter 2. As in Theorem 4.4.12 more than half the voters submit a height of h and therefore voter 2 cannot change the height of the facility. If voter 2 submits $\bar{\pi}'_2 = (x, y)$ then the facility will be selected uniformly at random between $(-1-\sqrt{2}, h)$ and (x, h). Let $X \sim Unif(-1, \sqrt{2}, x)$. Then voter 2's utility of the outcome is

$$u_2(\pi_2, r([\bar{\Pi}_{-2}, \bar{\pi}'_2])) = \sqrt[p_2]{E(|X-1|)^{p_2} + h^{p_2}} \tag{4.36}$$

which is minimized when $E(|X-1|)$ is minimized since h is a constant. Therefore the 2-dimensional game reduces to a 1-dimensional game. By Lemma 4.4.8, voter 2's minimally dishonest best response is to submit $x = 1 + \sqrt{2}$ which requires $y = h$. Therefore voter 2 is submitting a minimally dishonest best response and $\bar{\Pi}$ is a minimally dishonest Nash equilibrium. Thus the Price of Deception is ∞. $\square$

Theorems 4.4.12 and 4.4.12 show there is always a set of allowed locations $\mathcal{X}$ such that the Price of Deception is arbitrarily large. In the proof of Theorem 4.4.12, $\mathcal{X}$ is a triangle. In Theorem 4.4.13, $\mathcal{X}$ is a trapezoid and could easily be extended to a triangle without changing the proof. As mechanism designers, we may alter $\mathcal{X}$ such that we obtain more desirable results. Specifically, we show that if $\mathcal{X} = \{x \in \mathbb{R}^k : l \leq x \leq u\}$ then the Price of Deception of the λ-1-Median Problem and 1-Median Problem with random tie-breaking are 1 and $\sqrt{2}$ respectively.

Theorem 4.4.14. *Suppose $\mathcal{X} = \{x \in \mathbb{R}^k : l \leq x \leq u\}$. Then the Price of Deception of the λ-1-Median Problem is 1 for all $\lambda \in [0,1]^k$.*

Theorem 4.4.15. *Suppose $|N|$ is even. Suppose $\mathcal{X} = \{x \in \mathbb{R}^k : l \leq x \leq u\}$ and that ties are broken uniformly at random. Then the Price of Deception of the 1-Median Problem is $\sqrt{2}$.*

Unlike Example 4.4.11, the optimal selection of $\bar{\pi}_{vj}$ is independent of $\bar{\pi}_{vi}$ for all $i \neq j$. Thus, the k-dimensional case reduces to the 1-dimensional case and Theorems 4.4.14 and 4.4.15 hold by Theorems 4.4.6 and 4.4.10. Moreover, we also immediately obtain a variety of strategy-proof mechanisms regardless of $\mathcal{X}$.

Theorem 4.4.16. *Fix $\lambda \in \{0,1\}^k$ and $\mathcal{X} = \{x \in \mathbb{R}^k : l \leq x \leq u\}$. The λ-1-Median Problem is strategy-proof.*

Given the description of $\mathcal{X}$, determining a best a response is again separable. Therefore it suffices to show the result holds for the one dimensional case and Theorem 4.4.16 follows from Theorem 4.4.6.

Theorem 4.4.17. *Fix $\lambda \in \{0,1\}^k$ and let $\mathcal{X}$ be arbitrary. The λ-1-Median Problem is strategy-proof.*

Proof. Given the set $\mathcal{X}$ and preferences $\Pi \subseteq \mathcal{X}$, it suffices to show Π is a Nash equilibrium.

We begin by defining a second voting procedure by expanding $\mathcal{X}$. Let $l, u \in \mathbb{R}^k$ be such that $\mathcal{X} \subseteq \mathcal{X}' = \{x \in \mathbb{R}^k : l \leq x \leq u\}$. Consider the λ-1-Median Problem on the set $\mathcal{X}'$ with sincere preferences Π. By Theorem 4.4.16, Π is a Nash equilibrium implying $\pi_v \in \mathcal{X}$ is a best response for each voter v.

Since $\mathcal{X} \subseteq \mathcal{X}'$ and $\pi_v \in \mathcal{X}$ for all v, π_v must also be a best response in the original procedure with the smaller feasible set $\mathcal{X}$. Therefore Π is a Nash equilibrium in the original problem and the λ-1-Median Problem is strategy-proof regardless of $\mathcal{X}$. $\square$

Corollary 4.4.18. *Let $\mathcal{X}$ be arbitrary. When $|N|$ is odd, the* 1*-Median Problem is strategy-proof.*

As with Corollary 4.4.7, Corollary 4.4.18 immediately holds because there are never any ties when $|N|$ is odd.

4.4.2 1-Mean Problem

We now proceed to the 1-Mean Problem where the facility location $r(\Pi)$ is the unique minimizer of $\sum_{v=1}^{n} ||\pi_v - x||_2^2$. It is well known that $r(\Pi) = \sum_{v=1}^{n} \frac{\pi_v}{n}$.

Theorem 4.4.19. *Suppose voter v's cost of outcome x is $u_v(\pi_v, x) = ||\pi_v - x||_{p_v}$ for some $p_v \in (0, \infty)$ for each v. Then the Price of Deception of the 1-Mean Problem in $\mathbb{R}^k$ for ≥ 2 is ∞.*

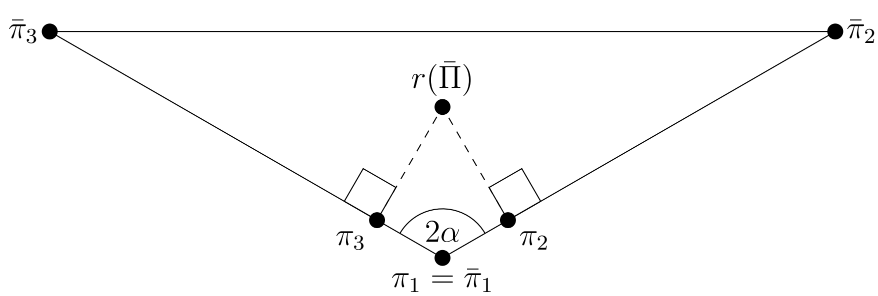


Figure 4.10: Preferences for Theorem 4.4.19

Proof. We show the result for $p_v = 2$ for all v and explain how to generalize the result at the end. Consider the three points $\bar{\pi}_1 = (0,0), \bar{\pi}_2 = (\frac{3}{2sin(\alpha)}, \frac{3}{2cos(\alpha)})$, and $\bar{\pi}_3 = (-\frac{3}{2sin(\alpha)}, \frac{3}{2cos(\alpha)})$ where $\alpha < \frac{\pi}{2}$. Define $\mathcal{X} = conv.hull(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$. Suppose the sincere facility locations are given by $\pi_1 = (0,0), \pi_2 = (cos(\alpha), sin(\alpha))$, and $\pi_3 = (-cos(\alpha), sin(\alpha))$

as shown in Figure 4.10. With respect to the sincere data, the facility would be placed at $r(\Pi) = (0, \frac{2sin(\alpha)}{3})$ with total distance (societal cost)

$$U(\Pi, r(\Pi)) = \sum_{i=1}^{3} ||r(\Pi) - \pi_i||_2^2 = \frac{2sin(\alpha)}{3} + 2\sqrt{1 - \frac{8sin^2(\alpha)}{9}} \leq 2 \tag{4.37}$$

Now consider the putative preferences given by $\bar{\Pi} = (\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$. With respect to $\bar{\Pi}$ the facility is placed at $r(\bar{\Pi}) = (0, \frac{1}{cos(\alpha)})$. With respect to the sincere preferences Π, this has a total cost of

$$U(\Pi, r(\bar{\Pi})) = \sum_{i=1}^{3} ||r(\bar{\Pi}) - \pi_i||_2^2 = \frac{1}{cos(\alpha)} + 2\sqrt{cos^2(\alpha) + \left(\frac{1}{cos(\alpha)} - sin(\alpha)\right)^2} \geq \frac{1}{cos(\alpha)}. \tag{4.38}$$

If $\bar{\Pi}$ corresponds to a minimally dishonest Nash equilibrium of Π, then the Price of Deception of this instance is

$$\frac{U(\Pi, r(\bar{\Pi}))}{U(\Pi, r(\Pi))} \geq \frac{1}{2cos(\alpha)} \to \infty \; as \; \alpha \to \frac{\pi}{2}. \tag{4.39}$$

It remains to show that $\bar{\Pi}$ is a minimally dishonest Nash equilibrium. To do this, it suffices to show that any change to $\bar{\pi}_i$ yields a worse outcome for individual i. Start with $i = 1$. Suppose player 1 changes her preferences to $\bar{\pi}_1' = \bar{\pi}_1 + d$ where $||d|| \neq 0$. The location $\bar{\pi}_1'$ is in $\mathcal{X}$ and therefore $d_2 > 0$. After updating her preferences the facility location is moved to $r([\bar{\Pi}_{-1}, \bar{\pi}_1']) = r(\bar{\Pi}) + \frac{1}{3}d$. This causes the facility to up and possibly to the left or the right. Regardless of p_1 this is worse for player 1 and therefore she is reporting a minimally dishonest best response.

The idea is similar to show players 2 and 3 are providing minimally dishonest best responses. By symmetry, it suffices to consider player 2. Suppose player 2 updates her preferences to $\bar{\pi}_2' = \bar{\pi}_2 - d$. Since $\bar{\pi}_2' \in \mathcal{X}$, $d \in cone\left((1,0),(\pi_2)\right)$. After updating her preferences, the facility will be located at $r([\bar{\Pi}_{-2}, \bar{\pi}_2']) = r(\bar{\Pi}) - \frac{1}{3}d$. Let $B_2 = \{x :$

$||x - \pi_2||_2 \leq ||\pi_2 - r(\bar{\Pi})||_2$ be the set of points player 2 prefers to $r(\bar{\Pi})$. By construction, $\{x \in \mathbb{R} : r(\bar{\Pi}) - \frac{1}{3}dx\}$ is tangent to B_2 as shown in Figure 4.11 hence for all d where $||d|| > 0$, reporting $\bar{\pi}_2 - d$ would yield a worse solution for player 2. Thus she has given a minimally dishonest best response. This implies that $\bar{\Pi}$ is a minimally dishonest Nash equilibrium and therefore the Price of Deception converges to ∞ as α approaches $\frac{\pi}{2}$.

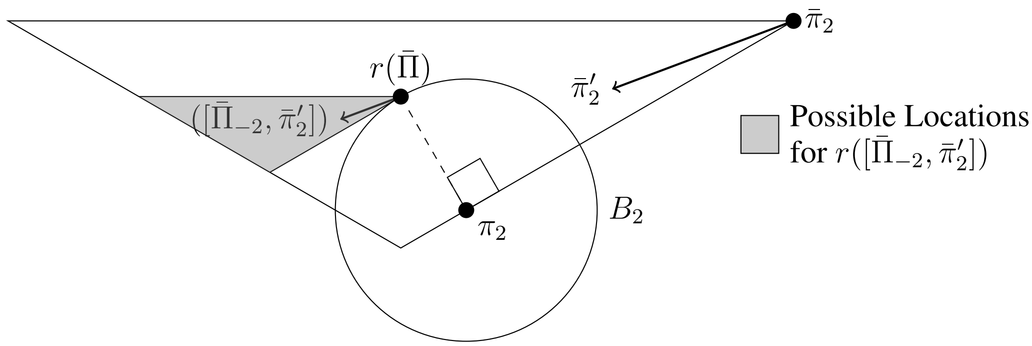


Figure 4.11: Possible Locations for $r([\bar{\Pi}_{-2}, \bar{\pi}'_2])$

We now consider other $p_v \in (0, \infty)$. As observed previously, player 1 is minimally dishonest. In the construction given in Figure 4.10, $\bar{\pi}_2$ is placed such $A = \{x \in \mathbb{R} : r(\bar{\Pi}) - \frac{1}{3}dx\}$ is tangent to B_2 ensuring that voter 2 cannot alter her preferences to get a better outcome. However, if $p_2 \neq 2$ then the non-euclidean ball $B = \{x : ||\pi_2 - x||_{p2} \leq ||\pi_2 - r(\bar{\Pi})||_{p2}\}$ may overlap with A. However, if Figure 4.11 is stretched horizontally, then the set A converges to a horizontal line. Therefore for any finite norm, we can stretch the set of allowed feasible locations such that no individual can alter their preferences to get a better outcome. $\square$

The proof of Theorem 4.4.19 suggests that the Price of Deception is dependent on the feasible region $\mathcal{X}$. To better understand the Price of Deception, we first aim to understand the possible locations for $r(\bar{\Pi})$.

Definition 4.4.20. Given $a \in \mathbb{R}^k$ and any compact $S \subseteq \mathcal{X}$, let $b(a, S) = max_{x \in S} a \cdot x$.

Definition 4.4.21. The vector $a \in \mathbb{R}^k$ is *limiting* for $S \subseteq \mathcal{X}$ if for each $x \in \mathcal{X}$ such that $a \cdot x > b(a, S)$ there an $\epsilon > 0$ such that $x - \epsilon a \in \mathcal{X}$.

Lemma 4.4.22. *Suppose voter v's cost of outcome x is $u_v(\pi_v, x) = ||\pi_v - x||_{p_v}$ for some $p_v \in (0, \infty)$ for each v. Let $\bar{\Pi}$ be a Nash equilibrium for Π in the 1-Mean Problem. If a is limiting for Π in $\mathcal{X}$, then $a \cdot r(\bar{\Pi}) \leq b(a, \Pi)$.*

Proof. Without loss of generality let a be a unit vector. For contradiction, suppose that $a \cdot r(\bar{\Pi}) = b(a, \Pi) + c$ for some $c > 0$. Then there must be at least one voter v such that $a \cdot \bar{\pi}_v > b(a, \Pi)$. By Definition 4.4.21 and by convexity of $\mathcal{X}$, there is an $\epsilon \in (0, c|N|]$ such that $\bar{\pi}'_v = \bar{\pi}_v - \epsilon a \in \mathcal{X}$. Moreover, if v instead submits $\bar{\pi}'_v$, then the facility location moves to $r = r(\bar{\Pi}) - \frac{\epsilon a}{|N|}$ where $b(a, \Pi) \leq a \cdot r < b(a, \Pi) + c$ by selection of ϵ. Therefore the new location r is strictly closer to every point in Π including π_v and therefore v is able to alter $\bar{\pi}_v$ to get a better outcome contradicting that $\bar{\Pi}$ is a Nash equilibrium. □

Example 4.4.23. Limiting Directions for a Rectangle

Suppose individuals must submit locations inside of the rectangle $\mathcal{X} = \{x : -2 \leq x_1 \leq 2, -1 \leq x_2 \leq 1\}$ and that $\Pi = \{(0,0), (1.5, 0), (0, .5)\}$ as shown in Figure 4.12.

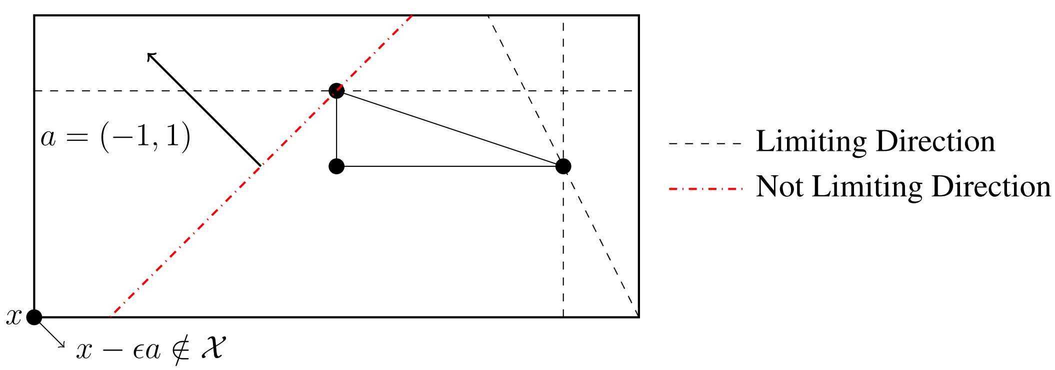


Figure 4.12: Voters' Sincere Locations and Some Limiting Directions for Example 4.4.23.

The direction $a = (-1, 1)$ is not limiting. Consider the point $x = (-2, -1)$. We cannot move from x orthogonally toward the dash-dotted line without leaving the set $\mathcal{X}$. Moreover, it is straightforward to verify that $(1, 0), (-1, 0), (0, 1)$ and $(0, -1)$ are limiting regardless of the of the sincere profile Π. In general, for any hyperrectangle $R = \{x \in \mathbb{R}^k : a_i \leq x_i \leq b_i \; \forall i = 1, ..., k\}$ has limiting directions $\pm e_i$ where e_i is the ith standard basis vector regardless of the sincere profile Π.

Lemma 4.4.22 is sufficient for us to find the Price of Deception for the most common set of feasible locations, the hyperrectangle.

Theorem 4.4.24. *Suppose voter v's cost of outcome x is $u_v(\pi_v, x) = ||\pi_v - x||_{p_v}$ for some $p_v \in (0, \infty)$ for each v and that the set of feasible facility locations is $\mathcal{X} = \{x \in \mathbb{R}^k : x_i \in [a_i, b_i]\ \forall i = 1, ..., k\}$ for all $j = 1, ..., n$. Then the Price of Deception of of the 1-Mean Problem with n individuals is $\Theta(n)$.*

Proof. First we show an upper bound of $2n$. Let $\bar{\Pi}$ be a Nash equilibrium for the sincere profile Π. Let $[a'_j, b'_j] \subseteq [a_j, b_j]$ be minimal such that $\mathcal{X}' = [a'_1, b'_1] \times ... \times [a'_k, b'_k]$ contains Π. Without loss of generality, we may assume $a'_j = 0$. As established, the decision mechanism selects $r(\Pi) = \sum_{i=1}^{n} \frac{\pi_i}{n}$ to minimize $\sum ||\pi_i - x||_2^2$ and therefore $r_j(\Pi) \in [0, b'_j]$ for all j. Denoting $r_j(\Pi)$ and π_{ij} as the jth coordinate of $r(\Pi)$ and π_i respectively, we have that

$$\sum_{i=1}^{n} ||\pi_i - r(\Pi)||_2^2 = \sum_{j=1}^{k} \sum_{i=1}^{n} (\pi_{ij} - r_j(\Pi))^2 \tag{4.40}$$

$$\geq \sum_{j=1}^{k} \frac{b_j'^2}{2} \tag{4.41}$$

and therefore the facility cost $\sum_{i=1}^{n} ||\pi_i - r(\Pi)||_2^2$ is at least $\sum_{j=1}^{k} \frac{b_j'^2}{2}$.

Given equilibrium $\bar{\Pi}$, the facility is located at $r(\bar{\Pi})$. In Example 4.4.23, we demonstrated that the standard basis vector e_j is limiting for Π and therefore $r_j(\bar{\Pi}) \in [0, b'_j]$ by Lemma 4.4.22. Therefore, the total cost for facility location $r(\bar{\Pi})$ is

$$\sum_{i=1}^{n} ||\pi_i - r(\bar{\Pi})||_2^2 = \sum_{j=1}^{k} \sum_{i=1}^{n} (\pi_{ij} - r_j(\bar{\Pi}))^2 \tag{4.42}$$

$$\leq \sum_{j=1}^{k} \sum_{i=1}^{n} b_j'^2 \tag{4.43}$$

$$= \sum_{j=1}^{k} n b_j'^2. \tag{4.44}$$

Therefore, the Price of Deception for sincere Π and submitted $\bar{\Pi}$ is

$$\frac{\sum_{i=1}^{n} ||\pi_i - r(\Pi)||_2^2}{\sum_{i=1}^{n} ||\pi_i - r(\Pi)||_2^2} \leq \frac{\sum_{i=1}^{k} nb_j'^2}{\sum_{i=1}^{k} \frac{b_j'^2}{2}} = 2n \tag{4.45}$$

We now present an instance with a Price of Deception n. Suppose we have n players and let $\mathcal{X}$ be the hypercube $\{x : 0 \leq x_i \leq n \ \forall i = 1, ..., k\}$. Let $\pi_1 = e_1$ and $\pi_i = \mathbf{0}$ for all $i > 1$. If everyone is sincere, then the facility is located at $r(\Pi) = (\frac{1}{n}, \mathbf{0})$ with a societal cost of

$$U(\Pi, r(\Pi)) = \sum_{i=1}^{n} ||\pi_i - r(\Pi)||_2^2 = \frac{(n-1)}{n} \tag{4.46}$$

Now consider the putative preferences $\bar{\Pi}$ where $\bar{\pi}_1 = ne_1$ and $\bar{\pi}_i = \pi_i$ for $i > 1$. With respect to these preferences, the facility is located at $r(\bar{\Pi}) = (1, \mathbf{0})$ for a sincere total cost of

$$U(\Pi, r(\bar{\Pi})) = \sum_{i=1}^{n} ||\pi_i - r(\bar{\Pi})||_2^2 = (n-1) \tag{4.47}$$

If $\bar{\Pi}$ is a minimally dishonest Nash equilibrium, then the Price of Deception of this instance is

$$\frac{U(\Pi, r(\bar{\Pi}))}{U(\Pi, r(\Pi))} = n \tag{4.48}$$

Showing that $\bar{\Pi}$ is a minimally dishonest Nash equilibrium follows in the same fashion as Theorem 4.4.19; if player i deviates from $\bar{\pi}_i$, then the facility moves further from π_i implying player i is minimally dishonest. Therefore the Price of Deception is between $\frac{n}{2}$ and n implying the Price of Deception is in $\Theta(n)$. $\square$

4.4.3 Minimizing L_2 Norm

Theorem 4.4.25. *Let $\mathcal{X} \subseteq \mathbb{R}$. The Price of Deception for facility location when minimizing the sum of L_2 norm distances is 1 when breaking ties deterministically and ∞ when breaking ties uniformly at random.*

Theorem 4.4.25 follows directly from Theorems 4.4.3 and 4.4.10 since the L_1 and L_2 norm are equivalent in $\mathbb{R}$.

Theorem 4.4.26. *Suppose voter v's outcome of x is $u_v = (\pi_v, x) = ||\pi_v - x||_{p_v}$ for some $p \in (0, \infty)$. Let $\mathcal{X} \subseteq \mathbb{R}^k$ for $k \geq 2$. The Price of Deception for facility location when minimizing the sum of L_2 norm distances is ∞ when breaking ties by selecting the center point or uniformly at random.*

Proof. It suffices to show this is true in $\mathbb{R}^2$. Let $n = 2c$ and suppose $\pi_i = (0, 0)$ for all i. If everyone is honest then the facility is located at $r(\Pi) = (0, 0)$ with a total cost of 0. Suppose instead that the players submit $\bar{\Pi}$ where $\bar{\pi}_i = (-1, 1)$ for $i = 1, ..., c$ and $\bar{\pi}_i = (1, 1)$ for $i = c - 1, ..., 2c$. With respect to these preferences, the facility is either located at $(0, 1)$ or uniformly at random between $(-1, 1)$ and $(1, 1)$ with a sincere cost of at least 1. Furthermore, if $\bar{\Pi}$ is a minimally dishonest Nash equilibrium, then the Price of Deception of this instance is ∞.

We now show $\bar{\Pi}$ is a minimally dishonest Nash equilibrium. It suffices to examine player 1. If player 1 instead submits $\bar{\pi}'_1$ where $\bar{\pi}'_{12} \neq \bar{\pi}_{12}$, then $r([\bar{\Pi}_{-1}, \bar{\pi}'_1]) = (1, 1)$ yielding a worse outcome for player 1. If $\bar{\pi}'_1$ is directly to the left of $\bar{\pi}_1$, then the outcome does not change and player 1 is less honest. If player 1 submits $\bar{\pi}'_1 = (w, 1)$ for some $w \in (-1, 1)$, then the facility is located either at $(\frac{1+w}{2} \geq 0, 1)$ or uniformly at random between $(w, 1)$ and $(1, 1)$. Both correspond to worse outcomes for player 1. Finally, if player 1 submits $\bar{\pi}'_1 = (w, 1)$ for some $w \geq 1$, then $r([\bar{\Pi}_{-1}, \bar{\pi}'_1]) = (1, 1)$ yielding a worse outcome for player 1. All possibilities yield either worse outcomes for player 1 or cause player 1 to be less honest without any benefit. Thus player 1 is giving the unique minimally dishonest best response and $\bar{\Pi}$ is a minimally dishonest Nash equilibrium. Therefore the Price of Deception when minimizing the L_2 norm is ∞. $\square$

5

PRICE OF DECEPTION IN ELECTIONS

5.1 Introduction

Arrow's possibility theorem [4] and its many descendants (see e.g. [5, 47]) tell us that to choose a voting rule is to trade off some desirable rationality properties against others because they are not mutually attainable. The Gibbard-Satterthwaite [28, 65], Gardenfors [27] and related theorems tell us that every election rule that one would consider to be reasonable is manipulable by a strategic voter. Therefore, we always sacrifice non-manipulability in lieu of other properties. But is it wise to take such an all-or-nothing position with respect to manipulability, especially since we always settle for nothing? To quantify the effect of manipulation in voting, we examine the game of deception known as the Strategic Voting Game.all

We first prove (Theorem 5.3.2) that the straightforward definition of the Game of Deception is not adequate because it permits arbitrarily bad but absurd outcomes. Branzei *et al.* [13] eliminate these spurious outcomes by examining only equilibria that can be obtained through an iterative process where each player selects their best response with respect to the previous iteration given that everyone was truthful in the first iteration. However, we prove (Theorem 5.3.4) that as the number of voters increases, the probability converges to 1 that this iterative process terminates at the first iteration. Hence that refinement cannot provide an accurate model since strategic behavior is known to occur in large populations.

To better understand human behavior we again incorporate our minimal dishonesty refinement, a plausible and experimentally supported refinement (see Section 2.3), into the players' Game of Deception. We then analyze the Price of Deception of several standard voting rules, including Borda count, approval voting, and plurality voting. Different rules turn out to have significantly different Prices of Deception. The results therefore support our proposal that the Price of Deception can help discriminate among different voting rules. For instance, plurality voting has the worst Price of Deception of all the voting rules we analyze, despite its widespread use in U.S.A. elections.

5.1.1 Lexicographic Tie-Breaking

We begin by analyzing voting procedures when ties are broken lexicographically. Table 5.1 summarizes our results for several voting mechanisms given $m > 2$ candidates.

Table 5.1: Prices of Deception for Various Voting Mechanisms.

	Price of Deception	
Voting Mechanism	Standard Scores	Normalized Scores
Dictatorship	1	1
Veto	$1 + \frac{1}{m-1}$	$1 + \frac{m-1}{m^2-m+1}$
Approval	2	$\frac{2m}{m+1}$
Borda Count	m	$[\frac{m+2}{3}, \frac{m^2}{2m-1}]$
Majority Judgment	$u - 1$	$\frac{um-2m+1}{u-1}$
α-Copeland	$[m-2, \infty]$	$[\frac{m-1}{2}, m]$
Plurality	∞	$\frac{2m+1}{3}$

The Price of Deception for each mechanism is given in the "Standard Scores" column of Table 5.1. For instance, in approval voting, if candidate c would win the election with 30 points when everyone is sincere, then the winner at any minimally dishonest equilibrium will have at least 15 points with respect to the sincere preferences. We also normalize each social utility for each voting rule so that $\min_{\Pi \in \mathcal{P}} \min_{c \in C} U(\Pi, c) = 1$ and $\max_{\Pi \in \mathcal{P}} \max_{c \in C} U(\Pi, c) = m$ for each pair of profile Π and candidate c via an affine transformation. Normalization enables us to make comparisons between otherwise dissimilar voting procedures.

With the exception of plurality voting, our results show a striking relationship between the use of information and the Price of Deception. Dictatorship employs one datum, the candidate most preferred by the dictator. Of the procedures analyzed, it ignores the most voter preference data and has the least Price of Deception. Veto voting relies on one datum per voter, the least preferred candidate. It uses the next smallest amount of information, and has the next lowest Price of Deception. Approval voting uses one bit of information per candidate from each voter, and has the next lowest Price of Deception. The next three procedures in Table 5.1 use full information and have substantially larger Prices of Deception, all of order m.

Plurality uses the same amount of information as veto voting, but has the largest Price of Deception. And indeed, plurality voting has received heavy criticism for encouraging strategic voting [16]. (It has also been criticized for limiting the number of political parties [21], wasted votes resulting in lower voter turnout, and being vulnerable to the spoiler effect (i.e. grossly violating independence of irrelevant alternatives (IIA)). We find it confirmatory that, of the voting mechanisms we have analyzed, plurality has the worst price of deception - a candidate is able to win at an equilibrium even if the candidate would receive no votes when everyone is honest. We also find it both ironic and confirmatory that plurality voting was employed in many state primary elections in the U.S. in 2016, amid many reports of strategic voting.

On the other hand, approval voting has been used by organizations such as the Mathematical Association of America, the Institute for Operations Research and the Management Sciences, the American Statistical Association, the Institute of Electrical and Electronics Engineers and the United Nations [56]. In addition, it is believed that approval voting improves voter turnout, deters the spoiler effect and reduces negative campaigning [12]. Of the voting mechanisms analyzed, approval voting has one of the lowest Prices of Deception.

Majority Judgment has been proposed by Balinski [7] as an alternative to voting in the sense that voters submit cardinal data rather than ordinal rankings. He shows that Majority

Judgment avoids many of the axiomatic pitfalls, such as IIA, that plague classical voting. It is pertinent, therefore, to evaluate its Price of Deception to see how it compares, in this sense, with classical voting rules. The result is that Majority Judgment has Price of Deception of the same order of magnitude as Borda and Copeland voting. Hence, by this measure, Majority Judgment is similar to some classic voting rules, but substantially poorer than approval voting.

5.1.2 Random Tie-Breaking

We established that mechanisms can have significantly different Prices of Deception and that we are able to discriminate between different mechanisms based upon how badly they can be manipulated. We also establish that making even a small change to a decision mechanism such as changing how ties are handled can significantly impact the Price of Deception.

For mechanisms such as veto, approval, Borda count with a risk-neutral society, and α-Copeland the Price of Deception remains unchanged when we switch from a lexicographic tie breaking rule to a random tie breaking rule. For each of these mechanisms the best lower bound for the Price of Deception was achieved when there was a unique candidate with a high score at equilibrium – no tie breaking rule was needed to determine the winner. However, as shown in Table 5.2, the Price of Deception can change depending on the tie breaking rule.

Table 5.2: Prices of Deception May Vary Depending on how Ties are Broken. $*$ Assumes Society is Risk-Neutral. † Denotes Normalized Price of Deception.

	Tie Breaking Rule		
Voting Mechanism	Lexico	Random	Unique Winner
Majority Judgment	$u-1$	$\max\{\frac{um-m}{u+m-2}, \frac{u-1}{2}\}^*$	$\frac{u-1}{2}$
Majority Judgment†	$\frac{um-2m+1}{u-1}$	$\max\{\frac{um^2-2m^2+m}{2um-u-3m+2}, \frac{um-2m+1}{u+m-2}\}^*$	$\frac{um-2m+1}{u+m-2}$
Plurality	∞	∞	∞
Plurality†	$\frac{2m+1}{3}$	m	$\frac{2m+1}{3}$

For sufficiently small m ($m \leq \frac{u-2}{\alpha-1}$), Majority Judgment when breaking ties lexicographically has a Price of Deception that is at least α times as bad when compared to breaking ties randomly. It is not surprising that lexicographic tie breaking has a worse Price of Deception; by definition, it is a biased mechanism that violates either neutrality or anonymity. Voters may be able to take advantage of the bias in the mechanism to yield a more extreme result.

In contrast, the Price of Deception for a plurality election is significantly better when ties are broken lexicographically. If ties are broken randomly, a candidate can lose the election even if the candidate is sincerely the unanimous favorite.

Not only does the type of tie breaking rule potentially have an impact on the Price of Deception, this impact is not always same. In one case (Majority Judgment) we see that lexicographic tie breaking leads to a worse Price of Deception. However, for plurality election, the exact opposite occurs – random tie breaking leads to a worse Price of Deception.

5.2 The Model

As in Chapters 3 and 4 we reinforce the notation and definitions from Chapter 2 by redefining them in context of the Voting Problem.

An instance of the Voting Problem consists of

- Set $V = \{v_1, ..., v_n\}$ and $C = \{c_1, ..., c_m\}$ of voters and candidates, respectively.
- For each $v \in V$, a total ordering π_v on C, representing voter v's preferences on the candidates. Denote by $\mathcal{P}_i$ as the set of all possible preferences. For the procedures Dictatorship, Veto, Borda Count, α-Copeland, and Plurality the preferences π_v are strict. For Approval voting, π_v is a strict ordering over a subset of C where $c \in \pi_v$ if v approves of candidate c. For Majority Judgment voting, $\pi_v = \{\pi_{v1}, ..., \pi_{vm}\}$ where $\pi_{vc} \in \{1, ..., u\}$ for some predefined u is voter v's valuation of candidate c.
- Denote by Π the collection of π_v over all $v \in V$. Π is called the *preference profile*.

- The profile Π is submitted to a publically known procedure r. The outcome $r(\Pi)$ corresponds to a single candidate $c \in C$ or a distribution over the set of candidates.

We begin by defining a large family of voting procedures that includes Borda, approval, plurality, and veto. These are procedures in which each voter assigns an integer score within a bounded range to each candidate, and a candidate with largest sum of scores wins the election.

Definition 5.2.1. For candidate set C, a **score voting procedure** bounded by K is a social choice procedure to select a winning candidate as follows: For each voter $v \in V$ there exists a score function $g_v : C \mapsto \{0, 1, \ldots, K-1\}$. The social choice is a member of the set

$$\arg\max_{c \in C} S_c \equiv \arg\max_{c \in C} \sum_{v \in V} g_v(c). \tag{5.1}$$

Denote by $\mathcal{G}$ the set of possible score functions[1]. Hence $|\mathcal{G}| \leq K^{|C|}$.

With exception of the α-Copeland voting procedure, every algorithm we analyze in this chapter is a score voting procedure. The α-Copeland winner is determined from a tournament graph where the weights on edges are determined by plurality, a score voting procedure.

Example 5.2.2. Borda count as a score voting procedure.

For Borda count, $g_v(c) = |C| - k$ *if* v*'s* k*th favorite candidate. For example, if voter* v*'s preference list is* $\pi_v = (c_1, c_2, c_3)$ *then* v*'s favorite candidate is* c_1 *followed by* c_2 *and then* c_3*. In this case,* $g_v(c_1) = 2, g_v(c_2) = 1$ *and* $g_v(c_3) = 0$*. Furthermore, if we have five voters,* v_1, v_2, v_3, v_4

[1]The score function g is not the same as the individual utility function u. We assume that $g_v(c) \geq g_v(c')$ if and only if $u_v(\pi_v, c) \geq u_v(\pi_v, c')$. However, g does not specify individual v's utility for a probability distribution over a subset of the candidates.

and v_5 with the following sincere preferences:

$$\pi_{v_1} = (c_1, c_2, c_3) \tag{5.2}$$

$$\pi_{v_2} = (c_1, c_2, c_3) \tag{5.3}$$

$$\pi_{v_3} = (c_3, c_1, c_2) \tag{5.4}$$

$$\pi_{v_4} = (c_3, c_1, c_2) \tag{5.5}$$

$$\pi_{v_5} = (c_3, c_1, c_2) \tag{5.6}$$

Then the score for candidate c_1 is

$$\sum_{i=1}^{5} g_{v_i}(c_1) = 2 + 2 + 1 + 1 + 1 = 7. \tag{5.7}$$

Similarly, the scores for candidates c_2 and c_3 are 2 and 6 respectively. Therefore if Π is submitted to a decision mechanism using Borda count, then c_1 is selected. In this example each voter assigns each candidate a value between 0 and $|C| - 1$. Therefore, in Definition 5.2.1, $K > |C| - 1 = 2$.

It is well known that every reasonable voting procedure is manipulable [28, 65, 27]. For instance in Example 5.2.2, voter v_4 could instead submit $\bar{\pi}_{v_4} = (c_3, c_2, c_1)$ causing candidates c_1 and c_3 to tie. Under certain tie-breaking rules, this yields a better outcome for voter v_4 since voter 4 prefers c_3 to c_1. As a result, we cannot expect that voters are submitting truthful information. The voters are actually submitting strategic preferences in the Strategic Voting Game.

Strategic Voting Game

- Each voter v has information $\pi_v \in \mathcal{P}_v$ describing their preferences. The collection of all information is the (sincere) profile $\Pi = \{\pi_v\}_{v=1}^{n}$.

- To play the game, voter v submits putative preference data $\bar{\pi}_v \in \mathcal{P}_v$. The collection of all submitted data is denoted $\bar{\Pi} = \{\bar{\pi}_v\}_{v=1}^n$.

- It is common knowledge that a central decision mechanism will select candidate $r(\bar{\Pi}, \omega) \in C$ when given input $\bar{\Pi}$ and random event ω.

- The random event $\omega \in \Omega$ is selected according to μ. We denote $r(\bar{\Pi})$ as the distribution of outcomes according to Ω and μ.

- Voter v evaluates $r(\bar{\Pi})$ according to v's sincere preferences π_v. Specifically, voter v's utility of the set of outcomes $r(\bar{\Pi})$ is $u_v(\pi_v, r(\bar{\Pi}))$.

Voter utilities are consistent with Π; $u_i(\pi_i, c) > u_i(\pi_i, c')$ only if c appears prior to c' in the preferences π_i. For single-valued voting rules (i.e. $r(\bar{\Pi}) = r(\bar{\Pi}, \omega) = r(\bar{\Pi}, \omega')$ for all $\omega, \omega' \in \Omega$) it suffices to consider only the preferences π_i and not the utilities u_i since they are consistent. However, when r is not single valued, u_i is necessary to determine a voter's valuation of probability distribution over a set of candidates. We begin by considering r single-valued.

By the definition of the strategic voting game, a set of submitted preferences $\bar{\Pi}$ forms a pure strategy Nash equilibrium if no individual v would obtain an outcome they sincerely prefer to $r(\bar{\Pi})$ (with respect to π_v) by altering $\bar{\pi}_v$.

Example 5.2.3. A Nash equilibrium of the Strategic Voting Game.

Suppose there are five voters, three candidates and the Borda count is used to select the winning candidate. Suppose further that the preferences from Example 5.2.2 are the sincere preferences of *the five voters. Then the following submitted preferences correspond to a Nash equilibrium:*

$$\bar{\pi}_{v_1} = (c_1, c_2, c_3) \tag{5.8}$$

$$\bar{\pi}_{v_2} = (c_1, c_2, c_3) \tag{5.9}$$

$$\bar{\pi}_{v_3} = (c_3, c_2, c_1) \tag{5.10}$$

$$\bar{\pi}_{v_4} = (c_3, c_2, c_1) \tag{5.11}$$

$$\bar{\pi}_{v_5} = (c_3, c_1, c_2) \tag{5.12}$$

Only voters v_3 and v_4's submitted preferences deviate from their sincere preferences. As a result, the scores for candidates c_1, c_2 and c_3 are 5, 4, and 6 respectively and candidate c_3 wins the election with respect to the submitted preferences. Voters v_3, v_4, and v_5 receive their sincere first choice and therefore have no further incentive to alter their submitted preferences. Though voters v_1 and v_2 receive their last choice, neither voter can alter the results of the election and the submitted preferences form a Nash equilibrium.

5.3 Strategic Voting Equilibria without Minimal Dishonesty

In Example 5.2.2, using Borda count would cause c_1 to win the election if everyone is sincere. However, in Example 5.2.3, we see that we may obtain a different result when individuals behave strategically; the strategic voters select candidate c_3, the sincere Condorcet winner. This equilibrium makes sense in that a majority of voters prefer c_3 to c_1 and a majority prefer c_3 to c_2. Furthermore, if all voters were honest, then both c_1 and c_3 have similar Borda counts – 7 and 6 respectively. However, we see in the next example that not all equilibria make sense.

Example 5.3.1. A Second Nash equilibrium of the Strategic Voting Game.

Once again suppose the sincere preferences are given in Example 5.2.2. Then the following preferences are a Nash equilibrium where c_2 wins the election.

$$\bar{\pi}_{v_1} = (c_2, c_1, c_3) \tag{5.13}$$

$$\bar{\pi}_{v_2} = (c_2, c_1, c_3) \tag{5.14}$$

$$\bar{\pi}_{v_3} = (c_2, c_3, c_1) \tag{5.15}$$

$$\bar{\pi}_{v_4} = (c_2, c_3, c_1) \tag{5.16}$$

$$\bar{\pi}_{v_5} = (c_2, c_3, c_1) \tag{5.17}$$

With respect to the submitted preferences candidates c_1,c_2 and c_3 have Borda counts of 2, 10 and 3 respectively and c_2 wins the election. These preferences correspond to an equilibrium since each voter can change a candidate's score by at most two.

Unlike Example 5.2.3, the voters in Example 5.3.1 are lying spuriously to obtain poor results – voters v_3, v_4 and v_5 all indicate that they most prefer their least preferred candidate, c_2, causing c_2 to win the election. We see in the next theorem that this oddity is unique to neither the preferences in Example 5.2.2 nor the Borda count.

Theorem 5.3.2. *For each of the non-dictatorial voting rules, Kemeny, Borda, approval, Dodgson, STV, and any Condorcet-consistent rule, there exist instances of the Strategic Voting Game in which:*

(i) there is a single candidate c_1 (respectively c_m) who is most (respectively least) preferred by every voter;

(ii) there is a pure strategy Nash equilibrium that elects c_m.

Proof. Let $C = \{c_1, c_2, \ldots, c_m\}$. Let $\pi_v = (c_1, c_2, \ldots, c_m)\ \forall v \in N$. Property (i) holds by construction. For any $m \geq 2$ and $|N| \geq m + 1$ unanimity implies that the submitted preference profile $\bar{\Pi}$ where $\bar{\pi}_v = (c_m, c_{m-1}, \ldots, c_1)\ \forall v$ yields social choice $r(\bar{\Pi}) = c_m$. Moreover, if any one voter v alters $\bar{\pi}_v$, c_m remains the Condorcet winner, the Kemeny winner, the Borda winner, etc. Therefore, $\bar{\Pi}$ corresponds to a pure Nash equilibrium of the strategic voting game. Incidentally, by unanimity $r(\Pi) = c_1$ and hence Π corresponds to a pure Nash equilibrium of the game, though that is not needed for the proof. $\square$

The Nash equilibrium strategy set $\bar{\Pi}$ in the proof of Theorem 5.3.2 is absurd because the voters are dishonest to their own disbenefit. It is not plausible that voters would lie in order to achieve a less preferable outcome. The strategic voting game admits many other nonsensical pure strategy equilibria. For example, most submitted profiles that are unanimous yield Nash equilibria. These observations call for a refinement concept to remove such equilibria.

5.3.1 Dynamic Cournot Equilibria

Branzei *et al.* [13] propose to refine the set of equilibria by retaining only pure strategy equilibria reachable from sincere voting by a sequence of best responses of the voters. We refer to these as dynamic Cournot equilibria. This refinement does eliminate the absurd equilibrium in Example 5.3.1. However, it also eliminates the equilibrium in Example 5.2.3. With respect to the sincere preferences in Example 5.2.2, everyone is honest in the only dynamic Cournot equilibrium.

Example 5.3.3. A Dynamic Cournot Equilibrium

Consider the sincere preferences from Example 5.2.2.

$$\pi_{v_1} = (c_1, c_2, c_3) \tag{5.18}$$

$$\pi_{v_2} = (c_1, c_2, c_3) \tag{5.19}$$

$$\pi_{v_3} = (c_3, c_1, c_2) \tag{5.20}$$

$$\pi_{v_4} = (c_3, c_1, c_2) \tag{5.21}$$

$$\pi_{v_5} = (c_3, c_1, c_2) \tag{5.22}$$

Suppose further that we select the winning candidate using Borda count and break ties by selecting the best candidate with the lowest index. With respect to the sincere preferences the Borda counts for candidates c_1, c_2, and c_3 are 7, 2, and 6 respectively. Candidate c_1 wins the election and neither v_1 nor v_2 have a desire to change the outcome of the election. The remaining voters prefer c_3. However, no individual can change the result of the election; v_3 cannot increase the score of c_3 and can only decrease the score of c_1 by one which does not cause c_3 to win the election. Thus, voters have no incentive to deviate from their sincere preferences and the sincere preferences are the only dynamic Cournot equilibrium.

Example 5.3.3 demonstrates that the dynamic Cournot refinement is too stringent. We show in Theorem 5.3.4 and Corollary 5.3.5 that, with probability converging to 1, only one dynamic Cournot equilibrium exists, and in that equilibrium all voters are truthful. Hence the restriction to dynamic Cournot equilibria is, with high probability, tantamount to forbidding manipulation. Definition 5.2.1 puts no requirement on how ties are broken. To detect unavoidable ties, we say voter v *distinguishes* candidates c, d is iff $g_v(c) \neq g_v(d)$.

Theorem 5.3.4. *Let a score voting procedure bounded by K be employed to select from candidate set C. Let p be any probability distribution on $\mathcal{G}$, the set of possible score functions, such that every pair of candidates is distinguishable with positive probability. Let a population of n voters be sampled independently according to p. Then as $n \to \infty$ the probability converges to 1 that the only dynamic Cournot equilibrium is for all voters to be truthful.*

Proof: Let random variable S_c be the score of $c \in C$. By assumption $E[\frac{S_c}{n}] = \sum_{g \in \mathcal{G}} p(g)g(c)$. Hence if the candidate pair c, d is distinguishable there must exist $g \in \mathcal{G}$ such that $p(g) > 0$ and $g(c) \neq g(d)$. By Lemma 5.6.1 with $\epsilon = \frac{1}{2}$,

$$\forall k \ \lim_{n\to\infty} P(|S_c - S_d| < k) = 0. \tag{5.23}$$

The proof of the lemma is deferred to the appendix. The idea is that if $E[\frac{S_c}{n}] \neq E[\frac{S_d}{n}]$

then $|S_c - S_d] \to \infty$ with probability 1 by the law of large numbers. If the expected values are equal, $S_c - S_d$ is the sum of n i.i.d. non-trivial random variables with mean 0, which by the central limit theorem has probability tending to zero of being within order less than $1/\sqrt{n}$ of 0.

Set $k = 2K$ to conclude

$$\lim_{n\to\infty} P\left(\bigcup_{c_1 \neq c_2 \in C} |S_{c_1} - S_{c_2}| \geq k\right) \leq \sum_{c_1 \neq c_2 \in C} \lim_{n\to\infty} P(|S_{c_1} - S_{c_2}| \geq 2K) = 0. \qquad (5.24)$$

Therefore, if all voters are truthful, the probability converges to 1 that there is a unique winning candidate who wins by at least $2K$.

A single voter can change $S_c - S_d$ by at most $2K - 2$. (In fact w.p. 1 no coalition of order $n^{.5-\epsilon}$ voters could change the outcome.) Therefore, if all voters are truthful, the probability converges to 1 that the outcome is a Nash equilibrium, which therefore is the unique dynamic Cournot equilibrium. □

We now extend Theorem 5.3.4 to voting procedures that are based the outcomes of pairwise elections. A Condorcet winner is a candidate who is preferred to every other candidate by more than half of the voters. The principle is often accepted that the Condorcet winner, if it exists, should be the social choice. Several voting procedures take this principle as a point of departure and thereby operate on the tournament graph on vertex set C where a directed edge (c, d) exists iff c defeats d and an undirected edge $\{c, d\}$ exists iff c and d tie, i.e. are preferred by an equal number of voters. For example, in α-Copeland scoring candidate c gets 1 point for each outgoing edge and $0 \leq \alpha < 1$ points for each incident undirected edge.

Corollary 5.3.5. *Theorem 5.3.4 holds for social choice rules that select a winning subset of the candidates based only on the tournament graph, and perform tie-breaking within that subset either independent of V or based on a scoring function according to Definition 5.2.1.*

Proof: A pairwise election between $c, d \in C$ is equivalent to a score voting procedure with $K = 2$ and $g_v(c) + g_v(d) \leq 2 \, \forall v \in N$. By Lemma 5.6.1 $P(|S_c - S_d| \leq 2) \to 0$ as $n \to \infty$. Hence the probability of a pairwise tie goes to zero. Moreover, since a single voter can only change the difference of the scores within a pairwise election by 2, the probability goes to 0 that any voter can alter the tournament graph. If tie-breaking is based on a scoring function, then by Theorem 5.3.4 the probability goes to 0 that any voter can change the tie-breaking outcome. $\square$

5.3.2 Minimally Dishonest Voting and the Price of Deception

Theorem 5.3.2 shows that we must refine the set of equilibria of the strategic voting game. Theorem 5.3.4 and its corollary show that the dynamic Cournot refinement of Branzei *et al.* [13] is too restrictive. We propose a different refinement based on the hypothesis that although people may lie for a purpose, they do not lie spuriously. It follows that a voter would lie "just enough" to achieve a more-preferred outcome, but would not lie more than necessary to achieve it. We call such behavior *minimal dishonesty* as defined in Section 2.3. We begin by measuring honesty.

Definition 5.3.6 (Bubble Sort or Kendall Tau Distance)**.** For all $c_i, c_j \in C$, denote by $R(\pi, c_i, c_j)$ the relationship between c_i and c_j with respect to π. The the distance between π_1 and π_2 is defined as

$$d(\pi_1, \pi_2) := |\{i, j\} : 1 \leq i < j \leq |C|; R(\pi_1, c_i, c_j) \neq R(\pi_2, c_i, c_j)\}|. \tag{5.25}$$

Example 5.3.7. Calculating $d(\pi_1, \pi_2)$.

In Example 5.3.1, voter v_3's sincere prefererences were $\pi_{v_3} = (c_3, c_1, c_2)$. With respect to the sincere preferences voter v_3 prefers c_1 to c_2 (i.e. $c_1 \succ c_2$ and $R(\pi_{v_3}, c_1, c_2) = \{\succ\}$). However, with respect to the submitted preferences list $\pi'_{v_3} = (c_2, c_3, c_1)$, voter v_3 indicates they prefer c_2 to c_1 and $R(\pi'_{v_3}, c_1, c_2) = \{\prec\} \neq R(\pi_{v_3}, c_1, c_2)$. Similarly, $R(\pi'_{v_3}, c_2, c_3) \neq R(\pi_{v_3}, c_2, c_3)$ and

$d(\pi_{v_3}, \pi'_{v_3}) = 2$.

Definition 5.3.8. Let Π be the sincere preferences and let $\bar{\Pi}$ be a Nash equilibrium in the Strategic Voting Game. A voter $v \in V$ is *minimally dishonest* if $d(\bar{\pi}'_v, \pi_v) < d(\bar{\pi}_v, \pi_v)$ implies that v's utility of the more honest outcome, $u_v(\pi_v, r([\bar{\Pi}_{-v}, \bar{\pi}'_v]))$, is less than v's utility of the equilibrium outcome, $u_v(\pi_v, r(\bar{\Pi}))$.

A pure strategy Nash equilibrium is minimally dishonest if every voter is minimally dishonest.

Stated in the contrapositive, players are not minimally dishonest if a small change to their purported preference $\bar{\pi}$ is more truthful and yields at least as (truthfully) preferred an outcome.

We can now precisely define the Price of Deception in strategic voting games.

Definition 5.3.9. Let r be the single-valued voting rule of a strategic voting game. Let U be an associated real-valued function that, given a profile Π of voter preferences over C, outputs for each $c \in C$ a *societal utility* $U(\Pi, c)$ of candidate c respect to Π. The function r must be such that for all Π, $r(\Pi) \in \arg\max_{c \in C} U(\Pi, c)$. When voting is characterized as a score voting procedure (according to Definition 5.2.1) $U(\Pi, c)$ is taken to be the candidate score $S_c^{\Pi} \equiv \sum_{v \in V} g_v^{\Pi}(c)$.

Let $NE(\Pi)$ denote the set of purported preference profiles of the minimally dishonest equilibria of the strategic voting game. Then the Price of Deception is

$$\max_{\bar{\Pi} \in NE(\Pi)} \frac{U(\Pi, r(\Pi))}{U(\Pi, r(\bar{\Pi}))}. \tag{5.26}$$

Normalized Price of Deception

The Price of Deception as defined is not a fine enough measure to discriminate among scenarios with unbounded Price of Deception. For example, suppose that the sincere winner gets 40% of the votes in one plurality election and gets 20% of the votes in another. Suppose in both elections there is a strategic equilibrium in which the winner is no voter's top choice. The Price of Deception is equally bad – infinity – in both elections, yet we might wish it to be measured as worse in the first election. In Section 3.3 we handle this by bounding how close preferences can get to zero. However, our voting rules are integer valued and therefore such a trick will not work.

Another potential shortcoming of the Price of Deception is that its value can be greatly altered by changing a score's scale. For instance, in a plurality election it is possible for a winning candidate in the strategic voting game to be no voter's first choice. The Price of Deception can be $1/0 = \infty$. Create a new decision mechanism, plurality+, where a candidate receives 2 points for being most favored by a voter, and 1 point from the voter otherwise. Plurality and plurality+ always yield the same outcome, yet they have significantly different Prices of Deception. With respect to plurality+, every candidate receives between n and $2n$ votes and the Price of Deception is at most 2.

To address both of these issues, we can normalize the Price of Deception so that its value is at most m for all voting mechanisms. We apply an affine transformation to the scoring function such that each candidate receives at least 1 and at most m points. While Price of Deception without normalization gives a pure measure of how much a voting rule can be manipulated, we believe that normalization is appropriate for comparing Prices of Deception of different mechanisms. For example, plurality and plurality+ have the same Prices of Deception after normalization.

5.4 Prices of Deception for Voting Rules with Lexicographic-Tie Breaking

This section derives exact values or estimates of the Price of Deception of several well-known voting procedures, and also the Price of Deception of Balinski's recently proposed "majority judgement" procedure [7]. We include proofs of several lower and upper bounds that illustrate the ideas and types of reasoning employed. We defer to Section 5.6 proofs that are technical but require no additional concepts.

Since r is single-valued, the winning candidate, $r(\Pi) \in \operatorname{argmax}_{c \in C} U(\Pi, c)$, is uniquely selected even though their may be multiple candidates that maximize social utility. We assume that ties are broken lexicographically according to some predetermined list. For each voting mechanism the predetermined list is described and an example is given to demonstrate the lower bound for the Price of Deception. Throughout this section, $m = |C|$ denotes the number of candidates and n denotes the number of voters.

5.4.1 Dictatorship

Since dictatorship is immune to strategic voting, its Price of Deception equals 1.

5.4.2 Veto

The veto rule is a score voting procedure in which each voter gives zero points to their least preferred candidate and one point to all other candidates. The winning candidate has the highest number of points. Given n voters and m candidates, the score for each candidate is between 0 and n and each voter gives out $m-1$ points.

Theorem 5.4.1. *The Price of Deception for veto voting is* $1 + \frac{1}{m-1}$.

Proof: First, we show that the Price of Deception is at most $1 + \frac{1}{m-1}$. Suppose that at a minimally dishonest equilibrium candidate c_m is the winning candidate. Let A be the set of voters that would veto c_m if everyone was sincere. The sincere score for c_m is $n - |A|$.

For all $v \in A$, at the equilibrium, v must veto c_m. If not, v is not minimally dishonest since v can be more honest and veto c_m which either causes c_m to lose the election, a strict improvement for v, or has no result on the outcome.

Therefore, the equilibrium score for c_m is at most $n - |A|$. For c_m to win the election, c_m must receive at least the average number of votes. Hence $n - |A| \geq \frac{n(m-1)}{m}$ implying that $|A| \leq \frac{n}{m}$ and that c_m's sincere score is at least $n - \frac{n}{m}$. Since the sincere winner can receive at most n points, the Price of Deception is at most $\frac{n}{n - \frac{n}{m}} = 1 + \frac{1}{m-1}$.

Finally, we give an example showing that the Price of Deception is at least $1+\frac{1}{m-1}$. The predetermined list for tie breaking is such that if c_1 and $c_j \neq c_1$ both have the same score, then c_1 will not be the winner of the election. There are m groups of voters $A_1, ..., A_m$. Let k be a multiple of m and consider the following preference profiles for each group:

$$\pi_i = (c_2, ..., c_m, c_1) \quad \forall i \in A_1 \ where \ |A_1| = \frac{k}{m} \tag{5.27}$$

$$\pi_i^j = (c_1, c_2, ..., c_m) \quad \forall i \in A_j \ where \ |A_j| = \frac{k}{m} + 1 \ \forall j \in \{2, ..., m\}. \tag{5.28}$$

If everyone is sincere, c_2 through c_{m-1} tie for first place with $k + m - 1$ points and candidate c_1 would receive $\left(1 - \frac{1}{m}\right) k + m - 1$ points. We now give a set of submitted preferences describing a minimally dishonest equilibrium where c_1 is the unique winner giving a Price of Deception of $\frac{k+m-1}{\left(1-\frac{1}{m}\right)k+m-1}$ which approaches $1 + \frac{1}{m-1}$ as k grows large.

All voters in group i will veto candidate c_i. Since A_1 is the smallest set, c_1 is the winner in this equilibrium. Every voter in A_1 is honest and cannot cause c_1 to lose points and cannot cause any other candidate to gain points. Therefore no one in A_1 can alter their submitted preferences to be more honest or get a better result. Every voter in A_m is honest and their favorite candidate is the unique winner therefore no one in A_m has incentive to change their submitted preference. For all other i, their favorite candidate is the unique winner, so only honesty could motivate them to alter their vote. If voter $v \in A_i$ is more honest, then candidate c_i will tie with c_1 resulting in a worse outcome for v, who prefers c_1 to all other candidates. Therefore these preferences describe a minimally dishonest equilibrium and the Price of Deception can be arbitrarily close to $1 + \frac{1}{m-1}$. □

To normalize the score of any voting mechanism, we need to apply a affine transformation to the points given by the voters such that the ratio of the maximum to minimum number of total points a candidate can receive is m. For veto, this is achieved by having a voter give $\frac{1}{n}$ points to the candidate whom they veto and $\frac{m}{n}$ points to all other candidates.

Theorem 5.4.2. *The normalized Price of Deception for veto voting is* $1 + \frac{m-1}{m^2-m+1}$.

Proof: This proof follows in the same fashion as the proof for Theorem 5.4.1. As before, $|A| \leq \frac{n}{m}$ and, updating the scores in the upper bound, the Price of Deception is at most $\frac{n \cdot \frac{m}{n}}{(n-\frac{n}{m}) \cdot \frac{m}{n} + \frac{n}{m} \cdot \frac{1}{n}} = 1 + \frac{m-1}{m^2-m+1}$. Similarly, updating the scores in the lower bound, the Price of Deception is at least $\frac{m^2}{(m-1)^2(1+\frac{1}{k+m-1}) + m}$ which approaches $1 + \frac{m-1}{m^2-m+1}$ as k grows large. □

5.4.3 Approval

Approval voting is a score voting procedure in which candidate c receives one point from every voter who approves of c. As with veto voting, the winning candidate has the highest number of points. For n voters and m candidates, each candidates receives between 0 and n points. We assume that voter v has no preferences over the set of candidates whom v does not approve of.

Theorem 5.4.3. *The Price of Deception for approval voting is* 2*.*

Proof: We begin by giving the upper bound. Suppose that c_1 would get the most points if everyone were sincere and that c_m wins the election at a minimally dishonest equilibrium. Let A and B be the set of individuals who sincerely approve of c_1 and c_m respectively. The Price of Deception is $\frac{|A|}{|B|}$.

We claim that for all $v \in A \setminus B$, v declares approval of c_1 at the equilibrium. If not, consider the consequences of v being more honest by declaring approval of c_1. This either has no effect on the outcome or causes c_1 to win the election. Thus, by minimal dishonesty, the claim holds. By a symmetric argument, for all $v \notin B$, v declares disapproval of c_m at the equilibrium.

This implies that c_1 receives at least $|A \setminus B|$ points and c_m receives at most $|B|$ points at an equilibrium. Since c_m receives the most points at an equilibrium, $|B| \geq |A \setminus B| \geq |A| - |B|$ which implies that $|B| \geq \frac{|A|}{2}$ and that the Price of Deception is at most 2.

Finally, we show by construction that the Price of Deception is at least 2. Consider $k+1$ voters who prefer c_m to c_1 and who disapprove of all other candidates and k voters who only approve of c_1. If voters were sincere, c_1 would receive $2k$ points and c_m would receive k points. We now give an equilibrium where c_m wins the election yielding a Price of Deception of $\frac{2k+1}{k+1}$ which approaches 2 as k grows large.

Each of the first $k+1$ voters purports to approve only of c_m. The remaining k voters are honest. The first k voters receive their first choice and if they are more honest then they receive a strictly worse outcome. The remaining k voters are honest and cannot change the result of the election. Therefore these submitted preferences form a minimally dishonest equilibrium and the Price of Deception is at least 2. $\square$

Since the maximum and minimum scores for any candidate are n and 0 respectively, the scoring function for approval voting is not normalized. We normalize by giving $\frac{m}{n}$ points to candidate c_i for each voter who approves of c_i, and $\frac{1}{n}$ points for each voter who does not approve. Each candidate now receives between 1 and m points.

Theorem 5.4.4. *The normalized Price of Deception for approval voting is* $\frac{2m}{m+1}$.

Proof: Once again this proof follows in the same fashion as the previous proof. Updating the upper bound from Theorem 5.4.3, the Price of Deception is at most $\frac{2m}{m+1}$. Updating the lower bound, the Price of Deception is at least $\frac{(2k+1)m}{(k+1)m+k}$ which approaches $\frac{2m}{m+1}$ as k grows large. $\square$

5.4.4 Borda Count

Borda count is a score voting procedure where $g_v(c) = |C| - k$ if c is v's kth favorite candidate. The winning candidate is a candidate that receives the most points. Given n voters and m candidates, the points for each candidate are between 0 and $n(m-1)$ and each voter gives out $\binom{m}{2}$ points.

Theorem 5.4.5. *The Price of Deception for Borda count voting is* m.

Proof: Suppose $c_1 = r(\Pi)$ and that $c_2 = r(\bar{\Pi})$ is the winner at an equilibrium. Let x_{ij} be the number of voters who would give $i-1$ points to c_1 and $j-1$ points to c_2 if they were honest for $i, j \in \{1, ..., m\}$. The sincere scores for the candidates are then given by $S_{c_1}^{\Pi} = \sum_{i \neq j}(i-1)x_{ij}$ and $S_{c_2}^{\Pi} = \sum_{i \neq j}(j-1)x_{ij}$ and the Price of Deception is

$$\frac{S_{c_1}^{\Pi}}{S_{c_2}^{\Pi}} = \frac{\sum_{i \neq j}(i-1)x_{ij}}{\sum_{i \neq j}(j-1)x_{ij}}. \tag{5.29}$$

Due to minimal dishonesty, voter v is honest if c_2 is v's sincerely least preferred candidate. Because of this and Lemma 5.6.2 in Section 5.6.1, the following holds for the equilibrium scores, $S_{c_1}^{\Pi'}$ and $S_{c_2}^{\bar{\Pi}}$, for c_1 and c_2 respectively:

$$S_{c_1}^{\bar{\Pi}} \geq \sum_{i>j>1}(i-2)x_{ij} + \sum_{i=2}^{m}(i-1)x_{i1}, \tag{5.30}$$

$$S_{c_2}^{\bar{\Pi}} \leq \sum_{j>1}\sum_{i \neq j}(m-1)x_{ij}. \tag{5.31}$$

Since $S_{c_2}^{\bar{\Pi}} \geq S_{c_1}^{\bar{\Pi}}$, the following mathematical program gives an upper bound on the Price of Deception:

$$\max z = \frac{\sum_{i \neq j}(i-1)x_{ij}}{\sum_{i \neq j}(j-1)x_{ij}} \tag{5.32}$$

$$\text{subject to: } \sum_{i>j>1}(i-m-1)x_{ij} + \sum_{j>i}(1-m)x_{ij} + \sum_{i=2}^{m}(i-1)x_{i1} \leq 0 \tag{5.33}$$

$$x_{ij} \in \mathbb{Z}_{\geq 0}. \tag{5.34}$$

This mathematical program can be relaxed to the following linear program:

$$\max z' = \sum_{i \neq j} (i-1)x_{ij} \tag{5.35}$$

$$\text{subject to: } \sum_{i>j>1} (i-m-1)x_{ij} + \sum_{j>i} (1-m)x_{ij} + \sum_{i=2}^{m} (i-1)x_{i1} \leq 0 \tag{5.36}$$

$$\sum_{i \neq j} (j-1)x_{ij} = 1 \tag{5.37}$$

$$x_{ij} \in \mathbb{R}_{\geq 0}. \tag{5.38}$$

The dual to this linear program is

$$\min w' = y_2 \tag{5.39}$$

$$subject\ to: (i-m-1)y_1 \quad +(j-1)y_2 \geq i-1 \qquad \forall\, i > j > 1 \tag{5.40}$$

$$(1-m)y_1 \quad +(j-1)y_2 \geq i-1 \qquad \forall\, j > i \tag{5.41}$$

$$(i-1)y_1 \quad \geq i-1 \qquad \forall i = 2, ..., m \tag{5.42}$$

$$y_1 \in \mathbb{R}_{\geq 0}. \tag{5.43}$$

A feasible solution to the dual is $y_1 = 1$ and $y_2 = m$ with objective value m. By weak duality [66], for every feasible z' and w', $z' \leq w'$ and the Price of Deception is at most m.

To realize the lower bound we give three groups of voters, A_1 A_2 and A_3, with sincere preferences Π. The predetermined list for tie breaking is such that if c_2 and $c_j \neq c_2$ have the same score, then c_2 will not win the election.

$$\pi_i = (c_1, c_3, c_4, ..., c_m, c_2) \qquad \forall i \in A_1\ where\ |A_1| = n \tag{5.44}$$

$$\pi_i = (c_m, c_{m-1}, ..., c_1) \qquad \forall i \in A_2\ where\ |A_2| = n-1 \tag{5.45}$$

$$\pi_i = (c_2, c_1, c_m, c_{m-1}, ..., c_3) \qquad \forall i \in A_3\ where\ |A_3| = m \tag{5.46}$$

For sufficiently large n, $S^{\Pi}_{c_i} < S^{\Pi}_{c_m} = nm + m^2 - 4m + 1$ for all $i \neq m$ and $S^{\Pi}_{c_2} =$

$n - 1 + m^2 - m$. If c_2 is the winner at a Nash equilibrium, then the Price of Deception of Borda count voting is at least $\frac{nm+m^2-4m+1}{n-1+m^2-m}$ which approaches m as n grows large. We claim that the following preferences $\bar{\Pi}$ form a minimally dishonest equilibrium:

$$\bar{\pi}_i = (c_1, c_3, c_4, ..., c_m, c_2) \ (honest) \qquad \forall i \in A_1 \tag{5.47}$$

$$\bar{\pi}_i = (c_2, c_m, c_{m-1}, ..., c_3, c_1) \qquad \forall i \in A_2 \tag{5.48}$$

$$\bar{\pi}_i = (c_2, c_1, c_m, c_{m-1}, ..., c_3) \ (honest) \qquad \forall i \in A_3 \tag{5.49}$$

For all $i \geq 3$, $S^{\bar{\Pi}}_{c_i} = nm - n + 2 - i + (m-1)i - 3m \leq nm - n + 2 + m^2 - 4m$ and $S^{\bar{\Pi}}_{c_2} = S^{\bar{\Pi}}_{c_1} + 1 = nm + m^2 - 2m - n + 1$ and thus c_2 wins the election. Candidate c_2 defeats c_1 by one point and all other candidates by at least $2m - 1$ points. Since any single voter can cause a candidate's score to change by at most $m - 1$, no candidate can cause c_i to win for $i \in \{3, ..., m\}$. Furthermore, voters in A_1 can neither increase c_1's score nor decrease c_2's score. Therefore, no one can alter their preferences to get a strictly better result. Finally, if a voter $v \in A_2$ is more honest, then it will cause c_1 to at least tie which results in a worse outcome for v. Thus, the Price of Deception is at least m. $\square$

To normalize the Borda count, we let $\bar{g}_v(c) = |C| - k + 1$ if c is v's kth favorite candidate. The points for each candidate are now between n and nm and the Price of Deception is trivially at most m.

Theorem 5.4.6. *The normalized Price of Deception for Borda count voting is between* $\frac{m+2}{3}$ *and* $\frac{m^2}{2m-1}$.

When updating Theorem 5.4.5, the lower bound becomes $\frac{m+2}{3}$. When updating the proof for the upper bound, since all scores increase by one point, only the second column and right hand sides will change. They will both increase by one and a new dual feasible solution is $y_1 = \frac{m}{2m-1}$ and $y_2 = \frac{m^2}{2m-1}$ and the normalized Price of Deception is at most $\frac{m^2}{2m-1}$.

5.4.5 Majority Judgment

Majority Judgment [7] is a scoring social choice mechanism where each voter v has a sincere valuation of candidate c given by $g_v^{\Pi}(c) \in \{1, ..., u\}$. As before, voter v may act strategically and submit a value of $g_v^{\bar{\Pi}}(c) \in \{1, ..., u\}$ for candidate instead. The winning candidate is a candidate with the highest median score. The highest and lowest score for any candidate is u and 1 respectively.

Unlike other mechanisms analyzed in this paper, Majority Judgment does not rely on ordinal information from the voters. As a result, we must alter the definition of minimal dishonesty for this mechanism. A voter v is minimally dishonest if for every candidate c and for every $x \in \{1, ..., u\}$ where $|g_v^{\Pi}(c) - x| < |g_v^{\Pi}(c) - g_v^{\bar{\Pi}}(c)|$, v prefers the outcome obtained by submitting a value of $g_v^{\Pi'}(c)$ for candidate c over the outcome obtained by submitting a value of x for candidate c.

Theorem 5.4.7. *The Price of Deception for Majority Judgment is* $u - 1$.

Proof: If any voter has a sincere value of u for candidate c, then voter v will be honest about c. Similarly, if a voter has a sincere value of 1 for candidate c, then v will be honest about c. Thus, if candidate c has a sincere median score of u (1), then c will also have a median score of u (1 respectively) at equilibrium. Therefore if the sincere winner has a sincere score of u, then then equilibrium winner must have a sincere median score of at least 2 and the Price of Deception is at most $\frac{u}{2}$. Alternatively, the sincere winner may have a sincere score of $u - 1$, the equilibrium winner may have a score of 1 and the Price of Deception is at most $u - 1$.

We now show that the Price of Deception is at least $u - 1$. Let $g_v^{\Pi}(c)$ be v's sincere value for candidate c and $g_v^{\bar{\Pi}}(c)$ be v's submitted value for candidate c at an equilibrium. The predetermined list for tie breaking is such that if c_1 has the highest median score, then c_1 will win the election. Consider three groups of voters, A_1, A_2, and A_3. Their sincere and submitted preferences are as follows:

$$v \in A_1: \quad g_v^{\Pi}(c_1) = g_v^{\bar{\Pi}}(c_1) = 1, \quad g_v^{\Pi}(c_i) = g_v^{\bar{\Pi}}(c_i) = u, \tag{5.50}$$

$$v \in A_2: \quad g_v^{\Pi}(c_1) = g_v^{\bar{\Pi}}(c_1) = 1, \quad g_v^{\Pi}(c_i) = g_v^{\bar{\Pi}}(c_i) = 1, \tag{5.51}$$

$$v = A_3: \quad g_v^{\Pi}(c_1) = g_v^{\bar{\Pi}}(c_1) = u, \quad g_v^{\Pi}(c_i) = u - 1, \quad g_v^{\bar{\Pi}}(c_i) = 1, \tag{5.52}$$

where $|A_1| = n$, $|A_2| = n$ and $|A_3| = 1$.

For some $i \neq 1$, c_i wins the election if everyone is honest with a median score of $u - 1$ while c_1 only has a median score of 1. Due to the tie breaking rule c_1 is selected with the submitted preferences. Furthermore, no individual can change their preferences to get a better candidate and all individuals are minimally dishonest and the Price of Deception is $u - 1$. □

If voter v's score for candidate c is x, then v's normalized score for c is $\frac{m-1}{u-1}(x-1)+1$. After the transformation, the maximum score for any candidate is m and the minimum score for any candidate is 1. The bound for the Price of Deception can be obtained by updating Theorem 5.4.7.

Theorem 5.4.8. *The normalized Price of Deception for Majority Judgment is* $\frac{um-2m+1}{u-1}$.

5.4.6 α-Copeland

In α-Copland voting, each voter has a strict ordering over all candidates. For each pair of candidates c_i and c_j, if the majority prefers c_i to c_j, then c_i gets one point. In the event of a tie, each candidate receives α points where $0 \leq \alpha < 1$. Each candidate receives between 0 and $m - 1$ points in total.

Theorem 5.4.9. *The Price of Deception for α-Copland voting is at least $m - 2$.*

Proof: The predetermined list for tie breaking is such that if c_2 and $c_j \neq c_2$ have the same score, then c_2 does not win the election. Again let $S_{c_i}^{\Pi}$ be the sincere Copeland score for candidate c_i and let $S^{\bar{\Pi}}_{c_i}$ be the Copeland score for candidate c_i with respect to the submitted preferences. A set of sincere preferences to achieve this lower bound is as follows:

$$\pi_1 = \quad (c_3, c_4, c_5, ..., c_{m-2}, c_{m-1}, c_2, c_1, c_m) \tag{5.53}$$

$$\pi_2 = \quad (c_4, c_5, c_6, ..., c_{m-1}, c_m, c_2, c_1, c_3) \tag{5.54}$$

$$\vdots \tag{5.55}$$

$$\pi_{m-2} = \quad (c_m, c_3, c_4, ..., c_{m-3}, c_{m-2}, c_2, c_1, c_{m-1}) \tag{5.56}$$

$$\pi_i = \quad (c_1, c_3, c_4, ..., c_{m-3}, c_{m-2}, c_{m-1}, c_m, c_2) \qquad \forall i \in A \tag{5.57}$$

$$\pi_i = \quad (c_1, c_m, c_{m-1}, ..., c_6, c_5, c_4, c_3, c_2) \qquad \forall i \in B \tag{5.58}$$

where $|A| = |B| = \frac{m-3}{2}$ where m is odd. With these preferences, $S^{\Pi}_{c_1} = m-2$, $S^{\Pi}_{c_2} = 1$ and $S^{\Pi}_{c_i} = \frac{m-1}{2}$ for all other i. Candidate c_1 is the winner if everyone is sincere. A minimally dishonest equilibrium is as follows:

$$\bar{\pi}_1 = \quad (c_2, c_3, c_4, c_5, ..., c_{m-2}, c_{m-1}, c_1, c_m) \tag{5.59}$$

$$\bar{\pi}_2 = \quad (c_2, c_4, c_5, c_6, ..., c_{m-1}, c_m, c_1, c_3) \tag{5.60}$$

$$\vdots \tag{5.61}$$

$$\bar{\pi}_{m-2} = \quad (c_2, c_m, c_3, c_4, ..., c_{m-3}, c_{m-2}, c_1, c_{m-1}) \tag{5.62}$$

$$\bar{\pi}_i = \quad (c_1, c_3, c_4, ..., c_{m-3}, c_{m-2}, c_{m-1}, c_m, c_2) \qquad \forall i \in A \tag{5.63}$$

$$\bar{\pi}_i = \quad (c_1, c_m, c_{m-1}, ..., c_6, c_5, c_4, c_3, c_2) \qquad \forall i \in B \tag{5.64}$$

In this equilibrium, $S^{\bar{\Pi}}_{c_2} = S^{\bar{\Pi}}_{c_1} + 1 = m - 1$ and $S^{\bar{\Pi}}_{c_i} = \frac{m-3}{2}$ for all other i. Letting C be the directed cycle defined by $c_3 \rightarrow c_4 \rightarrow ... \rightarrow c_m \rightarrow c_3$, candidate c_i wins a majority over the $\frac{m-3}{2}$ candidates appearing before c_i and loses to the $\frac{m-3}{2}$ appearing after c_i. Furthermore, if c_j appears $k \leq \frac{m-3}{2}$ positions prior to c_i, then c_i would lose a majority election to c_j by $m - 2k$ votes. For i where $3 \leq i \leq m$, candidate c_i loses to c_2 by 1 vote and to c_1 by 1 vote. We now show that the solution is an equilibrium.

Voters in A and B, they cannot decrease c_2's score, cannot increase c_1's score and for all other i, can only increase c_i's score by 1 point by placing c_i ahead of c_1. Therefore, for any permutation, candidate c_2 still wins the election. Furthermore, voters in A and B are minimally dishonest.

By symmetry, it now suffices to show that voter 1 can do no better. The voter will only change their preferences if they can make c_i win where $3 \leq i \leq m-1$ or if they can be more honest and get at least as good a result. However, for i where $3 \leq i \leq m-1$, the voter can only increase the score of c_i by 1 since c_i already appears ahead of c_1. Since c_2 has a score of $m-1$ and c_i has a score of $\frac{m-3}{2}$, the voter cannot make c_i win. Furthermore, the voter will be more honest only if they switch c_i with c_2 for some i where $3 \leq i \leq m-1$. However, this will cause the score of c_2 to decrease by 1 and thus c_1 and c_2 will tie resulting in a strictly worse solution for the voter. Thus, the submitted preferences form a minimally dishonest equilibrium and the Price of Deception is at least $m-2$. □

To normalize 1st order Copeland scores, we add one to the score of every candidate. Every candidate receives between 1 and m points and the normalized Price of Deception is by definition at most m. After updating the proof for Theorem 5.4.9, we obtain the following bound on the Price of Deception.

Theorem 5.4.10. *The normalized Price of Deception for 1st Order Copeland voting is between* $\frac{m-1}{2}$ *and* m.

5.4.7 Plurality

Plurality voting is a scoring voter mechanism where each voter give only one point to their favorite candidate and zero points to all other candidates. The winning candidate is a candidate that receives the most points. The maximum and minimum number of points that a candidate may receive is n and 0 respectively.

Theorem 5.4.11. *The Price of Deception for plurality voting is* ∞.

Proof: The predetermined list for tie breaking is such that if c_2 and $c_j \neq c_2$ have the same score, then c_2 does not win the election. Suppose that the sincere and putative preferences are given by

$$\pi_i = \quad (c_1, ..., c_m) \qquad \bar{\pi}_i = \pi_i \qquad \forall i \in A_1 \tag{5.65}$$

$$\pi_i = \quad (c_1, ..., c_m) \qquad \bar{\pi}_i = (c_2, c_1, c_3, c_4, ..., c_m) \qquad \forall i \in A_2 \tag{5.66}$$

$$\pi_i = \quad (c_3, c_4, ..., c_m, c_1, c_2) \qquad \bar{\pi}_i = \pi_i \qquad \forall i \in A_3 \tag{5.67}$$

where $|A_1| = k - 2, |A_2| = k + 1$ and $|A_3| = k$. When everyone is honest, c_1 wins the election with $2k - 1$ points and c_2 receives zero points. With the submitted preferences c_2 wins the election with $k + 1$ points. Since c_2 defeats c_3 by one point and c_1 by 3 points, no voter can alter their preferences to receive a strictly better outcome. Furthermore, if anyone in A_2 is more honest, then c_3 would win the election. Thus the submitted preferences form a Nash equilibrium and the Price of Deception is ∞. □

To normalize the scoring for plurality voting candidate c_i would receive $\frac{m}{n}$ points from a voter that lists c_i as their favorite candidate and $\frac{1}{n}$ from all other voters.

Theorem 5.4.12. *The normalized Price of Deception for plurality voting is* $\frac{2m+1}{3}$.

Proof: Updating the proof of Theorem 5.4.11, the Price of Deception is at least $\frac{(2m+1)k-m}{3k-1}$ which converges to $\frac{2m+1}{3}$ as k grows large.

It remains to prove the upper bound. Suppose that c_1 is a winner if everyone is sincere and that c_2 is the winner at equilibrium. We may assume that c_2 is not the sincere winner since this would yield a Price of Deception of one.

First we claim that if voter v most prefers candidate c_i then v either indicates that they most prefer c_i or c_2. For contradiction, suppose that v indicates that they most prefer $c_j \notin \{c_2, c_i\}$. If v is completely honest, then c_j's score decreases and c_i's score increases. This cannot yield a worse solution for v contradicting that the submitted preferences formed a minimally dishonest equilibrium. Thus, if voter v most prefers candidate c_i then v indicates that they most prefer either c_i or c_2.

Next we claim there is a candidate c_i that loses to c_2 by at most one vote. For contradiction suppose that every candidate loses to c_2 by at least two votes. Since c_2 is not a sincere winner and by the previous claim, there must be a voter v who most prefers c_i for some $i \neq 2$ but indicates that they most prefer c_2. If v were honest then c_2 would lose one vote and c_i would gain one vote. Since c_2 wins by at least two votes, either c_2 will remain the unique winner or c_i and c_2 will tie. Neither outcome is worse for v contradicting that the preferences formed a minimally dishonest equilibrium. Thus there is a candidate c_i that loses to c_2 by exactly one vote.

Now we break the problem into two cases. In Case 1 c_1 comes in second at equilibrium and in Case 2 c_3 comes in second but c_1 does not.

Case 1: c_1 finishes in second. Let A be the set of voters that most prefer c_1, W be the set of voters that indicate that they most prefer c_2 and let $B = W \cap A$. Since c_1 loses to c_2 by at most one vote, $|B| = 0$. Thus $|W| \geq |A|$ and $|W| + |A| \leq n$, and $|A| \leq \frac{n}{2}$. If c_2 was sincerely most preferred by no one, then the Price of Deception is $\frac{|A|\frac{m}{n}+(n-|A|)\frac{1}{n}}{n\frac{1}{n}} \leq \frac{m+1}{2}$.

Case 2: c_3 finishes in second but c_1 does not. Let A be the set of voters that most prefer c_1, W be the set of voters that indicate that they most prefer c_2, $B = W \cap A$ and let D be the set of voters that indicate that they most prefer c_3. Since there are n voters, we have that $|W| + |D| + |A| - |B| \leq n$. Since c_3 is in second place, $|D| + 1 \geq |W| \geq |D|$. By definition, $|W| \geq |B|$. Furthermore, c_1 loses to c_3 by at least one vote and we have that $|D| \geq |A| - |B| + 1$. Using Fourier-Motzkin elimination, we obtain that $|A| \leq \frac{2}{3}n$. Once again if c_2 was sincerely most preferred by no one, then the Price of Deception is $\frac{|A|\frac{m}{n}+(n-|A|)\frac{1}{n}}{n\frac{1}{n}} \leq \frac{2m+1}{3}$. □

5.5 Prices of Deception for Voting Rules with Random-Tie Breaking

To this point, like Gibbard and Satterthwaite, we have examined decision mechanisms that are single-valued. Neutral (anonymous) voting mechanisms are such that relabeling the candidates (respectively voters) does not change the outcome of an election. In general, single-valued voting mechanisms cannot be both neutral and anonymous [68]. By breaking ties lexicographically according to a predetermined list of candidates in Section 5.4, we have violated neutrality. Alternatively every theorem in Section 5.4 holds if we break ties by violating only anonymity and select the most preferred candidate by a particular voter from the set of candidates with the highest score.

In this section, we consider neutral and anonymous decision mechanisms where ties are broken randomly. One option is to randomly select a tie breaking rule that violates neutrality or anonymity and report the rule to the voters. For example, in Section 5.4, instead of breaking ties by selecting the first possible candidate from a predetermined list, ties would be broken by selecting from a randomly chosen list. The selected list however would be reported to candidates prior to voting. Thus, from the perspective of the voters at the time that they cast their ballots the decision mechanism is single-valued and all upper bounds from Section 5.4 hold.

More interesting however is when voters do not know the result of the randomness prior to voting. We now consider voting mechanisms where the winner is selected uniformly at random from the set of candidates that have the highest score. Instead of being single valued, $f(\Pi) \subseteq C$ is the set of candidates that have positive probability of winning[2].

5.5.1 Valuation of Risk and the Price of Deception

An immediate consequence of allowing r to be set-valued is that ordinal preferences over C are no longer sufficient to characterize a voters' preferences. Ordinal preferences over C are insufficient to determine whether voter v prefers their 2nd most preferred candidate to an outcome yielding their most preferred candidate half the time and their 3rd most preferred candidate the rest of the time. Instead, each candidate needs to have ordinal

[2] For non-uniform distributions we should view $r(\Pi) \in [0,1]^{|C|}$ as a probability distribution where $r(\Pi)_i$ denotes the probability that c_i wins the election. However since we are selecting the candidate uniformly at random, giving only the set of candidates that have positive probability of winning is sufficient to characterize the outcome.

preferences over all $2^{|C|} - 1$ non-empty subsets of C[3].

Alternatively, as in Definition 5.2.1 for score voting procedures, we can always cast individual preferences as cardinal values. However, a voter v must have a value for $g_v(C')$ for all $C' \subseteq C$. An individual's expected utility, $\frac{1}{|r(\Pi)|}\sum_{c\in r(\Pi)} u_v(c)$, is not necessarily equal to the utility of the outcome, $u_v(r(\Pi))$. An individual is *risk-neutral* if the two are equal. An individual is *risk-prone* (*risk-averse*) if the utility of the outcome $r(\Pi)$ is more (respectively less) than the expected utility of $r(\Pi)$.

Let $P_v(C', k) = \sum_{c\in C'} \frac{1}{|C'|}\mathbb{1}_{\{u_v(c)\geq k\}}$ be the probability that the randomly selected candidate will have a utility is at least k given outcome C' for voter v. Voter v is *rational* only if $u_v(C') \geq u_v(C'')$ whenever $P_v(C', k) \geq P_v(C'', k)$ for all k. Individual valuations of risk can greatly impact the set of equilibria and in general may greatly alter the Price of Deception. However, all results in this section only require that voters are rational.

Society also has its own valuation of risk that may not correlate with the voters' valuations of risk. Unlike the voters, society's valuation of risk has no impact on the set of equilibria. It does, however, impact the Price of Deception. As society becomes more risk-averse (prone) the Price of Deception will not decrease (respectively increase). Generally, we only assume that society is rational.

Definition 5.3.8 for minimal dishonesty is unchanged when r becomes set-valued and Definition 5.3.9 for the Price of Deception is easily updated for set-valued functions given society's valuation of risk. We now analyze how the Prices of Deception from Section 5.4 change when ties are broken randomly. In addition, we find the Price of Deception when the winner is uniquely determined.

5.5.2 Random Dictator

Given that there is never a tie when the winning candidate is determined by a dictator, we instead analyze the random dictator rule. A voter is selected uniformly at random and the

[3] For non-uniform distributions, a voter may need to have preferences over all distributions over the set of candidates, $\{x \in [0,1]^{|C|} : ||x||_1 = 1\}$.

voter's most preferred candidate is selected as the winning candidate. Furthermore, since honesty is the only minimally dishonest response for any dictator, every individual will be honest and the Price of Deception is one.

5.5.3 Veto

The key step in Theorem 5.4.1 was showing that if c_m has the most points at an equilibrium, then the sincere score of for c_m is at least $n - \frac{n}{m}$. However, this holds for all candidates who have positive probability of winning in an election where ties are broken randomly and the upper bound for the Price of Deception remains unchanged assuming society is rational. Furthermore, the lower bound presented in Theorem 5.4.1 is still valid and the (normalized) Price of Deception remains $1 + \frac{1}{m-1}$ (respectively $1 + \frac{m-1}{m^2-m+1}$). The example in Theorem 5.4.1 has a unique candidate with high score. Thus, the Price of Deception remains unchanged even if there is no unique winner.

5.5.4 Approval

Like veto, the key step in Theorem 5.4.3 was showing if c_m has the most points at an equilibrium and c_1 was the sincere winner, then the sincere score for c_m is at least half the sincere score of c_1. Once again this holds for all candidates that have positive probability of winning and the upper bound for the Price of Deception remains unchanged assuming society is rational. The lower bound in Theorem 5.4.3 remains valid with random tie breaking and the (normalized) Price of Deception remains 2 (respectively $\frac{2m}{m+1}$). Once again the Price of Deception remains unchanged in the event of no unique winner.

5.5.5 Borda Count

The preferences given in Theorem 5.4.5 remain valid and the (normalized) Price of Deception is at least m (respectively $\frac{m+2}{3}$).

Theorem 5.5.1. *Given that two individuals tie at an equilibrium and that society is risk-neutral or risk-prone, the (normalized) Price of Deception for Borda count is at most* $\frac{2m^2-6m+4}{3m-7}$ *(respectively* $\frac{2m^2-4m}{5m-11}$*).*

The proof of Theorem 5.5.1 is similar to Theorem 5.4.5 and can be found in Section 5.6.1.

Theorem 5.5.2. *Given that at least three individuals tie at an equilibrium and that society is risk-neutral or risk-prone, the (normalized) Price of Deception for Borda count is at most* $m-1$ *(respectively* $\frac{m}{2}$*).*

Proof: Since society is risk-neutral or risk-prone the societal utility of the outcome $r(\bar{\Pi})$ is $U(\Pi, r(\bar{\Pi})) \geq \frac{(|r(\bar{\Pi})|-1)n+(|r(\bar{\Pi})|-2)n+\ldots+n+0}{|r(\bar{\Pi})|} \geq n$. Since a winning candidate can receive at most $n(m-1)$ points, the Price of Deception is at most $m-1$. The proof for normalized Price of Deception is similar. $\square$

It remains to establish the upper bound when there is a unique winner. The proof technique in Theorem 5.4.5 applies when there is a unique winner and the (normalized) Price of Deception is m (respectively in $[\frac{m+2}{3}, \frac{m^2}{2m-1}]$). However, when ties are broken randomly and there is a unique winner, we are able to establish that the normalized Price of Deception is at most $\frac{m^3-2m^2+3m-3}{2m^2-3m}$ giving us that the bound $\frac{m+2}{3}$ is tight for $m=3$.

Theorem 5.5.3. *If there is a unique winner, then the normalized Price of Deception for Borda count voting is* $\frac{m+2}{3}$ *for* $m=3$ *and between* $\frac{m+2}{3}$ *and* $\frac{m^3-2m^2+3m-3}{2m^2-3m}$ *for* $m>3$.

We replace the model from Theorem 5.4.5 with two new models. Let c be a sincere winner. In the first model, we assume that c is the only candidate to come in second place at a minimally dishonest equilibrium and obtain an upper bound of $\frac{m+1}{3}$. In the second model, we assume there is some $c' \neq c$ that comes in second place at a minimally dishonest equilibrium. In the second model we obtain a bound of $\frac{m+2}{3}$ for $m=3$ and $\frac{m^3-2m^2+3m-3}{2m^2-3m}$ for $m>3$. Both models can be found in the Section 5.6.1.

We also conjecture that the $\frac{m+2}{3}$ bound is tight. There are two reasons why $\frac{m^3-2m^2+3m-3}{2m^2-3m}$ is likely to be loose. In the proof of Theorem 5.4.6 we create a variable x_{ijk} representing the number of sincere voters that give i points to c_1, j points to c_2 and k points to c_3. In the case when $m = 3$, this enumerates all types of voters. However, we lose quite a bit of information when adding more candidates. In addition, we only consider two models. In this first model the sincere winner comes in second place and in the second model there is at least one other candidate that comes in second place. For $m = 3$, this describes all possible outcomes. However, when there are more candidates there are many more models to consider. Not only are there multiple possibilities for what place a sincere winner could come in, there are many possibilities for how many candidates tie for second, third, etc.

5.5.6 Majority Judgment

Theorem 5.5.4. *The (normalized) Price of Deception for Majority Judgment is $u - 1$ (respectively $\frac{um-2m+1}{u-1}$) when society is completely risk-averse.*

Proof: The proof for the standard score is symmetric to the normalized score. Thus we consider only the standard score. We consider two cases. In the first we assume that there is a sincere winner with median score u. In the second case we assume a sincere winner's median score is at most $u - 1$.

If the sincere winner has a sincere median score of u then by minimal dishonesty that candidate will have an equilibrium median score of u. Furthermore, if a candidate has a sincere median score of 1, they will also have an equilibrium median score of 1. As a result every candidate that comes in first at equilibrium must have a sincere median score of at least 2 and the Price of Deception is at most $\frac{u}{2}$.

If the sincere winner has a sincere median score of $u - 1$ or less, then it may be possible for for an equilibrium winner to have a sincere median score of one and the Price of Deception is at most $u - 1$.

The preferences given in the proof of Theorem 5.4.7 give a valid lower bound of $u - 1$

completing the proof of this theorem. $\square$

Theorem 5.5.5. *Given a unique winner, the Price of Deception for (normalized) Majority Judgment is* $\frac{u-1}{2}$ *(respectively* $\frac{um-2m+1}{u+m-2}$*).*

Proof. The argument for normalized Price of Deception is similar so we only examine the standard score. As shown in Theorem 5.5.4, if a candidate has a sincere median of 1, then the candidate will have a score of 1 at equilibrium. Since there are no ties and 1 is the smallest possible, the candidate cannot win the election. Similarly, if a candidate has a sincere median score of u, the candidate is the unique winner with a score of u at every equilibrium and the Price of Deception is one. Thus, if the Price of Deception is more than one, $u-1$ is the highest possible sincere score for the sincere winner and 2 is the lowest possible sincere score for the equilibrium winner and the Price of Deception is at most $\frac{u-1}{2}$. It remains to give a set of preferences achieving this value. Consider the following preferences

$$v \in A_1: \qquad g_v^{\Pi}(c_1) = g_v^{\Pi'}(c_1) = u-1, \qquad g_v^{\Pi}(c_2) = g_v^{\bar{\Pi}}(c_2) = u, \tag{5.68}$$

$$v \in A_2: \qquad g_v^{\Pi}(c_1) = g_v^{\Pi'}(c_1) = u-1, \qquad g_v^{\Pi}(c_2) = g_v^{\bar{\Pi}}(c_2) = 1, \tag{5.69}$$

$$v = A_3: \qquad g_v^{\Pi}(c_1) = g_v^{\Pi'}(c_1) = 1, \qquad g_v^{\Pi}(c_2) = 2, g_v^{\bar{\Pi}}(c_2) = u, \tag{5.70}$$

where $|A_1| = n$, $|A_2| = n$ and $|A_3| = 1$. All unlisted values are one. The sincere winner is c_1 with a score of $u-1$ while c_2 has a sincere score of only 2. The winner given the submitted preferences is c_2. Voters in A_1 and A_2 are honest and cannot alter their preferences to obtain a better outcome. If voter v' is more honest, then v' will obtain a strictly worse outcome (assuming ties are broken such that c_2 has positive probability of losing given that c_2 ties with c_1). $\square$

Theorem 5.5.6. *If society is risk-neutral, then the (normalized) Price of Deception is* $\max\{\frac{um-m}{u+m-2}, \frac{u-1}{2}\}$ *(respectively* $\max\{\frac{um^2-2m^2+m}{2um-u-3m+2}, \frac{um-2m+1}{u+m-2}\}$*).*

Proof. Once again we only consider the standard scores. The bound of $\frac{u-1}{2}$ comes from Theorem 5.5.5 when there is a unique winner. We now assume there is not a unique winner and the remainder of the proof follows in a similar fashion to Theorem 5.5.4. The lower bound $\frac{um-m}{u+m-2}$ comes directly from the example in Theorem 5.4.7.

Once again if a sincere winner has a sincere score of u, then the sincere winner will tie for first at equilibrium and every other candidate tied for first will have a sincere score of at least 2. Thus, society's value of the outcome is at least $\frac{1}{m}u + \frac{m-1}{m}2 = \frac{u+2m-2}{m}$ since there is at least one candidate tied for first at equilibrium with a score of u. Thus the Price of Deception is at most $\frac{um}{u+2m-2}$.

Alternatively, an equilibrium winner can have a sincere score of 1. This candidate will also have an equilibrium score of 1 and thus every candidate must tie for first place. Since every candidate has an equilibrium score of 1, there can be no candidates with a sincere score of u. Let $x \leq u-1$ denote the highest sincere score for any candidate. Society's value of the outcome is at least $\frac{1}{m}x + \frac{m-1}{m} = \frac{x+m-1}{m}$ yielding a Price of Deception of at most $\frac{xm}{x+m-1}$. This quantity is maximized when $x = u-1$ giving a Price of Deception of $\frac{um-m}{u+m-2}$. □

5.5.7 α-Copeland

There were no ties given in the proof of Theorem 5.4.9 and the lower bound given is still valid when ties are broken randomly. Thus the Price of Deception remains unchanged.

5.5.8 Plurality

The lower bound given in Theorem 5.4.11 remains valid and the Price of Deception is ∞ even if ties are broken randomly. However, the normalized Price of Deception changes.

Theorem 5.5.7. *The normalized Price of Deception for plurality voting is m.*

Proof: The upper bound is m by definition. To obtain the lower bound suppose that we have k voters with sincere preferences $(c_m, c_1, c_2, ..., c_{m-1})$ and k voters with sincere preferences

$(c_m, c_{m-1}, ..., c_1)$. Candidate c_m is the only candidate that would receive any points if candidates were honest.

We now describe an equilibrium where c_1 and c_{m-1} have a probability of $.5$ each of being the winning candidate. The first k voters submit the preference list $(c_1, c_m, c_2, ..., c_{m-1})$ while the second set of voters submit the preference list $(c_{m-1}, c_m, c_{m-2}, ..., c_1)$. No voter can cause c_m to win and if any voter is more honest, they would get a worse solution. Thus the Price of Deception is $\frac{m}{1} = m$. $\square$

Finally we note that the proof for Theorem 5.4.12 remains valid if there is a unique winner even when ties are broken randomly. Thus, given a unique winner, the normalized Price of Deception is at most $\frac{2m+1}{3}$.

5.6 Additional Proofs

Lemma 5.6.1. *Let $c, d \in C$ be a pair of candidates in an election with n random voters sampled from a probability distribution p on $\mathcal{G}(C, g)$. Suppose there exists $g \in \mathcal{G}(C, g)$ such that $g(c) \neq g(d)$ and $p(g) > 0$. Then for any constants k and $\epsilon > 0$,*

$$\lim_{n \to \infty} n^{.5-\epsilon} P(|S_c - S_d| \leq k) = 0. \tag{5.71}$$

Proof: For $i = 1, \ldots, n$ define the random variable $Y_i = g_{v_i}(c) - g_{v_i}(d)$, where v_i is the ith sample voter drawn according to μ. Let $Y = \sum_{i=1}^{n} Y_i$. On the one hand, $Y = S_c - S_d$. On the other hand, Y is a sum of n i.i.d. random variables each with nonzero mean or with nonzero variance σ^2. The variance σ^2 is finite because the range of $|Y_i|$ is bounded by $g(|C|)$.

If $E[Y_i] \neq 0$ then, by the strong law of large numbers, with probability 1

$$lim_{n \to \infty} \frac{Y}{n} \overset{a.s.}{=} E[Y_i]. \tag{5.72}$$

Then with probability 1,

$$\lim_{n\to\infty} |S_c - S_d| \overset{a.s.}{>} \frac{n}{2}(|E[Y_i]) \gg 2g(|C|). \tag{5.73}$$

If $E[Y_i] = 0$ then $\sigma^2 \neq 0$ and by the central limit theorem $Y/\sqrt{n\sigma^2}$ converges in distribution to the standard Gaussian (normal) distribution $N(0,1)$. The density of a standard Gaussian distribution variable X, $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$, attains its maximum value $\frac{1}{\sqrt{2\pi}}$ at $x = 0$. Therefore

$$P(|X| \leq k/n^\epsilon) \leq k\sqrt{\frac{2}{\pi}}/n^\epsilon \underset{n\to\infty}{0}. \tag{5.74}$$

□

5.6.1 Borda Count

Lemma 5.6.2. *Given a sincere profile Π and a profile $\bar{\Pi}$ forming a minimally dishonest equilibrium where $|r(\bar{\Pi})| \leq 2$, let $\omega_v = \max_{c\in r(\bar{\Pi})} g_v^{\Pi}(c) + 1$. If $g_v^{\Pi}(c) \geq \omega_v$, then $g_v^{\bar{\Pi}}(c) \geq g_v^{\Pi}(c) - 1$.*

Proof: For contradiction, assume there is a c such that $g_v^{\Pi}(c) \geq \omega_v$ but $g_v^{\bar{\Pi}}(c) \leq g_v^{\Pi}(c) - 2$. If $c \in r(\bar{\Pi})$, then the result immediately holds since v could move c up one position in their preference list and c would be the unique winner. We now split the problem into two cases and show that we have a contradiction in both cases.

Case 1: There is a $c' \notin r(\bar{\Pi})$ such that $g_v^{\bar{\Pi}}(c') \geq g_v^{\Pi}(c) - 1$ but $g_v^{\Pi}(c') \leq g_v^{\Pi}(c) - 2$. Voter v can be more honest by switching the location of c and c in $\bar{\pi}_v$ which either has no effect on the election, causes c to be added to $r(\bar{\Pi})$ or causes c to be the unique winner. All three possibilities yield at least as good an outcome for v, a contradiction. Thus, Case 1 cannot occur.

Case 2: If $|r(\bar{\Pi})| = 1$ then Case 1 must occur. Thus we may assume that $r(\bar{\Pi}) = \{c', c''\}$. Without loss of generality we assume that v prefers c' to c''.

Both $g_v^{\bar{\Pi}}(c') \geq g_v^{\Pi}(c) - 1$ and $g_v^{\bar{\Pi}}(c'') \geq g_v^{\Pi}(c) - 1$. Voter v can be more honest by switching the locations of c'' and c in π'_v. After the switch, either c' will win, c will win or c' and c will tie. All three possibilities result in a strict improvement for v since $g_v^{\Pi}(c) \geq \omega_v > g_v^{\Pi}(c') > g_v^{\Pi}(c'')$. This is a contradiction since Π' describes an equilibrium and the lemma holds. □

Lemma 5.6.3. *When ties are broken randomly, if* $r(\bar{\Pi}) \cap r(\Pi) = \emptyset$ *and* $\{c_2, c_3\} = r(\bar{\Pi})$, *then there is a candidate* $c \notin r(\bar{\Pi})$ *such that* $S_c^{\bar{\Pi}} = S_{c_2}^{\bar{\Pi}} - 1 = S_{c_3}^{\bar{\Pi}} - 1$.

Proof: Without loss of generality we examine only standard scores. For contradiction assume Lemma 5.6.3 is not true and that for all $c \notin r(\bar{\Pi})$, $S_c^{\bar{\Pi}} \leq S_{c_2}^{\bar{\Pi}} - 2$. Let A be the set of voters that prefer c_2 to c_3 and let B be everyone else. We claim that for all $v \in V$, $g_v^{\bar{\Pi}}(c_2) = m-1$ and $g_v^{\bar{\Pi}}(c_3) = 0$. If not, then v can either move c_2 up in their preference list causing c_2 to be the unique winner, or v can move c_3 down in their preference list causing c_2 to be the unique winner since all other candidates lose by at least 2. Thus the claim holds.

A symmetric claim holds for B and thus $|A| = |B|$. This implies that $S_{c_2}^{\bar{\Pi}} = \frac{n}{m}\binom{m}{2}$. Since this is the average score, every candidate must have this score contradicting that there were only two candidates with positive probability of winning completing the proof. □

Theorem 5.5.1. If ties are broken randomly, two individuals tie at an equilibrium and that society is risk-neutral or risk-prone, the (normalized) Price of Deception for Borda count is at most $\frac{2m^2-6m+4}{3m-7}$ (respectively $\frac{2m^2-4m}{5m-11}$).

Proof: If $r(\bar{\Pi}) \cap f(\Pi) \neq \emptyset$ the Price of Deception is at most 2 since society is either risk-neutral or risk prone. Thus we assume that $c_1 \in r(\Pi) \setminus r(\bar{\Pi})$ and that $r(\bar{\Pi}) = \{c_2, c_3\}$.

Let x_{ijk} be the number of voters that would give $i-1$ points to c_1, $j-1$ points to c_2 and $k-1$ points to c_3 if they were honest for $i, j, k \in \{1, ..., m\}$. Let M be the set $\{\{i, j, k\} \in \{1, ..., m\}^3 : i \neq j \neq k \neq i\}$. The Price of Deception is given by

$$\frac{\displaystyle\sum_{\{i,j,k\}\in M} 2(i-1)x_{ijk}}{\displaystyle\sum_{\{i,j,k\}\in M} (j+k-2)x_{ijk}} \tag{5.75}$$

We begin by claiming that $x_{i21} = x_{i12} = 0$ for all i. For contradiction, suppose that $x_{i21} > 0$ and let v be such a voter. Due to Lemma 5.6.3, there is a candidate $c \notin \{c_2, c_3\}$ such that $S_c^{\Pi'} = S_{c_2}^{\Pi'} - 1$. First, $g_v^{\bar{\Pi}}(c_2) = m-1$, since otherwise v can move c_2 up in their preference list resulting in a better solution. Similarly, since c loses by only one point, $g_v^{\bar{\Pi}}(c) = m-1$ implying that $c = c_2$, a contradiction. Thus we may omit x_{ijk} from the model when $\max\{j,k\} = 2$.

The remainder of this proof follows in the same fashion as Theorem 5.4.5. As before we can show that if voter v would sincerely give c_2 or c_3 no points, then v would also do this at an equilibrium. Furthermore, if v sincerely prefers c_2 over c_3, v's submitted preferences will also indicate this. Due to all this and Lemma 5.6.2, we know the following about the scores of each candidate:

$$S_{c_1}^{\bar{\Pi}} \geq \sum_{\substack{\{i,j,k\}\in M \\ i\geq\max\{j,k\}>2}} (i-2)x_{ijk} + \sum_{\substack{\{i,j,k\}\in M,\ j=1 \bigvee k=1 \\ i<\max\{j,k\}>2}} x_{ijk} \tag{5.76}$$

$$S_{c_2}^{\bar{\Pi}} \leq \sum_{\substack{\{i,j,k\}\in M \\ j>k,\ \max\{j,k\}>2}} (m-1)x_{ijk} + \sum_{\substack{\{i,j,k\}\in M \\ k>j>1,\ \max\{j,k\}>2}} (m-2)x_{ijk}, \tag{5.77}$$

$$S_{c_3}^{\bar{\Pi}} \leq \sum_{\substack{\{i,j,k\}\in M \\ j>k>1,\ \max\{j,k\}>2}} (m-2)x_{ijk} + \sum_{\substack{\{i,j,k\}\in M \\ k>j,\ \max\{j,k\}>2}} (m-1)x_{ijk}. \tag{5.78}$$

We generate a mathematical program maximizing the Price of Deception where $S_{c_i}^{\bar{\Pi}} \geq S_{c_1}^{\bar{\Pi}}$ for $i \in \{2,3\}$. After relaxing the problem into a linear program and taking the dual, we get that every feasible solution to the following linear program creates an upper bound on the Price of Deception:

$$
\begin{aligned}
&\min && w' = y_3 && && && (5.79)\\
&subject\ to: \\
&(i-m-1)y_1 && +(i-m)y_2 && +(j+k-2)y_3 \geq 2(i-1) && \forall\ \{i,j,k\} \in M,\ j > k > 1, i \geq \max\{j,k\} > 2 && (5.80)\\
&(1-m)y_1 && +(2-m)y_2 && +(j+k-2)y_3 \geq 2(i-1) && \forall\ \{i,j,k\} \in M,\ j > k > 1, i < \max\{j,k\} > 2 && (5.81)\\
&(i-m)y_1 && +(i-m-1)y_2 && +(j+k-2)y_3 \geq 2(i-1) && \forall\ \{i,j,k\} \in M,\ k > j > 1, i \geq \max\{j,k\} > 2 && (5.82)\\
&(2-m)y_1 && +(1-m)y_2 && +(j+k-2)y_3 \geq 2(i-1) && \forall\ \{i,j,k\} \in M,\ k > j > 1, i < \max\{j,k\} > 2 && (5.83)\\
&(i-2)y_1 && +(i-m-1)y_2 && +(1+k-2)y_3 \geq 2(i-1) && \forall\ \{i,1,k\} \in M,, i \geq \max\{1,k\} > 2 && (5.84)\\
&y_1 && +(3-m)y_2 && +(1+k-2)y_3 \geq 2(i-1) && \forall\ \{i,1,k\} \in M,, i < \max\{1,k\} > 2 && (5.85)\\
&(i-m-1)y_1 && +(i-2)y_2 && +(j+1-2)y_3 \geq 2(i-1) && \forall\ \{i,j,1\} \in M,, i \geq \max\{j,1\} > 2 && (5.86)\\
&(3-m)y_1 && +y_2 && +(j+1-2)y_3 \geq 2(i-1) && \forall\ \{i,j,1\} \in M,, i < \max\{j,1\} > 2 && (5.87)\\
&y_1, y_2 \in \mathbb{R}_{\geq 0}. && && && && (5.88)
\end{aligned}
$$

The coefficients for y_1 correspond to the constraint $S_{c_2}^{\bar{\Pi}} \geq S_{c_1}^{\bar{\Pi}}$, and the coefficients for y_2 correspond to the constraint $S_{c_3}^{\bar{\Pi}} \geq S_{c_1}^{\bar{\Pi}}$, the coefficients for y_3 correspond to the constraint created by setting the denominator of the Price of Deception to one. Finally, the lower bounds for each of the constraints come from the numerator of the Price of Deception.

A feasible solution to this linear program is $y_1 = y_2 = \frac{2m-2}{3m-7}$ and $y_3 = \frac{2m^2-6m+4}{3m-7}$ and thus the Price of Deception when there are exactly two individuals tied is at most $\frac{2m^2-6m+4}{3m-7}$.

When updating the linear program for normalized scores only the third column and right hand side of the dual linear program will change since the first two columns correspond to the differences of two candidates scores. Each entry in the third column and right hand sides will increase by two. A new feasible solution to the linear program is $y_1 = y_2 = \frac{2m}{5m-11}$ and $y_3 = \frac{2m^2-4m}{5m-11}$ and the normalized Price of Deception is at most $\frac{2m^2-4m}{5m-11}$. □

Theorem 5.5.3. *If ties are broken randomly and there is a unique winner, the normalized Price of Deception for Borda count voting is $\frac{m+2}{3}$ for $m = 3$ and between $\frac{m+2}{3}$ and $\frac{m^3-2m^2+3m-3}{2m^2-3m}$ for $m > 3$.*

Prior to proving Theorem 5.5.3, we establish that if the sincere winner does not win at equilibrium, then there must be another candidate that barely loses.

Lemma 5.6.4. *If $c' = r(\bar{\Pi})$ and $c' \notin r(\Pi)$ then there is a candidate $c \neq c'$ such that $S_c^{\bar{\Pi}} = S_{c'}^{\bar{\Pi}} - 1$.*

Proof: For contradiction, suppose that every candidate loses by at least two points. Since c' is not a sincere winner, there must be some voter v that moved c' up in their list of preferences. Let c be the candidate that appears immediately after c' in v's submitted preferences list. By Lemma 5.6.2, v prefers c to c'. As a result v is more honest if they switches the location of c and c'. Since c' wins by at least two votes, either c' will remain the unique winner or only c and c' will tie for first. Neither possibility is worse for v. This contradicts that the set of preferences formed an minimally dishonest equilibrium and the lemma holds. □

We now prove Theorem 5.5.3 in two parts.

Lemma 5.6.5. *The normalized Price of Deception for Borda count voting (breaking ties lexicographically) is at most $\frac{m+1}{3}$ if the sincere winner is the only candidate in second place.*

Proof: As with Theorem 5.4.5, let $c_1 = r(\Pi)$ and $c_2 = r(\bar{\Pi})$. Let x_{ij} be the number of voters that would give i points to candidate c_1 and j points to candidate c_2. The sincere scores for the candidates are given by $S^{\Pi}_{c_1} = \sum_{i \neq j} i x_{ij}$ and $S^{\Pi}_{c_2} = \sum_{i \neq j} j x_{ij}$. The Price of Deception is

$$\frac{S^{\Pi}_{c_1}}{S^{\Pi}_{c_2}} = \frac{\sum_{i \neq j} i \cdot x_{ij}}{\sum_{i \neq j} j \cdot x_{ij}} \tag{5.89}$$

Claim 1: $x_{ij} = 0$ for all i and j where $m > i > j$.

For contradiction, suppose that there is some voter v that would give c_1 i points and c_2 j points if sincere where $m > i > j$. By Lemma 5.6.2, v gives c_1 either i or $i-1$ points at equilibrium. However, since $i < m$, by Lemma 5.6.4 c_1 loses by at most one point and v can get a strictly better solution by moving c_1 up one position in their submitted preference list. This contradicts that the preferences formed an equilibrium and thus $x_{ij} = 0$ for all i and j where $m > i > j$.

Claim 2: If voter v would sincerely give m points to c_1 and j points to c_2, then voter v gives at most j points to candidate c_2 at equilibrium.

For contradiction, suppose this is not the case. Let c_3 ($c_3 \neq c_1$ in an argument similar to Claim 1) be the candidate appearing directly after c_2 in v's submitted preference list. By Lemma 5.6.2, v prefers c_3 to c_2. Voter v could move c_2 down one position in their submitted preference list causing candidate c_3 to move up one position. Since c_1 is the only candidate in second place, only c_1, c_2 and possibly c_3 can tie first after the move. However since v prefers both c_1 and c_3 to c_2, v will prefer the outcome obtained after moving c_2 down one position. This contradicts that that the preferences form an equilibrium. Thus, v gives c_2 at most j points.

When these two claims are combined with Lemma 5.6.2 the following holds for the equilibrium scores, $S_{c_1}^{\bar{\Pi}}$ and $S_{c_2}^{\bar{\Pi}}$, for c_1 and c_2 respectively:

$$S_{c_1}^{\Pi} \geq \sum_{j<m} m \cdot x_{mj} + \sum_{j>i}^{m} x_{ij} \tag{5.90}$$

$$S_{c_2}^{\bar{\Pi}} \leq \sum_{j<m} j \cdot x_{ij} + \sum_{j>i} m \cdot x_{ij} \tag{5.91}$$

As in Theorem 5.4.5, since $S_{c_1}^{\Pi} \leq S_{c_2}^{\Pi}$ the following mathematical program gives an upper bound on the Price of Deception:

$$\max z = \frac{\sum_{i \neq j} i \cdot x_{ij}}{\sum_{i \neq j} j \cdot x_{ij}} = \frac{\sum_{j<m} m \cdot x_{mj} + \sum_{j>i} i \cdot x_{ij}}{\sum_{j<m} j \cdot x_{mj} + \sum_{j>i} j \cdot x_{ij}} \tag{5.92}$$

$$\text{subject to: } \sum_{j<m} (m-j) x_{mj} + \sum_{j>i} (1-m) x_{ij} \leq 0 \tag{5.93}$$

$$x_{ij} \in \mathbb{Z}_{\geq 0}. \tag{5.94}$$

This problem can be relaxed to the following linear program:

$$\max z = \sum_{j<m} m \cdot x_{mj} + \sum_{j>i} i \cdot x_{ij} \tag{5.95}$$

$$\text{subject to: } \sum_{j<m} (m-j) x_{mj} + \sum_{j>i} (1-m) x_{ij} \leq 0 \tag{5.96}$$

$$\sum_{j<m} j \cdot x_{mj} + \sum_{j>i} j \cdot x_{ij} = 1 \tag{5.97}$$

$$x_{ij} \in \mathbb{R}_{\geq 0}. \tag{5.98}$$

The dual of this problem is given by

$$\min w = y_2 \tag{5.99}$$

$$\text{subject to: } (m-j)y_1 + j \cdot y_2 \geq m \; \forall j < m \tag{5.100}$$

$$(1-m)y_1 + j \cdot y_2 \geq i \; \forall j > i \tag{5.101}$$

$$y_1 \in \mathbb{R}_{\geq 0}. \tag{5.102}$$

A feasible solution to this problem is $y_1 = \frac{2m-1}{3m-3}$ and $y_2 = \frac{m+1}{3}$ with value $\frac{m+1}{3}$ completing the proof of the lemma. □

Lemma 5.6.6. *Given a unique winner, the normalized Price of Deception for Borda count voting is at most $\frac{m+2}{3}$ for $m = 3$ and $\frac{m^3-2m^2+3m-3}{2m^2-3m}$ for $m > 3$ if c_1 is a sincere winner and $c_3 \neq c_1$ comes in second at equilibrium.*

Proof: As with Theorem 5.4.5, let $c_2 = r(\bar{\Pi})$. Let x_{ijk} be the number of voters that would give i points to candidate c_1, j points to candidate c_2 and k points to candidate c_3. Let $M = \{\{i,j,k\} \in \{1,...,m\}^3 : i \neq j, j \neq k, i \neq k\}$. All sums include only elements from M. The sincere scores for the candidates are given by $S^{\Pi}_{c_1} = \sum_{\{i,j,k\}} i \cdot x_{ijk}$, $S^{\Pi}_{c_2} = \sum_{\{i,j,k\}} j \cdot x_{ijk}$ and $S^{\Pi}_{c_2} = \sum_{\{i,j,k\}} k \cdot x_{ijk}$. The Price of Deception is

$$\frac{S^{\Pi}_{c_1}}{S^{\Pi}_{c_2}} = \frac{\sum_{\{i,j,k\}} i \cdot x_{ijk}}{\sum_{\{i,j,k\}} j \cdot x_{ijk}} \tag{5.103}$$

Claim 1: $x_{ijk} = 0$ for all j and k where $m > i > j$. This claim follows in the same fashion as Claim 1 in Lemma 5.6.5.

Claim 2: If v would sincerely give candidate c_2 1 point, then voter v is honest. This is a trivial consequence of minimal dishonesty since v receives their worst outcome.

Claim 3: If v would sincere give c_3 m points, the v will give c_3 m points at an equilibrium.

For contradiction, suppose this is not the case. Since c_3 is currently in second places and only loses by one point by Lemma 5.6.4, v can receive a strictly better outcome by moving c_3 up one position. This contradicts that the preferences formed an equilibrium and the claim must hold.

When these three claims are combined with Lemma 5.6.2 the following holds for the equilibrium scores, $S_{c_1}^{\bar{\Pi}}$, $S_{c_2}^{\bar{\Pi}}$ and $S_{c_3}^{\bar{\Pi}}$, for c_1, c_2 and c_3 respectively:

$$\begin{aligned} S_{c_1}^{\Pi'} \geq \sum_{\substack{\{i,j,k\}: \\ i>j>k}} (i-1)x_{ijk} + \sum_{\substack{\{i,j,m\}: \\ m>i>j>1}} (i-1)x_{ijm} + \sum_{\substack{\{i,1,m\}: \\ m>i>1}} i \cdot x_{i1m} \\ + \sum_{\substack{\{i,j,m\}: \\ m>j>i}} x_{ijm} + \sum_{\substack{\{i,j,m\}: \\ j>\max\{i,k\}}} x_{ijk} \end{aligned} \tag{5.104}$$

$$\begin{aligned} S_{c_2}^{\Pi'} \leq \sum_{\substack{\{i,j,k\}: \\ i>j>k}} m \cdot x_{ijk} + \sum_{\substack{\{i,j,m\}: \\ m>i>j>1}} (m-1)x_{ijm} + \sum_{\substack{\{i,1,m\}: \\ m>i>1}} x_{i1m} \\ + \sum_{\substack{\{i,j,m\}: \\ m>j>i}} (m-1)x_{ijm} + \sum_{\substack{\{i,j,m\}: \\ j>\max\{i,k\}}} m \cdot x_{ijk} \end{aligned} \tag{5.105}$$

$$\begin{aligned} S_{c_3}^{\Pi'} \geq \sum_{\substack{\{i,j,k\}: \\ i>j>k}} x_{ijk} + \sum_{\substack{\{i,j,m\}: \\ m>i>j>1}} m \cdot x_{ijm} + \sum_{\substack{\{i,1,m\}: \\ m>i>1}} m \cdot x_{i1m} \\ + \sum_{\substack{\{i,j,m\}: \\ m>j>i}} m \cdot x_{ijm} + \sum_{\substack{\{i,j,m\}: \\ j>\max\{i,k\}}} x_{ijk} \end{aligned} \tag{5.106}$$

$$\begin{aligned} S_{c_3}^{\Pi'} \leq \sum_{\substack{\{i,j,k\}: \\ i>j>k}} (j-1)x_{ijk} + \sum_{\substack{\{i,j,m\}: \\ m>i>j>1}} m \cdot x_{ijm} + \sum_{\substack{\{i,1,m\}: \\ m>i>1}} m \cdot x_{i1m} \\ + \sum_{\substack{\{i,j,m\}: \\ m>j>i}} m \cdot x_{ijm} + \sum_{\substack{\{i,j,m\}: \\ j>\max\{i,k\}}} (j-1)x_{ijk} \end{aligned} \tag{5.107}$$

Since $S_{c_2}^{\bar{\Pi}} \geq S_{c_3}^{\bar{\Pi}} \geq S_{c_1}^{\bar{\Pi}}$, the following mathematical program gives us an upper bound on the Price of Deception:

$$\max\ z = \frac{\sum_{\{i,j,k\}} i \cdot x_{ijk}}{\sum_{\{i,j,k\}} j \cdot x_{ijk}} \tag{5.108}$$

subject to:

$$\sum_{\substack{\{i,j,k\}: \\ i>j>k}} (i-j) \cdot x_{ijk} + \sum_{\substack{\{i,j,m\}: \\ m>i>j>1}} (i-1-m)x_{ijm} + \sum_{\substack{\{i,1,m\}: \\ m>i>1}} (i-m)x_{i1m} \tag{5.109}$$

$$+ \sum_{\substack{\{i,j,m\}: \\ m>j>i}} (2-m)x_{ijm} + \sum_{\substack{\{i,j,m\}: \\ j>\max\{i,k\}}} (2-j)x_{ijk} \leq 0 \tag{5.110}$$

$$\sum_{\substack{\{i,j,k\}: \\ i>j>k}} (1-m) \cdot x_{ijk} + \sum_{\substack{\{i,j,m\}: \\ m>i>j>1}} x_{ijm} + \sum_{\substack{\{i,1,m\}: \\ m>i>1}} (m-1)x_{i1m} \tag{5.111}$$

$$+ \sum_{\substack{\{i,j,m\}: \\ m>j>i}} x_{ijm} + \sum_{\substack{\{i,j,m\}: \\ j>\max\{i,k\}}} (1-m)x_{ijk} \leq 0 \tag{5.112}$$

$$x_{ijk} \in \mathbb{Z}_{\geq 0}\ \forall \{i,j,k\} \in M \tag{5.113}$$

By relaxing integrality, fixing the denominator of the objective to be one by scaling, and by taking the dual, we obtain the following linear program:

$$\min\ w = y_3 \tag{5.114}$$

subject to:

$$(i-j)y_1 \quad +(1-m)y_2 + j \cdot y_3 \geq i \quad \forall \{i,j,k\} \text{ where } i>j>k \tag{5.115}$$

$$(i-1-m)y_1 \quad +y_2 + j \cdot y_3 \geq i \quad \forall \{i,j,m\} \text{ where } m>i>j>1 \tag{5.116}$$

$$(i-m)y_1 \quad +(m-1)y_2 + j \cdot y_3 \geq i \quad \forall \{i,1,m\} \text{ where } m>i>1 \tag{5.117}$$

$$(2-m)y_1 \quad +y_2 + j \cdot y_3 \geq i \quad \forall \{i,j,m\} \text{ where } m>j>i \tag{5.118}$$

$$(2-j)y_1 \quad +(1-m)y_2 + j \cdot y_3 \geq i \quad \forall \{i,j,k\} \text{ where } j>\max\{i,k\} \tag{5.119}$$

$$y_1, y_2 \in \mathbb{R}_{\geq 0} \tag{5.120}$$

A feasible solution to this problem is $y_1 = \frac{m^2+m-3}{2m^2-3m}$, $y_2 = \frac{m-1}{2m-3}$ and $y_3 = \frac{m^3-2m^2+3m-3}{2m^2-3m}$

and the Price of Deception is at most $\frac{m^3-2m^2+3m-3}{2m^2-3m}$. For $m = 3$, this bound evaluates to $\frac{5}{3} = \frac{m+2}{3}$. $\square$

6

STRATEGIC STABLE MARRIAGE

6.1 Introduction

Stable marriage has an illustrious history in both theory and application. Gale and Shapley introduced the concept of stability and invented the first algorithm that always finds a stable marriage, given the preferences of the men and women [25]. Knuth [39] also pioneered theoretical analysis of stable marriage. Applications of stable marriage including Roth's celebrated design of the resident matching and kidney-exchange markets [59, 3] are surveyed in Section 6.1.4.

In general, theoretical studies of stable marriage assume that individuals' preferences are known. But in many situations an individual's preferences are private information. This opens up the troublesome possibility that individuals may behave strategically. A putatively stable marriage based on strategically submitted preference data could be *sincerely unstable*, that is, unstable with respect to the true preferences of the individuals. Gusfield and Irving posed the question of strategic submission of preference data in their classic text on stable marriage [31]. However, little progress has been made and Manlove's recent text [45] reports that this question remains open.

A series of results in the literature provides a fairly satisfactory answer to the question for the Gale-Shapley algorithm. We describe these results in detail in Section 6.1.4. The main points are that the men have no incentive to lie, and, if the men are honest and the women are strategic, the women can collude to manipulate the algorithm to arrive at any given sincerely stable marriage. The question of strategic preference data is unanswered for all other algorithms. Applications of the Stable Marriage Problem and its variants tend to employ the Gale-Shapley algorithm, often citing the concern for sincere stability.

This paper addresses Gusfield and Irving's general question of achieving a sincerely stable marriage when individuals submit private preference data strategically. We identify what we believe to have been the primary obstacle to stability and propose a natural means of overcoming it.

To examine this problem, we consider the Strategic Stable Marriage (SSM) game. Each individual x has private information π_x which describes x's preferences for members of the opposite gender. The set of all private information is the profile Π. Each individuals submits a putative preference list $\bar{\pi}_x$ to a publicly known stable marriage algorithm r. The mechanism r then selects the marriage $r(\bar{\Pi})$ that is stable with respect to $\bar{\Pi}$. The outcome for x is the individual who is married to x in $r(\bar{\Pi})$ and is evaluated by x according x's sincere preferences π_x.

Our first result (Theorem 6.3.4), is that regardless of r, the set of pure strategy Nash equilibrium outcomes of SSM is the set of all individually rational marriages (with respect to the sincere preferences). This is discouraging because no matter how one designs the public mechanism r, there will be Nash equilibria with poor outcomes. For example, let $M^{\circlearrowleft}$ be a minimax stable marriage for Π, meaning that the highest ranked partner anyone gets has the lowest rank possible. $M^{\circlearrowleft}$ is a worst possible stable marriage in the sense that it makes the happiest person as unhappy as possible. Yet, for every r, there is a pure strategy Nash equilibrium with outcome $M^{\circlearrowleft}$.

But that is not the worst implication of the result. The worst thing is that M may not be stable with respect to the sincere preferences. The only marriages not attainable by pure Nash equilibria are those that are egregiously unstable by assigning to at least one person an individual whom they are unwilling to marry. Therefore, all of the unstable marriages, as long as they are rational, are outcomes of pure Nash equilibria of SSM, no matter what the stable marriage mechanism r. *We believe that this has been the principal obstacle to progress on the general question of strategic stable marriage.* From a normative point of view, this obstacle is deleterious because there is no mechanism r that guarantees a stable marriage when players are strategic about their private preference information. It is also unfortunate from a empirical point of view, for it inhibits predictive power. The equilibria include bizarre ones wherein people omit all but their least favored partner from their submitted preference list, and thereby get a poor outcome. To accommodate for the bizarre equilibria we once again apply the minimal dishonest refinement from Section 2.3 to SSM.

We also require that r, when cast as maximizing a societal value f, satisfy two rationality criteria, f must be monotonic and independent of non-spouses (INS). Monotonicity in this context means that the societal value of a marriage M does not decrease if an individual increases their preference for the spouse M assigns to them. INS in this context means that the societal value does not change if an individual exchanges two non-spouses in their preference list.

With these conditions, we show that stability is assured. Specifically, we show (Theorem 6.3.14) that if r is monotonic and INS representable, then for every minimally dishonest equilibrium $\bar{\Pi}$, the marriage $r(\bar{\Pi})$ is stable with respect to the sincere preferences Π.

To clarify the definition of a monotonic INS representable decision mechanism, it is only necessary that r has a *representation* that is monotonic and INS. The mechanism r must behave as if society has a value for all marriages and r chooses a stable marriage with maximum value. For example, in the case of the Gale-Shapley algorithm, there is no *a priori* defined societal value. However, the Gale-Shapley algorithm always selects the man-optimal marriage and there exists a function that the man-optimal marriage uniquely optimizes among the set of stable marriages [31]. Furthermore, this function is monotonic and INS and therefore the Gale-Shapley algorithm is a monotonic INS representable stable marriage mechanism. With the exception of the sex-equal decision mechanism that minimizes the absolute difference between the utility of men and women, we provide a monotonic INS representation for every stable marriage mechanism (Theorem 6.3.13) we have found to be commonly referenced in the literature [31, 36]. These include the Gale-Shapley algorithm, egalitarian stable marriage mechanisms, minimum regret stable marriage mechanisms and decision mechanisms that maximize the minimum of men and women's utilities.

We believe that Theorem 6.3.14 has the potential for significant impact on applications of stable marriage. As mechanism designers, we no longer need to be limited to only the man-optimal or woman-optimal marriage if we desire a sincerely stable outcome. Instead we now have a wide array of algorithms to choose from, algorithms that are not completely biased against either the women or the men. It is true that the man-optimal algorithm has the additional property that men will tend to reveal their sincere preferences. However, the goal of mechanism design should be a good societal outcome, rather than the elicitation of honesty. If the latter were the goal, mechanism design would be solved by the revelation principle.

In addition to raising the general issue of strategic submission of private information in Stable Marriage Problems, Gusfield and Irving [31] ask specifically for an algorithm that always selects a sincere egalitarian stable marriage even when individuals behave strategically. Here we prove a negative result: there is no monotonic INS representable r such that there is always a minimally dishonest equilibrium of SSM and where for every minimally dishonest equilibrium $\bar{\Pi}$ the marriage $r(\bar{\Pi})$ is an egalitarian stable marriage with respect to the sincere preferences (Theorem 6.3.15).

This means that in SSM, no r with these rationality criteria is guaranteed to always get an egalitarian stable marriage from a minimally dishonest equilibrium. Theorem 6.3.15 doesn't mean that no algorithm r ever yields a sincere egalitarian stable marriage. Indeed, we prove that there are mechanisms r for which the egalitarian marriage does come from a minimally dishonest equilibrium. But there may be other minimally dishonest equilibria that yield different marriages.

6.1.1 The Gale-Shapley Algorithm and Never-One-Sided Mechanisms

Next, we take a closer examination of the Gale-Shapley algorithm. When Gale and Sotomayor first showed that when men are honest any stable marriage can be obtained, they also considered the set of strong equilibrium points [26]. Strong equilibrium points are obtained from a refinement on the set of equilibria – in addition to no individual being able to alter their preferences to get a better outcome, there also cannot be a coalition that can collude to strictly improve the outcome for every coalition member.

Gale and Sotomayor show that the woman-optimal marriage is the outcome of a strong equilibrium when using the Gale-Shapley algorithm [26]. Given that honesty is a best response for men, and that when men are honest, the women can always work together to obtain the woman-optimal marriage, one could hope that *all* strong equilibria yield the woman-optimal marriage. However, Gale and Sotomayor prove an additional result which they remark is "unfortunate": the woman-optimal marriage is not the only marriage obtained from the set of strong equilibria.

Minimal dishonesty provides the perverse, yet satisfying result that Gale and Sotomayor desired. When both men and women are minimally dishonest, the Gale-Shapley algorithm produces the woman-optimal sincerely stable marriage at all equilibria (Theorem 6.4.2). Moreover, we show there is always a minimally dishonest equilibrium.

We also consider mechanisms that are fairer than the biased Gale-Shapley algorithm. A stable marriage mechanism r is *never-one-sided* if, when given a preference profile Π with at least two stable marriages, there is positive probability of avoiding the man-optimal marriage and positive probability of avoiding the woman-optimal marriage.

Let r be an arbitrary monotonic, INS and never-one-sided representable stable marriage algorithm. Let M be a marriage and let Π be a sincere preference profile. We show that M is sincerely stable if and only if there exists a minimally dishonest equilibrium $\bar{\Pi}$ in the strategic stable marriage game where $r(\bar{\Pi}) = M$ (Theorem 6.4.4).

6.1.2 Extensions to Student Placement and College Admissions

We consider extending our results to both the Admissions Problem and the Student Placement Problem.

The *College Admissions Problem* is a polyandrous generalization of the Stable Marriage Problem. Each man may be married to at most one woman, but each woman w may be married to any number of men up to her quota q_w. It is traditional to refer to the men as students and the women as colleges. Both have preferences over the other set and stability is still a necessary condition. The most famous application of the Admissions Problem is the Nobel Prize winning National Resident Matching Program (NRMP) [59] which assigns residents to hospitals.

The *Student Placement Problem* is a variant of the Admissions Problem where the colleges' preferences are publicly known, or, equivalently, colleges must be honest. Applications of the Student Placement Problem include assigning students to universities in Turkey [6], and to primary schools in New York [1], and Boston [2]. If players are honest, both the admissions and Student Placement Problems can readily be transformed into Stable Marriage Problems by making multiple copies of the colleges. The question at hand is whether sincerely stable marriages can be obtained if players are strategic.

We prove that Theorem 6.3.14 also holds for student placement. As long as a monotonic INS representable stable marriage mechanism is used, the outcome will be a sincerely stable marriage. In addition, the result still holds even if some of the students commit to telling the truth even when they could act strategically to get a better outcome. As a consequence, we urge the community to consider mechanisms besides the Gale-Shapley algorithm for assigning students to schools. At the very least, the desire for a sincerely stable marriage should not be the sole reason for selecting the Gale-Shapley algorithm. Rather, we should seek to obtain a sincerely stable marriage that optimizes societal utility in some sense.

Roth has long claimed that the Admissions Problem is not equivalent to the Stable

Marriage Problem [60]. He proves that, unlike in the Stable Marriage Problem, a single college may be able to alter its preferences to obtain a marriage that it prefers over the college-optimal marriage. Hence, when students and colleges are strategic, the Gale-Shapley algorithm may yield an equilibrium marriage that is not sincerely stable. We prove a considerably more general statement: *When students and colleges are strategic, no stable marriage algorithm always yields an equilibrium marriage that is sincerely stable, with or without minimal dishonesty refinement*. This result strongly confirms Roth's distinction between college admissions and the other Stable Marriage Problems.

6.1.3 Other Extensions

Finally, we describe some other extensions which generalize results or weaken assumptions. All of the results hold if there are unequal numbers of men and women. The minimal dishonesty refinement can be replaced by a weaker "locally minimal dishonesty" refinement without affecting the results. Individuals are minimally dishonest if every change to their putative preferences that increases honesty decreases utility. Individuals are *locally* minimally dishonest if every *small* change to their putative preferences that increases honesty decreases utility.

The results also extend to strategic stable marriage games that permit collusion. Permitting collusion is equivalent to the strong equilibrium refinement discussed in Section 6.1.1. We show that every minimally dishonest equilibrium is also a strong equilibrium. Thus, minimal dishonesty is sufficient to guarantee that a coalition cannot manipulate the outcome of an equilibrium.

6.1.4 Related Literature

The Stable Marriage Problem was originally examined by Gale and Shapley [25]. They proved the existence of stable marriages via the Gale-Shapley algorithm that completes in $O(n^2)$. The Stable Marriage Problem has since received much attention and has been applied to a variety of areas including the Nobel Prize winning work by Roth and Shapley on the theory of stable allocations and the practice of market design including the design of the National Resident Matching Program (NRMP) [51]. Other applications include:

Canadian Resident Matching Service (CVaRMS) [17],

Japan Residency Matching Program (JRMP) [37],

US Navy has a web-based multi-agent system for assigning sailors to ships [41],

Design of the kidney exchange market [3],

Assigning stundents to universities in Turkey [6],

Assigning students to primary schools in New York [1] and Boston [2].

Types of Stable Marriages

The Gale-Shapley algorithm [25] proves the existence of a stable marriage by finding the man-optimal stable marriage – the stable marriage that is simultaneously best for all men [25]. Roth later shows that the man-optimal stable marriage is also woman-pessimal [58] – the man-optimal marriage is the stable marriage that is simultaneously worst for all women.

This suggests that there may be more than one stable marriage. This is true and Knuth [39] motivates a study on the number of stable marriages. Irving and Leather [34] show that if there are n men and n women where $n = 2^k$ then there is an instance with at least $2.28^n/(1+\sqrt{3})$ stable marriages. Although the number of stable marriages can therefore be exponential, Pittel proves that the expected number of stable marriages is asymptotic to $e^{-1}n\ln n$ [55]. No non-trivial upper bounds are known.

With so many stable marriages, it is natural that we would look for a "best stable marriage". A natural partial ordering on the set of stable marriages has $M \geq M'$ if every man likes his M-partner at least as much as his M'-partner. Knuth proves that this partial ordering induces a distributive lattice [39]. Furthermore, Knuth shows that the lattice remains intact if $\geq$ is replaced with $\leq$ and if *man* is replaced with *woman*. Thus for two stable marriages M and M' if there is man who prefers M then there is a woman who prefers M' and there may be no universally agreed-upon "best" stable marriage.

Several types of stable marriages are often considered in the literature. Most mechanisms depend on the ordinal preferences $\bar{\pi}_y$ for each individual y. However, there are mechanisms that depend on a cardinal cost $c_y(z)$ individual y assigns to being married to individual z. Specific types of stable marriages are often defined as optimizers of an objective. Letting $\sigma(X, x) = i$ if x is in the i^{th} position in X, we give five common examples below.

Man-Optimal	$\min \sum_{\{m,w\} \in M} \sigma(\bar{\pi}_m, w)$
Minimum Regret	$\min \max_{\{m,w\} \in M} \{\max\{\sigma(\bar{\pi}_m, w), \sigma(\bar{\pi}_w, m)\}\}$
Egalitarian	$\min \sum_{\{m,w\} \in M} (\sigma(\bar{\pi}_m, w) + \sum_{\{m,w\} \in M} \sigma(\bar{\pi}_m, w))$
Optimal Marriage	$\min \sum_{\{m,w\} \in M} (c_m(w) + c_w(m))$
Sex-Equal	$\min \lvert \sum_{\{m,w\} \in M} \sigma(\bar{\pi}_m, w) - \sum_{\{m,w\} \in M} \sigma(\bar{\pi}_m, w) \rvert$

The man-optimal marriage minimizes the sum of the men's costs and is found by the Gale-Shapley algorithm in $O(n^2)$ time. A minimum regret stable marriage maximizes the happiness of the least happy individual. Gusfield shows that a minimum regret stable marriage can also be found in $O(n^2)$ time [30]. An egalitarian stable marriage places equal weight on every individual and minimizes the sum of everyone's cost. The optimal marriage is a stable marriage that minimizes the sum of all individual costs, where the costs are specified by the individuals' cardinal preferences. Irving, Leather and Gusfield give a polyhedral description of the set of stable marriages [35] that allows for a $O(n^4)$ ($O(n^4 \ln n)$) algorithm to find an egalitarian (optimal respectively) stable marriage. A sex-equal marriage is a stable marriage that minimizes the absolute difference between the cost for the men and the cost for the women. Gusfield and Irving proposed the problem of finding a polynomial time algorithm to give a sex-equal marriage as the 6th of 12 open problems for the Stable Marriage Problem in [31]. This problem has been resolved by Kato who proved that finding a sex-equal stable marriage is NP-hard [38].

Strategic Behavior in the Stable Marriage Game

Much of the literature related to the Stable Marriage Problems deals with the many-to-one marriages seen in the college admissions and Student Placement Problems. In these problems each woman w is allowed to marry up to q_w men. All decision mechanisms that select a stable marriage for the college Admissions Problem can be simulated by a one-to-one stable marriage mechanism after creating q_w copies of each woman. Thus it is unsurprising that many properties for the Stable Marriage Problem also apply to the many-to-one marriage problems [31]. For instance, Gale and Sotomayor show that if an individual is unmarried in one stable marriage then that individual is unmarried in every stable marriage [26]. This results translates to the many-to-one settings and guarantees that every college will be assigned the same number of applicants in every stable marriage [61]. While the college Admissions Problem can be simulated via the Stable Marriage Problem, the two are vastly different once we consider manipulation.

Regarding the Stable Marriage Problem, Roth shows that every stable marriage mechanism is manipulable [57]. Men have no incentive to lie if the decision mechanism always selects the man-optimal marriage [20] (Gale-Shapley algorithm). In addition, if the men are honest, the algorithm will select a sincerely stable marriage [26]. Furthermore, the women can collectively manipulate their preferences to obtain the woman-optimal (or any other stable) marriage. However, in general this is not the only equilibrium even if we only examine "strong" equilibria (where no coalition can collectively strategize to strictly improve the outcome for each coalition member) [IBID].

Roth's result showing that every stable marriage mechanism is manipulable also applies to the college Admissions Problem. However, it does not apply to to the Student Placement Problem. In the Student Placement Problem, women by definition are always honest. Since it is in the best interest of men to be honest when running the Gale-Shapley algorithm, there exists an incentive compatible mechanism.

For the Stable Marriage Problem, it is known that no subset of men can manipulate their preferences to obtain a result they all strictly prefer to the man-optimal marriage [31]. Once again, this result also applies to the Student Placement Problem. However, Roth showed that the result does not hold for the college Admissions Problem – a single college may be able to alter its preferences to obtain a marriage that it strictly prefers to the college-optimal marriage [60]. This raises the disquieting possibility that the results from [26] may not apply to the college Admissions Problem and that the decision mechanism may select a marriage that is sincerely unstable at an equilibrium even when we use the Gale-Shapley algorithm.

Other literature considers manipulation in a different way. In the computational social choice literature, strategic behavior is analyzed for its computational complexity, the worst-case amount of computational effort needed to find a beneficial manipulation [9]. Teo proves that the Gale-Shapley algorithm is computationally easy to manipulate [67]. Pini proves that certain other mechanisms are NP-hard to manipulate [54]. However, with the exception of the Gale-Shapley algorithm, it is not known what type of marriage is obtained at an equilibrium, or even if the marriage will be sincerely stable.

6.2 The Model

We redefine the Game of Deception and minimal dishonesty from Chapter 2 as they pertain to the Stable Marriage Problem.

Definition 6.2.1. An instance of the Stable Marriage Problem consists of

- Sets $V = \{m_1, m_2, ...m_n\}$ and $W = \{w_1, w_2, ...w_n\}$ of "men" and "women", respectively.
- For each $m \in V$, a total ordering π_m on a subset of W, representing man m's strict preferences on the subset of women he is willing to be married to; symmetrically, for

each $w \in W$, a preference list π_w.

For all $y \in V \cup W$, if α precedes β in π_y, we write $\alpha \succ_{\pi_y} \beta$ to indicate that y prefers α to β. Following tradition, we use u to represent "unmarried" and we extend π_y to a partial order on $W \cup \{u\}$ or $V \cup \{u\}$ as follows:

- $\alpha \succ_{\pi_y} u$ for all α whom y is willing to marry;
- $u \succ_{\pi_y} z$ for all z of the other gender whom y is not willing to marry;
- Enforce transitivity of preferences;
- Denote by Π the collection of π_y over all $y \in V \cup W$. π is called the *preference profile*.

Definition 6.2.2. A *marriage* is a partial injection $M \subset V \times W$. That is, M is a set of disjoint pairs $\{m, w\} \in V \times W$. If $\{m, w\} \in M$, we say M weds m to w, or equivalently M weds w to m, and we define the *M-partner* of w to be $s_M(w) = m$ and the M-partner of m to be $s_M(m) = w$. If for $y \in V \cup W$ there is no $z \in V \cup W$ such that $\{y, z\} \in M$ or $\{z, y\} \in M$, then we say y is *unmatched* or *unmarried* and we define the M-partner of y to be $s_M(y) = u$.

Definition 6.2.3. A marriage M is *individually rational* with respect to Π if $s_M(y) \succeq_{\pi_y} u$ for every individual y.

Definition 6.2.4. A pair $\{m, w\}$ is a blocking pair for M with respect to $\bar{\pi}$ if $w \succ_{\pi_m} s_M(m)$ and $m \succ_{\pi_w} s_M(w)$.

Definition 6.2.5. The marriage M is *stable* with respect to Π if M is rational and has no blocking pairs according to Π.

Stability is a necessary condition for any solution [39, 31, 45]. If either condition does not hold, the marriage is unstable. If the marriage is not rational, some y prefers being unmarried over their M-partner and would leave $s_M(y)$ to be unmarried. If there is a blocking pair $\{m, w\}$, then both m and w prefer each other to their respective M-partners and would leave their M-partners to be together.

A stable marriage mechanism takes a preference profile Π, selects a stable marriage M and assigns each individual their M-partner. The process for selecting M is not necessarily deterministic. Let S be the set of all marriages (not necessarily stable) between V and W. Let $\mathbb{P}$ be the set of all possible preference profiles for V and W. Every stable marriage mechanism that selects a stable marriage based on Π can be characterized by

M.1 A discrete sample space Ω comprising a finite set of atoms $\omega \in \Omega$;

M.2 A probability measure μ on Ω where $\mu(\omega) > 0$ for all atoms $\omega \in \Omega$,

M.3 A function $f : S \times \mathbb{P} \times \Omega \to \mathbb{R}$,

where the stable marriage mechanism selects $\omega \in \Omega$ according to μ and then selects a marriage stable with respect to Π that maximizes $f(M, \Pi, \omega)$. All tie-breaking can be incorporated into f, Ω and μ. Thus, without loss of generality, we require

M.4 $|\operatorname{argmax}_{M \in S}\{f(M, \Pi, \omega) : M \text{ } is \text{ } stable\}| = 1 \ \ \forall \omega \in \Omega.$

Let $r(\Pi)$ be the set of marriages that have positive probability of being selected when using the stable marriage mechanism $r = \{f, \Omega, \mu\}$. That is,

$$r(\Pi) = \bigcup_{\omega \in \Omega:} \left\{ \operatorname*{argmax}_{M \in S}\{f(M, \Pi, \omega) : M \text{ } is \text{ } stable\} \right\}. \tag{6.1}$$

As defined, r is not necessarily deterministic and $|r(\Pi)|$ may be greater than one. Corollary 6.3.3 establishes that $|r(\bar{\Pi})| = 1$ at every equilibrium. After completing the proof of Corollary 6.3.3, we treat $r(\bar{\Pi})$ as a singleton if $\bar{\Pi}$ is an equilibrium.

The two most commonly studied types of stable marriages are the man-optimal and woman-optimal stable marriages, and egalitarian stable marriages. Without loss of generality, by gender symmetry the Gale-Shapley algorithm always selects the man-optimal marriage with respect to the submitted preference profile.

Example 6.2.6. $r_1 = \{f_1, \Omega, \mu\}$ that selects the man-optimal marriage (Gale-Shapley algorithm).

Define

$$\sigma(X, x) = i \text{ if } x \text{ is in the } i^{th} \text{ position in } X. \tag{6.2}$$

The man-optimal marriage *with respect to* Π *is the stable marriage that maximizes* $f_1(M, \Pi, \omega) = -\sum_{\{m,w\}\in M} \sigma(\pi_m, w)$ *for all* $\omega \in \Omega$ *[57]. The man-optimal marriage is known to be unique and the decision mechanism* $r_1 = \{f_1, \Omega, \mu\}$ *selects the man-optimal marriage regardless of the selection of* Ω *and* μ*. It is also known [25] that in the man-optimal marriage, each man weds his most-preferred woman among all women he weds in the set of all stable marriages.*

Example 6.2.7. $r_2 = \{f_2, \Omega, \mu\}$ that selects a egalitarian marriage uniformly at random.

Define $\sigma(X, x)$ *as in Equation 6.2. An* egalitarian marriage *with respect to* Π *is a stable marriage that maximizes* $g(M, \Pi) = -\sum_{\{m,w\}\in M} \left(\sigma(\bar{\pi}_m, w) + \sigma(\bar{\pi}_w, m)\right)$ *[31]. Let* Ω *be the set of all permutations of all possible marriages, let* $f_2(M, \Pi, \omega) \equiv g(M, \Pi) + \frac{\sigma(\omega, M)}{|\omega|+1}$ *and let* $\mu(\omega) = \mu(\omega')$ *for all* $\omega, \omega' \in \Omega$*. We claim that* $r_2 = \{f_2, \Omega, \mu\}$ *selects an egalitarian marriage uniformly at random.*

Since $\frac{\sigma(\omega, M)}{|\omega|+1} < 1$*,* M *maximizes* $f_2(M, \Pi, \omega)$ *only if* M *also maximizes* $g(M, \Pi)$*. Hence* $r(\Pi)$ *is a subset of the set of egalitarian marriages. Since* $\sigma(\omega, M)$ *takes a unique value for each* M*,* r_2 *always selects a unique marriage given* $\omega \in \Omega$ *and* r_2 *is well-defined.*

Let $p(M)$ *be the probability that the marriage* M *is selected by* r *and* $\mathcal{M}$ *be the set of all stable marriages. By symmetry,* $p(M) = p(M')$ *for all* $M, M' \in \mathcal{M}$*. Therefore* $r_2 = \{f_2, \Omega, \mu\}$ *selects an egalitarian marriage uniformly at random.*

Individuals may act strategically and submit insincere preferences to r. We examine the Strategic Stable Marriage game to analyze the effect of strategic behavior.

Strategic Stable Marriage Game (SSM)

- Each individual i has information $\pi_i \in \mathcal{P}_i$ describing their preferences. The collection of all information is the (sincere) profile Π.
- To play the game, individual i submits putative preference data $\bar{\pi}_i \in \mathcal{P}_i$. The collection of all submitted data is denoted $\bar{\Pi}$.
- It is common knowledge that a central decision mechanism will select the marriage $r(\bar{\Pi}, \omega)$ that is stable with respect to $\bar{\Pi}$ when given the random event ω.
- The random event $\omega \in \Omega$ is selected according to μ. We denote $r(\bar{\Pi})$ as the distribution of outcomes according to Ω and μ.
- Individual i evaluates $r(\bar{\Pi}, \omega)$ according to i's partner in the marriage $r(\bar{\Pi}, \omega)$ and i's sincere preferences π_i.

Definition 6.2.8. A marriage M is a sincerely (putatively) stable marriage is if it is stable with respect to Π ($\bar{\Pi}$ respectively).

Example 6.2.9. Honesty can fail to be an equilibrium strategy in SSM

Let there be two men, m_1, m_2 and two women w_1, w_2. Let the sincere preferences π be

$$\pi_{m_1} = w_1 \succ_{\pi_{m_1}} w_2 \qquad \pi_{w_1} = m_2 \succ_{\pi_{w_1}} m_1 \tag{6.3}$$

$$\pi_{m_2} = w_2 \succ_{\pi_{m_2}} w_1 \qquad \pi_{w_2} = m_1 \succ_{\pi_{w_2}} m_2 \tag{6.4}$$

There are two sincerely stable marriages given by $M_1 = \{\{m_1, w_1\}, \{m_2, w_2\}\}$ and $M_2 = \{\{m_1, w_2\}, \{m_2, w_1\}\}$. The Gale-Shapley algorithm would yield M_1, which is the man-optimal marriage in the sense that each man is married to his most-preferred woman among all those to *whom he could be married to in some stable marriage. Were Gale-Shapley run with the roles of women and men reversed, it would yield M_2, which is the woman-optimal marriage. In this small example there are no other stable marriages.*

Suppose $M_1 \in r(\Pi)$. If all players are truthful (if $\bar{\Pi} = \Pi$) then there is positive probability of selecting M_1. However, $\bar{\Pi} = \Pi$ would not be a Nash equilibrium in the game SSM, because w_1 can improve her outcome by altering her putative preference data to $\bar{\pi}_{w_1} = m_2$. The only stable marriage with respect to the altered preference data would be M_2. The other marriages would not be stable by the following reasoning: M_1 is not rational because w_1 prefers u to m_1; $\{\{m_1, w_2\}\}$ is not stable because m_2 and w_1 both prefer each other to u; $\{\{m_2, w_2\}\}$ is unstable because $\{m_1, w_2\}$ is a blocking pair. Therefore r will select M_2, an outcome that w_1 prefers to M_1. A symmetric argument proves that if $r(\Pi) = M_2$ the SSM game is not at equilibrium at $\bar{\Pi} = \Pi$. Therefore no algorithm r makes SSM an incentive-compatible game.

We continue Example 6.2.9 by exhibiting a Nash equilibrium to this instance of SSM.

Example 6.2.10. An equilibrium strategy $\bar{\pi}$ in an instance of SSM

For the SSM game defined in Example 6.2.9, consider the strategically submitted putative preferences $\bar{\pi}$:

$$\bar{\pi}_{m_1} = w_1 \qquad \bar{\pi}_{w_1} = m_2 \succ_{\bar{\pi}_{w_1}} m_1 \tag{6.5}$$

$$\bar{\pi}_{m_2} = w_2 \qquad \bar{\pi}_{w_2} = m_1 \succ_{\bar{\pi}_{w_2}} m_2 \tag{6.6}$$

The only stable marriage with respect to $\bar{\pi}$ is M_1, which weds each man to his sincerely most-preferred woman. Therefore neither man can improve his outcome by submitting different preference data. If w_1 alters $\bar{\pi}_{w_1}$ to m_1 or $m_1 \succ_{\bar{\pi}_{w_1}} m_2$, M_1 continues to be the only putative stable marriage. Any algorithm r must select M_1 and so w_1 receives the same outcome. If w_1 were to alter $\bar{\pi}_{w_1}$ to m_2 or the empty list $\emptyset$, the only stable marriage would be $\{\{m_2, w_2\}\}$, which is a worse sincere outcome for w_1. By symmetry, w_2 cannot alter $\bar{\pi}_{w_2}$ to get a better outcome for herself. Hence $\bar{\Pi}$ is an equilibrium set of strategies for this instance of SSM.

6.3 Equilibria of the Strategic Stable Marriage Game

We begin by examining pure strategy equilibria of SSM. Given that it is well known that there is no strategy-proof stable marriage mechanism, it is unsurprising that there are equilibria with sincerely unstable outcomes.

Example 6.3.1. Strategic behavior can lead to unstable outcomes.

For three men a, b, c and three women w, x, y let the true preferences be given by Π and the putative preferences be be given by $\bar{\Pi}$ below.

$$\pi_{m_1} = w_1 \succ_{\pi_{m_1}} w_2 \succ_{\pi_{m_1}} w_3 \qquad \pi_{w_1} = m_3 \succ_{\pi_{w_1}} m_1 \succ_{\pi_{w_1}} m_2 \tag{6.7}$$

$$\pi_{m_2} = w_2 \succ_{\pi_{m_2}} w_3 \qquad \pi_{w_2} = m_2 \succ_{\pi_{w_2}} m_3 \succ_{\pi_{w_2}} m_1 \tag{6.8}$$

$$\pi_{m_3} = w_3 \succ_{\pi_{m_3}} w_1 \succ_{\pi_{m_3}} w_2 \qquad \pi_{w_3} = m_1 \succ_{\pi_{w_3}} m_3 \succ_{\pi_{w_3}} m_2 \tag{6.9}$$

$$\bar{\pi}_{m_1} = w_1 \succ_{\bar{\pi}_{m_1}} w_2 \qquad \bar{\pi}_{w_1} = m_3 \tag{6.10}$$

$$\bar{\pi}_{m_2} = w_3 \qquad \bar{\pi}_{w_2} = m_1 \tag{6.11}$$

$$\bar{\pi}_{m_3} = w_1 \qquad \bar{\pi}_{w_3} = m_2 \tag{6.12}$$

The stable marriages with respect to Π are the man-optimal $M_1 = \{\{m_1, w_1\}, \{m_2, w_2\}, \{m_3, w_3\}\}$ and woman-optimal $M_2 = \{\{m_1, w_3\}, \{m_2, w_2\}, \{m_3, w_1\}\}$. The only stable marriage with respect to $\bar{\Pi}$ is $M = \{\{m_1, w_2\}, \{m_2, w_3\}, \{m_3, w_1\}\}$. Hence, regardless of r, r will choose M even though $\{\{m_1, w_1\}, \{m_2, w_2\}, \{m_3, w_3\}\}$ dominates it with respect to Π.

Moreover, $\bar{\Pi}$ is a Nash equilibrium in SSM. No man can improve his outcome because he is now married to the only woman willing to marry him, and he prefers her to being unmarried. Similarly, neither w_2 nor w_3 can improve her outcome. Woman w_1 cannot improve her outcome because she is now married to her (sincerely) most-preferred man.

We show that at every equilibrium a marriage will be selected deterministically regardless of which stable marriage mechanism is used. We build on this determinism to prove that a marriage M is the outcome of an SSM equilibrium iff M is rational with respect to the sincere preferences.

Lemma 6.3.2. *Let r and Π be arbitrary. Suppose $\bar{\Pi}$ is an equilibrium of the Strategic Stable Marriage game (SSM). Then there cannot exist a marriage $M \in r(\bar{\Pi})$, another putatively stable marriage M' (i.e. M' is stable with respect to $\bar{\pi}$), and an individual $y \in V \cup W$ who prefers their M'-partner to their M-partner, i.e., such that $s_{M'}(y) \succ_{\pi_y} s_M(y)$. Furthermore, for all $M \in r(\bar{\Pi})$, M is individually rational with respect to Π.*

Proof. We prove the first statement by contradiction. Assume such $\bar{\Pi}$, $M \in r(\bar{\Pi})$, M' and y exist. Without loss of generality make the maximality assumption that M' is such that $s_{M'}(y)$ is y's most preferred partner among all putatively stable marriages.

As in previous chapters, let $[\bar{\Pi}_{-y}, \bar{\pi}'_y]$ denote the profile obtained after replacing $\bar{\pi}_y$ with $\bar{\pi}'_y$ in the profile $\bar{\Pi}$.

Suppose y were to strategically submit $\bar{\pi}'_y$ consisting only of $s_{M'}(y)$. The marriage M' would be stable with respect to $[\bar{\Pi}_{-y}, \bar{\pi}'_y]$ since deleting elements from a preference list cannot create a new blocking pair. By the Rural Hospitals Theorem [61] (Lemma 6.6.1), y is married in every marriage that is stable with respect to $[\bar{\Pi}_{-y}, \bar{\pi}'_y]$. Moreover, in each of these marriages y weds $s_{M'}(y)$ because, according to $\bar{\pi}'_y$, y is not willing to wed anyone else. If instead y submits $\bar{\pi}_y$, $M \in r(\bar{\Pi})$ and property M.2 imply that y weds the less preferred $s_M(y)$ with strictly positive probability. By the maximality assumption, $r(\bar{\Pi})$ never weds y to someone y prefers to $s_{M'}(y)$. Therefore, y strictly prefers to submit $\bar{\pi}'_y$ than to submit $\bar{\pi}_y$, contradicting $\bar{\Pi}$ being a Nash equilibrium.

To prove the second statement, suppose that $M \in r(\bar{\Pi})$ is not individually rational with respect to Π. Then for some individual y, M weds y to someone y is sincerely unwilling to marry. That is, $u \succ_{\pi_y} s_M(y)$. The first part of the lemma implies that for all $M' \in r(\bar{\Pi})$, $u \succ_{\pi_y} s_{M'}(y)$. In words, r always weds y to someone y is sincerely unwilling to marry. Therefore, y can get the strictly better outcome u by submitting an empty preference list, contradicting $\bar{\Pi}$ being a Nash equilibrium. $\square$

Corollary 6.3.3. *For every stable marriage mechanism r and every equilibrium $\bar{\Pi}$ of the SSM game, the putatively stable marriage is deterministically selected.*

Proof. Suppose to the contrary there are $M, M' \in r(\bar{\Pi})$ and an individual y where $s_M(y) \neq s_{M'}(y)$. By the second part of Lemma 6.3.2, M and M' are individually rational. Since preferences are strict, $s_M(y) \succ_{\pi_y} s_{M'}(y)$ or $s_{M'}(y) \succ_{\pi_y} s_M(y)$, a contradiction to the first part of Lemma 6.3.2. □

Theorem 6.3.4. *For every instance of SSM, regardless of the stable marriage mechanism r, there exists a pure strategy Nash equilibrium whose outcome is the marriage M if and only if M is individually rational (with respect to the sincere preferences π).*

Proof. Lemma 6.3.2 guarantees that every possible outcome M of r is individually rational. It remains to show existence. Let M be a rational marriage with respect to Π. For each married individual y, let $\bar{\pi}_y = s_M(y)$; for each unmarried individual y let $\bar{\pi}_y = \emptyset$. M is the only stable marriage with respect to $\bar{\pi}$ and hence, regardless of r, $r(\bar{\Pi}) = M$. If $s_M(y) = u$, the putative profile $\bar{\Pi}$ indicates that no one is willing to wed y. Hence y cannot improve by altering $\bar{\pi}_y$. If y is married in M, only $s_M(y)$ is putatively willing to marry y. Hence y can only alter his/her preferences to become unmarried. By rationality this is not an improvement for y. Hence $\bar{\Pi}$ is an equilibrium. □

6.3.1 Minimal Dishonesty

The proof of Theorem 6.3.4 can require individuals to lie in absurd ways to obtain results that they least prefer. For instance, Theorem 6.3.4 implies that everyone reporting that they would be prefer to be unmarried is a perfectly reasonable outcome regardless of the sincere profile. This is unsatisfactory in a predictive sense - any rational marriage can eventuate - and in a normative sense - as discussed in Section 2.3 individuals tend to lie only if they benefit. To improve the theory, we refine the set of equilibria to those where individuals are *minimally dishonest*. Empirical justification for this refinement was given in Section 2.3. Here we define a metric on the set of preference lists, and employ that metric to formally define minimal dishonesty (Definition 6.3.7). With this refinement, we will establish in Theorem 6.3.14 that every monotonic and INS stable marriage mechanism yields a sincerely stable marriage when individuals behave strategically.

As stated in the introduction, Gusfield and Irving specifically ask for a mechanism that selects an egalitarian stable marriage in the face of strategic behavior. We will give a negative answer, proving in Theorem 6.3.15 that no monotonic INS stable marriage mechanism always selects a sincere egalitarian stable marriage when strategic individuals are minimally dishonest.

According to minimal dishonesty, individuals prefer to be as little dishonest as possible without worsening their outcome. To apply this concept, one needs a way to measure dishonesty. Once again, we will use the Bubble Sort or Kendall Tau distance – the most common way to evaluate the distance between two ordered lists.

Definition 6.3.5 (Bubble Sort or Kendall Tau Distance.)**.** Let π_m^1 and π_m^2 be two preference lists over the set $W = \{w_1, w_2, \ldots, w_n\}$. Let $w_{n+1} \equiv u$ so that $\{W \cup \{u\}\} = \{w_1, w_2, \ldots, w_{n+1}\}$. For all $w_i, w_j \in \{W \cup \{u\}\}$ denote by $R(\pi_m, w_i, w_j)$ the relationship between w_i and w_j with respect to π_m. Then the *distance* between π_m^1 and π_m^2 is defined as

$$d(\pi_m^1, \pi_m^2) \equiv |\{i, j\} : 1 \leq i < j \leq n+1; R(\pi_m^1, w_1, w_2) \neq R(\pi_m^2, w_1, w_2)\}|. \tag{6.13}$$

Example 6.3.6. Calculating the distance between two preference lists.

Consider the following preference lists π_m and $\bar{\pi}_m$ for the set of women $W = \{w_1, w_2, w_3, w_4\}$:

$$\pi_m = w_1 \succ_{\pi_m} w_2 \succ_{\pi_m} w_3 \qquad\qquad \bar{\pi}_m = w_4 \succ_{\bar{\pi}_m} w_2 \tag{6.14}$$

With respect to the sincere preferences π_m, m sincerely prefers w_1 to all other partners and he is willing to marry w_1. However, in his submitted list $\bar{\pi}_m$, he indicates that he prefers w_4 and w_2 to w_1, he is indifferent between w_1 and w_3 and that he prefers being unmarried to w_1. Thus for each $z \in$

$\{w_2, w_3, w_4, u\}$, $\bar{\pi}_m$ and π_m describe different relationships between w_1 and z and $R_{\bar{\pi}_m}(w_1, z) \neq R_{\pi_m}(w_1, z)$. Similarly $\bar{\pi}_m$ and π_m indicate different relationships for each of the following pairs: $\{w_2, w_4\}, \{w_3, w_4\}, \{w_3, u\}$ and $\{w_4, u\}$. Thus there are 8 distinct pairwise disparities between $\bar{\pi}_m$ and π_m and $d(\bar{\pi}_m, \pi_m) = 8$.

The formal definition of minimal dishonesty can now be stated succinctly in the notation of Definition 6.3.5.

Definition 6.3.7. Let Π be the sincere preferences and let $\bar{\Pi}$ be an equilibrium in the Strategic Stable Marriage game. Individual y is *minimally dishonest* if $d(\bar{\pi}'_y, \pi_y) < d(\bar{\pi}_y, \pi_y)$ implies $r(\bar{\Pi}) = M \succ_{\pi_y} M'$ for some $M' \in r([\bar{\Pi}_{-y}, \bar{\pi}'_y])$.

If there is a $\bar{\pi}'_y$ such that $d(\bar{\pi}'_y, \pi_y) < d(\bar{\pi}_y, \pi_y)$ and y does not sincerely prefer $r(\bar{\Pi})$ to $r([\bar{\Pi}_{-y}, \bar{\pi}'_y])$, then y can obtain at least as good a result by submitting the more honest $\bar{\pi}'_y$. If an individual is able to be more honest and obtain at least as good a result, we assume the individual would do so since individuals prefer being honest. Thus we only examine *minimally dishonest equilibria* – equilibria where every individual is minimally dishonest.

Example 6.3.8. Not every equilibrium is a minimally dishonest equilibrium.

Recall the sincere preferences Π and putative preferences $\bar{\Pi}$ from Example 6.3.1.

$$\pi_{m_1} = w_1 \succ_{\pi_{m_1}} w_2 \succ_{\pi_{m_1}} w_3 \qquad \pi_{w_1} = m_3 \succ_{\pi_{w_1}} m_1 \succ_{\pi_{w_1}} m_2 \tag{6.15}$$

$$\pi_{m_2} = w_2 \succ_{\pi_{m_2}} w_3 \qquad \pi_{w_2} = m_2 \succ_{\pi_{w_2}} m_3 \succ_{\pi_{w_2}} m_1 \tag{6.16}$$

$$\pi_{m_3} = w_3 \succ_{\pi_{m_3}} w_1 \succ_{\pi_{m_3}} w_2 \qquad \pi_{w_3} = m_1 \succ_{\pi_{w_3}} m_3 \succ_{\pi_{w_3}} m_2 \tag{6.17}$$

$$\bar{\pi}_{m_1} = w_1 \succ_{\bar{\pi}_{m_1}} w_2 \qquad \bar{\pi}_{w_1} = m_3 \tag{6.18}$$

$$\bar{\pi}_{m_2} = w_3 \qquad \bar{\pi}_{w_2} = m_1 \tag{6.19}$$

$$\bar{\pi}_{m_3} = w_1 \qquad \bar{\pi}_{w_3} = m_2 \tag{6.20}$$

When using the Gale-Shapley algorithm with the putative preferences $\bar{\Pi}$, the marriage $M =$

$\{\{m_1, w_2\}, \{m_2, w_3\}, \{m_3, w_1\}\}$ *is selected. Example 6.3.1 established that* $\bar{\Pi}$ *is a Nash equilibrium. We now show that* $\bar{\Pi}$ *is a not minimally dishonest equilibrium, according to Definitions 6.3.5 and 6.3.7.*

Consider the preference list $\bar{\pi}'_{m_3} = w_3 \succ_{\bar{\pi}'_{m_3}} w_2 \succ_{\bar{\pi}'_{m_3}} w_1$*. Since* $d(\bar{\pi}_{m_3}, \pi_{m_3}) = 4$ *(the disparities are* $\{w_3, w_1\}, \{w_3, w_2\}, \{w_3, u\}, \{w_2, u\}$*) and* $d(\bar{\pi}'_{m_3}, \pi_{m_3}) = 1$ *(*$\{w_2, w_3\}$*), man* m_3 *would be more honest if he submitted* $\bar{\pi}'_{m_3}$*. The marriage* M *is the only stable marriage with respect to* $[\bar{\pi}_{-m_3}, \bar{\pi}'_{m_3}]$ *and* m_3 *obtains the same result if he is more honest. Thus* $\bar{\Pi}$ *is not a minimally dishonest equilibrium.*

In Example 6.3.8 we could equally well have let m_3 submit preference list $\bar{\pi}''_{m_3} = w_1 \succ_{\bar{\pi}''_{m_3}} w_3$, which also must result in the marriage M. This would disprove minimal dishonesty because $d(\bar{\pi}''_{m_3}, \pi_{m_3}) = 3 < 4 = d(\bar{\pi}_{m_3}, \pi_{m_3})$. One could argue that m_3 would more readily choose $\bar{\pi}''_{m_3}$ than $\bar{\pi}'_{m_3}$ because the former differs from $\bar{\pi}_{m_3}$ in a particularly simple way, namely by changing the location of one woman. Later, in Section 6.5, we define a weaker "local" version of minimal dishonesty, and show that our results extend to it.

We continue Example 6.3.8 by providing all minimally dishonest equilibria.

Example 6.3.9. Minimally dishonest equilibria can yield a sincerely stable marriage.

In Examples 6.2.9 and 6.3.8, we have seen both a non-equilibrium solution and an equilibrium solution that is not minimally dishonest. Let r *be the Gale-Shapley algorithm. Let the sincere preferences* Π *be as in Examples 6.3.1 and 6.3.8. By enumeration one can verify there are only two minimally dishonest equilibria. In both equilibria the men are honest. The two women's putative preference profiles are:*

$$\begin{aligned} \bar{\pi}_{w_1} &= m_3 \succ_{\bar{\pi}_{w_1}} m_1 \succ_{\bar{\pi}_{w_1}} m_2 && (6.21) \\ \bar{\pi}_{w_2} &= m_2 \succ_{\bar{\pi}_{w_2}} m_3 \succ_{\bar{\pi}_{w_2}} m_1 && (6.22) \\ \bar{\pi}_{w_3} &= m_1 \succ_{\bar{\pi}_{w_3}} m_2 && (6.23) \end{aligned}$$

and

$$\begin{aligned} \bar{\pi}_{w_1} &= m_3 \succ_{\bar{\pi}_{w_1}} m_2 && (6.24) \\ \bar{\pi}_{w_2} &= m_2 \succ_{\bar{\pi}_{w_2}} m_3 \succ_{\bar{\pi}_{w_2}} m_1 && (6.25) \\ \bar{\pi}_{w_3} &= m_1 \succ_{\bar{\pi}_{w_3}} m_3 \succ_{\bar{\pi}_{w_3}} m_2 && (6.26) \end{aligned}$$

both of which yield the woman-optimal sincerely stable $M_2 = \{\{m_1, w_3\}, \{m_2, w_2\}, \{m_3, w_1\}\}$. *We will later show in Theorem 6.4.2 that the woman-optimal outcome is not a coincidence.*

In what follows we will frequently consider swapping the locations of two members of a preference list. For convenience we employ the following notation:

$$\delta(\bar{\Pi}, y, \{z_{-1}, z_1\}) = \begin{cases} \bar{\Pi} \text{ after swapping } z_{-1} \text{ and } z_1 \text{ in } \bar{\pi}_y & \text{if } z_{\pm 1} \succ_{\bar{\pi}_y} u \\ \bar{\Pi} \text{ after removing } z_i \text{ from } \bar{\pi}_y & \text{if } z_i \succ_{\bar{\pi}_y} u, z_{-i} = u \\ \bar{\Pi} \text{ after adding } z_i \text{ to the end of } \bar{\pi}_y & \text{if } u \succ_{\bar{\pi}_y} z_i, z_{-i} = u \end{cases} \tag{6.27}$$

We require two more formal definitions, one for monotonicity, the other for INS.

Definition 6.3.10. A function f is monotonic if for every $\bar{\Pi} \in \mathbb{P}$, $\omega \in \Omega$, $M \in S$, and $\{y, z\} \in M$, then

$$f(\delta(\bar{\Pi}, y, \{z, z_1\}), M, \omega) \geq f(\bar{\Pi}, M, \omega) \text{ if } z_1 \succ_{\bar{\pi}_y} z. \tag{6.28}$$

Monotonicity means that if an individual y indicates that if they prefer their M-partner more than they previously indicated, the value of M should not decrease. Decision mechanisms are often used to maximize social welfare. Thus, it seems natural for the value of a marriage to respond positively (or at least not negatively) when an individual alters their preferences in this way.

Definition 6.3.11. A function f is independent of non-spouses (INS) if for every $\bar{\Pi} \in \mathbb{P}$, $\omega \in \Omega$, $M \in S$, and $\{y, z\} \in M$, then

$$f(\bar{\Pi}, M, \omega) = f(\delta(\bar{\Pi}, y, \{z_1, z_{-1}\}), M, \omega) \text{ if } z \neq z_{\pm 1} \text{ and } z_{\pm 1} \succ_{\bar{\pi}_y} u, \tag{6.29}$$

$$f(\bar{\Pi}, M, \omega) = f(\delta(\bar{\Pi}, y, \{z_1, u\}), M, \omega) \text{ if } z \neq z_1, u \succ_{\bar{\pi}_y} z_1. \tag{6.30}$$

INS here is an adaptation of the independence of irrelevant alternatives definition used in voting [4]. Essentially, it states that the value of a marriage should depend only on the spouses' valuations of each other. To be precise, INS requires (1) if y exchanges the positions of z_1 and z_{-1} then the value of any marriage that weds y to $z \notin \{z_1, z_{-1}\}$ doesn't change, and (2) if y adds z_1 onto the end of their preference list then the value of any marriage that does not wed y to $z \neq z_1$ remains unchanged. Although changing the position of z_1 and z_{-1} does not change the value of a marriage, it may alter the stability.

Definition 6.3.12. $r = \{f, \Omega, \mu\}$ is a monotonic INS representable stable marriage mechanism if f is both monotonic and INS.

Theorem 6.3.13. *Each of the following stable marriage selection criteria can be represented by a monotonic INS stable marriage mechanism:*

1. *Man-Optimal*
2. *Egalitarian*
3. *Maximize the minimum individual utility*
4. *Maximize the minimum of the women's total utility and the men's total utility*

when using any of the following tie-breaking rules:

a. *Uniformly at random*

b. *Lexicographically based on a predetermined list $\mathcal{L}$ of marriages*

c. *Lexicographically giving the lowest indexed m_i his most preferred partner available*

Proof. The proof of Theorem 6.3.13 presents a monotonic INS function g_i where the only stable marriages that maximize g_i are the stable marriages described by criterion i. Then a sample space Ω_j, a probability measure μ_j and a monotonic INS function t_j are given that discriminate between marriages according to tie-breaking rule j. Finally, $f_{ij} = g_i + t_j$ is such that $r_{ij} = \{f_{ij}, \Omega_j, \mu_j\}$ selects a stable marriage satisfying criterion i according to tie-breaking rule j.

Define $\sigma(X, x)$ as in Equation 6.2. Let $S(\Pi)$ be the set of marriages stable with respect to Π. By definition [31], the marriages satisfying the above criteria are described by the following sets respectively:

1. $\mathcal{S}^1 = \underset{M \in S(\Pi)}{\operatorname{argmax}} \left\{ g_1(\pi, M) = \sum_{\{m,w\} \in M} -\sigma(\pi_m, w) \right\}$

2. $\mathcal{S}^2 = \underset{M \in S(\Pi)}{\operatorname{argmax}} \left\{ g_2(\pi, M) = \sum_{\{m,w\} \in M} \Big(-\sigma(\pi_m, w) - \sigma(\pi_w, m) \Big) \right\}$

3. $\mathcal{S}^3 = \underset{M \in S(\Pi)}{\operatorname{argmax}} \Big\{ g_3(\pi, M) = \min_{\{m,w\} \in M} \min \Big\{ -\sigma(\pi_m, w), -\sigma(\pi_w, m) \Big\} \Big\}$

4. $\mathcal{S}^4 = \underset{M \in S(\Pi)}{\operatorname{argmax}} \left\{ g_4(\pi, M) = \min \left\{ \sum_{\{m,w\} \in M} -\sigma(\pi_m, w), \sum_{\{m,w\} \in M} -\sigma(\pi_w, m) \right\} \right\}$

We show that the tie-breaking rules are given by the following functions, sample spaces and probability measures:

a. $t_{\mathrm{a}}(M, \Pi, \omega) = \dfrac{\sigma(\omega, M)}{|\omega| + 1}$ *where* Ω_{a} *is the set of all permutations of* S *and* $\mu_{\mathrm{a}}(\omega) = \mu_{\mathrm{a}}(\omega')$ *for all* $\omega, \omega' \in \Omega_{\mathrm{a}}$.

b. $t_{\mathrm{b}}(M, \Pi, \omega) = \dfrac{|\mathcal{L}| - \sigma(M, \mathcal{L})}{|\mathcal{L}|}$ *where* Ω_{b} *and* μ_{b} *are arbitrary.*

c. $t_{\mathrm{c}}(M, \Pi, \omega) = \displaystyle\sum_{\{m_i,w\} \in M} \frac{|W| - \sigma(\pi_{m_i}, w)}{|W|^i}$ *where* Ω_{c} *and* μ_{c} *are arbitrary.*

For each selection criterion $i \in \{1, ..., 4\}$, we show $g_i(\Pi, M)$ is monotonic and INS. We then show the function $t_j(\Pi, M, \omega)$ is monotonic INS and discriminates between optimal marriages according to tie-breaking rule j for each $j \in \{a, b, c\}$ given Ω_j and μ_j. Monotonicity and INS are preserved through addition. Therefore $f_{ij}(\Pi, M, \omega) = g_i(\Pi, M) +$

$t_j(\Pi, M, \omega)$ is monotonic and INS. Thus the decision mechanism $r_{ij} = \{f_{ij}, \Omega_j, \mu_j\}$ selects a stable marriage according to criterion i and tie-breaking rule j.

Each g_i is monotonic and INS: Let f and g be two monotonic INS functions. First we show that $f + g$ and $\min\{f, g\}$ are also monotonic and INS. INS means that under certain conditions, f and g will not change. Therefore, functions of f and g such as $f + g$ and $\min\{f, g\}$ will will not change under those conditions, and are also INS. Both summing and taking the minimum are well known to be monotonic operators. Therefore both $f + g$ and $\min\{f, g\}$ are monotonic INS functions.

Next we show $-\sigma(\pi_m, w)$ and $-\sigma(\pi_w, m)$ are monotonic and INS for each $\{m, w\} \in M$. By symmetry, consider only $-\sigma(\pi_m, w)$. If m moves w up (or down) in π_m then $-\sigma(\pi_m, w)$ increases (respectively decreases) and therefore $-\sigma(\pi_m, w)$ is strictly monotonic. The value of $-\sigma(\pi_m, w)$ depends only on w's location and remains unchanged if m swaps two other women, adds a woman onto the end of $\bar{\pi}_m$ or if anyone else changes their preferences. Therefore $-\sigma(\pi_m, w)$ is both monotonic and INS.

Each g_i consists only of sums and minima of monotonic INS functions and therefore is monotonic and INS.

Each t_j is monotonic and INS: Both t_a and t_b are independent of Π and therefore are monotonic and INS. Like each g_i, t_c is also monotonic and INS.

Each t_j discriminates between optimal marriages according to tie-breaking rule j: Since each g_i is integer and $0 \leq t_j < 1$, a marriage that maximizes $f_{ij} = g_i + t_j$ also maximizes g_i. Therefore, the marriages that maximize f_{ij} for some $\omega \in \Omega_j$ are a subset of $\mathcal{S}^i$. Furthermore, for every $\omega \in \Omega_j$, $t_j(M, \Pi, \omega)$ has a unique value for each M implying f_{ij} has a unique optimizer and $r_{ij}(\Pi) \subseteq \mathcal{S}^i$ is well defined. It remains to show that that the probability of selecting M given criterion i and tie-breaking rule j is equal to the probability of selecting M given r_{ij}.

In Example 6.2.7, we demonstrated that t_a, Ω_a and μ_a correctly discriminate between optimal marriages. Thus $r_{ia} = \{f_{ia}, \Omega_a, \mu_a\}$ is a monotonic INS representable decision mechanism that selects a marriage according to criterion i while breaking ties uniformly at random.

Breaking ties lexicographically by a predetermined list of marriages is a deterministic tie-breaking rule. Given Π, let M_i be the marriage that is selected according to criterion i and tie-breaking rule b. Let $M \in r_{ib}(\Pi) \subseteq \mathcal{S}^i$ be any marriage selected with positive probability by r_{ib}. Since M_i satisfies criterion i, $M_i \in \mathcal{S}^i$ and $g_i(M_i, \Pi) = g_i(M, \Pi)$. Furthermore, since M_i is selected according to tie-breaking rule b, M_i appears no later than M in $\mathcal{L}$ and $\sigma(\mathcal{L}, M_i) \leq \sigma(\mathcal{L}, M)$. These two inequalities imply $f_{ib}(M_i, \Pi, \omega) \geq f_{ib}(M, \Pi, \omega)$ for all ω. Since every $M' \in r_{ib}(\Pi)$ maximizes f_{ib} for some ω, $r_{ib}(\Pi) = M_i$ as desired. Therefore r_{ib} is a monotonic INS decision mechanism that selects a marriage according to criterion i while breaking ties lexicographically according to $\mathcal{L}$.

Once again, breaking ties lexicographically by giving the lowest indexed m_i his most preferred partner available is a deterministic tie-breaking rule. Let M_i be the marriage selected according to criterion i and tie-breaking rule c. As before, let $M \in r_{ic}(\bar{\pi}) \subseteq \mathcal{S}^i$. Once again, $g_i(M_i, \Pi) = g_i(M, \Pi)$. Suppose $M \neq M_i$ and let j be the lowest index such that $s_M(m_j) \neq s_{M_i}(m_j)$. Thus, for all $k < j$, m_k has the same partner in M and M_i. Since M_i is selected according to tie-breaking rule c, m_j prefers his M_i-partner to his M-partner according to Π implying that $\sigma(\pi_{m_j}, s_{M_i}(m_j)) \leq \sigma(\pi_{m_j}, s_M(m_j)) - 1$. This implies

$$t_3(M,\Pi,\omega) = \sum_{\{m_k,w\}\in M} \frac{|W| - \sigma(\pi_{m_k}, w)}{|W|^k} \tag{6.31}$$

$$= \sum_{\substack{\{m_k,w\}\in M:\\ k<j}} \frac{|W| - \sigma(\pi_{m_k}, w)}{|W|^k} + \frac{|W| - \sigma(\pi_{m_j}, s_M(m_j))}{|W|^j} + \sum_{\substack{\{m_k,w\}\in M:\\ k>j}} \frac{|W| - \sigma(\pi_{m_k}, w)}{|W|^k} \tag{6.32}$$

$$\leq \sum_{\substack{\{m_k,w\}\in M:\\ k<j}} \frac{|W| - \sigma(\pi_{m_k}, w)}{|W|^k} + \frac{|W| - \sigma(\pi_{m_j}, s_M(m_j))}{|W|^j} + \frac{1}{|W|^j} \tag{6.33}$$

$$\leq \sum_{\substack{\{m_k,w\}\in M:\\ k<j}} \frac{|W| - \sigma(\pi_{m_k}, w)}{|W|^k} + \frac{|W| - \sigma(\pi_{m_j}, s_{M_i}(m_j))}{|W|^j} \tag{6.34}$$

$$= \sum_{\substack{\{m_k,w\}\in M_i:\\ k\leq j}} \frac{|W| - \sigma(\pi_{m_k}, w)}{|W|^k} \leq t_3(M_i,\Pi,\omega) \tag{6.35}$$

and $f_{ic}(M_i,\Pi,\omega) \geq f_{ic}(M,\Pi,\omega)$ for all ω. As in the previous instance, this implies $r_{ic}(\Pi) = M_i$ and r_{ic} is a monotonic INS representable decision mechanism that selects a marriage according to criterion i while breaking lexicographically by giving the lowest indexed m_i his most preferred partner available. □

We come now to our principal positive result.

Theorem 6.3.14. *If r is monotonic and INS representable, then for every minimally dishonest equilibrium $\bar{\Pi}$ of the strategic stable marriage game, the marriage $r(\bar{\Pi})$ is stable with respect to the sincere preferences Π.*

Proof. The bulk of the proof is a sequence of nine lemmas that gradually reveal the necessary structure of $\bar{\Pi}$ at a minimally dishonest equilibrium. These are stated and proved in 6.6.1. The early lemmas consider a marriage M that is stable with respect to a putative profile $\bar{\Pi}$, and identify changes to $\bar{\pi}_y$ that retain the stability of M by not interfering with the ranking of y's M-partner. These lemmas depend mainly on the transitivity of preferences and do not invoke monotonicity or INS. In the succeeding lemmas, the main ideas are essentially as follows:

- If y is unmarried in a putatively stable M, y can only be helped by adding people y is willing to marry to $\bar{\pi}_y$, and can not be hurt by removing unacceptable partners from $\bar{\pi}_y$. Therefore more honesty does not harm y.

- If y is married in M, the Rural Hospitals Theorem [61] and retention of stability from the early lemmas assure that y remains married when y makes certain alterations to $\bar{\pi}_y$. If y has falsely purported to prefer x less than z, altering $\bar{\pi}_y$ by swapping x and z in $\bar{\pi}_y$ keeps M stable, and by INS and monotonicity, the value of marriages that y likes at least as much as M will not decrease while the value of all other marriages will not increase and the mechanism still selects a marriage that y likes as least as much as M.

- Similarly, monotonicity and INS permit y to elevate y's M-partner to a more honest rank without losing M's optimality.

- These arguments imply that $\bar{\Pi}$ must be quite similar to Π at any minimally dishonest equilibrium.

- Given that $\bar{\Pi}$ is constrained to be similar to Π, Lemma 6.3.2 in turn constrains the set of stable marriages – not just $r(\bar{\Pi})$ – to be a singleton and have other useful properties.

For readability, we state three critical lemmas here.

Lemma 6.6.4. *If y is unmarried at a minimally dishonest equilibrium then $\bar{\pi}_y = \pi_y$.*

Lemma 6.6.8. *At any minimally dishonest equilibrium for a monotonic INS representable stable marriage mechanism, if y is married to the k^{th} member of π_y, then the first k elements of $\bar{\pi}_y$ are identical to and appear in the same order as the first k elements of π_y.*

Lemma 6.6.9. *For preference profiles* Π *and* $\bar{\Pi}$*, and a marriage* M *that is stable with respect to* $\bar{\Pi}$*, if* $\bar{\pi}_y$ *agrees with* π_y *up to* y*'s* M*-partner, then* M *is stable with respect to* Π.

Lemmas 6.6.4 and 6.6.8 show that $\bar{\Pi}$ meets the conditions of Lemma 6.6.9. Therefore, $\bar{\Pi}$ yields a marriage that is stable with respect to Π, proving the theorem. $\square$

However, there is no monotonic INS representable stable marriage mechanism that always selects an egalitarian stable marriage.

Theorem 6.3.15. *There is no monotonic INS representable* r *such that there is always a minimally dishonest equilibrium and where for every minimally dishonest equilibrium* $\bar{\Pi}$ *the marriage* $r(\bar{\Pi})$ *is an egalitarian stable marriage with respect to the sincere preferences.*

Proof. Suppose by contradiction there is a monotonic INS representable mechanism $r = \{f, \Omega, \mu\}$ that always selects a sincere egalitarian stable marriage. The proof begins by presenting a set of sincere preferences Π^1 with stable marriages M_1, M_2 and M_3. To ensure that a sincere egalitarian marriage is selected at an equilibrium, for $j \in \{1, 3\}$ there must exist a sample path ω_j, where a non-egalitarian stable marriage M_j has a higher value than the unique egalitarian stable marriage M_2 when everyone is honest and Π^1 is submitted.

We then analyze a second set of sincere preferences Π^2, obtained by slightly modifying Π^1. The set of stable marriages remains unchanged but every marriage that was egalitarian for Π^1 (respectively Π^2) is not egalitarian for Π^2 (respectively Π^1). By INS and monotonicity, M_j still has a higher value than M_2 for the sample path ω_j if individuals are honest and submit Π^2. However, this allows the construction of a minimally dishonest equilibrium $\bar{\Pi}$ for the sincere preferences Π^2 where M_2, a marriage that is not egalitarian with respect to Π^2, is selected. This implies there is no r that always selects a sincere egalitarian stable marriage.

Consider the preference profile Π^1. With respect to Π^1 the set of stable marriages is given in Table 6.1. The unique egalitarian stable marriage is M_2.

$$\pi^1_{m_1} = w_1 \succ_{\pi^1_{m_1}} w_2 \succ_{\pi^1_{m_1}} w_4 \succ_{\pi^1_{m_1}} w_3 \quad \pi^1_{w_1} = m_2 \succ_{\pi^1_{w_1}} m_3 \succ_{\pi^1_{w_1}} m_4 \succ_{\pi^1_{w_1}} m_1 \quad (6.36)$$

$$\pi^1_{m_2} = w_2 \succ_{\pi^1_{m_2}} w_3 \succ_{\pi^1_{m_2}} w_4 \succ_{\pi^1_{m_2}} w_1 \quad \pi^1_{w_2} = m_3 \succ_{\pi^1_{w_2}} m_1 \succ_{\pi^1_{w_2}} m_4 \succ_{\pi^1_{w_2}} m_2 \quad (6.37)$$

$$\pi^1_{m_3} = w_3 \succ_{\pi^1_{m_3}} w_1 \succ_{\pi^1_{m_3}} w_4 \succ_{\pi^1_{m_3}} w_2 \quad \pi^1_{w_3} = m_1 \succ_{\pi^1_{w_3}} m_2 \succ_{\pi^1_{w_3}} m_4 \succ_{\pi^1_{w_3}} m_3 \quad (6.38)$$

$$\pi^1_{m_4} = w_4 \succ_{\pi^1_{m_4}} w_1 \succ_{\pi^1_{m_4}} w_2 \succ_{\pi^1_{m_4}} w_3 \quad \pi^1_{w_4} = m_4 \succ_{\pi^1_{w_4}} m_1 \succ_{\pi^1_{w_4}} m_2 \succ_{\pi^1_{w_4}} m_3 \quad (6.39)$$

Table 6.1: Set of Stable Marriages With Respect to Π^1 and Π^2.

Marriage	Spouses			
M_1	$\{m_1, w_1\}$	$\{m_2, w_2\}$	$\{m_3, w_3\}$	$\{m_4, w_4\}$
M_2	$\{m_1, w_2\}$	$\{m_2, w_3\}$	$\{m_3, w_1\}$	$\{m_4, w_4\}$
M_3	$\{m_1, w_3\}$	$\{m_2, w_1\}$	$\{m_3, w_2\}$	$\{m_4, w_4\}$

Claim 1: For each $j \in \{1, 3\}$ there is $\omega_j \in \Omega$ such that $f(M_j, \Pi^1, \omega_j) > f(M_2, \Pi^1, \omega_j)$.

By symmetry let $j = 1$. Suppose instead $f(M_1, \Pi^1, \omega) \leq f(M_2, \Pi^1, \omega)$ for all $\omega \in \Omega$. Suppose the sincere preferences are given by Π^1 and that $\bar{\Pi}$ is a minimally dishonest equilibrium. Since r always selects a sincere egalitarian stable marriage, M_2 is stable with respect to $\bar{\Pi}$. Furthermore by Corollary 6.6.7, M_2 is the only stable marriage with respect to $\bar{\Pi}$. We now show that $\pi^1_{w_i} = \bar{\pi}_{w_i}$ for $i \in \{1, 2, 3\}$ implying that M_1 is also stable with respect to $\bar{\Pi}$, a contradiction to Lemma 6.3.2 since m_1 prefers M_1 to M_2.

Subclaim 1: If Claim 1 does not hold, then $m_4 \succ_{\bar{\pi}_w} u$ for each $w \in \{w_1, w_2, w_3\}$.

Suppose there is w such that $u \succ_{\bar{\pi}_w} m_4$. By Lemma 6.6.8, π^1_w and $\bar{\pi}_w$ agree on the first two elements. Woman w is more honest in the profile $\bar{\Pi}' = \delta(\bar{\Pi}, w, \{m_4, u\})$. Furthermore, by Lemma 6.6.3 M_2 remains stable and all new marriages wed w to m_4. However, by Lemma 6.6.8. π^1_j, $\bar{\pi}_j$ and $\bar{\pi}'_j$ agree on the first element for $j \in \{m_4, w_4\}$ and m_4 and w_4 must wed in every marriage. Therefore no new stable marriages are created when w becomes more honest by adding m_4 onto her preference list and w receives at least as good an outcome after becoming more honest, a contradiction to minimal dishonesty. This completes Subclaim 1.

Subclaim 2: If Claim 1 does not hold, then m_4 is the third element of $\bar{\pi}_w$ for each $w \in \{w_1, w_2, w_3\}$.

Suppose without loss of generality this does not hold for w_1. By Lemma 6.6.8, $\pi^1_{w_1}$ and $\bar{\pi}_{w_1}$ agree on the first two elements. By Subclaim 1, $\bar{\pi}_{w_1}$ is then given by $m_2 \succ_{\bar{\pi}_{w_1}} m_3 \succ_{\bar{\pi}_{w_1}} m_1 \succ_{\bar{\pi}_{w_1}} m_4$. Similar to the proof of Subclaim 1, w_1 can be more honest by swapping m_1 and m_4 without obtaining a worse result. Once again this contradicts minimal dishonesty and Subclaim 2 holds.

Subclaim 3: If Claim 1 does not hold, then $\bar{\pi}^1_w = \bar{\pi}'_w$ for each $w \in \{w_1, w_2, w_3\}$.

Once again, without loss of generality assume this subclaim does not hold for w_1. By Lemma 6.6.8 and Subclaims 1 and 2, $\bar{\pi}_{w_1}$ is given by $m_2 \succ_{\bar{\pi}_{w_1}} m_3 \succ_{\bar{\pi}_{w_1}} m_4$ and woman w_1 can only be more honest by adding m_1 onto $\bar{\pi}_{w_1}$ – by updating $\bar{\Pi}$ to $\bar{\Pi}' = \delta(\bar{\Pi}, w_1, \{m_1, u\})$. If adding m_1 onto w_1's list creates no new stable marriages, then w_1 violates minimal dishonesty. If adding m_1 creates a new stable marriage, then by Lemma 6.6.3 the only new stable marriage is M_1 and $\pi^1_w = \bar{\pi}_w$ for $w \in \{w_2, w_3\}$. Since f is INS, $f(M_i, \Pi^1, \omega) = f(M_i, \bar{\Pi}', \omega)$ for $i \in \{1, 2\}$. Furthermore, since Claim 1 is assumed to be false, $f(M_1, \bar{\Pi}', \omega) = f(M_1, \Pi^1, \omega) \leq f(M_2, \Pi^1, \omega) = f(M_2, \bar{\Pi}', \omega)$ for all $\omega \in \Omega$. Since M_1 and M_2 are the only two stable marriage and since given any ω the mechanism r always selects a unique marriage, $f(M_1, \bar{\Pi}', \omega) < f(M_2, \bar{\Pi}', \omega)$ implying that M_2 is still selected if w_1 is more honest, a contradiction to minimal dishonesty completing the proof of the subclaim.

By Subclaims 1, 2 and 3, if Claim 1 is false then $\bar{\pi}^1_{w_i} = \bar{\pi}_{w_i}$ for $i \in \{1, 2, 3\}$ and M_1 is stable with respect to $\bar{\Pi}$, a contradiction to Lemma 6.3.2. Therefore, Claim 1 holds. We have established the first part of the proof; for $j \in \{1, 3\}$ there is some sample path ω_j, where a non-egalitarian stable marriage M_j must have a higher value than the unique egalitarian stable marriage M_2 if everyone honest and π^1 is submitted.

Now consider a second set of sincere preferences Π^2 pbtained by moving w_4 (m_4) up one position in m_1, m_2 and m_3's preference lists (w_1, w_2 and w_3 respectively). Once again the marriages M_1, M_2 and M_3 in Table 6.1 are stable. However, unlike Π^1, M_1 and M_3 are the only egalitarian stable marriages.

$$\pi^2_{m_1} = w_1 \succ_{\pi^2_{m_1}} w_4 \succ_{\pi^2_{m_1}} w_2 \succ_{\pi^2_{m_1}} w_3 \quad \pi^2_{w_1} = m_2 \succ_{\pi^2_{w_1}} m_4 \succ_{\pi^2_{w_1}} m_3 \succ_{\pi^2_{w_1}} m_1 \tag{6.40}$$

$$\pi^2_{m_2} = w_2 \succ_{\pi^2_{m_1}} w_4 \succ_{\pi^2_{m_2}} w_3 \succ_{\pi^2_{m_2}} w_1 \quad \pi^2_{w_2} = m_3 \succ_{\pi^2_{w_1}} m_4 \succ_{\pi^2_{w_2}} m_1 \succ_{\pi^2_{w_2}} m_2 \tag{6.41}$$

$$\pi^2_{m_3} = w_3 \succ_{\pi^2_{m_1}} w_4 \succ_{\pi^2_{m_3}} w_1 \succ_{\pi^2_{m_3}} w_2 \quad \pi^2_{w_3} = m_1 \succ_{\pi^2_{w_1}} m_4 \succ_{\pi^2_{w_3}} m_2 \succ_{\pi^2_{w_3}} m_3 \tag{6.42}$$

$$\pi^2_{m_4} = w_4 \succ_{\pi^2_{m_1}} w_1 \succ_{\pi^2_{m_4}} w_2 \succ_{\pi^2_{m_4}} w_3 \quad \pi^2_{w_4} = m_4 \succ_{\pi^2_{w_1}} m_1 \succ_{\pi^2_{w_4}} m_2 \succ_{\pi^2_{w_4}} m_3 \tag{6.43}$$

Claim 2: For each $j \in \{1, 3\}$ there is an $\omega_j \in \Omega$ such that $f(M_j, \Pi^2, \omega_j) > f(M_2, \Pi^2, \omega_j)$.

By symmetry, assume $j = 1$. Since f is monotonic and INS, $f(M_1, \Pi^2, \omega) = f(M_1, \Pi^1, \omega)$ and $f(M_2, \Pi^1, \omega) \geq f(M_2, \Pi^2, \omega)$. By Claim 1, there is an $\omega_1 \in \Omega$ such that $f(M_1, \Pi^1, \omega_1) > f(M_2, \Pi^1, \omega_1)$. Thus ω_1 is such that $f(M_1, \Pi^2, \omega_1) = f(M_1, \Pi^1, \omega_1) > f(M_2, \Pi^1, \omega_1) \geq f(M_2, \Pi^2, \omega_1)$ and Claim 2 holds.

Given Claims 1 and 2, we can now construct a minimally dishonest equilibrium $\bar{\Pi}$ with respect to the sincere preferences Π^2 where M_2 is selected by r. This is a contradiction since r always selects an egalitarian stable marriage and thus no such r can exist.

The preference $\bar{\Pi}$ below has the unique stable marriage M_2 and corresponds to a minimally dishonest equilibrium with respect to the sincere profile Π^2, a contradiction since M_2 is not an egalitarian stable marriage with respect to Π^2.

$$\bar{\pi}_{m_1} = w_1 \succ_{\bar{\pi}_{m_1}} w_4 \succ_{\bar{\pi}_{m_1}} w_2 \qquad \bar{\pi}_{w_1} = m_2 \succ_{\bar{\pi}_{w_1}} m_4 \succ_{\bar{\pi}_{w_1}} m_3 \tag{6.44}$$

$$\bar{\pi}_{m_2} = w_2 \succ_{\bar{\pi}_{m_2}} w_4 \succ_{\bar{\pi}_{m_2}} w_3 \succ_{\bar{\pi}_{m_2}} w_1 \quad \bar{\pi}_{w_2} = m_3 \succ_{\bar{\pi}_{w_2}} m_4 \succ_{\bar{\pi}_{w_2}} m_1 \succ_{\bar{\pi}_{w_2}} m_2 \tag{6.45}$$

$$\bar{\pi}_{m_3} = w_3 \succ_{\bar{\pi}_{m_3}} w_4 \succ_{\bar{\pi}_{m_3}} w_1 \succ_{\bar{\pi}_{m_3}} w_2 \quad \bar{\pi}_{w_3} = m_1 \succ_{\bar{\pi}_{w_3}} m_4 \succ_{\bar{\pi}_{w_3}} m_2 \succ_{\bar{\pi}_{w_3}} m_3 \tag{6.46}$$

$$\bar{\pi}_{m_4} = w_4 \succ_{\bar{\pi}_{m_4}} w_1 \succ_{\bar{\pi}_{m_4}} w_2 \succ_{\bar{\pi}_{m_4}} w_3 \quad \bar{\pi}_{w_4} = m_4 \succ_{\bar{\pi}_{w_4}} m_1 \succ_{\bar{\pi}_{w_4}} m_2 \succ_{\bar{\pi}_{w_4}} m_3 \tag{6.47}$$

With respect to $\bar{\Pi}$, M_2 is the only stable marriage and therefore is both man and woman-optimal. By [20, Theorem 17] no man can alter his preferences to obtain a marriage he prefers to his man-optimal partner. Since the sincere and putative preferences agree up to each man's partner in M_2, no man can alter his putative preferences to obtain a marriage he sincerely prefers. A symmetric argument holds for the women and $\bar{\Pi}$ is an equilibrium.

It remains to show that everyone is minimally dishonest. Only m_1 and w_1 are dishonest. Examine m_1. To be more honest, man m_1 can only change $\bar{\Pi}$ to $\bar{\Pi}' = \delta(\bar{\Pi}, m_1, \{w_3, u\})$. However, this makes M_3 stable. By Claim 2 and since f is a monotonic and INS, there is an $\omega_3 \in \Omega$ such that $f(M_3, \bar{\Pi}', \omega_3) = f(M_3, \Pi^2, \omega_3) > f(M_2, \Pi^2, \omega_3) = f(M_2, \bar{\Pi}', \omega_3)$. Thus, the alteration would generate a worse solution for m_1. A symmetric argument holds for w_1 with M_1 and ω_1. Hence m_1 and w_1 are both minimally dishonest. The minimally dishonest equilibrium $\bar{\Pi}$ has the unique stable marriage M_2, a contradiction since M_2 is not an egalitarian stable marriage for Π^2. This completes the proof of Theorem 6.3.15. $\square$

6.4 The Gale-Shapley Algorithm and Never-One-Sided Mechanisms

In Section 6.3.1, we examined pure strategy minimally dishonest Nash equilibria of the SSM game. However, such an equilibrium may not exist. In this section, we first establish that not every mechanism has a minimally dishonest equilibrium. We then characterize the set of minimally dishonest equilibria for the Gale-Shapley algorithm and never-one-sided mechanisms and show both types of mechanisms always have minimally dishonest equilibria.

Proposition 6.4.1. *If r in SSM uniformly randomly selects an egalitarian stable marriage, then it is possible that no minimally dishonest equilibrium exists.*

Proof. In Theorem 6.3.13, r was shown to be monotonic and INS representable. Consider the set of sincere preferences Π.

$$\pi_{m_1} = w_1 \succ_{\pi_{m_1}} w_2 \succ_{\pi_{m_1}} w_3 \qquad \pi_{w_1} = m_3 \succ_{\pi_{w_1}} m_1 \succ_{\pi_{w_1}} m_2 \tag{6.48}$$

$$\pi_{m_2} = w_2 \succ_{\pi_{m_2}} w_3 \succ_{\pi_{m_2}} w_1 \qquad \pi_{w_2} = m_1 \succ_{\pi_{w_2}} m_2 \succ_{\pi_{w_2}} m_3 \tag{6.49}$$

$$\pi_{m_3} = w_3 \succ_{\pi_{m_3}} w_1 \succ_{\pi_{m_3}} w_2 \qquad \pi_{w_3} = m_2 \succ_{\pi_{w_3}} m_3 \succ_{\pi_{w_3}} m_1 \tag{6.50}$$

Π is perfectly symmetric between genders and among individuals. Each gender's set of preferences is a standard Condorcet cycle. By Theorem 6.3.14, one of the sincerely stable marriages, $M_1 = \{\{m_1, w_1\}, \{m_2, w_2\}, \{m_3, w_3\}\}$ and $M_2 = \{\{m_3, w_1\}, \{m_1, w_2\}, \{m_2, w_3\}\}$, is selected at equilibrium. Suppose $\bar{\Pi}$ is a minimally dishonest equilibrium. By the gender symmetry of Π, without loss of generality marriage M_1 is selected. By Lemma 6.6.8, $\bar{\pi}_{m_i}$ and π_{m_i} agree on the first element and $\bar{\pi}_{w_i}$ and π_{w_i} agree on the first two elements.

The idea of the proof is that minimal dishonesty leads the men to admit that they are willing to marry their second choice. They remain assured of outcome M_1 by dishonestly ranking a second choice third, which makes M_1 the unique egalitarian stable marriage, because the women are honest about their second choices. However, once $\bar{\Pi}$ shows that all three men are so willing, equilibrium is lost. Any woman can make M_2 the unique stable marriage, simply by purporting to be willing to wed only her first choice. By symmetry between the genders, no such equilibrium exists.

To show that each man admits he is willing to marry his second choice, we show $\bar{\pi}_{m_i}$ omits no w_j for each i. By symmetry, it suffices to only analyze $\bar{\pi}_{m_1}$.

Claim 1: $w_3 \succ_{\bar{\pi}_{m_1}} u$.

Suppose instead that $u \succ_{\bar{\pi}_{m_1}} w_3$. Consider $\bar{\Pi}' = \delta(\bar{\Pi}, m_1, \{w_3, u\})$. If any new stable marriages appear, then m_1 weds w_3 in the new marriage. However, by Lemma 6.6.8 m_3 (w_3) is honest about the first (respectively two) element(s) in the submitted profile $\bar{\Pi}$ and $\{m_3, w_3\}$ still blocks any marriage that weds m_1 to w_3 with respect to $\bar{\Pi}'$ and no new stable marriages appear implying m_1's partner does not change, a contradiction to minimal dishonesty. Thus $w_3 \succ_{\bar{\pi}_{m_1}} u$.

Claim 2: $w_2 \succ_{\bar{\pi}_{m_1}} u$.

Suppose that this is not the case and, by Claim 1, m_1's preferences are given by $w_1 \succ_{\bar{\pi}_{m_1}} w_3$. Consider $\bar{\Pi}'' = \delta(\bar{\Pi}, m_1, \{w_2, u\})$. The only new stable marriage that can appear in $\bar{\Pi}''$ is M_2. Using f_2 from Example 6.2.7, the value of M_1 is more than the value of M_2 (since w_3 appears prior to w_2 in $\bar{\pi}''_{m_1}$). Since M_1 and M_2 are the only marriages stable with respect to $\bar{\Pi}''$, m_1's outcome does not change contradicting minimal dishonesty. Thus, $w_2 \succ_{\bar{\pi}_{m_1}} u$.

Since Claims 1 and 2 hold, $\bar{\pi}_{m_i}$ contains every w_j. By Lemma 6.6.8, $\bar{\pi}_{w_j}$ agrees with π_{w_j} on the first two elements and M_2 is also stable with respect to $\bar{\pi}$. Thus, if there is an equilibrium $\bar{\Pi}$ then there are two stable marriages. This contradicts Corollary 6.6.7 and there can be no such equilibrium. □

We now consider two types of mechanisms that always have equilibria: the Gale-Shapley algorithm and more fair *never-one-sided* mechanisms.

6.4.1 Gale-Shapley Algorithm

In this section, we characterize the set of minimally dishonest equilibria obtained when r is the Gale-Shapley algorithm. We prove that the woman-optimal sincerely stable marriage will always be obtained.

Theorem 6.4.2. *If r always selects the man-optimal marriage (Gale-Shapley algorithm), there exists a minimally dishonest equilibrium for any sincere preference profile Π. Moreover, all minimally dishonest equilibria yield the woman-optimal sincerely stable marriage.*

Proof. To show existence of an equilibrium that yields the woman-optimal sincerely stable marriage $\overline{M}$ with respect to Π, consider the preference profile $\bar{\Pi}^{\overline{M}}$ obtained by having men be honest and having women truncate their sincere profiles after their woman-optimal partner. By construction, $\overline{M}$ will be the only stable marriage with respect to $\bar{\Pi}^{\overline{M}}$.

In 6.6.2, an algorithm is presented to find a minimally dishonest equilibrium $\bar{\Pi}$ given $\bar{\Pi}^{\overline{M}}$ where $r(\bar{\Pi}) = \overline{M}$. Each iteration of the algorithm corrects a violation to minimal dishonesty and decreases the distance between an individual's putative and sincere preferences. Since the distances are finite and integer, the algorithm terminates in finite time (Lemma 6.6.10). Next we show that if $\bar{\Pi}$ is the profile obtained at the end of an iteration, then $\bar{\Pi}$ is an equilibrium and $\overline{M}$ is the unique stable marriage with respect to $\bar{\Pi}$ (Lemma 6.6.11). This implies the algorithm outputs an equilibrium $\bar{\Pi}$ that yields the sincerely stable marriage $\overline{M}$. We show that the equilibrium is minimally dishonest by showing that the algorithm does not terminate until all violations to minimal dishonesty are corrected (Lemma 6.6.12). Thus there is at least one minimally dishonest equilibrium $\bar{\Pi}$ that yields the woman-optimal $\overline{M}$.

It remains to show that no other marriage can be obtained at an equilibrium. In Section 6.3.1 we showed that the Gale-Shapley algorithm is monotonic and INS representable and thus any equilibrium must yield a sincerely stable marriage. Suppose the sincerely stable M is obtained given the equilibrium $\bar{\Pi}$.

We begin by showing that the minimally dishonest refinement implies that all men will be honest at equilibrium. For contradiction suppose man m is not honest and that $\bar{\pi}_m \neq \pi_m$. By Lemma 6.6.8, $\bar{\pi}_m$ and π_m agree up to m's partner in M. Thus, any disparity between $\bar{\pi}_m$ and π_m must appear after $s_M(m)$. However, by Lemmas 6.6.2 and 6.6.3, the marriage M remains stable (and man-optimal) after m corrects the disparity. Since the Gale-Shapley algorithm always selects the man-optimal marriage, man m obtains the same result contradicting minimal dishonesty. Thus all men are honest at a minimally dishonest equilibrium.

Now since the men are honest and, by Lemma 6.6.8, every woman is honest up to her partner in M, the woman-optimal marriage is stable with respect to the putative preferences. By Corollary 6.6.7, there is only one marriage stable with respect to the putative preferences and M is the woman-optimal marriage. $\square$

6.4.2 Never-One-Sided Mechanisms

Fairness is often a legitimate concern in decision-making. The Gale-Shapley algorithm is arguably the most unfair way to select a stable marriage, because it exclusively favors one side – the group that is sexually identified as men. Therefore, until now, concern for sincere stability has in a sense maximized unfairness! Our next result offers a broad set of less unfair mechanisms that are sure to yield sincere stability, and do so in every minimally dishonest equilibrium.

Definition 6.4.3. r is *never-one-sided* if for all $\bar{\Pi}$ that admit more than one stable marriage, there exist $M_1, M_2 \in r(\bar{\Pi})$ (possibly $M_1 = M_2$) where M_1 is not man-optimal and M_2 is not woman-optimal.

Theorem 6.4.4. *Let r be an arbitrary monotonic, INS and never-one-sided representable stable marriage algorithm. Let M be a marriage and let Π be a sincere preference profile. Then M is sincerely stable if and only if there exists a minimally dishonest equilibrium $\bar{\Pi}$ in the strategic stable marriage game where $r(\bar{\Pi}) = M$.*

The proof of existence follows in the same fashion as Theorem 6.4.2. For a sincerely stable M, when given $\bar{\Pi}^M$ obtained by having every individual truncate $\bar{\Pi}$ after their M-partner, Algorithm 2 in 6.6.2 returns a minimally dishonest equilibrium $\bar{\Pi}$ where $r(\bar{\Pi}) = M$. Furthermore, every outcome is guaranteed to be stable by Theorem 6.3.14.

6.5 Extensions

In this section, we consider extensions to the Stable Marriage Problem. We first consider the college Admissions Problem. In this setting we label the women as "colleges" and men as "students". Unlike the Stable Marriage Problem, each college is allowed to marry multiple students. Each college has a quota indicating the maximum number of students that they are willing to marry. We show that our positive results fail to hold in this setting.

The Student Placement Problem is a special case of the college Admissions Problem where each college is assumed to be honest. Since the college Admissions Problem can be modeled by a Stable Marriage Problem by duplicating the colleges, the Student Placement Problem is also a special case of the Stable Marriage Problem where all women are honest. We consider a more general model of the Stable Marriage Problem where any subset of individuals are allowed to be honest. Furthermore, we show that the positive results still hold in this setting.

6.5.1 College Admissions Problem

Roth has long claimed that the Admissions Problem is significantly different than the Stable Marriage Problem by showing that unlike the Stable Marriage Problem, a single college is able to alter their preferences to obtain a marriage that they prefer to the college-optimal marriage [60]. Very few of our results extend to the college Admissions Problem. We provide the unsettling result that no stable marriage mechanism can guarantee a sincerely stable marriage is selected at a minimally dishonest equilibrium and, like Roth, we must emphasize that the college Admissions Problem is different than the Stable Marriage Problem.

Without knowing colleges' preferences between groups of students, we cannot cannot guarantee that a marriage is deterministically selected at every equilibrium (Corollary 6.3.3). Consider a set of preferences where a college has a 50% chance of obtaining their 1st and 4th choice in students and a 50% chance of obtaining their 2nd and 3rd choice in students. If the college is indifferent between these two outcomes, these preferences may correspond to an equilibrium. However, Corollary 6.3.3 does hold if every college has strict preferences between every mixture of students.

The most unsettling disparity between the Stable Marriage Problem and the College Admissions Problem is that we may obtain a marriage at an equilibrium that is not sincerely stable, as shown in Theorem 6.5.1. Mimicking Roth, Theorem 6.5.1 demonstrates there is a set of preferenes where a college can alter their preferences to obtain a marriage that they prefer to the college-optimal marriage even at a minimally dishonest equilibrium, regardless of the stable marriage mechanism used. This implies that no stable marriage mechanism guarantees that a sincerely stable marriage is selected.

Theorem 6.5.1. *There is a sincere profile* Π *for the college admissions strategic game where for every stable marriage mechanism* r*, no minimally dishonest equilibrium yields a sincerely stable marriage.*

Proof. Consider the sincere preferences

$$\pi_{c_1} = s_1 \succ_{\pi_{c_1}} s_2 \succ_{\pi_{c_1}} s_3 \succ_{\pi_{c_1}} s_4 \qquad \pi_{s_1} = c_3 \succ_{\pi_{s_1}} c_1 \succ_{\pi_{s_1}} c_2 \tag{6.51}$$

$$\pi_{c_2} = s_1 \succ_{\pi_{c_2}} s_2 \succ_{\pi_{c_2}} s_3 \succ_{\pi_{c_2}} s_4 \qquad \pi_{s_2} = c_2 \succ_{\pi_{s_2}} c_1 \succ_{\pi_{s_2}} c_3 \tag{6.52}$$

$$\pi_{c_3} = s_3 \succ_{\pi_{c_3}} s_1 \succ_{\pi_{c_3}} s_2 \succ_{\pi_{c_3}} s_4 \qquad \pi_{s_3} = c_1 \succ_{\pi_{s_3}} c_3 \succ_{\pi_{s_3}} c_2 \tag{6.53}$$

$$\pi_{s_4} = c_1 \succ_{\pi_{s_4}} c_2 \succ_{\pi_{s_4}} c_3 \tag{6.54}$$

where college c_1 has capacity for two students and colleges c_2 and c_3 have room for only one student. The only stable marriage with respect to these preferences is $M = \{\{c_1, s_3\}, \{c_1, s_4\}, \{c_2, s_2\}, \{c_3, s_1\}\}$.

For contradiction, suppose there is a minimally dishonest $\bar{\Pi}$ where $r(\bar{\Pi}) = M$. We now that show there are enough completely honest colleges and students to guarantee that c_1 is able to alter $\bar{\pi}_{c_1}$ to obtain a marriage that c_1 prefers to the college-optimal marriage, contradicting that $\bar{\Pi}$ is an equilibrium.

Claim 1: M is the only marriage stable with respect to $\bar{\Pi}$.

Suppose that M' is stable with respect to $\bar{\Pi}$. Lemma 6.3.2 still applies in this setting and every individual must sincerely like their M-partner(s) as much as their M'-partner(s). By Lemma 6.6.1, every college is married to the same number of students in every stable marriage. Thus, c_1 is always wed to two students. Students s_3 and s_4 are c_1's least preferred partners and $\{c_1, s_3\}, \{c_1, s_4\} \in M'$ since otherwise c_1 prefers M' to M. This implies that either $\{c_2, s_1\}$ or $\{c_2, s_2\}$ is in M'. Since c_2 sincerely prefers s_1 to $s_2 = s_M(c_2)$, $\{c_2, s_2\} \in M'$ implying that $M' = M$.

Claim 2: Colleges c_2 and c_3 and all students are honest up to their M-partner. Furthermore, c_1 is completely honest.

Suppose this isn't true for some individual $y \neq c_1$ (student or college). Lemma 6.6.5 and Lemma 6.6.8 describe a process for correcting disparities between $\bar{\pi}_y$ and π_y where (1) M remains stable, (2) any newly created stable marriage will be sincerely preferred by the individual updating their preferences, and (3) by monotonicity and INS, the values of marriages that y sincerely likes as much as M will not decrease while the values of all other marriages will not increase. These three properties together imply that each individual is honest up to their M-partner. In this setting we have assumed neither monotonicity nor INS and thus only (1) and (2) apply. However, by Claim 1, M is the only stable marriage with respect to $\bar{\Pi}$ and therefore after correcting a disparity, the mechanism will still select a marriage that y sincerely likes as much as M and we can still apply the results from Lemma 6.6.5 and Lemma 6.6.8. Thus each individual is honest up to their partner in M.

Unfortunately, Lemmas 6.6.5 and 6.6.8 only apply to individuals that are allowed only a single spouse. We now show that c_1 is honest. If c_1 is dishonest, it implies that c_1 receives a less preferred marriage M' if they submit the more honest π_{c_1}. However, this implies that c_1 is married to at most one student since s_3 and s_4 are their least preferred partners. Without loss of generality, suppose that c_1 is not married to s_3. By the first part of Claim 2, s_3 indicates that college c_1 is their first choice. This implies $\{c_1, s_3\}$ is a blocking pair contradicting that M' was obtained after c_1 updated their preferences to π_{c_1}. Therefore c_1 is honest and Claim 2 holds.

Claim 3: All college are honest.

Suppose that c_2 is not honest and submitting the more honest π_{c_2} yields a worse result for c_2. This implies there is a stable marriage M' where c_2 is married to neither s_1 nor s_2.

However by Claim 2, s_2 indicates c_2 is their most preferred partner and $\{c_2, s_2\}$ block M', a contradiction. Thus c_2 is honest. An identical argument using c_3 and s_1 shows c_3 is honest.

Claim 4: Students s_1 and s_3 are honest.

Once again suppose there is a stable M' when s_1 submits π_{s_1} that is worse for s_1. By Claim 3, college c_3 is honest and indicates that they only prefer s_3 to s_1. Thus, $\{c_3, s_3\} \in M'$ since otherwise $\{c_3, s_1\}$ blocks M'. Similarly using Claim 2, this implies that $\{c_1, s_1\}, \{c_1, s_2\} \in M'$ since otherwise $\{c_1, s_3\}$ blocks M'. However, this implies that $\{c_2, s_2\}$ blocks M', a contradiction implying that s_1 is honest. Similarly for s_3, if π_{s_3} yields a worse result M' for s_3, then $\{c_1, s_1\}, \{c_1, s_2\} \in M'$ and $\{c_2, s_2\}$ blocks M'. Therefore Claim 4 holds.

Claims 2-4 imply that $\bar{\Pi}$ is similar to

$$\bar{\pi}_{c_1} = s_1 \succ_{\bar{\pi}_{c_1}} s_2 \succ_{\bar{\pi}_{c_1}} s_3 \succ_{\bar{\pi}_{c_1}} s_4 \qquad \bar{\pi}_{s_1} = c_3 \succ_{\bar{\pi}_{s_1}} c_1 \succ_{\bar{\pi}_{s_1}} c_2 \tag{6.55}$$

$$\bar{\pi}_{c_2} = s_1 \succ_{\bar{\pi}_{c_2}} s_2 \succ_{\bar{\pi}_{c_2}} s_3 \succ_{\bar{\pi}_{c_2}} s_4 \qquad \bar{\pi}'_{s_2} = c_2 \tag{6.56}$$

$$\bar{\pi}_{c_3} = s_3 \succ_{\bar{\pi}_{c_3}} s_1 \succ_{\bar{\pi}_{c_3}} s_2 \succ_{\bar{\pi}_{c_3}} s_4 \qquad \bar{\pi}_{s_3} = c_1 \succ_{\bar{\pi}_{s_3}} c_3 \succ_{\bar{\pi}_{s_3}} c_2 \tag{6.57}$$

$$\bar{\pi}'_{s_4} = c_1 \tag{6.58}$$

where $\bar{\pi}_{s_2}$ ($\bar{\pi}_{s_4}$) and $\bar{\pi}'_{s_2}$ ($\bar{\pi}'_{s_4}$) agree on the first element. However, these preferences are not an equilibrium since college c_1 can update their preferences to $s_1 \succ s_4$ in order to obtain the marriage $M' = \{\{c_1, s_1\}, \{c_1, s_4\}, \{c_2, s_2\}, \{c_3, s_3\}\}$ – a marriage that c_1 prefers to M. This contradicts that $\bar{\Pi}$ is an equilibrium. Therefore there is no mechanism that guarantees a sincerely stable marriage for the college Admissions Problem. $\square$

6.5.2 Truth-Tellers and the Student Placement Problem

In practice, some individuals may prefer to be honest regardless of whether they can manipulate their preferences to obtain a partner they strictly prefer. Such individuals are often called *truth-tellers*, and there is experimental evidence that they exist [29, 32]. The Student

Placement Problem is an instance of the Stable Marriage (College Admissions) Problem where all women (colleges) are truth-tellers and where all the men (students) are strategic.

Corollary 6.3.3 does not hold in this setting and a marriage is not always selected deterministically at an equilibrium. If all individuals are truth-tellers and the decision mechanisms selects a marriage uniformly at random, then at the unique equilibrium $\bar{\pi} = \pi$ every sincerely stable marriage has positive probability of being selected. As a result, we do not treat $r(\bar{\Pi})$ as a singleton in this section. However, through the same proof technique in Corollary 6.3.3, we can prove that if an individual is strategic then their partner is selected deterministically at every equilibrium.

Corollary 6.6.7 also does not hold in this setting and there may be more than one marriage stable with respect to the equilibrium preferences. While the proof of Theorem 6.3.14 relies on Corollary 6.6.7, the positive result that every equilibrium stable marriage is also a sincerely stable marriage does apply to this setting.

Theorem 6.5.2. *Let a subset of players in SSM be truth-tellers. If r is monotonic and INS representable, then for every minimally dishonest equilibrium $\bar{\Pi}$, every marriage in $r(\bar{\Pi})$ is stable with respect to the sincere preferences Π.*

The proof of Theorem 6.5.2 can be found in 6.6.3. Theorem 6.5.2 immediately implies that a sincerely stable marriage will be selected in the the Student Placement Problem since the Student Placement Problem is an instance of the Stable Marriage Problem where all women are honest.

Without making conditions on the number of truth-tellers we still cannot guarantee that there is a mechanism that always selects a sincere egalitarian stable marriage or that there always exists a minimally dishonest equilibrium and our negative results still hold. The never-one-sided property still ensures that an equilibrium exists that selects a sincerely stable marriage but it is not guaranteed there is an equilibrium for each sincerely stable marriage. When running the Gale-Shapley algorithm an equilibrium exists and every strategic woman receives her sincere woman-optimal partner at every equilibrium. However, the selected marriage is not necessarily woman-optimal.

Theorem 6.5.3. *Let a subset of players in SSM be truth-tellers. Suppose r always selects the man-optimal marriage. Then there exists a minimally dishonest equilibrium for any sincere preference profile Π, and in every minimally dishonest equilibrium every strategic woman receives her woman-optimal partner.*

Theorem 6.5.4. *Let a subset of players in SSM be truth-tellers. If r is monotonic, INS and never-one-sided representable, then there exists at least one minimally dishonest equilibrium $\bar{\Pi}$. Moreover for every minimally dishonest equilibrium $\bar{\Pi}$ and every $M \in r(\bar{\Pi})$, M is sincerely stable.*

No additional techniques are required to prove these results and we defer the proofs to 6.6.3.

6.5.3 Coalitions

Gale and Sotomayor also specifically motivate the study of manipulation when collusion is allowed [31]. In this section we consider coalitions and *strong equilibria* – equilibria where no group of individuals can collude such that every member of the group obtains a strictly better outcome. We show that for any monotonic INS representable stable marriage mechanism, every minimally dishonest equilibrium is also a strong equilibrium. This implies that all the results from previous sections apply even when collusion is allowed. In addition, it implies that the core of the SSM game is non-empty when using the Gale-Shapley algorithm or a monotonic INS never-one-sided representable stable marriage mechanism even when we refine the set of equilibria to those where everyone is minimally dishonest.

Theorem 6.5.5. *Let r be a monotonic and INS representable stable marriage mechanism and let Π be arbitrary. Every minimally dishonest equilibrium $\bar{\Pi}$ is a strong equilibrium.*

Proof. By Corollary 6.6.7, $r(\bar{\Pi}) = M$ is the unique putatively stable marriage and therefore is putatively both man and woman-optimal. By [18], no coalition of men and women can alter $\bar{\Pi}$ such that every one of them prefers the outcome to M. By Lemma 6.6.8, each individual prefers M with respect to $\bar{\Pi}$ if and only if he/she prefers M with respect to Π. Therefore no coalition can alter $\bar{\Pi}$ to obtain a marriage they all sincerely prefer to M. Therefore $\bar{\Pi}$ is a strong equilibrium. $\square$

The converse does not necessarily hold, as discussed previously.

Lemma 6.5.6. *Let r and Π be arbitrary. For any strong equilibrium $\bar{\Pi}$, $r(\bar{\Pi})$ is a sincerely stable marriage.*

Proof. The profile $\bar{\Pi}$ is a Nash equilibrium and by Corollary 6.3.3 and Lemma 6.3.4, $r(\bar{\Pi}) = M$ is a sincerely rational marriage. If M is not stable with respect to Π then there is a blocking pair $\{m, w\}$. Moreover, m and w could obtain a strictly better solution by indicating that they are only willing to marry each other contradicting that $\bar{\Pi}$ is a strong equilibrium. Thus, M is sincerely stable. $\square$

Theorem 6.5.5 and Lemma 6.5.6 also allow us to construct a slightly different proof of Theorem 6.3.14. Since every minimally dishonest equilibrium of a monotonic INS representable r is a strong equilibrium and since every strong equilibrium yields a sincerely stable marriage, every minimally dishonest equilibrium of a monotonic INS representable r yields a sincerely stable marriage.

If a mechanism has a minimally dishonest equilibrium, then it also has a strong equilibrium and the core is non-empty. For instance, when using the Gale-Shapley algorithm or a monotonic INS never-one-sided representable r, the SSM game has a non-empty core.

Corollary 6.5.7. *Let Π be arbitrary and r be the Gale-Shapley algorithm or a monotonic INS never-one-sided representable stable marriage mechanism. The SSM game with r and the minimally dishonest refinement has a non-empty core.*

Corollary 6.5.7 follows immediately from Theorems 6.4.2, 6.4.4 and 6.5.5.

6.5.4 Locally Minimal Dishonesty

In this section we consider a weaker version of minimal dishonesty for which our results apply. To be *locally minimally dishonest* there may not be a pairwise change to the putative preference list that is more honest and results in as good an outcome.

Definition 6.5.8. Let Π be the sincere preferences and let $\bar{\Pi}$ be an equilibrium in the Strategic Stable Marriage game. Man m is locally minimally dishonest if a $\{w, z\} \in \times W \times \{W \cup \{u\}\}$ where $d(\delta(\Pi, m, \{w, z\})_m, \pi'_m) < d(\bar{\pi}_m, \pi'_m)$ implies $r(\bar{\Pi}) = M \succ_{\pi_m} M'$ for some $M' \in r(\delta(\bar{\Pi}, m, \{w, z\}))$.

Once again, if the condition fails to hold, then m could obtain at least as good a result when the more honest $\delta(\bar{\Pi}, m, \{w, z\})_m$ is submitted. Woman w is locally minimally dishonest if symmetric conditions hold for her.

A Nash equilibrium is a *locally minimally dishonest equilibrium* if every individual is locally minimally dishonest and a marriage M is a *locally minimally dishonest stable marriage* with respect to r and Π if there is a locally minimally dishonest equilibrium $\bar{\Pi}$ where $M = r(\bar{\Pi})$.

Recall Example 6.3.8. Both $\bar{\pi}'_{m_3}$ and $\bar{\pi}''_{m_3}$ suffice to show that m_3 is not minimally dishonest. However, $\bar{\pi}'_{m_3}$ does not indicate that m_3 is locally minimally dishonest because $\bar{\pi}'_{m_3}$ cannot be obtained from $\bar{\pi}_{m_3}$ by correcting a single discrepancy. Since $\bar{\pi}''_{m_3}$ is obtained by correcting a single discrepancy, it does show that m_3 is not locally minimally dishonest.

If an individual is not locally minimally dishonest then the individual is also not minimally dishonest. However, an individual may be locally minimally dishonest without being minimally dishonest. Thus every minimally dishonest equilibrium is also a locally minimally dishonest equilibrium but the converse does not necessarily hold. Now we explain why all of our results hold for both minimally dishonest equilibria and locally minimally dishonest equilibria. If a result describes a property of all minimally dishonest equilibria it suffices to show this when individuals are locally minimally dishonest. We have rigged the proofs of such results in this paper to rely only on pairwise changes to $\bar{\Pi}$. On the other hand, when we have shown that an equilibrium with a certain property exists, it sufficed to show this when individuals are minimally dishonest.

6.6 Additional Proofs

6.6.1 Equilibria of the Strategic Stable Marriage Game

Lemma 6.6.1 (Rural Hospitals Theorem [61])**.** *Let $\bar{\Pi}$ be a profile. If there exists a marriage M that is stable with respect to $\bar{\Pi}$ where $s_M(y) = u$, then y is unmarried in every stable marriage with respect to $\bar{\Pi}$.*

For the following proofs let $Q_y(z, \bar{\Pi}) := \{z' : z' \succ_{\bar{\pi}_y} z\}$, the set of individuals that y prefers over z with respect to the preference $\bar{\Pi}$. If $\{y, z\}$ is a blocking pair of marriage M with respect to $\bar{\Pi}$ then $y \in Q_z(s_M(z), \bar{\Pi})$ and $z \in Q_y(s_M(y), \bar{\Pi})$.

We begin by establishing how the set of stable marriages changes when an individuals swaps two individuals in their putative preference list $\bar{\Pi}$.

Lemma 6.6.2. *Suppose M is a marriage such that $s_M(y) = z_3$ where $z_1 \succ_{\bar{\pi}_y} z_2 \succ_{\bar{\pi}_y} u$ and either $z_2 \succ_{\bar{\pi}_y} z_3$ or $z_3 \succ_{\bar{\pi}_y} z_1$. Then M is stable with respect to $\bar{\Pi}$ if and only if M is stable with respect to $\bar{\Pi}' = \delta(\bar{\Pi}, y, \{z_1, z_2\})$.*

Proof. Consider individual rationality first. Since $\bar{\pi}_y$ and $\bar{\pi}'_y$ do not differ as to which individuals y is willing to marry, and $\bar{\Pi}'$ differs from $\bar{\Pi}$ only at y's preferences, M is individually rational with respect to $\bar{\Pi}$ iff M is rational with respect to $\bar{\Pi}'$.

Second, consider blocking pairs. The pair $\{y, z\}$ blocks M with respect to $\bar{\Pi}$ iff $z \in Q_y(z_3, \bar{\Pi})$ and $y \in Q_z(s_M(z), \bar{\Pi})$. With respect to $\bar{\Pi}$, by hypothesis either y strictly prefers both z_1 and z_2 to z_3, or y does not strictly prefer either z_1 or z_2 to z_3. Swapping z_1 and z_2 in $\bar{\pi}_y$ cannot not affect $Q_y(z_3, \bar{\Pi})$. Hence $Q_y(z_3, \bar{\Pi}) = Q_y(z_3, \bar{\Pi}')$. Also, $Q_z(s_M(z), \bar{\Pi}) = Q_z(s_M(z), \bar{\Pi}')$ since $\bar{\pi}_z = \bar{\pi}'_z$. Therefore, $\{y, z\}$ blocks M with respect to $\bar{\Pi}$ iff it blocks M with respect to $\bar{\Pi}'$. $\square$

Lemma 6.6.3. *Suppose M is a marriage such that $s_M(y) = z_3$ where $z_3 \succ_{\bar{\pi}_y} u$, and suppose $u \succ_{\bar{\pi}_y} z_1$. Then M is stable with respect to to $\bar{\Pi}$ iff M is stable with respect to $\bar{\Pi}' = \delta(\bar{\Pi}, y, \{z_1, u\})$.*

Proof. ($\Rightarrow$) Suppose to the contrary that M is stable with respect to $\bar{\Pi}$ but not with respect to $\bar{\Pi}'$. Since $\bar{\Pi}'$ is created by adding an element to $\bar{\Pi}$, if M is individually rational with respect to $\bar{\Pi}$ then it is also individually rational with respect to $\bar{\Pi}'$. Therefore there exists a blocking pair $\{y, z\}$ with respect to $\bar{\Pi}'$. However, as in Lemma 6.6.2, both $Q_y(z_3, \bar{\Pi}) = Q_y(z_3, \bar{\Pi}')$ and $Q_z(s_M(z), \bar{\Pi}) = Q_z(s_M(z), \bar{\Pi}')$. Therefore $\{y, z\}$ must also block M with respect to $\bar{\Pi}$, a contradiction.

($\Leftarrow$) Suppose by the contrapositive that M is not stable with respect to $\bar{\Pi}$. Then M is not individually rational or it admits a blocking pair with respect to $\bar{\Pi}$. Suppose the former. Since $\bar{\Pi}$ and $\bar{\Pi}'$ only differ as to whether or not z_1 is acceptable to y, and $s_M(y) = z_3 \succ_{\bar{\pi}_y} u \succ_{\bar{\pi}_y} z_1$ implies $z_3 \neq z_1$, M is not individually rational with respect to $\bar{\Pi}'$. If the pair $\{y, z\}$ blocks M with respect to $\bar{\Pi}$, then as above $Q_y(z_3, \bar{\Pi}) = Q_y(z_3, \bar{\Pi}')$ and $Q_z(s_M(z), \bar{\Pi}) = Q_z(s_M(z), \bar{\Pi}')$ imply that $\{y, z\}$ blocks M with respect to $\bar{\Pi}'$. Hence M is not stable with respect to $\bar{\Pi}'$. $\square$

Lemma 6.6.4. *At a minimally dishonest equilibrium if y is unmarried, then $\bar{\pi}_y = \pi_y$.*

Proof. We begin by showing there is no z such that $z \succ_{\bar{\pi}_y} u$ but $u \succ_{\pi_y} z$. For contradiction, let Z be the set of z such that $z \succ_{\bar{\pi}_y} u$ but $u \succ_{\pi_y} z$. Furthermore let $z^* \in Z$ be such that $z \succeq_{\bar{\pi}_y} z^*$ for all $z \in Z$. Suppose that y considers being more honest by removing z^* from their preference list and updating the preference profile to $\bar{\Pi}' = \delta(\bar{\Pi}, y, \{z, u\})$. Since y is minimally dishonest, this must generate a worse outcome for y. However, y was unmarried before the update and M must wed y to some individual $z' \in Z$ for some $M \in r(\bar{\Pi}')$. This implies there is a x where $\{y, x\}$ blocks M with respect to $\bar{\Pi}$ but not with respect to $\bar{\Pi}'$. Therefore $x \succ_{\bar{\pi}_y} z'$ and $y \succ_{\bar{\pi}_{z'}} s_M(z')$. However, since $z' \succ_{\bar{\pi}_y} z^*$, $x \succ_{\bar{\pi}'_y} z'$ and $y \succ_{\bar{\pi}'_{z'}} s_M(z')$ and $\{y, x\}$ blocks M with respect to $\bar{\Pi}'$, a contradiction.

Since stable marriages mechanisms select a marriage that is individaully rational and $\bar{\pi}_y$ contains no individual y is unwilling to marry, if y becomes more honest then y will prefer any newly created marriage to being unmarried. Thus y is completely honest. $\square$

Lemma 6.6.5. *At any minimally dishonest equilibrium $\bar{\Pi}$ for a monotonic INS representable stable marriage mechanism, if y is married to the k^{th} member of π_y, then the first k elements of $\bar{\pi}_y$ are a permutation of the first k elements of π_y.*

Proof. We show this in two steps. By Corollary 6.3.3, $r(\bar{\Pi}) = M$ is a singleton.

Claim 1: If $x \succ_{\pi_y} s_M(y)$ then $x \succ_{\bar{\pi}_y} u$.

For contradiction, assume this is not the case. Let $\mathcal{M}$ be the set of marriages stable with respect to $\bar{\Pi}$. Now suppose that y is more honest and adds x to the end of $\bar{\pi}_y$ updating $\bar{\Pi}$ to $\bar{\Pi}' = \delta(\bar{\Pi}, y, \{x, u\})$. By the "if" part of Lemma 6.6.3, for all $M' \in \mathcal{M}$, M' remains stable and since the decision mechanism is INS representable the value of M' remains unchanged. Furthermore, by Lemma 6.6.1 [61], y is married in every stable marriage with respect to $\bar{\Pi}'$. Then by the "only if" part of Lemma 6.6.3, appending x to $\bar{\pi}_y$ can only create new stable marriages that weds y and x. Hence, $r(\bar{\Pi}')$ can only consist of marriages that wed y to $s_M(y)$ or x. Since $x \succ_{\pi_y} s_M(y)$, y is not worse off, contradicting minimal dishonesty. Thus Claim 1 holds.

Claim 2: If $x \succeq_{\pi_y} s_M(y) \succ_{\pi_y} z$ then $x \succ_{\bar{\pi}_y} z$.

Assume to the contrary that $z \succ_{\bar{\pi}_y} x$, and select x and z so that they appear as close as possible in $\bar{\pi}_y$. This selection forces x and z to be adjacent in $\bar{\pi}_y$, for if $z \succ_{\bar{\pi}_y} \alpha \succ_{\bar{\pi}_y} x$ then either α can replace x or α can replace z.

Now suppose that y is more honest and switches x and z, updating $\bar{\Pi}$ to $\bar{\Pi}' = \delta(\bar{\Pi}, y, \{x, z\}$. For $x \neq s_M(y)$, since z and x are adjacent, Lemma 6.6.2 applies, hence M remains stable. If $x = s_M(y)$, similar to Lemma 6.6.2, M remains stable since $Q(s_M(y), \bar{\Pi}') \subset Q(s_M(y), \bar{\Pi})$.

By monotonicity and INS, the values of marriages that wed y to x (respectively z) do not decrease (respectively increase) and the values of all other marriages are unchanged. By Lemma 6.6.2 all new stable marriages wed y to x. Consequently, $r(\bar{\Pi}')$ can only consist of marriages that wed y to $s_M(y)$ or x. Since $x \succeq_{\pi_y} s_M(y)$, y is not worse off with $r(\bar{\Pi}')$, contradicting minimal dishonesty.

Since both claims hold, the first k elements of $\bar{\pi}_y$ are a permutation of the first k elements of π_y. $\square$

Lemma 6.6.6. *At any minimally dishonest equilibrium for a monotonic INS representable stable marriage mechanism, the man-optimal stable marriage $\tilde{M}$ with respect to the putative preferences $\bar{\Pi}$ is selected.*

Proof. Suppose this is not the case and there is a stable $M = r(\bar{\Pi})$, where $M \neq \tilde{M}$. Then there exists a man m such that $s_{\tilde{M}}(m) \succ_{\bar{\pi}_m} s_M(m)$. By Lemma 6.6.5, $s_{\tilde{M}}(m) \succ_{\pi_m} s_M(m)$, a contradiction to Lemma 6.3.2. $\square$

Corollary 6.6.7. *At any minimally dishonest equilibrium for a monotonic INS representable stable marriage mechanism, there is only one stable marriage with respect to the putative preferences.*

Proof. Symmetric to Lemma 6.6.6, the marriage selected must also be woman-optimal with respect to the submitted preferences. Since the man-optimal (woman-optimal) marriage is simultaneously the best (respectively worst) stable marriage for every man [25, 58], there can be only one stable marriage with respect to the putative preferences. $\square$

Lemma 6.6.8. *At any minimally dishonest equilibrium for a monotonic INS representable stable marriage mechanism, if y is married to the k^{th} member of π_y, then the first k elements of $\bar{\pi}_y$ are identical to and appear in the same order as the first k elements of π_y.*

Proof. Let $M = r(\bar{\Pi})$. For contradiction, suppose there are x and z such that $x \succ_{\pi_y} z \succeq_{\pi_y} s_M(y)$ but $z \succ_{\bar{\pi}_y} x$. As in the proof of Lemma 6.6.5, select x and z such that they are adjacent in $\bar{\pi}_y$. Suppose that y is more honest by switching x and z, updating $\bar{\Pi}$ to

$\bar{\Pi}' = \delta(\bar{\Pi}, y, \{x, z\})$. By Lemma 6.6.2, any newly created stable marriages wed y to x. By Corollary 6.6.7, there is only one stable marriage with respect to $\bar{\Pi}$. Therefore, $r(\bar{\Pi}')$ weds y to x or $s_M(y)$. This contradicts minimal dishonesty since y does no worse. □

Lemma 6.6.9. *For preference profiles Π and $\bar{\Pi}$, and a marriage M that is stable with respect to $\bar{\Pi}$, if π_y agrees with $\bar{\pi}_y$ up to y's M-partner, then M is stable with respect to Π.*

Proof. Assume to the contrary that M is stable with respect to $\bar{\Pi}$ but not with respect to Π. Either (1) there exists a married man or woman y such that $u \succ_{\pi_y} s_M(y)$ (M is not individually rational) or (2) there is an unmarried pair $\{m, w\}$ such that $m \succ_{\pi_w} s_M(w)$ and $w \succ_{\pi_m} s_M(m)$ (there is a blocking pair). If (1) then $u \succ_{\bar{\pi}_y} s_M(y)$ since π_y and $\bar{\pi}_y$ agree up to $s_M(y)$, contradicting that M was stable with respect to $\bar{\Pi}$. If (2) then $m \succ_{\bar{\pi}_w} s_M(w)$ and $w \succ_{\bar{\pi}_m} s_M(m)$ since π_y and $\bar{\pi}_y$ agree up to $s_M(y)$ for each y, contradicting that M was stable with respect to $\bar{\Pi}$. Thus, M must be stable with respect to Π. □

6.6.2 The Gale-Shapley Algorithm and Never-One-Sided Mechanisms

Let $Inv(r, \bar{\Pi}, \Pi)$ be the set of violations to the minimally dishonest criterion given equilibrium $\bar{\Pi}$. Formally,

$$Inv(r, \bar{\Pi}, \Pi) = \left\{ \{y, \bar{\pi}'_y\} : d(\bar{\pi}'_y, \pi_y) < d(\bar{\pi}_y, \pi_y) \text{ and } r([\bar{\Pi}_{-y}, \bar{\pi}'_y]) \succeq_{\pi_y} r(\bar{\Pi}) \right\}. \quad (6.59)$$

Let $Inv'(r, \bar{\Pi}, \Pi)$ be the set of $\{y, \bar{\pi}'_y\} \in Inv(r, \bar{\Pi}, \Pi)$ where $\bar{\pi}'_y$ agrees with π_y up to y's partner in $r([\bar{\Pi}_{-y}, \bar{\pi}'_y])$.

We now have sufficient definitions to give an algorithm than finds a minimally dishonest equilibrium that yields a sincerely stable marriage for the Gale-Shapley algorithm. Let $\overline{M}$ be the woman-optimal marriage with respect to the sincere preferences. For each man m, let $\bar{\pi}^{\overline{M}}_m = \pi_m$. For each woman w, if w is married in $\overline{M}$ then let $\bar{\pi}^{\overline{M}}_w$ be the formed by truncating π_w after $s_{\overline{M}}(w)$. If w is unmarried then let $\bar{\pi}^{\overline{M}}_w = \pi_w$. When given the Gale-Shapley algorithm r, sincere preferences Π and the putative preferences $\bar{\Pi}^{\overline{M}}$, Algorithm 2 will output a minimally dishonest equilibrium for Π that yields the since woman-optimal marriage $\overline{M}$.

Furthermore if r is instead monotonic, INS representable and never-one-sided then for any sincerely stable marriage M, Algorithm 2 can output a minimially dishonest equilibrium that yields M. For individual y let $\bar{\pi}_y^M$ be the preferences obtained after y truncates π_y after their M partner. This process was decribed above for the women and the woman-optimal marriage $\overline{M}$. Algorithm 2 will output a minimally dishonest equilibrium for Π that yields the marriage M.

Algorithm 2 Equilibrium Finding Algorithm for Gale-Shapley and never-one-sided Mechanisms

```
procedure EQUILIBRIUMFIND
    while Inv'(r, Π̄, Π) ≠ ∅ do
        Select {y, π̄'_y} ∈ Inv'(r, Π̄, Π)
        Π̄ ← [Π̄_{-y}, π̄'_y)]
    end while
    Output Π̄
end procedure
```

We first show that the algorithm terminates regardless of the input.

Lemma 6.6.10. *Algorithm 2 terminates.*

Proof. Consider the potential function counting the number of differences between $\bar{\Pi}$ and Π:

$$\phi(r, \bar{\Pi}, \Pi) = \sum_{m \in V} d(\bar{\pi}_m, \pi_m) + \sum_{w \in W} d(\bar{\pi}_w, \pi_w) \tag{6.60}$$

where $\phi(r, \bar{\Pi}, \Pi) \geq 0$ in each iteration of Algorithm 2. If individual y updates their preferences in an iteration then $d(\bar{\pi}_y, \pi_y)$ decreases by at least one and $d(\bar{\pi}_z, \pi_z)$ remains unchanged for $z \neq y$. Thus Algorithm 2 must terminate. Furthermore, $d(\bar{\pi}_m, \pi_m) \leq \binom{|W|+1}{2}$ and $d(\bar{\pi}_w, \pi_w) \leq \binom{|V|+1}{2}$ and Algorithm 2 terminates in $O(|V||W|^2 + |V|^2|W|)$ iterations. $\square$

Next we show that at the end of each iteration, $\bar{\Pi}$ is a equilibrium and that M is the only marriage stable with respect to $\bar{\Pi}$. Thus, when Algorithm 2 terminates, it outputs an equilibrium that yields the sincerely stable marriage M.

Lemma 6.6.11. *Suppose Algorithm 2 is given input* Π, $\bar{\Pi}^{\overline{M}}$ *where* $\overline{M}$ *is the woman-optimal sincerely stable marriage and the Gale-Shapley algorithm* r*. At the end of each iteration,* $\bar{\Pi}$ *is an equilibrium and* $\overline{M}$ *is the only marriage stable with respect to* $\bar{\Pi}$.

Proof. By construction of $\bar{\Pi}^{\overline{M}}$, $\overline{M}$ is the only stable marriage at the beginning of the first iteration. Thus $\overline{M}$ is both man and woman-optimal. No man can alter his preferences to obtain an outcome he prefers to the man-optimal marriage [20, Theorem 17]. Thus no man m can alter $\bar{\pi}_m^{\overline{M}}$ to obtain a marriage he prefers with respect to $\bar{\Pi}^{\overline{M}}$. By construction of $\bar{\Pi}^{\overline{M}}$, m also cannot alter $\bar{\pi}_m^{\overline{M}}$ to obtain a marriage he prefers with respect to Π. Symmetrically, no woman can alter her preference to obtain a better outcome and $\bar{\Pi}^{\overline{M}}$ is an equilibrium.

We now show that if $\bar{\Pi}$ is an equilibrium with the unique stable marriage $\overline{M}$ at the beginning of an iteration, then the condition holds at the end of the iteration completing the proof of the lemma.

Suppose that $\{y, \bar{\pi}'_y\} \in Inv'(r, \bar{\Pi}, \Pi)$ is updated in an iteration. Let $\bar{\Pi}' = [\bar{\Pi}_{-y}, \bar{\pi}'_y]$ be the preference profile at the end of the iteration. Since $\bar{\Pi}$ is an equilibrium, y does not receive a better partner with respect to $\bar{\Pi}'$. Since $\{y, \bar{\pi}'_y\} \in Inv'(r, \bar{\Pi}, \Pi)$, y does not obtain a worse partner with respect to $\bar{\Pi}'$ and y is married to $s_M(y)$ in every marriage in $r(\bar{\Pi}')$.

Let M' be any marriage stable with respect to $\bar{\Pi}'$. We first claim that $s_{M'}(y) = s_{\overline{M}}(y)$. Since every man is honest, y is a woman. By construction, both $\bar{\pi}_y$ and $\bar{\pi}'_y$ match π_y up to $s_{\overline{M}}(y)$. Since $\bar{\Pi}$ is an equilibrium that weds y to $s_{\overline{M}}(y)$, $s_{\overline{M}}(y)$ is y's woman-optimal partner with respect to $\bar{\Pi}'$. Thus if there was a M' such that $s_{M'}(y) \neq s_{\overline{M}}(y)$ then y would would obtain worse result since r always selects a marriage that is not woman-optimal[1].

[1]Lemma 6.6.11 also applies to a never-one-sided r for any stable marriage M (not just the woman-optimal $\overline{M}$). However, the individual y who is updating $\bar{\pi}_y$ can be either a man or a woman. Like the Gale-Shapley algorithm, if any new marriage is created when an individual becomes more honest, the resulting marriage will be worse for the individual updating their preferences. Since r is never-one-sided, it will yield a worse outcome for the individual updating their preferences. Therefore no new stable marriages can be created in an iteration of Algorithm 2.

Therefore $s_{M'}(y) = s_{\overline{M}}(y)$ for any stable M'.

We now claim that $\overline{M}$ is the only marriage stable with respect to $\bar{\Pi}'$. Suppose instead $M' \neq \overline{M}$ is stable with respect to $\bar{\Pi}'$ but not $\bar{\Pi}$. Individual y is married to $s_{\overline{M}}(y)$ in M'. Since only y alters his/her preferences, M' is individually rational with respect to $\bar{\Pi}$ if and only if M' is rational for $\bar{\Pi}'$. Thus assume there is a pair $\{m, w\}$ that blocks M' with respect to $\bar{\Pi}$ but not $\bar{\Pi}'$. Once again since y is the individual that alters her preferences, $y = w$. However, by construction of Inv', $\{m, y\}$ must also block M' with respect to $\bar{\Pi}'$ since the set of individuals y prefers to $s_{\overline{M}}(y)$ is the same given $\bar{\Pi}$ and $\bar{\Pi}'$. This is a contradiction and $\overline{M}$ is the only stable marriage with respect to $\bar{\Pi}'$.

It remains to show that $\bar{\Pi}'$ is an equilibrium. This again follows directly from [20, Theorem 17] and the lemma holds. □

It only remains to show that each individual is minimally dishonest. All individuals are minimally dishonest if and only if $Inv(r, \bar{\Pi}, \Pi) = \emptyset$. Thus it suffices to show that in each iteration $Inv(r, \bar{\Pi}, \Pi) = \emptyset$ if and only if $Inv'(r, \bar{\Pi}, \Pi) = \emptyset$.

Lemma 6.6.12. *Suppose Algorithm 2 is given input* Π, $\bar{\Pi}^{\overline{M}}$ *and a monotonic INS representable* r. *In each iteration,* $Inv(r, \bar{\Pi}, \Pi) = \emptyset$ *if and only if* $Inv'(r, \bar{\Pi}, \Pi) = \emptyset$.

Proof. Since $Inv'(r, \bar{\Pi}, \Pi) \subseteq Inv(r, \bar{\Pi}, \Pi)$, one direction holds immediately. Suppose that $\{y, \bar{\pi}'_y\} \in Inv(r, \bar{\Pi}, \Pi)$ and let $\bar{\Pi}' = [\bar{\Pi}_{-y}, \bar{\pi}'_y]$.

Let k be such that y's M-partner is the k^{th} individual in π_y. First we claim there is a $\{y, \bar{\pi}''_y\} \in Inv(r, \bar{\Pi}, \Pi)$ such that the first k elements $\bar{\pi}''_y$ are a permutation of the first k elements of π_y. This claim follows in the same fashion as Lemma 6.6.5. Suppose the first k elements of $\bar{\pi}'_y$ are not a permutation of the first k elements π_y, then one of the two disparities in the proof of Lemma 6.6.5 can be corrected. For each disparity corrected y becomes more honest and, since r is monotonic and INS representable, any new stable marriages created yield at least as good an outcome (with respect to the sincere preferences) for y. Let $\bar{\pi}''_y$ be the preference list obtained from $\bar{\pi}'_y$ after correcting all disparities described in Lemma 6.6.5. Since $\bar{\pi}''_y$ is more honest than $\bar{\pi}'_y$ and yields the same outcome for y, $\{y, \bar{\pi}''_y\} \in Inv(r, \bar{\Pi}, \Pi)$ as desired. Now without loss of generality we may assume the first k elements of $\bar{\pi}'_y$ are a permutation of the first k elements of π_y.

Next we claim we there is a $\{y, \bar{\pi}^*_y\} \in Inv(r, \bar{\Pi}, \Pi)$ such that the first k elements $\bar{\pi}^*_y$ appear in the same order as the first k elements of π_y. Similar to Lemma 6.6.11, $\overline{M}$ is unique marriage stable with respect to $\bar{\pi}'_y$. The remainder of this claim follows in the same fashion as Lemma 6.6.8. By correcting the disparities highlighted in Lemma 6.6.8, the preferences $\bar{\pi}'_y$ can be altered to generate a more honest $\bar{\pi}^*_y$ that produces the same outcome for y as $\bar{\pi}'_y$. Moreover, this implies that $\{y, \bar{\pi}^*_y\} \in Inv'(r, \bar{\Pi}, \Pi)$ completing the proof of the lemma. $\square$

As mentioned in the Section 6.4, Lemmas 6.6.10, 6.6.11 and 6.6.12 imply that there is a minimally dishonest equilibrium that yields the woman-optimal sincerely stable marriage $\overline{M}$.

Only Lemma 6.6.11 is specific to the Gale-Shapley algorithm. However, as mentioned in the proof of Lemma 6.6.11, the result can easily be modified for a never-one-sided r and any stable marriage M. Instead of starting with $\bar{\Pi}^{\overline{M}}$, Algorithm 2 is given the putative preference profile $\bar{\Pi}^{\overline{M}}$ where $\bar{\pi}^{\overline{M}}_y$ is obtained from truncating π_y after y's M-partner. Algorithm 2 then outputs a minimally dishonest $\bar{\Pi}$ that yields the sincerely stable marriage M.

6.6.3 Truth-Tellers and the Student Assignment Problem

Theorem 6.5.2. *Let a subset of players in SSM be truth-tellers. If r is monotonic and INS representable, then for every minimally dishonest equilibrium $\bar{\Pi}$, every marriage in $r(\bar{\Pi})$ is stable with respect to the sincere preferences Π.*

Proof. Let $\bar{\Pi}$ be the preferences at equilibrium and let Π correspond to the sincere preferences. For contradiction, assume there is a $M \in r(\bar{\Pi})$ such that M is not sincerely stable. For each individual y we may assume y is willing to marry $s_M(y)$. If this was not the case for some individual y then y indicates that he/she is willing to marry $s_M(y)$. This implies that y is strategic. Similar to Corollary 6.3.3, y's partner is deterministically selected and y could obtain a better result by reporting an empty preference list. Thus M is sincerely individually rational.

Since M is individually rational but not sincerely stable there is a man m and a woman w such that $w \succ_{\pi_m} s_M(m)$ and $m \succ_{\pi_w} s_M(w)$. The pair $\{m, w\}$ remains a blocking pair regardless of all other individuals' sincere and putative preferences. Thus, without loss of generality assume all other individuals are truth-tellers and thus honest. We now break the problem into three cases.

Case 1: m and w are both truth-tellers. Since $\bar{\Pi} = \Pi$, M must be a sincerely stable marriage completing this case.

Case 2: m and w are both strategic. Since both individuals are strategic, their partner must be selected deterministically and thus for all $M' \in r(\bar{\Pi})$, m and w are married. Assume that w is the k^{th} element in π_m. Lemma 6.6.5 still applies in this setting and the first k elements of $\bar{\pi}_m$ are a permutation of the first k elements of π_m. Thus, similar to Lemma 6.6.6, man m is assigned his man-optimal partner. Similarly w is assigned her woman-optimal partner and, similar to Corollary 6.6.7, m and w must be married in every putatively stable marriage. Thus, similar to Lemma 6.6.8, the first k elements of $\bar{\pi}_m$ must appear in the same order as the first k elements of π_m and by Lemma 6.6.9 the marriage M is stable with respect to the sincere preferences.

Case 3: m is strategic but w is a truth-teller. Man m is the only individual that is dishonest and he can obtain his man-optimal w' by submitting a preference list consisting only of w'. As a result, m must like $s_M(m)$ at least as much as w'. By [20, Theorem 17], m cannot alter his preference to get someone he prefers to w' and thus $s_M(m) = w'$ and M is sincerely stable.

In all three cases we determine that M is sincerely stable and the theorem holds. $\square$

Theorem 6.5.3. *Let a subset of players in SSM be truth-tellers. Suppose r always selects the man-optimal marriage. Then there exists a minimally dishonest equilibrium for any sincere preference profile Π, and in every minimally dishonest equilibrium every strategic woman receives her woman-optimal partner.*

Proof. The proof that every strategic woman receives her woman-optimal partner at every equilibrium is identical to the proof of Theorem 6.4.2. To determine at least one equilibrium exists, we start with the preference $\bar{\Pi}$ obtained from Π by having every strategic woman truncate her list after her woman-optimal partner. In the same fashion as Theorem 6.4.2, when given $\bar{\Pi}$ Algorithm 2 returns a minimally dishonest equilibrium. $\square$

Theorem 6.5.4. *Let a subset of players in SSM be truth-tellers. If r is monotonic, INS representable and never-one-sided, then there exists at least one minimally dishonest equilibrium $\bar{\Pi}$. Moreover for every minimally dishonest equilibrium $\bar{\Pi}$ and $M \in r(\bar{\Pi})$, M is sincerely stable.*

Proof. Let M be a sincerely stable marriage and let $Q(\pi_y, M)$ be the preference list obtained by truncating π_y after y's M-partner. Let $\bar{\pi}^M$ be such that $\bar{\pi}_y^M = \pi_y$ for every truth-teller y and such that $\bar{\pi}_y^M = Q(\pi_y, M)$ for all other y. We now claim that we can select a sincerely stable marriage M such that $s_M(y) = s_{M'}(y)$ for each strategic individual y and each marriage M' that is stable with respect to $\bar{\Pi}^M$.

Let $k > 0$ be the minimum integer for which this does not always hold for k strategic individuals. Let y be such an individual and let M' be such a marriage. Furthermore, assume M' is such that $s_{M'}(y)$ is y's optimal partner with respect to $\bar{\Pi}^M$. By Lemma 6.6.9, M' is sincerely stable. By construction of $\bar{\Pi}^M$ and $\bar{\Pi}^{M'}$, every marriage stable with respect to $\bar{\Pi}^{M'}$ is also stable with respect to $\bar{\Pi}^M$. Furthermore, by selection of M' individual y is married to $s_{M'}(y)$ in every marriage stable with respect to $\bar{\Pi}^{M'}$ and no new stable marriages are created by the additional truncation. However this contradicts the minimality of k and thus we can select a sincerely stable marriage M such that for each strategic individual y and for each marriage M' that is stable with respect to $\bar{\pi}_y^M$ that $s_M(y) = s_{M'}(y)$.

In the same fashion as Theorem 6.4.4, when given the input $\bar{\Pi}^M$ Algorithm 2 finds a minimally dishonest equilibrium $\bar{\Pi}$ such that every $M' \in r(\bar{\Pi})$ is sincerely stable. Furthermore, for every $M' \in r(\bar{\Pi})$ and every strategic individual y, y is married to their M-partner in M'.

The second part of the theorem statement follows directly from Theorem 6.5.2. $\square$

REFERENCES

[1] A. Abdulkadirolu, P. A. Pathak, and A. E. Roth, "The New York City high school match," *American Economic Review*, vol. 95, no. 2, pp. 364–367, 2005.

[2] A. Abdulkadirolu, P. A. Pathak, A. E. Roth, and T. Sönmez, "The Boston public school match," *American Economic Review*, vol. 95, no. 2, pp. 368–371, 2005.

[3] T. S. Alvin E. Roth and M. Unver, "Efficient kidney exchange: Coincidence of wants in a market with compatibility-based preferences," *American Economic Review* **97**, pp. 828–851, 2007.

[4] K. J. Arrow, *Social choice and individual values*. Yale university press, 2012, vol. 12.

[5] K. J. Arrow, A. Sen, and K. Suzumura, *Handbook of Social Choice & Welfare*. Elsevier, 2010, vol. 2.

[6] M. Balinski and T. Sonmez, "A tale of two mechanisms: Student placement," *Journal of Economic Theory*, vol. 84, no. 1, pp. 73–94, 1999.

[7] M. Balinski and R. Laraki, *Majority Judgment: Measuring, Ranking, and Electing*. MIT Press, 2010, ISBN: 9780262015134.

[8] J. Bartholdi, T. Seeley, C. Tovey, and J. VandeVate, "The pattern and effectiveness of forager allocation among food sources in honey bee colonies," *Journal of Theoretical Biology*, vol. 160, pp. 23–40, 1993.

[9] J. J. Bartholdi III, C. A. Tovey, and M. A. Trick, "The Computational Difficulty of Manipulating an Election," *Social Choice and Welfare*, vol. 6, no. 3, pp. 227–241, 1989.

[10] P. Battigalli, G. Charness, and M. Dufwenberg, "Deception: The role of guilt," *Journal of Economic Behavior & Organization*, vol. 93, pp. 227–232, 2013.

[11] C. Berge, "Two theorems in graph theory," *Proceedings of the National Academy of Science*, vol. 43, pp. 842–844, 1957.

[12] S. J. Brams and D. R. Herschbach, "The science of elections," *Science*, vol. 292, no. 5521, pp. 1449–1449, 2001.

[13] S. Branzei, I. Caragiannis, J. Morgenstern, and A. D. Procaccia, “How bad is selfish voting?” In *AAAI*, M. desJardins and M. L. Littman, Eds., AAAI Press, 2013, ISBN: 978-1-57735-615-8.

[14] G. Charness and M. Dufwenberg, “Promises and partnership,” *Econometrica*, vol. 74, no. 6, pp. 1579–1601, 2006.

[15] ——, “Bare promises: An experiment,” *Economics Letters*, vol. 107, no. 2, pp. 281–283, 2010.

[16] G. Cox, *Making Votes Count: Strategic Coordination in the World’s Electoral Systems*. Cambridge University Press, Cambridge, 1997.

[17] CVaRMS. (). Canadian resident matching service. *http://www.carms.ca/*.

[18] G. Demange, D. Gale, and M. Sotomayor, “A further note on the stable matching problem,” *Discrete Applied Mathematics*, vol. 16, no. 3, pp. 217 –222, 1987.

[19] P. Dubey, “Inefficiency of nash equilibria,” *Mathematics of Operations Research*, vol. 11, no. 1, pp. 1–8, 1986.

[20] L. E. Dubins and D. A. Freedman, “Machiavelli and the gale-shapley algorithm,” *The American Mathematical Monthly*, pp. 485–494, 1981.

[21] M. Duverger, “Factors in a two-party and multiparty system,” *Party Politics and Pressure Groups*, pp. 23–32, 1972.

[22] B. Escoffier, L. Gourvès, N. K. Thang, F. Pascual, and O. Spanjaard, “Strategy-proof mechanisms for facility location games with many facilities,” in *Proceedings of the Second International Conference on Algorithmic Decision Theory*, ser. ADT’11, Piscataway, NJ: Springer-Verlag, 2011, pp. 67–81, ISBN: 978-3-642-24872-6.

[23] D. Fotakis and C. Tzamos, “Strategy-proof facility location for concave cost functions,” *CoRR*, vol. abs/1305.3333, 2013.

[24] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991, Translated into Chinesse by Renin University Press, Bejing: China.

[25] D. Gale and L. Shapley, “College admissions and the stability of marriage,” *American Math Monthly* **69**, pp. 9–15, 1962.

[26] D. Gale and M. Sotomayor, “Ms. machiavelli and the stable matching problem,” *The American Math Monthly*, vol. 92, pp. 261–268, 1985.

[27] P. Gärdenfors, "Manipulation of social choice functions," *Journal of Economic Theory*, vol. 13, no. 2, pp. 217–228, 1976.

[28] A. Gibbard, "Manipulation of voting schemes: A general result," *Econometrica: Journal of the Econometric Society*, pp. 587–601, 1973.

[29] U. Gneezy, "Deception: The role of consequences," *The American Economic Review*, vol. 95, no. 1, pp. 384–394, 2005.

[30] D. Gusfield, "Three fast algorithms for four problems in stable marriage," *SIAM Journal on Computing*, vol. 16, no. 1, pp. 111–128, 1987.

[31] D. Gusfield and R. Irving, *The Stable Marriage Problem: Structure and Algorithms*. The MIT Press, 1989.

[32] S. Hurkens and N. Kartik, "Would i lie to you? on social preferences and lying aversion," *Experimental Economics*, vol. 12, no. 2, pp. 180–192, 2009.

[33] A. Hylland and R. Zeckhauser, "The efficient allocation of individuals to positions," *Journal of Political Economy*, vol. 87, no. 2, pp. 293–314, 1979.

[34] R. Irving and P. Leather, "The complexity of counting stable marriages," *SIAM J. Comput.*, vol. 15, no. 3, pp. 655–667, Aug. 1986.

[35] R. Irving, P. Leather, and D. Gusfield, "An efficient algorithm for the "optimal" stable marriage," *Journal of the Association for Computing Machinery*, vol. 34, no. 3, pp. 532–543, Jul. 1987.

[36] K. Iwama and S. Miyazaki, "A survey of the stable marriage problem and its variants," in *Informatics Education and Research for Knowledge-Circulating Society, 2008. ICKS 2008. International Conference on*, 2008, pp. 131–136.

[37] JRMP. (). Japan residency matching program. *http://www.jrmp.jp/*.

[38] A. Kato, "Complexity of the sex-equal stable marriage problem," *Japan Journal of Industrial and Applied Mathematics*, vol. 10, no. 1, pp. 1–19, 1993.

[39] D. Knuth, "Mariages stables," *Les presses de L'Université do Montréal*, 1976.

[40] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *STACS 99*, Springer, 1999, pp. 404–413.

[41] J. Liebowitz and J. Simien, "Computational efficiencies for multi-agents: A look at a multi-agent system for sailor assignment," *Electronic Government, an International Journal*, vol. 2, no. 4, pp. 384–402, 2005.

[42] R. López-Pérez and E. Spiegelman, "Why do people tell the truth? experimental evidence for pure lie aversion," *Experimental Economics*, vol. 16, no. 3, pp. 233–247, 2013.

[43] P. Lu, X. Sun, Y. Wang, and Z. A. Zhu, "Asymptotically optimal strategy-proof mechanisms for two-facility games," in *Proceedings of the 11th ACM Conference on Electronic Commerce*, ser. EC '10, Cambridge, Massachusetts, USA: ACM, 2010, pp. 315–324, ISBN: 978-1-60558-822-3.

[44] T. Lundquist, T. Ellingsen, E. Gribbe, and M. Johannesson, "The aversion to lying," *Journal of Economic Behavior & Organization*, vol. 70, no. 1, pp. 81–92, 2009.

[45] D. Manlove, *Algorithmics of Matching Under Preferences*. World Scientific Publishing Co., 2013.

[46] N. Mazar, O. Amir, and D. Ariely, "The dishonesty of honest people: A theory of self-concept maintenance," *Journal of Marketing Research*, pp. 633–644, 2008.

[47] H. Moulin, *Axioms of cooperative decision making*. Cambridge University Press, 1991.

[48] H. Moulin, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[49] R. B. Myerson, *Game theory*. Harvard university press, 2013.

[50] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*. New York, NY, USA: Cambridge University Press, 2007, ISBN: 0521872820.

[51] NRMP. (). National residency matching program. *http://www.nrmp.org*.

[52] C. Papadimitriou, "Algorithms, games, and the internet," in *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, ACM, 2001, pp. 749–753.

[53] S. Pàpai, "Strategyproof and nonbossy multiple assignments," *Journal of Public Economic Theory*, vol. 3, no. 3, pp. 257–271, 2001.

[54] M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh, "Manipulation complexity and gender neutrality in stable marriage procedures," *Autonomous Agents and Multi-Agent Systems*, vol. 22, no. 1, pp. 183–199, 2011.

[55] B. Pittel, "The average number of stable matchings," *SIAM J. Discret. Math.*, vol. 2, no. 4, pp. 530–549, Nov. 1989.

[56] M. Regenwetter and B. Grofman, “Approval voting, borda winners, and condorcet winners: Evidence from seven elections,” *Management Science*, vol. 44, no. 4, pp. 520–533, 1998.

[57] A. E. Roth, “The economics of matching: Stability and incentives,” *Mathematics of Operations Research* **7**, pp. 617–628, 1982.

[58] ——, “Stability and polarization of interests in job matching,” *Econometrica*, vol. 52, no. 1, pp. 47–57, 1984.

[59] ——, “The evolution of the labor market for medical interns and residents: A case study in game theory,” *Journal of Political Economy* **92**, pp. 991–1016, 1984.

[60] ——, “The college admissions problem is not equivalent to the marriage problem,” *Journal of Economic Theory*, vol. 36, no. 2, pp. 277–288, 1985.

[61] ——, “On the allocation of residents to rural hospitals: A general property of two-sided matching markets,” *Econometrica: Journal of the Econometric Society*, pp. 425–427, 1986.

[62] T. Roughgarden, *Selfish routing and the price of anarchy*, 2005.

[63] T. Roughgarden and É. Tardos, “How bad is selfish routing?” *Journal of the ACM (JACM)*, vol. 49, no. 2, pp. 236–259, 2002.

[64] M. A. Satterthwaite and H. Sonnenschein, “Strategy-proof allocation mechanisms at differentiable points,” *The Review of Economic Studies*, vol. 48, no. 4, pp. 587–597, 1981.

[65] M. A. Satterthwaite, “Strategy-proofness and arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions,” *Journal of economic theory*, vol. 10, no. 2, pp. 187–217, 1975.

[66] A. Schrijver, *Theory of Linear and Integer Programming*. John Wiley & Sons, Chichester, 1986.

[67] C. Teo, J. Sethuraman, and W. Tan, “Gale-Shapley stable marriage problem revisited: Strategic issues and applications,” *Management Science*, vol. 47, pp. 1252–1267, 1999.

[68] G. Weiss, Ed., *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1999, ISBN: 0-262-23203-0.

[69] L. Zhou, “On a conjecture by gale about one-sided matching problems,” *Journal of Economic Theory*, vol. 52, no. 1, pp. 123–135, 1990.