

Manipulability of Stable Mechanisms[†]

By PETER CHEN, MICHAEL EGESDAL, MAREK PYCIA, AND M. BUMIN YENMEZ*

We study the manipulability of stable matching mechanisms and show that manipulability comparisons are equivalent to preference comparisons: for any agent, a mechanism is more manipulable than another if and only if this agent prefers the latter to the former. One important implication is that when agents on one side of the market have unit demand, no stable matching mechanism is less manipulable than another for all agents. (JEL C78, D82)

In this paper, we study manipulability of stable matching mechanisms in a general market: there are two sides, such as firms and workers. Each agent can sign a set of contracts with agents on the other side of the market, and each agent has strict preferences over sets of contracts. Each contract specifies a firm, a worker, and the terms of matching between these two agents, which can involve many components such as wages, benefits, etc. The prominent solution concept in such matching markets is *stability*. If a matching is stable, then each agent is willing to keep all of her contracts and there are no additional contracts that agents would like to sign, possibly by dropping some of their current contracts.¹

We say that mechanism ψ is *as manipulable as* mechanism ϕ if, whenever agent a can gain from misreporting her preferences at mechanism ϕ and achieve a certain outcome, she can also gain and achieve this outcome by misreporting her preferences at mechanism ψ . Theorem 1 shows that this can happen if and only if stable mechanism ϕ is always as good as stable mechanism ψ for the agent. Furthermore, we say that mechanism ψ is *more manipulable than* mechanism ϕ for an agent if it is as manipulable as mechanism ϕ for this agent and, in addition, there exists an instance of the market in which she can gain by manipulating mechanism ψ but cannot replicate this gain by manipulating mechanism ϕ . Theorem 2 shows that this happens if and only if stable mechanism ϕ is always as good as stable mechanism ψ for the agent, and stable mechanism ϕ is strictly better than stable mechanism ψ for

*Chen: Booth School of Business, University of Chicago, 5807 S. Woodlawn Avenue, Chicago, IL 60637 (e-mail: peter_chen@chicagobooth.edu); Egedal: Department of Economics, Harvard University, 1805 Cambridge Street, Cambridge, MA 02138 (e-mail: egedal@fas.harvard.edu); Pycia: Department of Economics, University of California-Los Angeles, 8283 Bunche Hall, Los Angeles, CA 90095 (e-mail: pycia@ucla.edu); Yenmez: Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: byenmez@andrew.cmu.edu). This paper subsumes the manipulability results of Chen et al. (2012). For useful comments, we would like to thank two anonymous referees, Federico Echenique, Utku Ünver, and Peng Wang. Yenmez gratefully acknowledges financial support from National Science Foundation grant SES-1326584.

[†]Go to <http://dx.doi.org/10.1257/mic.20150035> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

¹Stable matchings exist for many-to-many matching markets when contracts are *substitutes* (Roth 1984, Fleiner 2003, Klaus and Walzl 2009, Hatfield and Kominers 2014, Chambers and Yenmez 2014, Pycia and Yenmez 2014). Contracts are substitutes when a contract that is chosen from a larger set is also chosen from a smaller set including that contract.

the agent for some instances of the problem. An important corollary of our results is that no two stable matching mechanisms can be ranked in terms of manipulability when agents on one side of the market have unit demand (Corollary 1). More explicitly, no stable mechanism is as manipulable as (or more manipulable than) another stable mechanism for all agents. While we present the results in a general framework of many-to-many matching with contracts, the results are new in the context of standard one-to-one matching as well.

Our study of manipulability comparisons builds on the prior work of Day and Milgrom (2008) and Pathak and Sönmez (2013). As we show, our manipulability concept is the ordinal equivalent of the concepts they define in auction settings in which agents assign utility values to outcomes.² Translated into the utility-based setting, our definition of as-manipulability takes the form of the as-intense-and-strong manipulability of Pathak and Sönmez, and of the incentive-profile comparisons of Day and Milgrom (for details see our Theorem 3 and the subsequent discussion).

While we study a manipulability concept that takes into account how much agents can gain from manipulation, our results are most closely related to Pathak and Sönmez's results on a weaker manipulability notion, which does not take the size of gains into account, and which they call as-strong manipulability. For this less demanding manipulability concept, they show in the context of one-to-one matching that if mechanism ϕ is always as good as mechanism ψ for an agent, then mechanism ψ is as manipulable as mechanism ϕ by this agent. This corresponds to one of the implications in our Theorem 1.

The manipulability comparisons extend earlier results on strategy-proofness of stable mechanisms. For one-to-one matching markets without contracts, Dubins and Freedman (1981) show that workers cannot gain by misreporting their preferences in the worker-proposing deferred acceptance algorithm (Gale and Shapley 1962). Hatfield and Milgrom (2005) extend this analysis to many-to-one matching markets with contracts in which contracts are substitutes and satisfy the law of aggregate demand for firms. On the other hand, Gale and Sotomayor (1985) point out that even though workers cannot gain by misreporting their preferences, firms may have incentives to misreport their preferences for the same algorithm. Pathak and Sönmez (2013) strengthen this insight and show that any other stable mechanism is more manipulable (in the sense that it does not take the size of gains into account) than the deferred acceptance algorithm for the proposing side, and that the deferred acceptance is more manipulable for the accepting side. In contrast, we take the size of gains into account, and we compare all stable mechanisms.

Manipulability presupposes that agents' preferences belong to some domain of preference rankings. We assume that the preferences satisfy the standard assumptions of substitutes and the law of aggregate demand but we do not restrict attention to the domain of all such preferences.³ We allow any domain of preferences satisfying these two conditions that is *closed* in the following sense: When we

²Since we derive our results in the general model of matching with contracts, our results are applicable both to classical matching problems and auctions.

³These terms are commonly used in the matching literature. We define them, and responsiveness, in the next section.

restrict the set of acceptable contracts for an agent to a stable matching outcome, we remain in the domain. We think that the concept of a closed domain is well suited to the study of incentive properties of matching mechanisms and that it might find applications beyond the analysis of our note. Most of the preference domains used in the literature—e.g., the domain of responsive preferences, the domain of substitutable preferences, and the domain of preferences that satisfy substitutes and the law of aggregate demand—are closed.

Our manipulability notions are applied in Chen et al. (2014). They show the existence of quantile stable mechanisms and compare them in terms of manipulability.

I. Model

There are two sets of agents: the set of firms \mathcal{F} , and the set of workers \mathcal{W} . The set of all agents is denoted by $\mathcal{A} \equiv \mathcal{F} \cup \mathcal{W}$. Each contract x is bilateral and specifies the relationship between a firm-worker pair. The firm and worker associated with contract x are represented by $x_{\mathcal{F}} \in \mathcal{F}$ and $x_{\mathcal{W}} \in \mathcal{W}$, respectively. The set of all contracts is finite and denoted by \mathcal{X} . For a set of contracts $\mathcal{X}' \subseteq \mathcal{X}$, $\mathcal{X}'_a \equiv \{x \mid x \in \mathcal{X}', a \in \{x_{\mathcal{F}}, x_{\mathcal{W}}\}\}$ denotes the set of contracts that agent a is associated with. A set of contracts $\mathcal{X}' \subseteq \mathcal{X}$ is **feasible** if for every firm-worker pair (f, w) , $|\mathcal{X}'_f \cap \mathcal{X}'_w| \leq 1$, i.e., each firm-worker pair can sign at most one joint contract. A **matching** is a feasible set of contracts.

Each agent a is endowed with a strict preference relation \succ_a over sets of contracts that involve agent a , i.e., over $2^{\mathcal{X}_a} \equiv \{\mathcal{X}' \mid \mathcal{X}' \subseteq \mathcal{X}_a\}$. Similarly, agent a 's weak preference relation is denoted by \succeq_a , so for all $\mathcal{Y}, \mathcal{Y}' \subseteq \mathcal{X}_a$, $\mathcal{Y} \succeq_a \mathcal{Y}'$ if and only if $\mathcal{Y} \succ_a \mathcal{Y}'$ or $\mathcal{Y} = \mathcal{Y}'$. Given \succ_a and $\mathcal{X}' \subseteq \mathcal{X}$, let $C_a(\mathcal{X}' \mid \succ_a)$ denote agent a 's most preferred subset of contracts involving agent a from \mathcal{X}' . More formally, $C_a(\mathcal{X}' \mid \succ_a) \subseteq \mathcal{X}'_a$ and for all $\mathcal{Y}_a \subseteq \mathcal{X}'_a$, $C_a(\mathcal{X}' \mid \succ_a) \succeq_a \mathcal{Y}_a$. To ease notation, we suppress the dependence on \succ_a throughout the paper when this does not lead to any ambiguity and denote the set of chosen contracts from subset $\mathcal{X}' \subseteq \mathcal{X}$ by $C_a(\mathcal{X}')$. Similarly, for any set of contracts \mathcal{X}' , let $C_{\mathcal{W}}(\mathcal{X}') \equiv \cup_w C_w(\mathcal{X}')$ and $C_{\mathcal{F}}(\mathcal{X}') \equiv \cup_f C_f(\mathcal{X}')$ be the sets of chosen contracts for the set of workers and firms, respectively. We say that a contract x is **acceptable** to agent a if there exists a set of contracts \mathcal{X}'_a containing x such that $\mathcal{X}'_a \succ_a \emptyset$; otherwise, we say that contract x is **unacceptable** to agent a .

Given agents and their preferences, we would like to find a matching that no set of agents would like to deviate from. This is formalized in the following definition of stability.

DEFINITION 1: Given a preference profile \succ , a matching \mathcal{Y} is **stable** if

- (i) for all a , $C_a(\mathcal{Y}) = \mathcal{Y}_a$ (**individual rationality**) and
- (ii) there does not exist a nonempty set of contracts \mathcal{Z} such that $\mathcal{Z} \cap \mathcal{Y} = \emptyset$ such that for all a , $\mathcal{Z}_a \subseteq C_a(\mathcal{Y} \cup \mathcal{Z})$ (**no blocking**).

Stability for a matching entails two things: individual rationality requires that each agent is better off by holding all of the contracts assigned rather than rejecting some of them. On the other hand, no blocking states that there is no subset of contracts \mathcal{Z} such that every agent a would prefer \mathcal{Z}_a if \mathcal{Z} were available to them. This is the standard definition of stability for many-to-many matching with contracts: see the setwise stability of Klaus and Walzl (2009).

We make the following assumptions on agents' preferences in our analysis:

DEFINITION 2: *Contracts are **substitutes** in the preferences of agent a if for any sets of contracts $\mathcal{Y}, \mathcal{Y}' \subseteq \mathcal{X}$ such that $\mathcal{Y} \subseteq \mathcal{Y}'$ and a contract x*

$$x \in C_a(\mathcal{Y}' \cup \{x\}) \Rightarrow x \in C_a(\mathcal{Y} \cup \{x\}).$$

Contracts are substitutes if a contract that is chosen from a larger set is still chosen from a smaller set including that contract. Substitutability is standard in matching: see Kelso and Crawford (1982) and Roth (1984). It guarantees the existence of a stable matching in our setup (Fleiner 2003).⁴

DEFINITION 3: *Contracts satisfy the **law of aggregate demand** in the preferences of agent a if for all $\mathcal{Y}, \mathcal{Y}' \subseteq \mathcal{X}$ such that $\mathcal{Y} \subseteq \mathcal{Y}'$*

$$|C_a(\mathcal{Y})| \leq |C_a(\mathcal{Y}')|.$$

The law of aggregate demand requires that the number of contracts chosen from a set is bigger than the number of contracts chosen from a subset of that set. The law of aggregate demand was introduced in Alkan (2002), Alkan and Gale (2003), and Fleiner (2003), and its implications were thoroughly analyzed by Hatfield and Milgrom (2005) (see also Kojima 2007).

An important class of preferences that satisfies both substitutability and the law of aggregate demand is the class of responsive preferences. Contracts are **responsive** in the preferences of agent a , if this agent has a ranking of all contracts and some capacity, and if his choice function selects the most preferred contracts in the set without going over the capacity. Responsive preferences are routinely used in matching, including the seminal work of Gale and Shapley (1962).

To study matching mechanisms, we need to specify the preference domains of all agents. For each agent, let us fix a subset \mathcal{P}_a of the set of all preference relations over sets of contracts that include agent a . We refer to \mathcal{P}_a as the **preference domain** of agent a . We refer to $\mathcal{P} \equiv \times_{a \in \mathcal{W}} \mathcal{P}_a$ simply as the **preference profile domain**.

⁴ Aygün and Sönmez (2013) show that substitutability alone does not guarantee the existence of stable matchings when choice rules are taken as primitives and prove that an axiom called *irrelevance of rejected contracts* is needed for the existence. This axiom is satisfied in our setup since choice rules are constructed using strict preferences over sets of contracts.

We need the following definition of closedness for preference profile domains:

DEFINITION 4: *Preference profile domain \mathcal{P} is closed if for all $\succ \in \mathcal{P}$ and for all matchings \mathcal{Y} that are stable with respect to \succ , if the preference relation \succ'_a ranks sets of contracts in the same way as \succ_a except that only contracts in \mathcal{Y}_a are acceptable to agent a , then $\succ'_a \in \mathcal{P}_a$.*

For example, suppose that $\mathcal{Y}_a = \{x_1, x_2\}$ and $\succ_a : \{x_1, x_3\} > \{x_1, x_2\} > \{x_3\} > \{x_1\} > \emptyset$. Closedness implies that when we restrict \succ_a to sets of contracts that only include those in \mathcal{Y}_a , then we remain in the domain. This means that the preference relation $\succ'_a : \{x_1, x_2\} > \{x_1\} > \emptyset$ is in the preference domain \mathcal{P}_a .

Closedness for a preference profile domain means that when we restrict the set of acceptable contracts for an agent to a stable matching outcome, we remain in the domain. There are many such domains:

- the domain of all preferences in which contracts are substitutes,
- the domain of all preferences in which contracts are substitutes and satisfy the law of aggregate demand, and
- the domain of all preferences in which contracts are responsive.

We explain why the domain of all preferences in which contracts are substitutes is closed. Take a stable matching \mathcal{Y} with respect to a substitutable preference profile \mathcal{P} . Since \mathcal{Y} is stable, $C_a(\mathcal{Y}_a | \succ_a) = \mathcal{Y}_a$. Furthermore, substitutability implies that for any subset \mathcal{Y}' of \mathcal{Y} , $C_a(\mathcal{Y}'_a | \succ_a) = \mathcal{Y}'_a$. The preference relation \succ'_a that ranks sets of contracts in the same way as \succ_a except that only contracts in \mathcal{Y}_a are acceptable have the following structure: The only sets that are ranked above the empty set are subsets of \mathcal{Y}_a , and the relative comparison of two such subsets in \succ_a and \succ'_a are the same. As a result, for any set \mathcal{Z} , $C_a(\mathcal{Z} | \succ'_a) = \mathcal{Z}_a \cap \mathcal{Y}_a$. To check substitutability, let $x \in C_a(\mathcal{Z}' \cup \{x\} | \succ'_a)$ and $\mathcal{Z} \subseteq \mathcal{Z}'$. This means $x \in (\mathcal{Z}'_a \cup \{x\}) \cap \mathcal{Y}_a$, so $x \in (\mathcal{Z}_a \cup \{x\}) \cap \mathcal{Y}_a = C_a(\mathcal{Z} \cup \{x\} | \succ'_a)$. Therefore, \succ'_a is in the substitutable preference domain, which shows that the domain is closed. One can similarly see that the other two domains are also closed.

For what follows, we fix a preference profile domain \mathcal{P} that is closed and contracts are substitutes and satisfy the law of aggregate demand in the preferences of agents. This allows us to use the **rural hospitals theorem**, which states that any agent signs the same number of contracts in any stable matching when contracts are substitutes and satisfy the law of aggregate demand in the preferences of agents (see Roth 1986, Hatfield and Milgrom 2005, and Hatfield and Kominers 2014).⁵

A **mechanism** ϕ maps preference profiles from \mathcal{P} to matchings. With a slight abuse of notation, we denote by $\phi(\succ)(a)$ the set of contracts that ϕ allocates to agent a under preference profile \succ . Mechanism ϕ is **stable** if for all preference profiles $\succ \in \mathcal{P}$ the matching $\phi(\succ)$ is stable with respect to \succ .

⁵See also Fleiner (2003) who analyzes a general model encompassing matching with contracts; his Corollary 26 is a version of the rural hospitals theorem.

The following property of closed domains of substitutable preferences satisfying law of aggregate demand will be useful:

LEMMA 1: *If mechanisms ϕ and ψ are stable, then for every agent a and every preference profile $\succ \in \mathcal{P}$, we have*

$$\{\phi(\succ'_a, \succ_{-a})(a) \mid \succ'_a \in \mathcal{P}_a\} = \{\psi(\succ''_a, \succ_{-a})(a) \mid \succ''_a \in \mathcal{P}_a\}.$$

PROOF:

Consider any $\succ'_a \in \mathcal{P}_a$. By the symmetry of the lemma, it is enough to construct $\succ''_a \in \mathcal{P}_a$ such that $\psi(\succ''_a, \succ_{-a})(a) = \phi(\succ'_a, \succ_{-a})(a)$. Let \succ''_a be the same as \succ'_a except that only contracts in $\phi(\succ'_a, \succ_{-a})(a)$ are acceptable to agent a . Since \mathcal{P} is a closed domain, $\succ''_a \in \mathcal{P}_a$.

We first show that $\phi(\succ'_a, \succ_{-a})$ is stable under (\succ''_a, \succ_{-a}) . Since all agents except a have the same preference relation and $\phi(\succ'_a, \succ_{-a})$ is individually rational under (\succ''_a, \succ_{-a}) , it is still individually rational under (\succ''_a, \succ_{-a}) for all agents except a . Furthermore, it is individually rational for agent a under (\succ''_a, \succ_{-a}) because $C_a(\phi(\succ'_a, \succ_{-a})(a) \mid \succ''_a) = \phi(\succ'_a, \succ_{-a})(a)$ by construction of \succ''_a . Furthermore, there cannot exist any blocking set under (\succ''_a, \succ_{-a}) since the same set would also be a blocking set under (\succ'_a, \succ_{-a}) by construction of \succ''_a .

Thus, both $\psi(\succ''_a, \succ_{-a})$ and $\phi(\succ'_a, \succ_{-a})$ are stable matchings under (\succ''_a, \succ_{-a}) ; the rural hospitals theorem implies that $|\phi(\succ'_a, \succ_{-a})(a)| = |\psi(\succ''_a, \succ_{-a})(a)|$. Since only contracts in $\phi(\succ'_a, \succ_{-a})(a)$ are acceptable under \succ''_a , we obtain $\phi(\succ'_a, \succ_{-a})(a) = \psi(\succ''_a, \succ_{-a})(a)$, completing the proof of the lemma. ■

This lemma shows that the set of outcomes that is achievable for an agent in two stable matchings is the same. This lemma is useful in establishing our results below.

A mechanism ϕ is **strategy-proof** for agent a if for every $\succ \in \mathcal{P}$ and $\succ'_a \in \mathcal{P}_a$, $\phi(\succ)(a) \succeq_a \phi(\succ'_a, \succ_{-a})(a)$. In other words, when a mechanism is strategy-proof for an agent, the agent cannot gain from misreporting her preferences.

II. Main Results: Manipulability of Matching Mechanisms

In this section, we study stable matching mechanisms and compare them in terms of their manipulability. Fix a mechanism ϕ . The key auxiliary concept in defining manipulability notions is the set of improvements,

$$M(a, \phi, \succ) = \{\phi(\succ'_a, \succ_{-a}) \mid \succ'_a \in \mathcal{P}_a \text{ and } \phi(\succ'_a, \succ_{-a}) \succ_a \phi(\succ)\},$$

which is comprised of improvements that agent a can obtain by misreporting to mechanism ϕ when the preference profile is \succ . When mechanism ϕ is strategy-proof, then $M(a, \phi, \succ) = \emptyset$ for every $a \in \mathcal{A}$ and $\succ \in \mathcal{P}$.

DEFINITION 5: Mechanism ψ is **as manipulable as** mechanism ϕ for agent a if

$$M(a, \psi, \succ) \supseteq M(a, \phi, \succ)$$

for every preference profile $\succ \in \mathcal{P}$. Mechanism ψ is **more manipulable than** mechanism ϕ for agent a if ψ is as manipulable as ϕ for agent a and there exists a preference profile $\succ \in \mathcal{P}$ such that

$$M(a, \psi, \succ) \supsetneq M(a, \phi, \succ).$$

In other words, mechanism ψ is as manipulable as mechanism ϕ for agent a whenever any gain from misreporting her preferences at mechanism ϕ is also a possible gain under ψ .⁶ More manipulability is the strict ordering counterpart of as-manipulability. These manipulability concepts are natural ordinal counterparts of the cardinal notions of “incentives to misreport” of Day and Milgrom (2008) and of the “intense and strong manipulability” of Pathak and Sönmez (2013). We provide a detailed discussion in the next section.

Our results establish a tight relationship between manipulability and consistent preference for a mechanism: the better the mechanism is for an agent, the less manipulable it becomes.

THEOREM 1: Let a be an agent. Suppose that ϕ and ψ are two stable mechanisms. Then $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ for every $\succ \in \mathcal{P}$ if and only if mechanism ψ is as manipulable as mechanism ϕ for agent a .

A weaker version of the “only if direction” was proven by Pathak and Sönmez (2013) under the assumptions that each worker has unit demand (i.e., for all $w \in \mathcal{W}$ and $\mathcal{Y} \subseteq \mathcal{X}$, $|C_w(\mathcal{Y})| \leq 1$), each firm-worker pair uniquely defines a contract, and contracts are responsive. The “if direction” is new (as is Theorem 2, which we state and prove next).

PROOF:

(\Rightarrow direction) Assume that $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ for every $\succ \in \mathcal{P}$. Let preference profile $\succ \in \mathcal{P}$ and preference relation $\succ'_a \in \mathcal{P}_a$ be such that

$$\phi(\succ'_a, \succ_{-a})(a) \succ_a \phi(\succ)(a).$$

Since ϕ and ψ are stable mechanisms, Lemma 1 implies that there exists a preference relation $\succ''_a \in \mathcal{P}_a$ such that $\psi(\succ''_a, \succ_{-a})(a) = \phi(\succ'_a, \succ_{-a})(a)$. This

⁶Equivalently, mechanism ψ is as manipulable as mechanism ϕ for agent a if for any preference profile $\succ \in \mathcal{P}$, the following holds: if there exists agent a and preference relation $\succ'_a \in \mathcal{P}_a$ such that $\phi(\succ'_a, \succ_{-a})(a) \succ_a \phi(\succ)(a)$, then there exists a preference relation $\succ''_a \in \mathcal{P}_a$ such that $\psi(\succ''_a, \succ_{-a})(a) = \phi(\succ'_a, \succ_{-a})(a)$ and $\psi(\succ''_a, \succ_{-a})(a) \succ_a \psi(\succ)(a)$.

equation, the above-displayed strict preference comparison, and the hypothesis that $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$, imply that

$$\psi(\succ''_a, \succ_{-a})(a) = \phi(\succ'_a, \succ_{-a})(a) \succ_a \psi(\succ)(a).$$

Therefore, mechanism ψ is as manipulable as mechanism ϕ for agent a .

(\Leftarrow direction) Now assume that mechanism ψ is as manipulable as mechanism ϕ for agent a . Suppose, for contradiction, that there exists a preference profile $\succ \in \mathcal{P}$ such that $\psi(\succ)(a) \succ_a \phi(\succ)(a)$. Let \succ'_a be the preference ranking identical to \succ_a except that only contracts in $\psi(\succ)(a)$ are acceptable; since \mathcal{P} is closed, $\succ'_a \in \mathcal{P}_a$. As in the proof of Lemma 1, the stability of ϕ and ψ and the rural hospitals theorem imply that

$$\phi(\succ'_a, \succ_{-a})(a) = \psi(\succ)(a).$$

Consequently, $\phi(\succ'_a, \succ_{-a})(a) \succ_a \phi(\succ)(a)$. This, together with the assumption that ψ is as manipulable as mechanism ϕ for agent a , implies the existence of $\succ''_a \in \mathcal{P}_a$ such that

$$\psi(\succ''_a, \succ_{-a})(a) = \phi(\succ'_a, \succ_{-a})(a) \quad \text{and} \quad \psi(\succ''_a, \succ_{-a})(a) \succ_a \psi(\succ)(a).$$

But the above equality implies $\psi(\succ''_a, \succ_{-a})(a) = \psi(\succ)(a)$, contradicting the inequality. ■

Theorem 1 provides a complete characterization of as-manipulability. Next, we consider the notion of more-manipulability.

THEOREM 2: *Let a be an agent. Suppose that ϕ and ψ are two stable mechanisms. Then (i) $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ for every $\succ \in \mathcal{P}$, and (ii) $\phi(\succ)(a) \succ_a \psi(\succ)(a)$ for some preference profile $\succ \in \mathcal{P}$ are satisfied if and only if mechanism ψ is more manipulable than ϕ for agent a .*

PROOF:

(\Leftarrow direction) Let us first assume that ψ is more manipulable than ϕ for agent a . Then, Theorem 1 implies condition (i). Furthermore, if condition (ii) fails for all preference profiles, then for every $\succ \in \mathcal{P}$, $\psi(\succ)(a) \succeq_a \phi(\succ)(a)$. This inequality and condition (i) imply $\psi(\succ)(a) = \phi(\succ)(a)$. This contradicts the assumption that ψ is more manipulable than ϕ for agent a .

(\Rightarrow direction) For the remainder of the proof, assume conditions (i) and (ii). Theorem 1 implies that $M(a, \psi, \succ) \supseteq M(a, \phi, \succ)$, and it remains to show that $M(a, \psi, \succ) \neq M(a, \phi, \succ)$ for some $\succ \in \mathcal{P}$. To prove this last statement, take $\succ \in \mathcal{P}$ to be a preference profile satisfying (ii). Since ϕ and ψ are stable mechanisms, Lemma 1 implies that there exists a preference relation $\succ'_a \in \mathcal{P}_a$ such that $\psi(\succ'_a, \succ_{-a})(a) = \phi(\succ)(a)$. Thus, $\phi(\succ) \in M(a, \psi, \succ)$. Since $\phi(\succ) \notin M(a, \phi, \succ)$, we infer that $M(a, \psi, \succ) \neq M(a, \phi, \succ)$ and conclude the proof. ■

For the special case of workers with unit demand, that is, when for all $w \in \mathcal{W}$ and $\mathcal{Y} \subseteq \mathcal{X}$ we have $|C_w(\mathcal{Y})| \leq 1$, one corollary of our results is that no two distinct stable mechanisms can be compared in terms of manipulability for all agents in the market.

COROLLARY 1: *If workers have unit demand, then no stable mechanism ψ is as manipulable as another stable mechanism ϕ for all agents in the market.*

In light of our Theorem 1, this corollary follows from Pareto efficiency of stable matchings.⁷ The relation between Pareto efficiency and stability has been analyzed by Haake and Klaus (2009) who showed that in the setting of the corollary every correspondence that is Maskin monotonic, individually rational, and Pareto efficient contains all stable matchings. See Appendix B for the proof of Corollary 1.

The corollary suggests that manipulability cannot be used as a criterion to choose from stable matchings when both sides of the market are strategic agents as in the medical match. However, when only one side of the market is strategic as in school choice then we can use the deferred acceptance algorithm since it is strategy-proof for students (Abdulkadiroğlu and Sönmez 2003).

III. Discussion

Our definitions of as-manipulability and more-manipulability take the size of potential gains into account. In particular, our as-manipulability notion is the ordinal counterpart of Day and Milgrom's (2008) concept of "incentives to misreport" and Pathak and Sönmez's (2013) concept of "intense and strong manipulability," both defined in the setting with cardinal values.

To see the relationship between our ordinal concept and their cardinal ones, we introduce cardinal utilities. For the sake of this characterization, we say that a profile of cardinal utilities $\{u_a\}_{a \in \mathcal{A}}$ that maps the sets of contracts \mathcal{X}_a to utility values $u_a(\mathcal{X}_a) \in \mathbb{R}$ for each agent a is **consistent with** the preference profile \succ if we have $\mathcal{X}_a \succ_a \mathcal{X}'_a \Leftrightarrow u_a(\mathcal{X}_a) > u_a(\mathcal{X}'_a)$ for each agent a .

We have the following equivalence result:

THEOREM 3: *Suppose that ϕ and ψ are two stable mechanisms. Mechanism ψ is as manipulable as mechanism ϕ for agent a if and only if for each $\succ \in \mathcal{P}$, there is a profile u of cardinal utilities consistent with \succ and such that*

$$(1) \quad \max_{\succ'_a \in \mathcal{P}_a} [u_a(\psi(\succ'_a, \succ_{-a})(a)) - u_a(\psi(\succ)(a))] \\ \geq \max_{\succ'_a \in \mathcal{P}_a} [u_a(\phi(\succ'_a, \succ_{-a})(a)) - u_a(\phi(\succ)(a))].$$

⁷We would like to thank an anonymous referee for making this point.

The cardinal manipulability comparison captured in the displayed inequality (1) is what Pathak and Sönmez (2013) call ψ being *as intensely and strongly manipulable as* ϕ .⁸ Furthermore, each side of the above inequality is what Day and Milgrom (2008) call “incentive profile” of the core-selecting auctions they study; the core-selecting auctions are a special case of the stable mechanisms we study in the general matching with contracts framework.

Before proving Theorem 3, let us consider any $\succ \in \mathcal{P}$ and notice that the existence of a utility profile consistent with \succ and satisfying (1) is equivalent to (1) being satisfied for all utility profiles consistent with \succ . Indeed, the equivalence follows easily from the definition of consistency.

PROOF:

Let us fix \succ and any utility profile u consistent with \succ . Since ϕ and ψ are stable, Lemma 1 implies that the maximum utility achievable by an agent under ϕ and ψ is the same:

$$(2) \quad \max_{\succ'_a \in \mathcal{P}_a} u_a(\psi(\succ'_a, \succ_{-a})(a)) = \max_{\succ'_a \in \mathcal{P}_a} u_a(\phi(\succ'_a, \succ_{-a})(a)).$$

This equality yields the theorem as follows:

(\Rightarrow direction) Suppose that mechanism ψ is as manipulable as mechanism ϕ for agent a , and fix \succ and any utility profile u consistent with \succ . By Theorem 1, $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$, and consistency implies $u_a(\phi(\succ)(a)) \geq u_a(\psi(\succ)(a))$. This inequality and equation (2) imply (1).

(\Leftarrow direction) The hypothesis (1) and equation (2) imply $u_a(\phi(\succ)(a)) \geq u_a(\psi(\succ)(a))$. Therefore, $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$, and Theorem 1 implies that mechanism ψ is as manipulable as mechanism ϕ . ■

IV. Conclusion

Studying natural intensity-based manipulability notions for stable matching mechanisms, we have shown that manipulability comparison for an agent is equivalent to preference comparison. One important implication is that when agents on one side of the market have unit demand, then no two stable mechanisms are comparable in terms of manipulability when agents on both sides of the market are strategic as in the medical match.

⁸They define their concept in the allocation setting with a continuum of outcomes, which requires them to take care of tie-breaking. This complication does not arise in our setting, hence our simpler formalization of their concept.

APPENDIX A. ANALOGUES OF OUR RESULTS FOR MANIPULABILITY
WITHOUT COMPARABILITY OF GAINS

While we focus on manipulability concepts that take into account the size of gains (as in Day and Milgrom (2008) and Pathak and Sönmez (2013)), there are also natural manipulability concepts that do not require comparability of gains, but only that manipulations are possible, irrespective of achieved gains. Pathak and Sönmez (2013) study such a weaker concept of manipulability comparison and call it *strong manipulability*. Since their as-strong manipulability concept is more permissive than the concept we call simply “manipulability,” and which corresponds to the “intense and strong manipulability” of Pathak and Sönmez, we refer to the weaker concept as “non-intense manipulability.”

DEFINITION A1: *Mechanism ψ is as non-intensely manipulable as mechanism ϕ for agent a if for any preference profile $\succ \in \mathcal{P}$, the following holds: if there exist agent a and preference relation $\succ'_a \in \mathcal{P}_a$ such that $\phi(\succ'_a, \succ_{-a})(a) \succ_a \phi(\succ)(a)$, then there exists a preference relation $\succ''_a \in \mathcal{P}_a$ such that $\psi(\succ''_a, \succ_{-a})(a) \succ_a \psi(\succ)(a)$.*

For this weaker manipulability concept, the analogue of one of the implications of Theorem 1 is immediately implied by our theorem. However, only a weak form of the reverse direction continues to hold. Let us first show in an example that the full analogue of the reverse direction does not have to hold, and then establish the weak form of the reverse direction that holds.

EXAMPLE A1: *Suppose there are three or more agents on each side of the market and fix an agent a . Consider any two stable mechanisms ϕ and ψ that always choose one of agent a 's most preferred stable matchings except if there are exactly three distinct stable outcomes for agent a . In the latter case, mechanism ϕ chooses one of the worst stable matchings for agent a while mechanism ψ chooses one of the stable matchings that match agent a with his median stable outcome.*

With three or more agents on each side of the market, there is a preference profile with exactly three distinct stable outcomes for agent a . Thus, the two mechanisms are different and mechanism ϕ does not preference-dominate mechanism ψ for agent a . At the same time, mechanism ψ is as non-intensely manipulable as mechanism ϕ .

The analogue of Theorem 1 then takes the following form.

THEOREM A1: *Let a be an agent. Suppose that ϕ and ψ are two stable mechanisms. If $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ for every $\succ \in \mathcal{P}$, then ψ is as non-intensely manipulable as ϕ for agent a . Conversely, suppose that ψ is as non-intensely manipulable as ϕ for agent a and one of these two mechanisms is strategy-proof for agent a , then $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ for every $\succ \in \mathcal{P}$.*

PROOF:

If $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ for every $\succ \in \mathcal{P}$, then ψ is as manipulable as ϕ for agent a by Theorem 1. Thus, ψ is as non-intensely manipulable as ϕ for agent a .

To prove the second statement, we need the following:

Claim: Let φ be a stable mechanism, which is strategy-proof for agent a . Then $\varphi(\succ)(a)$ is the best outcome among all stable matchings with respect to \succ for a .

Suppose not. Let $\mathcal{X}' \subseteq \mathcal{X}_a$ be the best outcome for agent a among stable matchings with respect to \succ . Let \succ'_a be the preference ranking identical to \succ except that only contracts in \mathcal{X}' are acceptable. Since \mathcal{P} is a closed domain, $\succ'_a \in \mathcal{P}_a$. As we have shown in the proof of Theorem 1, the rural hospitals theorem and stability of φ imply $\varphi(\succ'_a, \succ_{-a})(a) = \mathcal{X}'$. Therefore, $\varphi(\succ'_a, \succ_{-a})(a) \succ_a \varphi(\succ)(a)$, a contradiction, since φ is strategy-proof for agent a . This completes the proof of the claim.

Now, there are two cases to consider. First, suppose that ψ is strategy-proof for agent a . Since ψ is as non-intensely manipulable as ϕ for agent a , mechanism ϕ is strategy-proof for agent a . Thus, the claim above implies $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ since ψ is a stable mechanism. Second, suppose that ϕ is strategy-proof for agent a . Then the claim above implies $\phi(\succ)(a) \succeq_a \psi(\succ)(a)$ since ϕ is a stable mechanism. Thus, in both cases the second statement of Theorem 3 holds true. ■

A similar analogue of Theorem 2 also holds true.

APPENDIX B. PROOF OF COROLLARY 1

PROOF:

Suppose, for contradiction, that stable mechanism ψ is as manipulable as another stable mechanism ϕ for all agents in the market. Since ψ and ϕ are different, there exists $\succ' \in \mathcal{P}$ such that $\psi(\succ') \neq \phi(\succ')$. In particular, there exists a firm \hat{f} such that $\psi(\succ')(\hat{f}) \neq \phi(\succ')(\hat{f})$.

By Theorem 1, for every $\succ \in \mathcal{P}$, $\psi(\succ)(a) \succeq_a \phi(\succ)(a)$ for every agent a . In particular, for every worker w , $C_w(\psi(\succ')(w) \cup \phi(\succ')(w)) = \psi(\succ')(w)$ since workers have unit demand.

Let $x \in \psi(\succ')(\hat{f}) \setminus \phi(\succ')(f)$. Such a contract exists since $\psi(\succ')(\hat{f}) \neq \phi(\succ')(\hat{f})$ by construction and $|\psi(\succ')(\hat{f})| = |\phi(\succ')(\hat{f})|$ by the rural hospitals theorem. If $x \in C_{\hat{f}}(\psi(\succ')(\hat{f}) \cup \phi(\succ')(\hat{f}))$, then $x \in C_{\hat{f}}(\{x\} \cup \phi(\succ')(\hat{f}))$ by substitutability. But this means that $\{x\}$ is a blocking set for $\phi(\succ')$, which cannot hold since ϕ is a stable mechanism. As a result, for every $x \in \psi(\succ')(\hat{f}) \setminus \phi(\succ')(f)$, $x \notin C_{\hat{f}}(\psi(\succ')(\hat{f}) \cup \phi(\succ')(\hat{f}))$. Therefore, $C_{\hat{f}}(\psi(\succ')(\hat{f}) \cup \phi(\succ')(\hat{f})) = C_{\hat{f}}(\phi(\succ') \times (\hat{f})) = \phi(\succ')(\hat{f})$. Since choice functions are constructed from preferences, we get $\phi(\succ')(\hat{f}) \succeq_{\hat{f}} \psi(\succ')(\hat{f})$. This contradicts $\psi(\succ')(\hat{f}) \succeq_a \phi(\succ')(\hat{f})$ and $\psi(\succ')(\hat{f}) \neq \phi(\succ')(\hat{f})$. ■

REFERENCES

- Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School Choice: A Mechanism Design Approach." *American Economic Review* 93 (3): 729–47.
- Alkan, Ahmet. 2002. "A class of multipartner matching models with a strong lattice structure." *Economic Theory* 19 (4): 737–46.
- Alkan, Ahmet, and David Gale. 2003. "Stable schedule matching under revealed preference." *Journal of Economic Theory* 112 (2): 289–306.

- Aygün, Orhan, and Tayfun Sönmez.** 2013. "Matching with Contracts: Comment." *American Economic Review* 103 (5): 2050–51.
- Chambers, Christopher P., and M. Bumin Yenmez.** 2014. "Choice and matching." Unpublished.
- Chen, Peter, Michael Egedal, Marek Pycia, and M. Bumin Yenmez.** 2012. "Ranking by Manipulability and Quantile Stable Mechanisms." Unpublished.
- Chen, Peter, Michael Egedal, Marek Pycia, and M. Bumin Yenmez.** 2014. "Quantile Stable Mechanisms." <http://pycia.bol.ucla.edu/chen-egedal-pycia-yenmez-quantiles.pdf>.
- Day, Robert, and Paul Milgrom.** 2008. "Core-selecting package auctions." *International Journal of Game Theory* 36 (3): 393–407.
- Dubins, L. E., and D. A. Freedman.** 1981. "Machiavelli and the Gale-Shapley Algorithm." *American Mathematical Monthly* 88 (7): 485–94.
- Fleiner, Tamás.** 2003. "A Fixed-Point Approach to Stable Matchings and Some Applications." *Mathematics of Operations Research* 28 (1): 103–26.
- Gale, D., and L. S. Shapley.** 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69 (1): 9–15.
- Gale, David, and Marilda Sotomayor.** 1985. "Ms. Machiavelli and the Stable Matching Problem." *American Mathematical Monthly* 92 (4): 261–68.
- Haake, Claus-Jochen, and Bettina Klaus.** 2009. "Monotonicity and Nash implementation in matching markets with contracts." *Economic Theory* 41 (3): 393–410.
- Hatfield, John William, and Scott Duke Kominers.** 2014. "Contract Design and Stability in Many-to-Many Matching." Unpublished.
- Hatfield, John William, and Paul R. Milgrom.** 2005. "Matching with Contracts." *American Economic Review* 95 (4): 913–35.
- Kelso, Alexander S., Jr., and Vincent P. Crawford.** 1982. "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica* 50 (6): 1483–1504.
- Klaus, Bettina, and Markus Walzl.** 2009. "Stable many-to-many matchings with contracts." *Journal of Mathematical Economics* 45 (7–8): 422–34.
- Kojima, Fuhito.** 2007. "The law of aggregate demand and welfare in two-sided matching." *Economics Letters* 99 (3): 581–84.
- Pathak, Parag A., and Tayfun Sönmez.** 2013. "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation." *American Economic Review* 103 (1): 80–106.
- Pycia, Marek, and M. Bumin Yenmez.** 2014. "Matching with Externalities." Unpublished.
- Roth, Alvin E.** 1984. "Stability and Polarization of Interests in Job Matching." *Econometrica* 52 (1): 47–58.
- Roth, Alvin E.** 1986. "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets." *Econometrica* 54 (2): 425–27.