DEFERRED ACCEPTANCE AND REGRET-FREE TRUTH-TELLING

MARCELO ARIEL FERNANDEZ*

ABSTRACT. The deferred acceptance mechanism has been widely adopted across centralized matching markets, despite the fact that it provides participants with opportunities to "game the system." Accounting for the lack of information that participants typically have in these markets in practice, I introduce a new notion of behavior under uncertainty that captures participants' aversion to experience regret. I show that participants optimally choose not to manipulate the deferred acceptance mechanism in order to avoid regret. Moreover, the deferred acceptance mechanism is the unique mechanism within an interesting class (quantile-stable) to induce honesty from participants in this way.

KEYWORDS: Market Design; Deferred Acceptance; Regret; Manipulation; Strategy-proof; Stable mechanisms.

JEL CODES: D47, C78, D81, D82, D91.

1. INTRODUCTION

The deferred acceptance mechanism (DA) occupies a central place in the practice of market design. Among its various applications it is used in two-sided markets to assign rabbis to congregations (one-to-one matching), and to assign graduating medical students to their first position as residents in the U.S. (many-to-one matching).[1] Its success has been largely attributed to the fact that it is a stable mechanism.[2] That is, it takes as input the preferences of participants over their potential partners in the form of ranked order lists, and outputs a matching such that there is no applicant and program pair who prefer each other over their assigned partner. Since the DA

Department of Economics, Johns Hopkins University. Email: fernandez@jhu.edu.

*ACKNOWLEDGEMENTS. I am indebted to Federico Echenique, John Ledyard, and Leeat Yariv for their guidance and support. I am particularly thankful towards Laura Doval, Bhaskar Dutta, Guillaume Haeringer, Thayer Morrill, Marek Pycia, Tayfun Sönmez, Alexander Teytelboym, Ütku Unver, and Bumin Yenmez for discussions which helped improved the paper significantly. For their helpful comments I also thank Atila Abdulkadiroğlu, Moshe Babaioff, Hector Chade, Chris Chambers, David Delacrétaz, Umut Dur, Matt Elliott, Lars Ehlers, Amanda Fridenberg, Nicole Immorlica, Onur Kesten, Bettina Klaus, Flip Klijn, Fei Li, Fabio Maccheroni, Jordi Massó, Tatiana Mayskaya, Antonio Nicolò, Alejandro Neme, Lucas Núñez, Esteban Peralta, Luciano Pomatto, John Quah, Pietro Ortoleva, Wolfgang Pesendorfer, Antonio Romero-Medina, Assaf Romm, Jean-Laurent Rosenthal, Alvin Roth, Chew Soo Hong, Marciano Siniscalchi, Fernando Tohmé, Peter Troyan, Rodrigo Velez, Kyle Woodward, as well as seminar participants at Boston College, Caltech, CMU, Georgetown, HSE, JHU, NCSU, NUS, UBC, and at various conferences. This article is a revised version of the first chapter of my dissertation at Caltech.

[1]For a history of the medical residency match in the U.S. and a list of labor markets that adopted the deferred acceptance see Roth (2008, Table 1). The placement of graduating rabbinical students from the Hebrew Union College-Jewish Institute of Religion is described in Bodin and Panken (2003).

[2]A summary of the evidence is reported in Roth (2002).

produces a matching that is stable with respect to the *reported* preferences, truthful elicitation of preferences is paramount. However, it is known that every stable mechanism is manipulable (Dubins and Freedman, 1981; Roth, 1982). That is, there are configurations of reports that would make it profitable for an agent to misrepresent her preferences when submitting her ranked order list.

This paper reconciles the deferred acceptance mechanism with truth-telling behavior by its participants by leveraging the presence of incomplete information found in most real-world markets, and understanding behavior through the lenses of regret avoidance. To do so, I introduce a new notion of regret: An agent suffers *regret* if she takes an action (e.g. submits a ranked ordered list), and she finds it to be dominated given the information she has ex-post. A mechanism is regret-free truth-telling if no agent ever regrets reporting their preferences truthfully.[3] Crucially, whether an agent regrets a report depends both on her private information as well as on the feedback that she receives from the mechanism. In a typical application, participants are uncertain about others' preferences and reports. Moreover, the feedback received by participants is limited to the outcome of the mechanism (the matching), while true preferences, and submitted reports remain private, even ex-post.

In the context of one-to-one matching, both the applicant- and program-proposing deferred acceptance mechanisms (DA) provide agents on both sides of the market with incentives to report their preferences *truthfully* if they wish to avoid regret, when the information structure presents the features of a typical matching market discussed above. Moreover, truth is the *unique* report that is guaranteed to be free of regret in a market that uses the DA. Consequently, the unique prediction under regret-free behavior, incomplete information, and limited feedback, is that the matching that results from the DA is stable with respect to the true preferences.

The paper also provides a rationalization of the salience of the DA over other stable mechanisms. Regret-free truth-telling characterizes the DA among a class of stable mechanisms known as quantile-stable mechanisms, introduced and studied in Teo and Sethuraman (1998); Klaus and Klijn (2006); Chen et al. (2014, 2016b), among others. Although DA is not unique in the unrestricted domain of all possible stable mechanisms, no stable mechanism that differs from DA whenever possible can be regret-free truth-telling.[4,5]

[3]As a notion, regret-free truth-telling is weaker than strategy-proofness, and ex-post incentive compatibility, while stronger than undominated and worst-case minimizing (truth-telling) behavior; section 6.

[4]The only other stable and regret-free truth-telling mechanisms that I have found seem to be arbitrary (e.g. which stable matching is selected depends on how participants rank the partners that they find unacceptable), and in all cases they assign to every agent the partner that they would get in (either) the applicant- or program-proposing DA.

[5]Outside the class of stable matchings, a companion paper - in preparation - shows that neither the Boston

In the context of many-to-one matching, the applicant-proposing deferred acceptance mechanism continues to provide regret-free truth-telling incentives to agents on both sides of the market. However, the program-proposing deferred acceptance is not regret-free truth-telling for programs. The result further supports the decision made by the Board of Directors of the National Residency Matching Program (NRMP) to switch from the program-proposing to the applicant-proposing deferred acceptance over concerns of strategic manipulations.[6]

The intuition behind the results relates the rules of the mechanism to (i) the inferences that agents can make based on observables; (ii) the shape that profitable misrepresentations take, when profitable; and (iii) how such misrepresentations fare when it is not profitable to manipulate. The shape that misrepresentations must take to manipulate the DA when profitable, are such that they typically lead to detrimental outcomes when it is not profitable to manipulate. For the other stable mechanisms considered in this paper, this is not the case. Having observed a matching, they may regret truth-telling, since submitting an honest but shorter list (that includes the observed partner) would have yielded better outcomes; see section 5.2 for details.

The paper contributes to several strands of literature, reviewed in detail in section 7. Notably, it complements the insightful literature that rationalizes the success of stable mechanisms by showing that the gains from manipulations can be small (in an appropriate sense) when markets are large (Roth and Peranson, 1999; Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Lee, 2017; Azevedo and Budish, 2019, among others) or unbalanced (Ashlagi et al., 2017). The large and unbalanced markets approaches address the empirical ubiquity of small cores, and establish that stable outcomes can be supported in an (not necessarily unique) approximate equilibrium. The small cores argument is not suitable to distinguish among different stable mechanisms. In the case of many-to-one matchings, the arguments do not carry over straightforwardly since the size of the core does not per se limit the ability of programs to manipulate a stable mechanism.[7] In contrast, this paper is silent with regard to the empirical ubiquity of small cores, but supports a stable matching as the unique outcome of the DA when agents avoid regret. Additionally, it distinguishes

mechanism, nor the Top Trading Cycles (TTC) mechanism satisfy regret-free truth-telling for all participants.

[6]Roth and Peranson (1999) reports on the redesign of the NRMP, the reasons behind it, and an evaluation of the effect of these changes.

[7]Kojima and Pathak (2009) directly addresses the many-to-one environment, but relies on the assumption that participants submit short lists (limited acceptability). This assumption implies that a significant fraction of the market would remained unmatched, which does not fit the empirical evidence of the medical match. Regarding the unbalanced markets approach to many-to-one matching, Ashlagi et al. (2017) state: “ We note that these results do not imply that colleges cannot gain from manipulation in unbalanced matching markets (...) We therefore conjecture that when the imbalance is larger than a college’s capacity, even colleges have a limited scope for manipulation, but this conjecture does not directly follow from our analysis, which only establishes that the core is small.”

DA from other stable mechanisms in terms of the incentives to report truthfully, and its insights can be applied to the many-to-one context. The paper also contributes to the analysis of stable matching under incomplete information under different behavioral notions (Barberà and Dutta, 1995; Roth and Rothblum, 1999; Ehlers and Massó, 2007, 2015; Troyan and Morrill, 2020), and to the understanding of what distinguishes the deferred acceptance from other mechanisms (e.g. Kojima and Manea, 2010, among others).[8]

The structure of the paper is as follows: Section 2 presents an illustrative example of the notions of manipulability and regret in the DA. Section 3 introduces basic definitions in matching, as well as quantile- and interior-stable mechanisms. Section 4 introduces the notion of regret. Section 5.1 and 5.2 present the results for one-to-one and many-to-one matching environments respectively, while section 5.3 discusses the informational limits of the results. Section 6 discusses the relation between regret-free truth-telling and other notions of incentive compatibility. Section 7 frames the paper in the context of the literature. All proofs are relegated to the appendix.

2. ILLUSTRATIVE EXAMPLE

Before presenting the formal framework, it is useful to illustrate the notions of manipulability and regret through the following example in the context of the medical residency match.

There are two doctors (Alice and Bob) and two hospitals (City and General), with one vacancy each. Suppose that the true preferences are given by:[9]

$$\succ_{Alice}\text{: General, City} \qquad \succ_{General}\text{: Bob, Alice}$$
$$\succ_{Bob}\text{: City, General} \qquad \succ_{City}\text{: Alice, Bob}$$

All participants submit their preferences to the clearinghouse in the form of ranked order lists (which may differ from their true preferences). The clearinghouse uses the hospital-proposing DA, or H-DA, to generate the matching. It does so by simulating a sequence of proposals and rejections as follows: Hospitals simultaneously make offers to the doctor at the top of their lists. Doctors tentatively hold the best ranked offer among those received, and reject the rest. Hospitals that were

[8]Gale and Shapley (1962) characterize DA through constrained efficiency subject to stability; Balinski and Sönmez (1999) characterize DA as the unique stable mechanism that respects improvements; Alcalde and Barberà (1994) characterize DA by stability and strategy-proofness under the domain restriction of preferences satisfying top-dominance. Kojima and Manea (2010) characterize DA as the unique stable weakly Maskin monotonic mechanism.

[9]Following the convention that if Alice prefers General to City, it is listed as $\succ_{Alice}$: General, City. If an alternative is unacceptable to the agent, it is simply not listed.

rejected make new offers to their top ranked doctors that have not rejected them yet. The process iterates until there are no more rejections, with the last tentative match becoming the final output.

Suppose that Alice knows that all other participants are reporting their preferences truthfully. She decides to try to game the system, and lists only General, with the implication that she is not willing to work for City. Given the reported preferences of others and the rules of H-DA, this strategy is in fact successful. Had Alice report her preferences truthfully, she would have been matched to City. By misrepresenting her preferences, in this scenario, she is matched to General which is her top choice. This strategy is known as a *truncation*. It is a salient strategy in the literature and, under complete information, it is sufficient to consider misrepresentations of the form of truncation strategies (see, for example, Roth and Rothblum, 1999; Ehlers, 2008; Coles and Shorrer, 2014).

However, in real-world applications the participants have *incomplete information* both ex-ante and ex-post; they do not know the preferences of others nor their reports. Privacy concerns limit the amount of information that is revealed even after the matching is implemented. Meaning that while participants observe the resulting matching, the reports remain private even ex-post.

Suppose that Alice is not certain about the preferences and reports of others. Alice performs the truncation in which she only lists General, and observes that the resulting match is:

$$\begin{pmatrix} Alice & Bob & \cdot \\ \cdot & General & City \end{pmatrix}$$

Alice can ask herself which reports are consistent with the observed match. By the stability of the DA, she can conclude that General prefers Bob over her. Otherwise, herself and General would constitute a blocking pair. However, Alice cannot distinguish whether she is acceptable or not for General. Nor can she distinguish whether Bob's preferences coincide with hers or not. That is, by observing the resulting outcome, she can rule out some sets of reports, but is still not able to uniquely pin down what are the reports that everyone sent.

Can Alice come up with an alternative report that, given the information she now has, would have yield her a better outcome than remaining unemployed? The answer is *Yes*. From the stability of DA, it follows that had she told the truth, she could not be worse off than remaining unemployed. Moreover, she knows that there exist scenarios consistent with her observation (e.g. Bob sharing her preferences) where City wanted to hire her, but where her truncation prevented her from being hired. In any such scenario, had she told the truth, she would have been matched to City which she strictly prefers to being unemployed.

Hence, having observed the matching, Alice knows that she would have done better by being honest, and therefore we say in this case that Alice *regrets* having truncated her preferences.

3. FRAMEWORK

This section presents the basic definitions of one-to-one matching markets (section 3.1), as well as domains of stable mechanisms that are considered throughout the paper: quantile- and interior-stable (section 3.2).

3.1. *Basic definitions*

A one-to-one matching (or marriage) market is a triple $(M, W, \succ)$. M is a finite set of men, and W a finite set of women. Each man m is endowed with a strict preference relation, denoted $(\succ_m)$, over the set of women and the possibility of remaining single/unmatched ($W \cup \{m\}$). Woman w is *acceptable* to m whenever $w \succ_m m$, otherwise w is *unacceptable* to m. Similarly $(\succ_w)$ is woman w's strict preference on $M \cup \{w\}$. For an agent $i \in M \cup W$, the weak order associated with $\succ_i$ is denoted $\succeq_i$, and the set of all possible linear orderings for i is denoted by $\mathcal{P}_i$. The preferences of all agents constitute a preference profile, $\succ = ((\succ_m)_{m \in M}, (\succ_w)_{w \in W})$.[10]

A *matching* $\mu : M \cup W \to M \cup W$ assigns to each man m either a woman or himself, $\mu(m) \in W \cup \{m\}$; to each woman w either a man or herself, $\mu(w) \in M \cup \{w\}$; and does so in a consistent fashion, $\mu(m) = w \iff \mu(w) = m$. The set of all matchings for a fixed marriage market is denoted $\mathcal{M}$, and $\mu(m)$ is m's partner under μ.

A matching μ is *individually rational* if every agent prefers their assigned partner to remaining single; that is, $\mu(i) \succeq_i i, \ \forall i \in M \cup W$. A matching μ is *blocked* by a pair (m, w) at $\succ$ if they prefer each other over their assigned partners; that is, $m \succ_w \mu(w)$ and $w \succ_m \mu(m)$. A matching is *stable* if it is individually rational at $\succ$ and it is not blocked by any pair (m, w) at $\succ$. $S(\succ)$ is the set of all stable matchings under preference profile $\succ$.

A *centralized matching mechanism* is an institution that receives reports of preferences from all agents in the economy and produces a matching; formally, it is a mapping $\phi : \mathcal{P} \to \mathcal{M}$. The notation $\phi(\succ)(i) = j$ means that j is i's partner under mechanism ϕ when the reported preferences are $\succ$. The mechanism ϕ is commonly known.

A matching mechanism is *stable* if $\forall \succ \in \mathcal{P}$, $\phi(\succ) \in S(\succ)$. Gale and Shapley (1962) showed that the set of stable matchings $S(\succ)$ is non-empty for any one-to-one matching market. In doing so, they introduced the deferred acceptance algorithm (DA) informally described below.

[10]For any $i \in M \cup W$, $\succ_{-i} = (\succ_j)_{j \neq i}$ denotes the preferences of all agents except i.

- *Step 1*. Every man makes an offer to their most preferred (acceptable) woman. Each woman who receives more than one offer, tentatively holds on to her favorite (acceptable) one, and rejects the rest.
- *Step t*. Each man who was rejected in step $t-1$ makes an offer to his favorite (acceptable) woman who has not rejected him yet. Each woman holds on to her favorite (acceptable) offer among the ones received and the offer tentatively held if any, and rejects the rest.

The above description corresponds to the men-proposing DA (M-DA). The algorithm stops in finitely many steps and the resulting outcome is the men-optimal stable matching. That is, every man (weakly) prefers their assigned partner under this algorithm to the partner they would get in any other stable matching. In an analogous manner one can define the women-proposing deferred acceptance (W-DA), which outputs the women-optimal stable matching. The set of stable matchings is known to present an opposition of interest across sides of the market; i.e. if men unanimously agree that their outcome in stable matching μ, is preferable to another stable outcome μ', then women agree that they prefer μ' over μ. Moreover, the set of unmatched agents is the same across all stable matchings (McVitie and Wilson, 1970).

A matching mechanism ϕ is *strategy-proof* if it is a dominant strategy for every agent to report their preferences truthfully in the direct revelation induced game. This means that every agent cannot do better than to be honest, regardless of the actions of others, or beliefs. Formally, ϕ is *strategy-proof* if $\forall \succ \in \mathcal{P}$ and $\forall i \in M \cup W$ it holds that $\phi(\succ_i, \tilde{\succ}_{-i}) \succeq_i \phi(\succ_i', \tilde{\succ}_{-i}), \forall \succ_i', \forall \tilde{\succ}_{-i}$. Strategy-proof *for men* requires the condition to hold only for men. The M-DA is strategy-proof for men. That is, no matter what other men and women are reporting, a man cannot achieve a better partner by misrepresenting his preferences than he gets by reporting them truthfully.

3.2. *Domains*

In the paper, the following two classes of stable mechanisms are considered: quantile-stable mechanisms, and interior-stable mechanisms. The family of quantile-stable mechanism (also known as generalized-median mechanisms) is introduced and studied by Teo and Sethuraman (1998); Klaus and Klijn (2006); Chen et al. (2014, 2016b), among others. The family of interior-stable mechanism is introduced in this paper.

Definition 1 (Chen et al., 2014, 2016b)**.** *Let $q \in [0,1]$. The q-quantile-stable matching mechanism is the mapping $\{\phi^q : \mathcal{P} \to \mathcal{M} | \mu : \forall m \in M,\ \mu(m)$ is man m's partner in his $\lceil kq \rceil$-th best stable* *matching according to order $\succ_m$, where $k = |S(\succ)|\}$.* [11]

An easy way to interpret quantile-stable mechanisms is to think about the stable set as the size of a pie to be distributed between the set of men and set of women, and $q \in [0, 1]$ as the share women are going to get. Broadly speaking, by choosing q, the designer anchors the ex-post distribution of payoffs across sides of the market, making it constant, regardless of other details such as the number of participants in the market, or how they rank unacceptable alternatives.

The distribution of payoffs being constant across markets also implies that quantile-stable mechanisms are "easy to write," since they can be completely described with one parameter q. This is in the spirit of Wilson's critique (Wilson, 1987), posing as a desideratum for a mechanism not to depend on the fine details of the economy.[12] A particular case is that of the median-stable matching mechanism, ($q = 1/2$) which assigns each individual the partner they have in the median-preferred stable matching. The median-stable matching mechanism appears as a compromise solution between the two side-optimal stable mechanisms. Median-stable matchings have been found to be salient in decentralized two-sided matching problems (Echenique and Yariv, 2011). The family of quantile-stable mechanism is the family of all such compromises.

Next, the family of interior-stable mechanisms is introduced. The denomination "interior" refers to the fact that the mechanism selects a stable matching that is neither the side-optimal nor side-pessimal stable matching, whenever it is possible.[13] Thus, it makes a selection from the "interior" of the lattice of stable matchings, with respect to the side-unanimous ordering.

Definition 2. *A matching mechanism ϕ is interior-stable if for every preference profile $\succ$ where it is possible the mechanism selects a stable matching that is not the M-optimal nor the M-pessimal stable matching; i.e.,* $(\forall \succ\colon |S(\succ)| > 2)\ \left[\phi(\succ) \in S(\succ) \setminus \{\phi^M(\succ), \phi^W(\succ)\}\right]$.

Note that, in general, the class of quantile-stable mechanism and interior-stable mechanism are logically independent. That is, neither one implies the other. Hence, interior-stable mechanisms are picking up on a different type of compromise among stable matchings, across the sides of the

[11] For simplicity of exposition, and w.l.o.g., take $\lceil 0 \rceil = 1$ such that $\phi^0(\cdot) = \phi^M(\cdot)$; that is M-DA.

[12] To illustrate this point consider a mechanism that partitions the set of possible preference profiles into sets B and C, such that $\phi(\succ) = \phi^M(\succ)$ for $\succ \in B$ and $\phi(\succ) = \phi^W(\succ)$ for $\succ \in C$. This mechanism is stable since it always coincides with one of the allocations that a DA would assign. However one might find this mechanism undesirable since inessential changes in the reported preferences (e.g. rearranging the order among the alternatives in an agent's unacceptable set) can lead to large jumps in the distribution of payoffs across men and women.

[13] Interior-stable mechanisms coincide with the allocation prescribed by a DA mechanism only when the set of stable matchings has one or two elements.

market. While not quantifying it formally, the class of interior-stable mechanisms can be thought of as quite a large family since the only restriction (not to coincide with DA when possible) is weak.

4. REGRET AND REGRET-FREE TRUTH-TELLING

In this section I introduce the notion of regret, and define what it means for a mechanism to be regret-free truth-telling. I defer discussing the relation of regret-free truth-telling to existing incentive compatibility notions to section 6, after having presented the main results of the paper.

Regret has been defined as "the emotion that we experience when realizing or imagining that our current situation would have been better, if only we had decided differently" (Zeelenberg and Pieters, 2007). There is evidence in the psychology and neuroscience literature that people have regret, that fear of regret affects behavior, and that the effect of anticipated regret is related to the information the subject knows will be revealed to her.[14] Regret considerations, under different definitions, have been used in other domains to explain overbidding in first price auctions (Filiz-Ozbay and Ozbay, 2007), and to analyze robust monopoly pricing under uncertainty as well as under ignorance (Bergemann and Schlag, 2008, 2011), among others.[15] An action or strategy being "regret-free" has also served as a justification for other equilibrium concepts such as ex-post equilibria (Bergemann and Morris, 2008), and posterior equilibria (Green and Laffont, 1987).

In what follows I assume that each agent knows their own preference, but not that of others. After the mechanism has generated a matching, the entire matching is observable to all agents, but the reports given to the mechanism remain private. No ex-ante restriction is imposed on the possible preferences of others. Alternative assumptions on the information that agents have ex-ante and what they observe ex-post are discussed in section 5.3, together with the robustness of the results in these environments.

Given a mechanism ϕ, and a matching market $(M, W, \succ)$, suppose agent i reports $\succ_i'$, and observes matching μ. Then i's *inference set*, denoted $\mathcal{I}(\mu; \succ_i', \phi) = \{\succ_{-i} \in \mathcal{P}_{-i} : \phi(\succ_i', \succ_{-i}) = \mu\}$, identifies the preference reports that are consistent with the observed matching, given her report, and the known rules of the mechanism.[16] Player i knows that the reported preference profile must be in this set.

[14] See Zeelenberg (1999, 2018), Connolly and Butler (2006), Bourgeois-Gironde (2010), and references therein.

[15] A wide ranging list of applications can be found in Stoye (2009), and Zeelenberg and Pieters (2007).

[16] If participants had any additional information ex-ante or ex-post, then one would refine the inference set appropriately to reflect it.

Definition 3. *Given a mechanism ϕ and an observed matching μ, agent i regrets reporting $\succ_i'$ if there is an alternative report $\succ_i''$ such that*

(i) for each $\succ_{-i} \in \mathcal{I}(\mu; \succ_i', \phi)$ it holds that $[\phi(\succ_i'', \succ_{-i}) \succeq_i \mu]$; and,

(ii) for some $\tilde{\succ}_{-i} \in \mathcal{I}(\mu; \succ_i', \phi)$ it holds that $[\phi(\succ_i'', \tilde{\succ}_{-i}) \succ_i \mu]$.

Agent i regrets reporting $\succ_i'$ because she knows *ex-post* that her report is weakly dominated; i.e. agent i knows that the reported preference profile lies within the inference set and, furthermore, there exists an alternative report that would have resulted in either matching her to the same or a strictly preferred partner.

Definition 4. *Given a mechanism ϕ, a report $\succ_i'$ is regret-free for agent i, if it is not possible for i to regret $\succ_i'$; i.e. there is no matching μ that i could observe after reporting $\succ_i'$ such that there is an alternative report $\succ_i''$, that makes her regret reporting $\succ_i'$.*

A regret-free report guarantees agents that they will never face regret. As such, it appears as a stringent criterion, since it does not take into account how "likely" these matchings where regret obtains are to occur. In this regard, two remarks are in order. First, this is precisely one of the strengths of the notion, since it is does not depend on fine details of the beliefs, in the spirit of Wilson (1987); particularly given the difficulty in practice of establishing objective probabilities in the type of two-sided environments under consideration. Moreover, standard and familiar notions use the same type of criteria. For example, dominant strategy incentive compatibility requires that reporting truthfully is weakly better than any other report for *every* possible report others could make (regardless of the likelihood that they will effectively use such report). Second, section 5.3 shows that the results do not hinge on knife-edge cases; i.e. on matchings or reports that are "unlikely" or implausible to obtain in practice.

Definition 5. *A mechanism ϕ is regret-free truth-telling if for every market, and every agent, truth-telling is regret-free.*

5. RESULTS

Section 5.1 discusses the results pertaining to one-to-one matching markets. It first establishes that DA is regret-free truth-telling (section 5.1.1). It then goes on to show that DA is characterized by regret-free truth-telling in the context of quantile-stable mechanisms, and that no interior-stable mechanism can satisfy regret free-truth-telling (section 5.1.2). The analysis pertaining to many-to-one markets is presented in section 5.2, where it is shown that the program-proposing DA does not satisfy regret-free truth-telling. Lastly, section 5.3 discusses the informational limits of the results.

The issues underlying these results relate the matching mechanism to: (i) the inferences agents can make based on what they can observe; (ii) the shape that profitable misrepresentations must take when it is profitable to misrepresent; and (iii) how those same misrepresentations fare when it is not profitable to manipulate.

5.1. *One-to-one matching*

5.1.1. *Deferred acceptance*

In this section I show that the DA (both men- and women-proposing) provides incentives to report truthfully to agents on *both* sides of the market if they want to avoid regret. The incentives are strict in the sense that truth is found to be the *unique* regret-free report in DA (Proposition 1). Proofs are relegated to appendix A.1.

Theorem 1. *The deferred acceptance mechanism is regret-free truth-telling.*

The reasons why no agent ever regrets truth-telling to a clearinghouse that uses DA are different across the sides of the markets, as the discussion of many-to-one markets (section 5.2) makes clear. For agents on the proposing side, the DA provides dominant-strategy incentives to report their preferences truthfully (Dubins and Freedman, 1981; Roth, 1982). It follows immediately that truth cannot become dominated ex-post, and that they cannot regret truth-telling. For an agent on the receiving side of the mechanism, the result stems from the following two observations. First, the way that DA is manipulated when possible requires the agent to use reports (e.g. truncations) that typically lead to detrimental outcomes when her observed partner is her best achievable partner. Second, given the matching that the agent observes after truthfully reporting her preferences, she cannot *ever* reject the hypothesis that her observed partner is her best achievable partner. Together, these observations imply that the information revealed by observing the outcome of DA mechanism is not sufficient for an agent to to pin down a misrepresentation that would make them regret truth-telling.

I briefly illustrate the argument by considering the medical match example from section 2, in which there are two doctors Alice and Bob, and two hospitals City and General with one vacancy each, and where the clearinghouse uses H-DA. Alice prefers General to City, to remaining unemployed. Suppose that she reports her preferences truthfully. The set of matchings that she can expect to result can be divided into the ones that Alice is matched with her top choice (General), those in which she matches to her last acceptable choice (City), and those in which she is unmatched.

Clearly if, after being honest, Alice matches to her top choice (General), there is no room for her to regret truth-telling. Similarly, when Alice does not match after reporting her preferences truthfully, there is no other report that she could have provided in which she could have matched her to an acceptable partner. Thus, no matching that fits those descriptions can ever lead to Alice regretting reporting her preferences truthfully.

The only possible circumstance in which Alice could regret truth-telling is one in which she matches to her worst acceptable choice (City). Moreover, the only alternative report that could (potentially) lead her to regret truth-telling in this example is a truncation, where she would only list General as acceptable.

In this circumstance, since Alice is honest and does not match with her top choice (General), she infers (due to the stability of DA) that either General finds her unacceptable or that General prefers Bob over her. However, she is unable to determine whether Bob has reported to have the same preferences as her, or the reverse preferences. As argued in section 2, while in the latter case the truncation could have potentially led to a strictly better outcome for her, it is also possible (in the former) that it would have led to a strictly detrimental outcome for her. Therefore, after being honest and matching to City, Alice cannot conclude that a truncation dominates truth-telling ex-post.

Thus, having ruled out all possible ex-post matchings and alternative reports that could make Alice regret truth-telling, it follows that Alice does not regret truth-telling.

The following proposition shows that in the DA truth-telling is the *unique* regret-free report. Thus obtaining truthful reports from all agents is not a consequence of making the behavioral criterion (regret) arbitrarily coarse (for instance compared to strategy-proofness). If so, not only truth-telling, but other reports would also be regret-free.

Proposition 1. *Truth is the essentially unique regret-free report in the DA mechanism. Moreover, an agent regrets any other report through truth.*

The qualifier "essentially" refers to the existence of regret-free reports that are essentially equivalent to truth-telling; i.e. those that differ from it only in how they rank the alternatives in the unacceptable set.

When the centralized clearinghouse uses DA, for any agent, and any possible misrepresentation, there are (non-knife-edge) scenarios where the agent regrets misrepresenting her preferences, and does so because reporting truthfully dominates the misrepresentation. As previously mentioned, any misrepresentation that can be a profitable manipulation of the DA, is such that it typically leads to detrimental outcomes when her best achievable partner is the partner she would obtain under truth-telling. When such detrimental outcome of a misrepresentation obtains the agent posses sufficient information to pin down that truth-telling dominates the misrepresentation ex-post. Thus, in order to regret a misrepresentation, the agent only needs to bear in mind truth-telling, and not necessarily some other complex reports.

The example in section 2 illustrates these forces. The example shows that if an agent performs a truncation, and observes an outcome where she is unmatched, then the agent regrets the truncation.[17] Particularly, she regrets truncating by considering the outcomes that would have been generated if she had told the truth. Ruling out truncations is not sufficient in the present environment. However, the basic forces behind the result are the same for other types of misrepresentations. The arguments are presented in appendix A.1.

5.1.2. *Quantile-stable and interior-stable mechanisms.*

Having argued that the DA is regret-free truth-telling, I now show that no other quantile stable mechanism satisfies regret-free truth-telling (Theorem 2). Thus regret-free truth-telling characterizes the DA among quantile-stable mechanisms (Corollary 1). Moreover, it is shown that no interior-stable mechanism (see Definition 2) satisfies regret-free truth-telling (Theorem 3). Although the class of interior- and quantile-stable mechanisms are logically independent, the results are presented in immediate succession since the same reasoning underlies them. The proofs are relegated to appendix A.2.

Theorem 2. *Let ϕ^q be the q-quantile-stable mechanism, where $q \in (0,1)$. Then ϕ^q is not regret-free truth-telling.*

Corollary 1. *A mechanism is a quantile-stable regret-free truth-telling mechanism if and only if it is (either the men- or women-proposing) deferred acceptance mechanism.*

Theorem 3. *Let ϕ be an interior-stable mechanism. Then ϕ is not regret-free truth-telling.*

[17]It is also worth noting that the argument in general does not depend on the detrimental outcome being unemployment.

Given an observed matching that has been generated by a stable mechanism, it continues to be true that an agent can never reject the hypothesis that her observed partner is her best achievable partner. However, for these classes of mechanisms (quantile- and interior-stable), the shape that some specific misrepresentations (referred to as *soft-truncations*) take when profitable is such that these misrepresentations do not ever lead to detrimental outcomes when the observed partner is the best achievable partner. This means that observing the resulting matching provides sufficient information for agents to find reports that dominate truth-telling ex-post, and hence can lead to an agent to regret truth-telling.

In these soft-truncations the agent reports truthfully her preferences among alternatives in the weak upper-contour set of observed matched partner, and declares all other alternatives as unacceptable. While a soft-truncation is of no consequence for the agent when the clearinghouse uses DA (i.e. would return the same match as being honest always), this is not the case for other stable mechanisms, such as quantile- and interior-stable.

To illustrate the argument I provide an example where an agent regrets truth-telling in the context of the median-stable mechanism. The proof for general quantile-stable, and interior-stable mechanisms can be found in the appendix A.2.

Example 1 (regretting truth in the median-stable mechanism)**.** Consider a matching market with $|M| = |W| = 5$, and let w_1's preferences be

$$\succ_{w_1} : m_1, m_2, m_3, m_4, m_5.$$

Suppose that w_1 reports her preferences honestly to a clearinghouse that uses the median stable mechanism ($q = 1/2$) to generate matchings, and that she observes the matching:

$$\phi(\succ_{w_1}, \cdot) = \begin{pmatrix} w_1 & w_2 & w_3 & w_4 & w_5 \\ m_3 & m_2 & m_1 & m_4 & m_5 \end{pmatrix} = \mu.$$

I claim that w_1 regrets truth-telling when she observes μ, and she does so through the alternative report:

$$\succ'_{w_1} : m_1, m_2, m_3.$$

I proceed in two steps. In step 1, I argue that, under the alternative report $\succ'_{w_1}$, w_1 would obtain a matching that is *weakly* better than her observed one from being honest (m_3) for *any* reports of others in w_1's inference set; thus satisfying condition (i) of the regret definition. In step 2, I argue that, under the alternative $\succ'_{w_1}$, w_1 would obtain a matching that is *strictly* better than her observed one from being honest (m_3) for *some* reports of others in w_1's inference; thus satisfying condition (ii) of the regret definition.

Step 1. Since a median-stable mechanism depends on the entire set of stable matchings, it is important to see how the latter would change if w_1's report would be the soft-truncation $\succ'_{w_1}$.

The following lemma says that any stable matching when w_1 reports the soft-truncation $\succ'_{w_1}$ is also a stable matching when w_1 reports her preferences truthfully $\succ_{w_1}$, for any set of reports of others in w_1's inference set. This provides crucial information about what the stable set would look like under each type of report, and consequently about what matching would be implemented under each report. In fact, the stable matchings under the soft-truncation are exactly those that are stable when she reports truthfully *and* that are weakly preferred by w_1 to μ.

Lemma 1. $\forall \succ_{-w_1} \in \mathcal{I}(\mu; \succ_{w_1}, \phi^q)$,

$$S(\succ'_{w_1}, \succ_{-w_1}) = \{\mu' \in S(\succ_{w_1}, \succ_{-w_1}) : \mu' \succeq_{w_1} \mu = \phi^q(\succ)\}.$$

Corollary 2. $\phi^q(\succ'_{w_1}, \succ_{-w_1}) \succeq_{w_1} \phi^M(\succ'_{w_1}, \succ_{-w_1}) = \phi^q(\succ_{w_1}, \succ_{-w_1})$.

The women-pessimal stable matching that M-DA would produce under w_1's soft-truncation $\succ'_{w_1}$ must be the same as the one the median-stable mechanism produces when w_1 reports truthfully. In terms of the example, it means w_1 must be getting m_3 as her assigned partner in M-DA when she reports $\succ'_{w_1}$, which coincides with her assigned partner in the median matching when reporting $\succ_{w_1}$. Thus, the lemma has the key implication that ex-post w_1 knows she would not have obtained a detrimental outcome by using the soft-truncation $\succ'_{w_1}$ instead of reporting her preferences truthfully $\succ_{w_1}$.

Step 2. It remains to be shown that for some report of others in the inference set, reporting $\succ'_{w_1}$ would generate a matching that is *strictly* preferred by w_1 to μ; i.e. $\succ'_{w_1}$ satisfies condition (ii) from the regret definition at μ.

By performing a soft-truncation w_1 forces the mechanism to calculate the median with respect to a set that contains only weakly preferred matchings to the one obtained through truth-telling. I show that in fact the median over this set selects a partner for w_1 that she strictly prefers.

$$\begin{aligned}
&\succ^*_{m_1}: w_2, w_3, w_4, w_5, w_1 && \succ^*_{w_1}: m_1, m_2, m_3, m_4, m_5 \\
&\succ^*_{m_2}: w_3, w_4, w_5, w_1, w_2 && \succ^*_{w_2}: m_2, m_3, m_4, m_5, m_1 \\
&\succ^*_{m_3}: w_4, w_5, w_1, w_2, w_3 && \succ^*_{w_3}: m_3, m_4, m_5, m_1, m_2 \\
&\succ^*_{m_4}: w_5, w_1, w_2, w_3, w_4 && \succ^*_{w_4}: m_4, m_5, m_1, m_2, m_3 \\
&\succ^*_{m_5}: w_1, w_2, w_3, w_4, w_5 && \succ^*_{w_5}: m_5, m_1, m_2, m_3, m_4
\end{aligned} \tag{$\succ^*$}$$

Under profile $\succ^*$ there are exactly five stable matchings, linearly ordered according to the unanimous preferences of either side, and such that in every stable matching every agent is assigned a different stable partner.[18] One can verify that the preference profile $\succ^*$ is in w_1's inference set when she observes μ being the outcome of the median-stable mechanism. By lemma 1 it follows that, under the soft-truncation, the median-stable mechanism selects the partner associated with each man/woman's 2-nd best stable matching, therefore

$$\phi^q(\succ'_{w_1}, \succ^*_{-w_1}) \succ_{w_1} \phi^q(\succ_{w_1}, \succ^*_{-w_1}).$$

Then, $\succ'_{w_1}$ satisfies condition (ii) of the definition of regret at μ.

Step 1. and 2. together imply that, given her inference set after reporting truthfully and observing μ, w_1 knows she would have done at least as well by reporting $\succ'_{w_1}$ as by reporting $\succ_{w_1}$; additionally, she knows there exists reports consistent with the observed matching where the report $\succ'_{w_1}$ would have yielded a strictly preferred matching to the one obtained by truth-telling. Thus, when the clearinghouse uses the median-stable mechanism, w_1 regrets truth-telling $\succ_{w_1}$ when she observes matching μ, by considering the outcomes that would have resulted from the soft-truncation $\succ'_{w_1}$. The same argument can be extended to quantile-stable mechanisms in general, as shown in appendix A.2.

Theorems 2 and 3 show that a mechanism being regret-free truth-telling is not a necessary consequence of stability, but moreover, they help distinguish DA from other stable matching mechanism. In this sense, they help rationalize the exclusive adoption of DA mechanisms (among

[18] These are referred to as Latin square preferences following Hwang (1978, 1986). They have the property that no two women rank the same man in the same position, and men rank women in their mirror image; i.e. if w_1's top choice is m_1, then m_1 ranks w_1 last.

stable mechanisms) in practice. The feature that DA provides incentives to agents on both sides of the market to report their preferences truthfully, where other stable matching mechanisms do not, contrasts with the results of the large/unbalanced markets approach, where all stable mechanisms appear to provide the same incentives to agents, as well as the insightful contributions by Pathak and Sönmez (2013) and Chen et al. (2016a) where the authors define notions to compare the vulnerability of stable mechanisms to manipulation, and show that no two stable mechanisms can be ranked in terms of manipulability for all agents. In a loose sense, the key distinction with respect to the latter approach is that their manipulability comparisons across mechanisms allows agents to tailor their manipulations to the true preference profile, which would be akin to letting them leverage an underlying complete information structure.

5.2. *Many-to-one matching*

In the context of many-to-one matching markets, which is ubiquitous in practice, the next theorem shows that only the doctor-proposing deferred acceptance algorithm satisfies regret-free truth-telling for all agents.[19] The result strengths the negative conclusion of Roth (1985) which showed that the H-DA is not strategy-proof for hospitals by establishing that in the H-DA hospitals regret truth-telling. On the other hand, the result also presents a positive counterpart to Roth (1985)'s impossibility result regarding the existence of a strategy-proof mechanism for hospitals, due to the weakening of the incentive constraints to be regret-free truth-telling. The result also complements the strands of literature regarding large/unbalanced markets.[20]

The intuition behind the result can be traced back to the asymmetry of inferences that agents on each side of the market can make based on the observed matching, and the known fact that a hospital with multi-unit capacity may be competing excessively within side, or competing against themselves through their multiple positions (see e.g. Sönmez, 1997, 1999).

Theorem 4. *ϕ is a quantile-stable and regret-free truth-telling mechanisms if and only if ϕ is the doctor-proposing deferred acceptance.*

Given that preferences of hospitals are assumed to be responsive, the logic and proof of why truth-telling is regret-free in the doctor-proposing deferred acceptance for all agents follows the

[19]The many-to-one environment under consideration corresponds to the classic college admissions problem of Gale and Shapley (1962), and is therefore omitted. Hospitals are assumed to have responsive preferences (Roth, 1985), meaning that preferences over sets of doctors are consistent with how they rank doctors individually. In practice, the NRMP elicits responsive reports of preferences.

[20]See footnote 7.

same steps as Theorem 1, and is thus omitted. Adapting the example from Roth (1985) and Sönmez (1997) to account for the information structure, we obtain the negative result.

Example 2 (Hospital regrets truth-telling in H-DA)**.** There are three hospitals $\{1,2,3\}$ with respective capacities $(2,1,1)$, and four doctors $\{A,B,C,D\}$. Hospital 1's (responsive) preferences are given by $\succ_1$: A,B,C,D. Consider the case where hospital 1 reports its preferences truthfully, and observes that the outcome of the H-DA is:

$$\phi^H(\succ_1,\cdot) = \begin{pmatrix} 1 & 2 & 3 \\ \{C,D\} & B & A \end{pmatrix} = \mu.$$

The following preference profile $\hat{\succ}_{-1}$ is in 1's inference set.

$$\begin{array}{llll} q & & & \\ 2 & \succ_1: & A,B,C,D & \hat{\succ}_A: \; 3,1,2 \\ 1 & \hat{\succ}_2: & A,B,C,D & \hat{\succ}_B: \; 2,1,3 \\ 1 & \hat{\succ}_3: & C,A,B,D & \hat{\succ}_C: \; 1,3,2 \\ & & & \hat{\succ}_D: \; 1,2,3 \end{array}$$

Under μ, hospital 1 is filling both slots with their least desirable candidates. If the reports of others are $\hat{\succ}_{-1}$, then the outcome comes about due to an excessive within-side competition. Hospital 1's second vacancy generates a chain of rejections, that makes hospital 1 lose the ability to land doctor A, as well as doctor B.[21] Had hospital 1 reported $\succ_1'$: B,D,C,A then the result would have been

$$\phi^H(\succ_1',\hat{\succ}_{-1}) = \begin{pmatrix} 1 & 2 & 3 \\ \{B,D\} & A & C \end{pmatrix}$$

which is strictly better according to 1's true preferences. This preferable outcome follows from reducing within-side competition, thus avoiding the rejection chain. Crucially, the misrepresentation of hospital 1's preferences $\succ_1'$ does not decrease the ranking of doctors that hospital 1 matched to under μ.

Since hospital 1 is matched to its last acceptable candidates C,D under truth-telling, the only way

[21]Hospital 1's first vacancy displaces hospital 2 from matching with doctor A, while its second vacancy eventually displaces hospital 3 from matching with doctor C. These in turn displace hospital 1 from obtaining either doctor A or B.

that the misrepresentation $\succ_1'$ could leave hospital 1 worse is by producing a match in which one (or both) positions remain vacant. However, given that $\succ_1'$ lists all doctors as acceptable any such matching would be blocked by the either doctors C or D and hospital 1, which would contradict the stability of H-DA. Hence, the outcome under the misrepresentation $\succ_1'$ is weakly preferred by hospital 1 to the observed matching μ for any set of report of others $\succ_{-1}$ in 1's inference set.

Thus, hospital 1 regrets truth-telling ($\succ_1$) after observing matching μ by considering the alternative report $\succ_1'$.

The preceding analysis tackled the case of manipulation via preferences. However, Sönmez (1997) showed that stable mechanisms are also vulnerable to manipulation via capacities in the many-to-one matching environment. For the sake of brevity, I state the following result on regret-free truth-telling in regards to manipulation via capacities, and defer its discussion and proof to the appendix A.3.

Theorem 5. *If ϕ is the doctor-proposing deferred acceptance mechanism, then ϕ is regret-free truth-telling for doctors, and reporting the true capacities is regret-free for hospitals. On the other hand, if ϕ is the hospital-proposing deferred-acceptance mechanism, a hospital can regret reporting the true capacity.*

5.3. *Information Limits*

The concept of regret introduced in this paper needs to account to the structure of information and feedback given to players, since the behavior supported as regret-free changes with these. The following examples address the extent to which the results are robust to different informational assumptions. Example 3 shows that both incomplete information, and limited feedback play an important role in making the deferred acceptance regret-free truth-telling. Example 4 illustrates that reasonable degrees of uncertainty suffice to make truth-telling regret-free in DA.

Example 3 (Complete information ex-ante)**.** Consider a clearinghouse that uses M-DA. There are two men and two women with preferences given by:

$$\succ_{m_1}: \quad w_1, w_2 \qquad\qquad \succ_{w_1}: \quad m_2, m_1$$
$$\succ_{m_2}: \quad w_2, w_1 \qquad\qquad \succ_{w_2}: \quad m_1, m_2$$

In this case, there is no uncertainty regarding others' preferences, but only strategic uncertainty regarding what they report to the clearinghouse. Suppose that w_1 reports her preferences truthfully, and observes the outcome

$$\mu = \begin{pmatrix} m_1 & m_2 \\ w_1 & w_2 \end{pmatrix}.$$

Assuming w_1 knows that men do not play weakly dominated strategies, then w_1's inference set is degenerate, and coincides with the true preference profile. Since it presents multiple stable matchings, it follows that w_1 has (and knows that it has) a profitable manipulation. Thus w_1 would regret truth-telling when observing $\mu = \phi^M(\succ)$ by considering the outcome that would result from the *truncation* $w_1' : m_2$.[22]

Remark (Complete information ex-post)**.** If the mechanism were to reveal the entire set of messages that gave rise to the observed matching, again, the inference set of agents would be degenerate, even if there was ex-ante incomplete information. Thus, whenever the set of reports is associated with a non-singleton stable set, there is an agent that regrets their report.

Theorem 1 shows that DA is regret-free truth-telling in an environment where agents have incomplete information about others' preferences. In contrast, example 3 shows that, under complete information, the DA can lead to agents (on the receiving side) to regret truth-telling. The natural question that follows is: How much uncertainty is needed to achieve the result? I partially answer this question in the context of the previous example, in the form of sufficient conditions.

Example 4 (Sufficient uncertainty)**.** If, in addition to the preferences described in example 3, w_1 believes that it is possible that either (i) woman w_2 shares her preference ($\succ'_{w_2}$: m_2, m_1); or that, (ii) either man only find their favorite woman as acceptable ($\succ'_{m_1}$: w_1; $\succ'_{m_2}$: w_2), then truth-telling is regret-free for w_1.

In order to manipulate the M-DA when profitable, w_1 would have to truncate her preferences, by not listing m_1. However, in any of the other cases that w_1 believes to be possible there is a unique stable matching, and in this matching w_1's partner is m_1. Hence, in any such case, the truncation would lead to a strictly worse outcome, by leaving w_1 unmatched.

Truth-telling is regret-free for w_1, as long as she believes it to be possible that her observed partner is her best attainable partner.

[22]One should also note that truth-telling can be regret-free under complete information, whenever there is a unique stable matching with respect to the true preference profile.

6. DISCUSSION

The notion of regret-free truth-telling introduced in this paper is a weaker than that of strategy-proofness. It is also weaker than ex-post incentive compatibility, given the known equivalence between the latter and strategy-proofness in revelation games under private values. Still, one can readily see that the notion of regret-free is in the spirit of robust mechanism design. Concepts like ex-post and posterior equilibria have been justified on the grounds of satisfying some (undefined) notion of "no regret" or being "regret-free." Ex-post incentive compatibility requires truth-telling to be optimal ex-post, that is even if the agent learns the true type and actions of all other agents. Posterior equilibria imposes a similar yet weaker requirement, that truth-telling remains optimal even if the agent is to learn only the actions of others; but not necessarily their types. Regret-free truth-telling requires the action to remain *undominated* given the information that the agent has ex-post, which includes their ex-ante information, as well as all the informational feedback that the mechanism provides.

The definition of regret presented in this paper is by no means the only possible one. Since the literature on regret is too vast to survey here, instead, I address the prominent notion of minimizing maximum regret in the sense of Savage (1951).[23,24] For example, the literature on robust mechanism design has used minimizing maximum regret to study robust monopoly pricing, and monopoly pricing under ignorance (Bergemann and Schlag, 2008, 2011). Under this notion an individual computes her regret as the difference in outcomes between what would have been the optimal action at the true state of the world and the action she takes (see Bell, 1982). This requires a knowledge of the true preference profile at the end of the game which is absent in the setup analyzed in this paper, since in most matching applications privacy concerns prevent this sort of information revelation. Moreover, minimizing maximum regret requires cardinal information, both absent in this paper's setup, as well as in the data that is elicited from participants of these matching markets. The report that would minimize an agent's maximum regret in the sense of Savage (1951) does not generally coincide with a regret-free report for that agent.

It should also be noted that regret and the implied regret-minimization presented in this paper is different from worst-case minimization. For example, in the context of matching markets, under

[23]Milnor (1954) provides a prior-less axiomatization, while Hayashi et al. (2008) and Stoye (2011) deal with the multi-priors as well as other criteria. Sarver (2008) axiomatizes anticipated regret in the context choice over menus of lotteries.

[24]This notion is widely used both in statistical decision theory, treatment choice, and computer science; see Manski (2007); Stoye (2009); Halpern and Pass (2012) and references therein.

worst-case minimization, the agent expects to remain unmatched irrespective of her report in a stable mechanism. Therefore, she is indifferent among all reports that do not list unacceptable options as acceptable.

Worst-case minimization is regularly considered in the context of decision making under ignorance. Other related concepts are leximin preferences, protective strategies, and prudent strategies (Barberà and Jackson, 1988; Moulin, 1981). Barberà and Dutta (1995) introduce the notion of protective strategies and show that truth-telling is the unique protective strategy equilibrium in the revelation game associated with the deferred acceptance, both with and without transfers. The concept of protective strategies is a refinement of worst-case minimization. Two strategies that have the same worst outcome are distinguished by their 2nd-worst outcome; if the latter coincides, then they are that distinguished by the 3rd-worst outcome, etc. Under certain information structures, the regret-free criterion and protective strategies coincide, however, in general these two criteria are distinct. For instance, in a complete information environment (such as the one depicted in Example 3) truth-telling is a protective strategy for any participant of the DA, but it is not regret-free.

7. RELATED LITERATURE

This paper contributes to several strands of literature. First, it contributes to understanding the incentives DA provides for truth-telling, and distinguishing mechanisms by their susceptibility they are to manipulation. Second, it contributes to the literature on robust mechanism design and decision making under ignorance. Third, it contributes to the recent and emerging literature on market design with non-standard preferences and behavioral considerations.

Dubins and Freedman (1981), and Roth (1982) show that there exists no stable strategy-proof mechanism. Moreover, Pathak and Sönmez (2013) and Chen et al. (2016a) define notions of a mechanism being more manipulable than another, and show that stable mechanisms cannot be ranked by their manipulability *for all agents*. Based on these observations it would seem we are left with multiple stable mechanisms and no clear way of choosing among them in terms of their incentives for truthful reporting. This paper identifies incentives that DA provides agents to report truthfully that are not provided by any other quantile- or interior-stable mechanism.

Roth (1989), Roth and Rothblum (1999), and Ehlers (2008) look at the problem of incentives in DA from the Bayesian point of view, where expected utility maximizing participants have (common prior) beliefs over each other's preferences. Roth (1989) shows that it is possible that no Bayes-Nash equilibrium of the DA induced revelation game supports a matching outcome that is stable with respect to the true preferences. Looking to provide advice to participants, Roth and Rothblum (1999) show that truncation strategies first order stochastically dominate other misrepresentations under a symmetry conditions on priors. However, no clear order between truncation strategies and truth-telling emerges. Ehlers (2008) shows that the symmetry condition of Roth and Rothblum (1999) is stringent.

Using data from the NRMP, Roth and Peranson (1999) show that only a small set of doctors and hospitals receive different assignments under D-DA and H-DA; often referred to as a "small core." Aided by simulation results, they conjecture that the large size of the NRMP market explains why the core is small. Building on these observations, an insightful literature emerged that show that as the market grows large or unbalanced, the core of the market shrinks (in an appropriate sense). Under the assumption of limited acceptability, Immorlica and Mahdian (2005), and Kojima and Pathak (2009) show that only a vanishing fraction of the market has incentives to deviate from a truth-telling equilibrium, for one-to-one and many-to-one matching markets respectively. Lee (2017) dispenses with limited acceptability by using techniques from random matrices, and shows that truth-telling by all agents can be supported as an approximate equilibrium. On the other hand, Coles and Shorrer (2014) establishes that, absent bounds on the (random) cardinal utilities, agents on the receiving side of the DA may still have large incentives to truncate their preferences significantly, even in large markets. Azevedo and Budish (2019) introduce the concept of strategy-proofness in the large, and showed many well-known mechanisms (the DA among them) across different environments satisfy this property.[25] In a striking result, Ashlagi et al. (2017) show that small imbalances on the number of agents in each side, lead to the core shrinking rapidly, and show that a truth-telling equilibria can be supported as an approximate optimal strategy for all agents.

The present paper complements this literature in several ways. First, the main results hold irrespective of the size of the market.[26] Second, a stable matching outcome is sustained as the unique prediction under regret-free behavior when the clearinghouse uses DA.[27,28] Third, the approach in this paper recognizes incentives being provided by the DA that are not provided by other

[25]Strategy-proofness in the large is a weakening of strategy-proofness in two respects: first, it requires truth-telling to be approximately optimal; and second, to do so with respect to all full-support i.i.d distributions of play.

[26]Through simulations, Kadam (2014) shows that the number of participants needed for the results of Kojima and Pathak (2009) to hold can be unreasonably large in practice. Moreover, not all markets that use DA are large. For instance, while the NRMP has participants in the tens of thousands, the number of graduating rabbinical students is in the hundreds.

[27]Fernandez et al. (2020) show that, in the context of incomplete information matching from a Bayesian perspective, even a singleton core state-by-state, is not sufficient to guarantee that only the stable matching is supported in Bayes-Nash equilibria.

[28]In the implementation literature this is referred to as full implementation.

stable mechanisms. Forth, the insights do carry over to many-to-one markets in a straightforward fashion.[29] Fifth, the incentives provided do not depend on the specification of cardinal utilities, nor on details about beliefs or priors, beyond their support.[30,31]

The treatment in this paper maintains the ordinal preference approach tradition in matching and looks for incentives in the *spirit* of the "Wilson doctrine" and robust mechanism (Ledyard, 1977; Bergemann and Morris, 2008).[32] As discussed, regret-free truth-telling depends on the support of beliefs (and not their distribution). The notion only considers ordinal information, since mechanisms in practice elicit ordinal, and not cardinal preferences. The notion further accounts for the structure of information that participants typically have in practice, both ex-ante as well as ex-post. This is in line with the arguments supporting ex-post and posterior equilibria (see section 6).

This paper also contributes to the recent and emerging literature incorporating non-standard preferences and behavioral aspects to realms of mechanism and market design.[33] Specifically within matching, Meisner and von Wangenheim (2019), and Dreyfuss et al. (2019) analyze parents' behavior in the context of school choice through the lenses of loss aversion; Antler and Bachi (2019) look at two-sided search and matching when agents reasoning is coarse; Pan (2019) studies matching with overconfident agents in the context of school choice; Troyan and Morrill (2020) develop a notion of (non-)obviously manipulable and show that several known mechanism, satisfy this property.

REFERENCES

Alcalde, J. and S. Barberà (1994): "Top dominance and the possibility of strategy-proof stable solutions to matching problems," *Economic Theory*, 4, 417–435.

[29]Recall footnote 7.

[30]For example, in the context of school choice and using both survey and placement data, Kapor et al. (2020) show that parents' beliefs about admissions probabilities in a manipulable mechanism systematically differ from rational expectations. They show that failing to account for subjective beliefs would lead to a reversal in the comparative statics regarding the welfare implications of switching from a manipulable mechanism to DA.

[31]Ehlers and Massó (2007, 2015) study incentives in the form of Ordinal Bayesian incentive compatibility. They show that a unique stable matching state-by-state is required for a stable mechanisms to be Bayesian incentive compatible regardless of the cardinal utility representation of the ordinal preferences.

[32]Note that mechanism like the NRMP only elicit ordinal information, in the form of ranked order lists. This is contrast to settings such as auctions where an agent's bid does convey cardinal information.

[33]In the mechanism design context, Lopomo et al. (2009) tackle mechanism design in the context on Knightian uncertainty; De Clippel et al. (2019) studies implementation under level-k reasoning; Börgers and Li (2019) study define and study strategically simple mechanisms, that only depend on agents' first-order beliefs and first-order knowledge of rationality. Specifically using regret-related considerations Renou and Schlag (2010, 2011) introduce the notion of minimax regret equilibrium in the context of complete information, that accommodates multiple priors, and accounts for agents' (strategic) uncertainty about the rationality and conjectures of others; Halpern and Pass (2012) proposes iterated regret minimization.

Antler, Y. and B. Bachi (2019): “Searching forever after,” *Working Paper*.

Ashlagi, I., Y. Kanoria, and J. D. Leshno (2017): “Unbalanced random matching markets: The stark effect of competition,” *Journal of Political Economy*, 125, 69–98.

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

Balinski, M. L. and T. Sönmez (1999): “A tale of two mechanisms: student placement,” *Journal of Economic Theory*, 84, 73–94.

Barberà, S. and B. Dutta (1995): “Protective behavior in matching models,” *Games and Economic Behavior*, 8, 281–296.

Barberà, S. and M. Jackson (1988): “Maximin, leximin, and the protective criterion: characterizations and comparisons,” *Journal of Economic Theory*, 46, 34–44.

Bell, D. E. (1982): “Regret in decision making under uncertainty,” *Operations Research*, 30, 961–981.

Bergemann, D. and S. Morris (2008): “Ex-post implementation,” *Games and Economic Behavior*, 63, 527–566.

Bergemann, D. and K. Schlag (2011): “Robust monopoly pricing,” *Journal of Economic Theory*, 146, 2527–2543.

Bergemann, D. and K. H. Schlag (2008): “Pricing without priors,” *Journal of the European Economic Association*, 6, 560–569.

Bodin, L. and A. Panken (2003): “High tech for a higher authority: The placement of graduating rabbis from hebrew union college—jewish institute of religion,” *Interfaces*, 33, 1–11.

Börgers, T. and J. Li (2019): “Strategically simple mechanisms,” *Econometrica*, 87, 2003–2035.

Bourgeois-Gironde, S. (2010): “Regret and the rationality of choices,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 249–257.

Chen, P., M. Egesdal, M. Pycia, and M. B. Yenmez (2014): “Quantile Stable Mechanisms,” *Working Paper*, 1–22.

Chen, P., M. Egesdal, M. Pycia, and M. B. Yenmez (2016a): “Manipulability of stable mechanisms,” *American Economic Journal: Microeconomics*, 8, 202–14.

Chen, P., M. Egesdal, M. Pycia, and M. B. Yenmez (2016b): “Median stable matchings in two-sided markets,” *Games and Economic Behavior*, 97, 64–69.

Coles, P. and R. Shorrer (2014): “Optimal truncation in matching markets,” *Games and Economic Behavior*, 87, 591–615.

Connolly, T. and D. Butler (2006): “Regret in economic and psychological theories of choice,” *Journal of Behavioral Decision Making*, 19, 139–154.

De Clippel, G., R. Saran, and R. Serrano (2019): “Level-k mechanism design,” *The Review of Economic Studies*, 86, 1207–1227.

Dénes, J. and A. D. Keedwell (1991): *Latin squares: New developments in the theory and applications*, volume 46, Elsevier.

Dreyfuss, B., O. Heffetz, and M. Rabin (2019): “Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms,” *Working paper*.

Dubins, L. E. and D. Freedman (1981): “Machiavelli and the Gale-Shapley Algorithm,” *American Mathematical Monthly*, 88, 485–494.

Echenique, F. and L. Yariv (2011): “An Experimental Study of Decentralized Matching,” *Working Paper*.

Ehlers, L. (2008): “Truncation Strategies in Matching Markets,” *Mathematics of Operations Research*, 33, 327–335.

Ehlers, L. (2010): "Manipulation via capacities revisited," *Games and Economic Behavior*, 69, 302–311.

Ehlers, L. and J. Massó (2007): "Incomplete information and singleton cores in matching markets," *Journal of Economic Theory*, 136, 587–600.

Ehlers, L. and J. Massó (2015): "Matching markets under (in)complete information," *Journal of Economic Theory*, 157, 295–314.

Fernandez, M. A., K. Rudov, and L. Yariv (2020): "Centralized matching with incomplete information," *mimeo*.

Filiz-Ozbay, E. and E. Y. Ozbay (2007): "Auctions with anticipated regret: Theory and experiment," *American Economic Review*, 97, 1407–1418.

Gale, D. and L. S. Shapley (1962): "College admissions and the stability of marriage," *American Mathematical Monthly*, 69, 9–15.

Green, J. R. and J.-J. Laffont (1987): "Posterior implementability in a two-person decision problem," *Econometrica: Journal of the Econometric Society*, 69–94.

Halpern, J. Y. and R. Pass (2012): "Iterated regret minimization: A new solution concept," *Games and Economic Behavior*, 74, 184–207.

Hayashi, T. et al. (2008): "Regret aversion and opportunity dependence," *Journal of economic theory*, 139, 242–268.

Hwang, J. (1978): "Stable permutations in latin squares," *Soochow J. Math*, 4, 63–72.

Hwang, J. (1986): "The algebra of stable marriages," *International journal of computer mathematics*, 20, 227–243.

Immorlica, N. and M. Mahdian (2005): "Marriage, honesty, and stability," in *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, 53–62.

Kadam, S. V. (2014): "On the large market core convergence result for two-sided matching markets," *Working paper*.

Kapor, A. J., C. A. Neilson, and S. D. Zimmerman (2020): "Heterogeneous beliefs and school choice mechanisms," *American Economic Review*, 110, 1274–1315.

Klaus, B. and F. Klijn (2006): "Median stable matching for college admissions," *International Journal of Game Theory*, 34, 1–11.

Kojima, F. and M. Manea (2010): "Axioms for Deferred Acceptance," *Econometrica*, 78, 633–653.

Kojima, F. and P. A. Pathak (2009): "Incentives and stability in large two-sided markets," *American Economic Review*, 99, 608–627.

Ledyard, J. O. (1977): "Incentive compatible behavior in core-selecting organizations," *Econometrica*, 45, 1607–1623.

Lee, S. (2017): "Incentive Compatibility of Large Centralized Matching Markets," *The Review of Economic Studies*, 84, 444–463.

Lopomo, G., L. Rigotti, and C. Shannon (2009): "Uncertainty in mechanism design," *Working paper*.

Manski, C. F. (2007): "Minimax-regret treatment choice with missing outcome data," *Journal of Econometrics*, 139, 105–115.

McVitie, D. G. and L. B. Wilson (1970): "Stable marriage assignment for unequal sets," *BIT Numerical Mathematics*, 10, 295–309.

Meisner, V. and J. von Wangenheim (2019): "School choice and loss aversion," *Working paper*.

Milnor, J. (1954): "Games against nature," in R. Thrall, C. Coombs, and R. Davis, eds., *Decision Processes*, Wiley, New York.

Moulin, H. (1981): "Prudence versus sophistication in voting strategy," *Journal of Economic Theory*, 24, 398–412.

Pan, S. (2019): "The instability of matching with overconfident agents," *Games and Economic Behavior*, 113, 396–415.

Pathak, P. A. and T. Sönmez (2013): "School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation," *American Economic Review*, 103, 80–106.

Renou, L. and K. H. Schlag (2010): "Minimax regret and strategic uncertainty," *Journal of Economic Theory*, 145, 264–286.

Renou, L. and K. H. Schlag (2011): "Implementation in minimax regret equilibrium," *Games and Economic Behavior*, 71, 527–533.

Roth, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7, 617–628.

Roth, A. E. (1985): "The college admissions problem is not equivalent to the marriage problem," *Journal of economic Theory*, 36, 277–288.

Roth, A. E. (1989): "Two sided matching with incomplete information about others preferences," *Games and Economic Behavior*, 1, 191–209.

Roth, A. E. (2002): "The economist as engineer: Game theory, experimentation, and computation as tools for design economics," *Econometrica*, 70, 1341–1378.

Roth, A. E. (2008): "Deferred acceptance algorithms: history, theory, practice, and open questions," *International Journal of Game Theory*, 36, 537–569.

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

Roth, A. E. and U. G. Rothblum (1999): "Truncation strategies in matching markets - in search of advice for participants," *Econometrica*, 67, 21–43.

Roth, A. E. and M. Sotomayor (1990): *Two-sided matching: A study in game theoretic modeling and analysis*, Cambridge University Press.

Sarver, T. (2008): "Anticipating regret: Why fewer options may be better," *Econometrica*, 76, 263–305.

Savage, L. J. (1951): "The theory of statistical decision," *Journal of the American Statistical association*, 46, 55–67.

Sönmez, T. (1997): "Manipulation via capacities in two-sided matching markets," *Journal of Economic theory*, 77, 197–204.

Sönmez, T. (1999): "Can pre-arranged matches be avoided in two-sided matching markets?" *Journal of Economic theory*, 86, 148–156.

Stoye, J. (2009): "Minimax regret," *The New Palgrave Dictionary of Economics*.

Stoye, J. (2011): "Axioms for minimax regret choice correspondences," *Journal of Economic Theory*, 146, 2226–2251.

Teo, C.-P. and J. Sethuraman (1998): "The Geometry of Fractional Stable Matchings and Its Applications," *Mathematics of Operations Research*, 23, 874–891.

Thurber, E. G. (2002): "Concerning the maximum number of stable matchings in the stable marriage problem," *Discrete Mathematics*, 248, 195–219.

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

Wilson, R. (1987): "Game-Theoretic Analysis of Trading Processes," in T. Bewley, ed., *Advances in Economic Theory: Fifth World Congress*, Cambridge: Cambridge University Press, 33–70.

Zeelenberg, M. (1999): "Anticipated regret, expected feedback and behavioral decision making," *Journal of behavioral* *decision making*, 12, 93–106.

Zeelenberg, M. (2018): “Anticipated regret,” *The Psychology of Thinking about the Future*, 276.

Zeelenberg, M. and R. Pieters (2007): “A theory of regret regulation 1.0,” *Journal of Consumer psychology*, 17, 3–18.

APPENDIX A: PROOFS

A.1. *One-to-one matching: Deferred acceptance*

Theorem 1. The deferred acceptance mechanism is regret-free truth-telling; i.e. if $\phi \in \{\phi^M, \phi^W\}$ then $\forall(M, W, \succ)$, $\forall i \in N = M \cup W$, $\succ_i$ is regret-free.

The result is established through four claims: Claim 1 shows that truth-telling is regret-free for the proposing side, which is a consequence of dominant strategy incentive compatibility (Dubins and Freedman, 1981; Roth, 1982). Consequently the rest of the proof focuses only on the receiving side. Claim 2, 3 and 4 show that there does not exist a report through which i regrets truth-telling. Each claim deals with reports that differ from the truth-telling in a specific manner. Claim 2 shows that changing the order of alternatives that are preferred to the observed match is inconsequential; it would not change the resulting matching. Claim 3 and 4 show that any report that differs from truth-telling in a consequential manner is never a "safe deviation" compared truth-telling, in the sense that whenever it may result in a more preferable match it may also result in a less preferable as well. Claim 3 deals with those deviations where an alternative which is less preferred to the observed match by the true preference profile is reported as preferred to the said match. Claim 4 deals with deviations where the relative order between two alternatives that are less preferred to the observed match is reversed. All reports that differ from the truth in a consequential manner have to fit into the conditions of (at least) one of these claims.

PROOF: Let $(M, W, \succ)$ be an arbitrary matching market, and let the mechanism be the M-DA. Given an arbitrary agent i and an arbitrary matching μ, define

$$UC^{\succ_i}_{\phi(\succ)(i)} = \{j \in J : j \succ_i \phi(\succ)(i)\},$$

$$LC^{\succ_i}_{\phi(\succ)(i)} = \{j \in J : \phi(\succ)(i) \succ_i j\}.$$

That is, $UC^{\succ_i}_{\phi(\succ)(i)}$ denotes the (strict) upper contour set with respect to $\phi(\succ)(i)$ under $\succ_i$, that is the set of partners that player i considers strictly preferable (according to his/her true preference) to the partner under $\phi(\succ)$; analogously interpret $LC^{\succ_i}_{\phi(\succ)(i)}$.

Claim 1. *Truth-telling is regret-free for every agent in the proposing side.*

Claim 1 follows directly from strategy-proofness for men of the M-DA. Consequently we can focus on the receiving side (so $i \in W$ from now on). We must show that for an arbitrary agent on the receiving side and for an arbitrary matching that may result from her reporting her true preference, there is no alternative report through which that agent regrets truth-telling. We start by showing (Claim 2) that a report that differs from truth only in the way that it orders elements that are preferred to the observed matching cannot yield a better matching for the agent.

Claim 2. *Let $\mu \in \{\mu' : (\exists \succ_{-i} \in \mathcal{P}_{-i}) [\phi(\succ_i, \succ_{-i}) = \mu']\}$ be an arbitrary matching that results from i truth-telling. For any report $\succ'_i$ such that*

1. $UC^{\succ_i}_{\phi(\succ)(i)} = UC^{\succ'_i}_{\phi(\succ)(i)}$*; and,*
2. $\left(\forall a, b \in LC^{\succ_i}_{\phi(\succ)(i)} \cup \{\phi(\succ)(i)\}\right) [a \succ'_i b \Leftrightarrow a \succ_i b]$

it holds that $\phi(\succ_i, \succ_{-i}) = \phi(\succ'_i, \succ_{-i})$

The agent does not reject an offer under $\succ'_i$ that was accepted under $\succ_i$, she cannot affect the set of offers that are made to her, and consequently cannot affect the outcome favorably.[34]

Claim 3. *Suppose $\exists(\succ'_i, \hat{\succ}_{-i})$ such that*

1. $\exists \hat{\succ}_{-i} \in \mathcal{I}(\mu; \succ_i, \phi^M)$
2. $\succ'_i$: $\exists \tilde{j} \in LC^{\succ_i}_{\phi(\succ)(i)}$ *and* $\tilde{j} \in UC^{\succ'_i}_{\phi(\succ)(i)}$
3. $\phi(\succ'_i, \hat{\succ}_{-i}) \succ_i \phi(\succ_i, \hat{\succ}_{-i}) = \mu$

then $\exists \tilde{\succ}_{-i} \in \mathcal{I}(\mu; \succ_i, \phi^M)$ such that $\phi(\succ_i, \tilde{\succ}_{-i}) = \mu \succ_i \phi(\succ'_i, \tilde{\succ}_{-i})$.

Case 1. Truncation ($\tilde{j} = i$) It is enough to consider the following preference profile to see that the truncation can leave the agent worse off than telling the truth

$$\tilde{\succ}_k : \phi(\succ)(k), k, \ldots \;\; \forall k \neq i$$

Notice that in profile $\tilde{\succ}_k$ everyone other than i considers their assigned partner as their unique acceptable partner, this profile belongs to i's inference set. However under $\succ'_i$ that match is no longer acceptable for i. Since ϕ is stable with respect to the reported preferences (in particular individually rational) it leaves i unmatched under $\succ'_i$, which is a strictly worse off outcome from the point of view of i's the true preferences.

[34]Agent i did not receive any offer from a member of $UC^{\succ_i}_{\phi(\succ)(i)}$ in $\phi(\succ)$ by construction of the DA. The first round offers made by men are unaffected by i's report. Since agent i's offers are the same (and belong to $LC^{\succ_i}_{\phi(\succ)(i)} \cup \{\phi(\succ)(i)\}$) and her preferences over those alternatives did not change, the set of active agents in the next round is the same. The set of active agents in the second round is the same, again she faces choices on $LC^{\succ_i}_{\phi(\succ)(i)} \cup \{\phi(\succ)(i)\}$, and makes the same choice. The argument follows by induction.

Case 2. A non-truncation ($\tilde{j} \neq i$). Similar to the truncation case it is enough to consider the following profile, which is in i's inference set

$$\tilde{\succ}_{\tilde{j}} : i, \phi(\succ)(\tilde{j}), \tilde{j}, \ldots$$
$$\tilde{\succ}_{k} : \phi(\succ)(k), k, \ldots \ \forall k \neq \{\tilde{j}, i\}$$

Under the alternative report, everyone except i has the same preference profile as before, so first round proposals are the same. However, under $\succ_i'$ agent $\tilde{j}$ is declared as preferred to $\phi(\succ)(i)$, which means that agent i accepts $\tilde{j}$'s offer. Since there are no more active players the algorithm stops and matches agent i to $\tilde{j}$ which is a strictly worse outcome under i's true preference profile, that is $\phi(\succ_i, \tilde{\succ}_{-i})(i) = \phi(\succ)(i) \succ_i \tilde{j} = \phi(\succ_i', \tilde{\succ}_{-i})(i)$. Consequently, i cannot regret truth-telling ($\succ_i$) through a deviation ($\succ_i'$) consistent with the claim, at the observed matching (μ) in DA.

Claim 4. *Suppose* $\exists(\succ_i', \hat{\succ}_{-i})$ *such that*

1. $\phi(\succ_i, \hat{\succ}_{-i}) = \mu$
2. $\succ_i'$: $\exists u, v \in LC^{\succ_i}_{\phi(\succ)(i)}$ *such that* $u \succ_i v$ *and* $v \succ_i' u$
3. $\phi(\succ_i', \hat{\succ}_{-i}) \succ_i \phi(\succ_i, \hat{\succ}_{-i})$

then $\exists \tilde{\succ}_{-i}$ *such that* $\phi(\succ_i, \tilde{\succ}_{-i}) = \mu$ *and* $\phi(\succ_i, \tilde{\succ}_{-i}) \succ_i \phi(\succ_i', \tilde{\succ}_{-i})$.

The structure of the argument follows the same lines as the previous claim. For any report $\succ_i'$ that satisfies the conditions of the claim, the following preference profile ($\tilde{\succ}_{-i}$) is in i's inference set:

$$\tilde{\succ}_{v} : i, \phi(\succ)(v), \ldots$$
$$\tilde{\succ}_{u} : i, \phi(\succ)(u), \ldots$$
$$\tilde{\succ}_{\phi(\succ)(v)} : v, \phi(\succ)(i), \ldots$$
$$\tilde{\succ}_{\phi(\succ)(i)} : \phi(\succ)(v), i, \ldots$$
$$\tilde{\succ}_{k} : \phi(\succ)(k), k, \ldots \ \forall k \neq \{i, u, v, \phi(\succ)(i), \phi(\succ)(v)\}$$

If i reports $\succ_i'$ instead, the resulting allocation matches i to v, which is a worse outcome for agent i according to her true preference profile, that is $\phi(\succ_i, \tilde{\succ}_{-i})(i) \succ_i v = \phi(\succ_i', \tilde{\succ}_{-i})(i)$. Consequently, i cannot regret truth-telling ($\succ_i$) through a deviation ($\succ_i'$) consistent with the claim, at the observed matching (μ) given that the clearinghouse uses M-DA.

The analysis shows that there is no misrepresenting report through which agent i can regret truth-telling, given μ. Since this was done for an arbitrary μ, it holds for all such matchings that may result from truth-telling. Consequently, there is no μ at which i regrets truth-telling which means truth-telling is regret-free for agent i. Since this conclusion holds for an arbitrary agent in either side of the market (proposing or receiving), putting together the previous claims the theorem is established. ■

Let $T_i = \{\succ_i'' \in \mathcal{P}_i : (\forall a, b \in A_i(\succ_i) \cup \{i\}) \, [A_i(\succ_i'') = A_i(\succ_i) \text{ and } a \succ_i'' b \Leftrightarrow a \succ_i b]\}$ denote the set of all preferences for i that only differ from the true one in how they rank unacceptable choices between themselves. Note that, by construction, the DA does not take into account the relative ranking among alternatives in the unacceptable set of any agent; these reports differ from the truth only in an inessential manner. Thus, the matching generated by truth-telling and by any report in T_i is the same; corollary 3 follows.

Corollary 3. *Any report that differs from the truth only in how it ranks the elements of the unacceptable set among themselves ($\succ_i' \in T_i$) is also regret-free.*

Proposition 1. Truth is the essentially unique regret-free report in the DA mechanism. Moreover, i regrets any other report *through truth*.

PROOF: Any meaningful misrepresentation of an agent i's preferences $\succ_i' \in \mathcal{P}_i \setminus T_i$ must belong to one of the following cases:

1. $\succ_i' \in \mathcal{P}_i$ such that $\exists k \in A(\succ_i)$ and $k \in U(\succ_i')$.
2. $\succ_i' \in \mathcal{P}_i$ such that $\exists j \in U_i(\succ_i)$ and $j \in A_i(\succ_i')$.
3. $\succ_i'$ involves a permutation among the acceptable set.

To see that any misreport of the form of 1 can lead to regret, it is enough to consider a resulting matching $\phi(\succ_i', \cdot)$ where i remains unmatched. In any such circumstance, the following preference profile is in i's inference set:

$$
\begin{aligned}
i :& \cdots \succ' i \succ' k \\
k :& i \succ'' \phi(\succ', \succ_{-i})(k) \underset{(\sim)}{\succ''} k \\
j :& \phi(\succ', \succ_{-i})(j) \underset{(\sim)}{\succ''} j (\succ'' i)
\end{aligned}
$$

Individual rationality of the mechanism guarantees that i could not have been worse off by truth-telling, for any set of reports in i's inference set. On the other hand, if the reported preferences are indeed $\succ''_{-i}$, then stability (not IR) implies that $\phi(\succ_i, \succ''_{-i})(i) = k \succ_i i$, since otherwise (i, k) would be a blocking pair. To see that no misreport of the form of 2 can be regret-free, suppose $\succ'_i$ is such that $\exists j \in U_i(\succ_i)$ and $j \in A_i(\succ'_i)$. Then, $\exists \mu \in \mathcal{M}|_{\succ'_i} : \mu(i) = j$. Since $\phi(\cdot)$ is individually rational, $\phi(\succ_i, \tilde{\succ}_{-i}) \succ_i j$.

For any misrepresentation that fits the conditions of 3, the following algorithm finds a matching μ at which i regrets reporting $\succ'_i$ through $\succ_i$ for an arbitrary i in the receiving side. An analogous argument works for an agent in the proposing side.[35]

Let $|J|$ denote the cardinality of the agents on the proposing side that are acceptable to i with respect to her true preference profile. Relabel agents such that their index reflects their ranking according to $\succ_i$; i.e. j_1 is the $\succ_i$-maximal element (agent) on $A_{i,1}(\succ_i) = A_i(\succ_i)$, j_2 the $\succ_i$-maximal element on $A_{i,2}(\succ_i) = A_{i,1}(\succ_i) \setminus \{j_1\}$, etc.

Step 1. If the index of the $\succ'_i$-maximal element on $A_{i,|J|-1}(\succ_i)$ is smaller than the index of the $\succ'_i$-maximal element on $A_{i,|J|}(\succ_i)$, then go to step 2.

Otherwise, set $\mu \in \mathcal{M}|_{\succ'_i} : \mu(i) = \{\succ'_i$-maximal element on $A_{i,|J|-1}(\succ_i)\}$ and $\mu(k) = k \; \forall k \neq \{i, \mu(i)\}$.[36] *Break.*

Step $k \in \{2, \ldots, |J| - 1\}$. If the index of the $\succ'_i$-maximal element on $A_{i,|J|-k}(\succ_i)$ is smaller than the index of the $\succ'_i$-maximal element on $A_{i,|J|-(k-1)}(\succ_i)$, then go to step $k + 1$.

Otherwise, set $\mu \in \mathcal{M}|_{\succ'_i} : \mu(i) = \{\succ'_i$-maximal element on $A_{i,|J|-k}(\succ_i)\}$ and $\mu(k) = k \; \forall k \neq \{i, \mu(i)\}$. *Break.*

Given that $\succ'_i$ is a permutation of $\succ_i$ on $A_i(\succ_i)$ it cannot be the case that $\forall j \in A_i(\succ_i) \; j_k \succ'_i j_l$ whenever $k < l$. Therefore the algorithm necessarily sets a μ. Next we explain why at such μ i regrets $\succ'_i$ through $\succ_i$.

First, consider a case where the algorithm stops after step 1, setting $\mu(i) = x = \{\succ'_i$-maximal element on $A_{i,|J|-1}(\succ_i)\}$. By construction of DA, i can only have received offers from x and $y = \{\succ'_i$-maximal element on $A_{i,|J|}(\succ_i)\}$, necessarily so from x since it is matched to him under the observed matching. The preference profiles $\tilde{\succ}_{-i} \in \mathcal{M}|_{\succ'_i}$ are divided into those cases in which i received an offer from y and those in which it did not; there always exist preference profiles that

[35]In the case of the receiving side described in the text the algorithm looks for the first switch in the preference relation from least- to most-preferred acceptable alternative. In the case of the proposing side the search is done from most- to least-preferred acceptable partner.

[36]The essential part of the matching found by the algorithm is to whom agent i is matched, the choice of leaving everyone else unmatched is arbitrary and not unique.

satisfy each condition. If she did not, then she only observed an offer from x and consequently, $\phi(\succ_i, \tilde{\succ}_{-i}) = \phi(\succ'_i, \tilde{\succ}_{-i}) = x$ since i does not reject or accept any offer differently under $\succ'_i$ than under $\succ_i$. On the other hand, if i received an offer from y it means at some point she decided between y and x in favor of x. However, since the algorithm stopped to produce μ it means that $y \succ_i x$, consequently $\phi(\succ_i, \tilde{\succ}_{-i}) \succ_i \phi(\succ'_i, \tilde{\succ}_{-i})$.

The same logic extends to the case where the algorithm stops at a step k: i cannot have received offers from any $z \succ'_i \mu(i)$. For any $s, t \in J : \mu(i) \succ'_i s$ and $\mu(i) \succ'_i t$ it is the case that $s \succ'_i t \iff s \succ_i t$. That is, the binary relation between the options that can potentially have made an offer to i is the same under $\succ'_i$ than under $\succ_i$, which means that any offer that did not involve $\mu(i)$ is accepted or rejected in the same manner under both $\succ'_i$ and $\succ_i$. The only cases in which they differ are in those where $\mu(i)$ was chosen over some $s \in J : \mu(i) \succ'_i s$ and $s \succ_i \mu(i)$. Consequently $\phi(\succ_i, \tilde{\succ}_{-i}) \succ_i \phi(\succ'_i, \tilde{\succ}_{-i})$.

■

A.2. *One-to-one matching: Quantile- and interior-stable mechanisms*

Theorem 2. In any non-extreme q-quantile-stable matching mechanism is not regret-free truth-telling; that is, $\forall q \in (0,1)$ there is a market $(M, W, \succ)$ such that $\exists i \in N = M \cup W$ and a $\mu = \phi^q(\succ)$ where i regrets $\succ_i$ at μ through some $\succ'_i \in \mathcal{P}_i$

Proof: Fix $q \in (0,1)$. Without loss of generality, consider $i = m_1$. Define the function $\succ'_{m_1}: \mathcal{P}_i \to \mathcal{P}_i$ which will serve as the misreport that leads m_1 to regret truth-telling; as anticipated in section 5.1.2, this misrepresentation takes the form of a soft-truncation.

$$\succ'_{m_1}: \begin{cases} w \succ'_{m_1} w' \iff w \succ_{m_1} w' & \forall (w, w') \in W^2 \\ w \succeq'_{m_1} m_1 \iff w \succeq_{m_1} \phi^q(\succ_{m_1}, \succ_{-m_1})(m_1) & \forall w \in W \end{cases}$$

Although $\succ'_{m_1}$ is a function of $\succ$ we suppress it from the notation. Also notice that: $A(\succ'_{m_1}) \subseteq A(\succ_{m_1})$.

Lemma 1. $S(\succ'_{m_1}, \succ_{-m_1}) = \{\mu \in S(\succ_{m_1}, \succ_{-m_1}) : \mu \succeq_{m_1} \phi^q(\succ)\} \quad \forall \succ \in \mathcal{P}$

Proof of Lemma: The lemma follows from the following two claims:

Claim 5. $S(\succ'_{m_1}, \succ_{-m_1}) \subseteq S(\succ_{m_1}, \succ_{-m_1})$

Proof of claim: Suppose not, then there exists $\mu \in S(\succ'_{m_1}, \succ_{-m_1})$ and $\mu \notin S(\succ_{m_1}, \succ_{-m_1})$. It must be the case that μ is either blocked by a pair or by an individual under $\succ$. If μ is not individually rational under $(\succ_{m_1}, \succ_{-m_1})$ then it is not individually rational under $(\succ'_{m_1}, \succ_{-m_1})$. This is straightfoward since $\succ'_{-m_1} = \succ_{-m_1}$, and for m_1 it is a consequence of $A(\succ'_{m_1}) \subseteq A(\succ_{m_1})$ by construction of $\succ'_{m_1}$. Suppose it is blocked by a pair $(m_j, w) : m_j \neq m_1$, then since $\succ'_{-m_1} = \succ_{-m_1}$, (m_j, w) also blocks μ under preference profile $(\succ'_{m_1}, \succ_{-m_1})$. Then it has to be the case that the blocking pair involves m_1. If $\mu(m_1) = w'$, then it must hold that $w' \succ_{m_1} w$ and $w \succ'_{m_1} w'$, but it contradicts the construction of $\succ'_{m_1}$ since it does not permute binary relations that do not involve alternative m_1, which denotes remaining single. Lastly, consider the case $\mu(m_1) = m_1$ then it means m_1 must be single in every stable matching under $(\succ'_{m_1}, \succ_{-m_1})$. If μ is blocked by a pair (m_1, w), then it must be that $\forall \mu'' \in S(\succ_{m_1}, \succ_{-m_1}) : \mu''(m_1) \in W$.[37] Next we note that $A(\succ'_{m_1}) \supseteq \{\mu_M(\succ_{m_1}, \succ_{-m_1})\}$ since, by construction of $\succ'_{m_1}$, all women w that satisfy $w \succeq_{m_1} \phi^q(\succ_{m_1}, \succ_{-m_1})$ are listed as acceptable in the soft-truncation; $w \in A(\succ'_{m_1})$. Again, $\mu_M(\succ_{m_1}, \succ_{-m_1}) = m_1$ contradicts the hypothesis (for the same argument as in the previous footnote).On the other hand, $\mu_M(\succ_{m_1}, \succ_{-m_1}) \neq m_1$ implies $\mu_M(\succ'_{m_1}, \succ_{-m_1}) \neq m_1$, which holds since the M-DA follows the same steps; if at some point $\mu_M(\succ'_{m_1}, \succ_{-m_1})(m_1)$ rejected him at any step, then it should have rejected him in $\mu_M(\succ_{m_1}, \succ_{-m_1})(m_1)$. ■

By the construction of $\succ'_{m_1}$ if follows that $S(\succ'_{m_1}, \succ_{-m_1}) \subseteq \{\mu \succeq_{m_1} \phi^q(\succ)\}$.

Claim 6. $\mu \in S(\succ_{m_1}, \succ_{-m_1})$ *and* $\mu \succeq_{m_1} \phi(\succ_{m_1}, \succ_{-m_1}) \implies \mu \in S(\succ'_{m_1}, \succ_{-m_1})$.

Proof of claim: Suppose not, so $\mu \notin S(\succ'_{m_1})$, then either it is not individual rational or it is blocked by a pair. If $j \succ_j \mu(j)$ for $j \neq \{m_1\}$ then $\mu \notin S(\succ_{m_1})$, on the other hand if $m_1 \succ'_{m_1} \mu(i)$ then $\mu \not\succeq_{m_1} \phi^q(\succ) \succeq_{m_1} m_1$. If it is blocked by a pair (m_j, w) $j \neq \{1\}$ then they are also a blocking pair to μ under $(\succ_{m_1}, \succ_{-m_1})$. Lastly consider the blocking pairs (m_1, w). If $\mu(m_1) \neq m_1$, since by construction $\succ'_{m_1}$ does not change the binary relations not involving the alternative of being single

[37]Suppose not, so that $\mu''(m_1) = m_1 \ \forall \mu'' \in S(\succ_{m_1}, \succ_{-m_1})$; since (m_1, w) is the blocking pair, it follows that $w \succ_{m_1} m_1$ and $m_1 \succ_w \mu(w)$. Now let

$$\tilde{\mu} = \begin{cases} \mu(j) & \forall j \notin \{m_1, w, \mu(w)\} \\ w & \text{for } m_1 \\ m_1 & \text{for } w \\ \mu(w) & \text{for } \mu(w) \end{cases}$$

$\tilde{\mu}$ is an individually rational matching. Either, $\tilde{\mu}$ is stable, in which case $\tilde{\mu}(m_1) = w \in W$ which contradicts $\nexists \mu \in S(\succ_{m_1}, \succ_{-m_1}) : \mu(m_1) \neq m_1$, or $\tilde{\mu}$ is unstable. If the latter holds, still it must be individually rational (since it was stable for $(\succ'_{m_1}, \succ_{-m_1})$), then by the strong stability property (Roth and Sotomayor, 1990, Theorem 3.4, p. 56) there exists $\bar{\mu} \in S(\succ_{m_1}, \succ_{-m_1}) : \bar{\mu} \succeq_{m_1} \tilde{\mu}$ and $\bar{\mu} \succeq_w \tilde{\mu}$. Since $\tilde{\mu}(m_1) = w$ then $\bar{\mu}(m_1) \in W$ which is a contradiction.

m_1, they would also be a blocking pair to (m_1, w). Lastly, if $\mu(m_1) = m_1$, $w \succ'_{m_1} m_1$ and since $A(\succ'_{m_1}) \subseteq A(\succ_{m_1})$, $w \succ_{m_1} m_1$ which contradicts $\mu \in S(\succ_{m_1}, \succ_{-m_1})$. ■

■

Remark 1. For any $q \in (0,1)$, $\phi^q(\succ'_{m_1}, \succ_{-m_1}) \succeq_{m_1} \phi^W(\succ'_{m_1}, \succ_{-m_1}) = \phi^q(\succ_{m_1}, \succ_{-m_1})$.

Theorem (Chen et al., 2014, Theorem 4). For any $q, q' \in (0,1] : q \neq q'$ there exists a matching market such that ϕ^q is different than $\phi^{q'}$.

The key for this result is to find a market with enough stable matchings such that the non-extreme quantile mechanism and the extreme one result in different matches; i.e. $k(q' - q) > 1$ where $k = |S(\succ)|$. Note that a priori we need a little more, since it could be the case that the matches are different but m_1 is matched to the same partner in both. Putting together the remark, the theorem, and taking into account the construction of $\succ'_{m_1}$ we get the following corollary,

Corollary 4. *Let $q \in (0,1)$, and define $k^\star(q) = \min\{k \in \mathbb{N} : k(1-q) \geq 1\}$. If $\exists(\succ'_{m_1}, \succ_{-m_1})$ such that (1) $|S(\succ'_{m_1}, \succ_{-m_1})| \geq k^\star(q)$; and, (2) $\mu(m_1) \neq \mu'(m_1)$ for all $\mu, \mu' \in S(\succ'_{m_1}, \succ_{-m_1})$ then $\phi^q(\succ'_{m_1}, \succ_{-m_1}) \succ_{m_1} \phi^W(\succ'_{m_1}, \succ_{-m_1}) = \phi^q(\succ_{m_1}, \succ_{-m_1})$.*

Let $\hat{k} := \inf\{k \in \mathbb{N} : \lceil kq \rceil \geq k^\star(q)\}$, and consider the following preferences:

$$
\begin{array}{ll}
\succ^\star_{m_1}: w_1, w_2, \ldots, w_{\hat{k}-1}, w_{\hat{k}} & \succ^\star_{w_{\hat{k}}}: m_1, m_2, \ldots, m_{\hat{k}-1}, m_{\hat{k}} \\
\succ^\star_{m_2}: w_2, w_3, \ldots, w_{\hat{k}}, w_1 & \succ^\star_{w_{\hat{k}-1}}: m_{\hat{k}}, m_1, \ldots, m_{\hat{k}-2}, m_{\hat{k}-1} \\
\vdots & \vdots \\
\succ^\star_{m_{\hat{k}}}: w_{\hat{k}}, w_1, \ldots, w_{\hat{k}-2}, w_{\hat{k}-1} & \succ^\star_{w_1}: m_2, m_3, \ldots, m_{\hat{k}}, m_1
\end{array}
\tag{$\star$}
$$

This is a fairly standard way of generating a matching with $|S(\succ_{m_1}, \succ_{-m_1})| = \hat{k}$ (see Thurber (2002) and Chen et al. (2014)) namely, to construct preferences such that they form a Latin square marriage of order $\hat{k}$.[38] But moreover, each individual gets a different partner in each stable matching. As a consequence of Lemma 1, if $|S(\succ_{m_1}, \succ_{-m_1})| = \hat{k}(q)$ and $\{\hat{k} \in \mathbb{N} : \lceil \hat{k}q \rceil \geq k^\star(q)\}$ it follows

[38] Dénes and Keedwell (1991): A Latin square of order n is an $n \times n$ matrix L whose entries are taken from a set S of n symbols and which has the property that every symbol from S occurs exactly once in each row and exactly once in each column.

that $|S(\succ'_{m_1}, \succ_{-m_1})| \geq k^\star(q)$, then corollary 4 applies and we get that

$$\phi^q(\succ'_{m_1}, \succ_{-m_1}) \succ_{m_1} \phi^q(\succ_{m_1}, \succ_{-m_1})$$

which contradicts truth being regret-free.

Consequently, for any non-extreme ($q \in (0,1)$) $q-$quantile stable matching mechanism, we can find a market $(M, W, \succ)$ where an agent $i \in N$ regrets truth $\succ_i$ through some other report $\succ'_i$.[39,40]
■

Theorem 3. Let ϕ be an interior-stable mechanism, then ϕ is not regret-free truth-telling.

PROOF: Consider a market with $|M| = |W| = 4$, and Latin square preferences $(\star)$.[41] The instance presents four stable matchings, $S(\succ^*) = \{\mu_1 = \mu^M, \mu_2, \mu_3, \mu^W = \mu_4\}$. In each stable matching every agent gets a different stable partner. Man m_1's preferences are such that

$$\mu^M \succ^*_{m_1} \mu_2 \succ^*_{m_1} \mu_3 \succ^*_{m_1} \mu^W.$$

By assumption the mechanism ϕ is interior-stable, so $\phi(\succ^*) \notin \{\mu^W, \mu^M\} \implies \phi(\succ^*) \in \{\mu_2, \mu_3\}$. Suppose, wlog, that $\phi(\succ^*) = \mu_3$.

Consider the case where m_1's preferences are according to $\succ^*$, he reports truthfully and observes matching μ_3. Necessarily, $\succ^*_{-m_1}$ is in m_1's inference set. Suppose he considers the soft-truncation

$$\succ'_{m_1}: w_1, w_2, w_3, m_1, w_4$$

By virtue of the lemma 1,

$$S(\succ'_{m_1}, \succ^*_{-m_1}) = \{\mu_1, \mu_2, \mu_3\}.$$

Corollary 2 implies that m_1 would not have been strictly worse off by providing the soft-truncation report $\succ'_{m_1}$,

$$\phi^W(\succ'_{m_1}, \succ^*_{-m_1}) = \mu_3 = \phi(\succ^*).$$

[39] The theorem holds for every $(M, W, \succ) : M \geq M^\star(q) = \hat{k}(q)$ and $W \geq W^\star(q) = \hat{k}(q)$ the reason being that it will be a Latin rectangle which can be completed into a Latin square, this is a consequence of Hall's theorem.

[40] This is a maximal domain result; for any q it gives us an instance where someone regrets and tells us that for any instance greater than that it also will; however, this does not mean that it is the smallest instance at which an agent would regret truth in the $q-$quantile mechanism

[41] The market need not be balanced, its size can be arbitrarily large as long as it has four agents on each side. It can also be thought to be embedded in a larger market.

We still need to argue that there is an instance in the inference set where the soft-truncation would have yielded a strictly better outcome. Since the mechanism is interior-stable, it follows that

$$\phi(\succ'_{m_1}, \succ^*_{-m_1}) = \mu_2 \ \succ_{m_1} \ \mu_3.$$

Hence, if the stable mechanism selects matching μ_3 when the preferences are $\succ^*$, then any man (in particular m_1), regrets truth-telling in ϕ through a soft-truncation.

Lastly, note that the argument is wlog. If $\phi(P^*) = \mu_2$ the same argument would hold for any woman.

The argument holds beyond the instance of $\succ^*$. It makes use of the fact that some agent can potentially have (at least) four different stable partners in consecutive stable matchings. This matters because (i) it means that the soft-truncation is actually binding (i.e. cutting out some stable partner); (ii) it has enough stable matchings to select from. ■

A.3. *Many-to-one matching: capacity misrepresentation*

Theorem 5. If ϕ is the hospital-proposing deferred-acceptance mechanism, a hospital can regret reporting its true capacity. On the other hand, if ϕ is the doctor-proposing deferred acceptance mechanism, then ϕ is regret-free truth-telling for doctors, and reporting the true capacities is regret-free for hospitals.

The following example adapted from Sönmez (1997) shows that, when the clearinghouse uses H-DA, reporting capacities truthfully is not regret-free for hospitals.

The market presents two hospitals $\{1, 2\}$ with two vacancies each, and three doctors $\{A, B, C\}$. Hospital 2's responsive preferences are:

$$\succ_2 : \{B, C\}, \{A, C\}, C, \{A, B\}, B, A.$$

Consider the case where hospital 2 reports its capacity truthfully, and observes the matching

$$\mu_1 = \begin{pmatrix} 1 & 2 \\ \{B, C\} & A \end{pmatrix}.$$

Notice that in matching μ_1 hospital 2 fills only one vacancy. Had hospital 2 reported a capacity equal to one (the number of vacancies filled under μ_1, it could have done at least as well for any preferences and capacities in its inference set, and strictly better for some of them, thus making hospital 2 regret revealing its capacity truthfully.

To see that hospital 2 could not have done worse by reporting only one vacancy, recall that we can think of a multi-unit hospital, as two copies with unit capacity and with the same preferences over individuals. Then, subreporting capacity is equivalent to one of these copies increasing the ranking of the option of remaining unmatched; which is a monotonic transformation of the original preferences. By Kojima and Manea (2010), the H-DA is weak Maskin monotonic in the auxiliary market with copies, and the matched copy gets a weakly better assignment.

On the other hand, given the profile $\hat{\succ}$ below, by reporting only one vacancy, the H-DA would have matched hospital 2 to doctor C which they strictly prefer to doctor A.

$$\hat{\succ}_1 : \quad \{A,B\}, \{A,C\}, A, \{B,C\}, B, C \qquad\qquad \begin{array}{ll} \hat{\succ}_A : & 2,1 \\ \hat{\succ}_B : & 1,2 \\ \hat{\succ}_C : & 1,2 \end{array}$$

$$\phi^H(\succ_2, \hat{\succ}_{-2}, q_1 = 2, q_2 = 2) = \mu = \begin{pmatrix} 1 & 2 \\ \{B,C\} & A \end{pmatrix},$$

$$\phi^H(\succ_2, \hat{\succ}_{-2}, q_1 = 2, q_2 = 1) = \begin{pmatrix} 1 & 2 \\ \{A,B\} & C \end{pmatrix}.$$

Therefore, H-DA is not regret-free truth-telling for hospitals when reporting capacities.

The second half of Theorem 5 follows from a result by Ehlers (2010) that states that the D-DA is non-manipulable-through-capacities by those hospitals whose capacity was not filled. Hence, the only hospitals that could potentially manipulate are those that filled their capacity. In order to regret truth-telling they would have to subreport capacity, but they cannot guarantee that they will fill their positions only with weakly and strictly preferable candidates.