CLEARINGHOUSES FOR TWO-SIDED MATCHING: AN EXPERIMENTAL STUDY

FEDERICO ECHENIQUE, ALISTAIR J. WILSON, AND LEEAT YARIV

Abstract. We experimentally study the Gale and Shapley (1962) mechanism, utilized in a wide set of applications, most prominently the National Resident Matching Program (NRMP). Several insights come out of our analysis: First, only 48% of our observed outcomes are stable, and among those a large majority culminates at the receiver-optimal stable matching. Second, receivers rarely truncate their true preferences; it is the proposers who do not make offers in order of their preference, frequently skipping potential partners. Third, market characteristics affect behavior: both the cardinal representation and core size influence whether laboratory outcomes are stable. We conclude by using our controlled results and a behavioral model to shed light on a number of stylized facts we derive from new NRMP survey and outcome data, and to explain the small cores previously documented for the NRMP.

1. Introduction

Many two-sided matching markets function through centralized clearinghouses: medical residents to hospitals, children to schools, commissioned officers to military posts, college students to dorms, etc. All use highly structured procedures to generate matches. In principle, clearinghouses have the advantage that they can be designed to implement desirable outcomes at the market level. In particular, many extant clearinghouses aim at implementing *stable* outcomes.[1]

Among centralized clearinghouses, the best known is the National Resident Matching Program (NRMP), which matches physicians and residency programs in the United States. The algorithm used by the NRMP to match the participants is a form of what is commonly called *deferred acceptance* (DA, henceforth), first described in the literature in Gale and Shapley (1962). All participants, on both sides of the match, provide rankings of the other

Date: August, 2015.
F. Echenique: Division of the Humanities and Social Sciences, California Institute of Technology. A. Wilson: Department of Economics, University of Pittsburgh, alistair@pitt.edu. L. Yariv: Division of the Humanities and Social Sciences, California Institute of Technology
We are thankful to the following for their help and advice: Caterina Calsamiglia, Clayton Featherstone, Guillaume Fréchette, Andy Schotter, Emanuel Vespa, Walter Yuan, the Co-editor Karl Schmedders, and three anonymous referees. We gratefully acknowledge financial support from the National Science Foundation (grant SES 0963583), the Gordon and Betty Moore Foundation (grant 1158), and the Lee Center at Caltech.
[1]Stable matchings are characterized by two conditions: i) no agents prefer to remain by themselves over their allocated match; and ii) no two agents prefer to match to one another over their allotted partners.

side. DA functions by assigning market sides to the roles of proposers (physicians in the NRMP) and receivers (hospitals/programs), and then automating a sequence of proposals and conditional acceptances by receivers according to the submitted rankings until a final outcome is reached, which is then instituted. The DA process has the property that if both sides rank the other truthfully, the resulting outcome will be stable. Moreover, for one side of the market, the proposers, submitting a true ranking to the algorithm is incentive compatible. For the other side, the receivers, straightforward ranking is not in general incentive compatible.

In contrast with the theoretical predictions, data from the NRMP's physician survey suggests that physicians do not rank hospitals or programs truthfully. Many participants state that concerns over the likelihood of matching within the algorithm lead them to modify their ranking, while a smaller fraction explicitly state that they rank programs solely based on the likelihood of matching. The published data on NRMP outcomes reveals that close to a half of all matches are of physicians to their top-ranked program. If physicians rank programs truthfully, this would imply physicians' most-preferred programs have strong negative correlations—in opposition to classic assortative models where participants have common agreement on programs' desirability.

Our paper is an experimental investigation of behavior in a dynamic variant of the DA mechanism. We study the effects of an array of market features—including the number of stable matchings, the cardinal representation of preferences in the market, the fragility of the stable matchings to simple manipulations, etc.—on both behavior and outcomes. In our laboratory experiments, subjects on two sides of a matching market with known payoffs go through each step of the matching algorithm (paralleling the rhetoric used in Roth and Sotomayor 1990 to introduce the DA mechanism to readers). In each step, unmatched proposers first choose the identity of their next proposal, sequentially revealing their preference over those they have not proposed to. As in the DA algorithm, receivers take the next step and choose among any new proposals for the one they like best, similarly revealing their preferences at each step.

The behavior we observe in our experiment mirrors NRMP survey responses. Participants in physicians' (proposers') roles "skip" down their preference rankings. If the participant on the opposite side of the market who provides them their best-case payoff does not rank them highly in return, subjects skip down, proposing instead to a lower-payoff participant on the other side who does rank them highly. On the receiving side of the market, which in the context of the NRMP would correspond to the hospitals, we do not observe substantial departures from truth-telling, and the deviations we do observe are not those often suggested in the literature as simple and useful. These observed behaviors lead to some stark outcomes; in particular, *half of our experimental markets produce unstable outcomes.* Moreover, for

those experimental markets with multiple stable outcomes, where the experimental outcome is stable *the specific matching selected is not the outcome associated with truth-telling*—which leads to smaller gains for receivers when deviating from a straightforward response.

In more detail, our experimental markets are each comprised of 16 individuals, with 8 subjects on each market side, where all participants have complete information on everyone else's payoffs.[2] Our subjects participate in a variety of markets, differing over several theoretically motivated characteristics: i) *market complexity*, as captured by the number of stable matchings (either one, two, or four), and the number of turns required for the DA algorithm to converge under truth-telling; ii) *incentives to manipulate or report non-straightforwardly*, captured through the number of stable matchings and the degree of manipulation required by the receiving side to produce their preferred matching; and iii) the markets' *cardinal representation of preferences*, controlled by the payoff differences between differing matches.

Several findings emerge from our analysis: First, as mentioned above, stable matchings are not the norm, with only one-half of our markets ending at a stable outcome. Moreover, in markets with multiple stable outcomes, a large majority of stable outcomes we observe (71 percent) are not those associated with truth telling. Which specific stable matching is selected in a clearinghouse is of particular importance for applications. For instance, the NRMP initially used the DA algorithm with hospitals in the role of proposers, which results in the hospital-optimal stable matching under truth-telling. The algorithm was then modified in 1998 to have residents serve as proposers (among other changes). Our findings challenge the notion that the receiving side (the residents in the original version of the NRMP) were disadvantaged in terms of the selected matching, and suggest instead that changes to the algorithm might have made the residents worse off.

Second, *market characteristics are important in determining outcomes*. For instance, the cardinal representation has a significant effect on whether outcomes are stable, and the overall distance of the observed outcomes from the core. Where incentives are weak, the outcomes are far less likely to be stable, so instability is more likely to be an issue when different match partners are closer substitutes. Similarly, the degree of truncation required by receivers is highly predictive of which stable matching is chosen.

Third, individual behavior exhibits consistent patterns. *Proposers are not straightforward, and receivers do not optimally truncate*. We find that proposers "skip down" their preference

[2]Having complete information serves as a natural first step in understanding participants' response to incentives, void of issues pertaining to belief updating and learning that would arise in environments with incomplete information. While in reality information frictions are likely, we suspect that participants have some information about the "segment" of the market that is relevant to them (for example, highly ranked hospitals may have knowledge about similarly ranked hospitals and the top students in the market). Furthermore, the underlying theoretical framework is well understood when information is complete, while the theoretical literature on matching with frictions (informational and other) is arguably in its inception. Indeed, most extant theoretical work assumes agents possess complete information (a recent exception is Liu, Mailath, Postlewaite, and Samuelson (2014), see also our literature review below).

lists: for example, a proposer might propose to their third-best receiver, skipping the favorite and second-favorite one; then, if rejected by their third favorite, the proposer might skip down to their fifth favorite; and so on. This behavior is clearly at odds with the theory, but tallies with NRMP survey responses. In contrast, for receivers, we do not observe substantial deviations from straightforward play; they typically choose the best alternative out of any set of proposals. They do not strategically reject proposals, but instead reject fairly consistently those offers from proposers with the lowest payoff, with little reaction to market structure.

Our analysis does suggest that *proposers are sophisticated in their "skipping" behavior*. Proposers consider how a target receiver perceives them—their position in that receiver's preference list—when making a proposal decision. For example, if a proposer's first-best receiver ranks them as largely undesirable, that proposer is less likely to propose to them. Proposers are therefore much more likely to skip receivers who are not stable matches, and in some cases skip the most-preferred stable match receiver, tending instead toward the least-preferred stable match receiver, who receives a relatively higher payoff from matching to them.

One might wonder about the importance of our dynamic implementation of DA for our results. Though some clearinghouses have scramble components that are inherently dynamic, many are like the NRMP, where participants make a single static decision: a ranking of all potential matches. This ranking is then used by the algorithm to simulate a sequence of proposals and acceptances/rejections, terminating in the final matching. Rather than eliciting the entire ranking and running the algorithm, our experiment instead asks subjects to make choices as needed at each step. If they do not have a current partner, subjects on the proposing side are asked who they would like to make an offer to; while those on the receiving side are asked which (if any) of their received proposals they would like to accept. Under some fairly standard assumptions, we show that these dynamic and static implementations are theoretically equivalent. Our choice to use the dynamic implementation makes the mechanics of the DA algorithm clearer—and by making the connection between choices and outcomes more transparent, we hope to give the theory its best chance. Furthermore, especially in complete information matching markets as ours, a static implementation of DA would essentially require us to provide participants with preferences and then ask them to report back these preferences. Such an implementation is likely to suffer from experimental demand and results in the treatments would be difficult to interpret. Our dynamic implementation circumvents this hurdle.

The differences between the DA and our dynamic implementation is motivated by the concerns we have laid out. We argue that three types of evidence suggest our results are not driven merely by the dynamic implementation in our experiments, but describe more general phenomena present in other implementations of the DA. First, we introduce a static

"bounded-rationality" model (quantal response equilibrium) to the matching literature. Calibrating the model to a single parameter, and using only the final full matching in each experimental market, we validate the model using several non-fitted facets of our data. The model therefore formalizes an entirely static explanation for our experimental findings.

Second, we examine evidence derived from field data on the NRMP. Here we show that our behavioral model not only provides a good fit to our experimental data, but it also matches behavioral statements from NRMP surveys (response to the likelihood of matching in submitted rankings). Furthermore, its predicted outcomes strongly mirror a pattern in the available data on NRMP outcomes: a surprisingly high frequency of proposer matches to their top-ranked program. When measured according to the submitted ranking, 42 percent of simulated outcomes are to the top-ranked match (in comparison with 16 percent that would be induced by truthful behavior and stable matchings being implemented). In addition, studies that had access to the rankings submitted to the NRMP indicate that, given the physicians rankings, hospitals have little to gain by deviations from truth telling (that is, markets seem to have small-cores, see Roth and Peranson 1999). Simulations of our behavioral model lead to similar conclusions.

Third, in our review of the literature, we show that elements consistent with our results have appeared in static implementations of DA (albeit mostly in contexts somewhat different than ours). Similar findings with static implementations suggest the distinction between dynamic and static implementations might not be the first-order reason for the observed departures from theory. Our paper, with its larger variation across market characteristics, serves to systematize and enhance these prior findings by allowing us to identify a channel for what might be driving the deviations from stability in centralized clearinghouses.

Taken together—our experimental results, the behavioral model and its parallelism to the field, and the prior literature's results—there exists the case for a persistent heuristic in matching problems: the conflation of ex ante likelihoods of matching and the preferences for a particular match.

1.1. **Related Literature.** Laboratory experiments focusing on two-sided matching have been relatively scarce. Haruvy and Ünver (2007) studied repeated interactions between receivers and proposers, and inspected the predictive power of the DA (rather than strategic behavior *within* the DA algorithm). They ran a similar version to our sequential game in 4×4 markets. However, in their design: i) proposers were allowed to repeat offers, thereby creating a larger wedge between the game played and the DA algorithm; ii) proposers and receivers were paid for the results in every turn of the sequence (not only the ultimate matching); and, iii) in some sessions there were automated respondents, who automatically accepted the best offer. They found a substantial number of repeat offers (that most centralized clearinghouses do not allow) and significantly less "skipping" by proposers than we find.

Harrison and McCabe (1992) implemented the preference-revelation DA mechanism in one 3×3 (3 proposers and 3 receivers) market and one 4×4 market. They had subjects play a market repeatedly, and replaced many market roles with computers programmed to play truthfully. In their environment, outcomes are more in line with the theoretical predictions than ours. However, they do observe a small degree of "skipping," as well as receivers failing to successfully manipulate the mechanism.

A number of experimental papers seek to compare the different centralized mechanisms that are used in practice. Chen and Sönmez (2006) compare DA with the Boston and the Top Trading Cycle mechanisms. Their focus is on the school-choice problem, hence they have strategic agents on only one side of the market. Chen and Sönmez implemented a preference-revelation design, in which agents knew their own preferences, but not those of other participants (not even statistically).[3] In terms of manipulation, they find that proposers do misrepresent their preferences in DA, but less so than in the other mechanisms. Featherstone and Niederle (2011) also compare DA with the Boston mechanism.[4] They too find that proposers do not necessarily follow their dominant strategy to truthfully reveal, and skip highly-ranked potential matches that are very unlikely to accept them. However, they attribute the effect to weak market-specific incentives for the skipping player. Our own experiments indicate that this effect is more systematic across a larger range of markets. Additionally, by having the subjects engage with the DA mechanism more directly, we show that the effect is less likely to be due to confusion as to how the algorithm works, and more likely due to heuristics and beliefs that subjects bring to this type of matching problem.

Pais and Pintér (2008) test DA, Boston, and the Top Trading Cycles mechanisms in the laboratory under incomplete information. Automating the proposing side of the market to reveal truthfully, they also find greater manipulation by subjects in the Boston mechanism. Furthermore, the Top Trading Cycle mechanism dominates the other two procedures when assessed over both truth-telling and the efficiency of matches. Krishna and Ünver (2008) compare the DA mechanism with the bidding mechanisms commonly used for allocating students to courses in business schools. They show the superiority of the DA mechanism in terms of efficiency (and get frequencies of truthful revelation by proposers, which they focus on, comparable to those observed in our data). Wang and Zhong (2012) examine the random serial dictator and Boston mechanisms for school choice. Their results indicate a large degree of skipping behavior in the serial dictator mechanism, where participants rank schools they perceived as 'safer' higher up their lists than theory would predict.[5]

[3]Using a similar design, Calsamiglia, Haeringer, and Klijn (2010) experimentally examine school-choice, where the submitted preference lists are constrained in length.

[4]Also see Featherstone and Mayefsky (2011) who examine this comparison under incomplete information on the preferences of others.

[5]Eriksson and Strimling (2009) test a new matching game, which they term the *Cocktail Game* and also report deviations from truth-telling under their rules of interactions.

Pais, Pintér, and Veszteg (2011a) may be the closest to the current paper in that it considers two-sided matching through the Gale and Shapley algorithm. However, they consider school and teacher matching. Their experimental design entails 5 teachers and 3 schools, where two of the schools both have two positions available. In other words, some subjects who represent these two schools are to be matched with two teachers each (and teachers are indifferent between which of the two positions they receive in those schools). This generates a rather different strategic setting than the one we study. The paper also considers only one constellation of preferences.They also find limited truth-telling even when information is complete.

Our paper provides two important methodological innovations for the centralized matching literature: First, we consider a rich set of markets that allows us to study the selection of stable matchings that emerge organically as well as the impacts different market attributes have on outcomes: core size, cardinal presentations, number of stages required for the DA algorithm to end, and sensitivity to truncations. Second, we introduce a behavioral model to the matching literature, and show that its predictions are consistent not only with behavior in our experiments, but also with data from the field.

Finally, a few papers experimentally examine decentralized markets. Echenique and Yariv (2013) examine behavior in decentralized markets and find that outcomes are in most cases stable.[6] Their study focuses on selection, and they find that the median stable matching tends to emerge. Featherstone and Mayefsky (2011) and Kagel and Roth (2000) analyze the transition from decentralized matching to centralized clearinghouses, when the market features lead to inefficient matching through unraveling. Nalbantian and Schotter (1995) analyze several procedures for matching with transferable utility, decentralized matching among them, where agents are informed of their own payoffs, but not of anyone else's.[7]

2. Dynamic Design of Centralized Matching

Our paper is an experimental study of strategic behavior in centralized matching markets. Our design has subjects going through the steps of the DA algorithm, instead of submitting preferences. It has the advantage of being more transparent for the subjects, and of alleviating concerns over experimenter demand. The game we will induce in the laboratory is described in Roth and Sotomayor (1990, page 79):

(1) Actions in the market are organized in stages. Each stage is divided into two periods. Within each period, each proposer and receiver must make decisions without knowing the decisions of other proposers and [receivers] in that period.

[6]Also see Pais, Pintér, and Veszteg (2011b) who show that search costs and imperfect information can impede this finding.

[7]There is also some recent theoretical work analyzing matching in decentralized markets, see for example Haeringer and Wooders (2011), Hoffman, Moeller, and Paturi (2013), and Niederle and Yariv (2011).

(2) In the first period of the first stage, each proposer may make at most one proposal to any receiver he chooses (and is also free to make no proposal). Proposals can only be made by proposers.
(3) In the second period of the first stage, each receiver that has received any proposals may freely reject any or all of them immediately. A receiver may also keep at most one proposer "engaged" by not rejecting their proposal.
(4) In the first period of any stage, any proposer who was rejected in the preceding stage may make at most one proposal to any receiver he has not previously proposed to (and been rejected by). In the second period, each receiver may reject any or all of these proposals, including that of any proposer who proposed in an earlier stage and was kept engaged. A receiver may keep at most one proposer engaged by not rejecting his proposal.
(5) If, at the beginning of any stage, no proposer makes a proposal, then the market ends, and each proposer is matched to the receiver he is currently engaged with. Proposers who are not engaged to any receivers, and receivers who are not engaged to any proposers, remain unmatched.[8]

The game imitates the steps within the DA algorithm (see the Appendix for a description). In most centralized matching markets, proposers and receivers submit preferences to a central matching authority (as is the case in the National Residents Matching Program). The authority then uses the submitted preferences as inputs to the DA algorithm, instituting the resulting matching. In contrast, in the game above, proposers and receivers decide on proposals at each step; a matching emerges sequentially through their actions.

Roth and Sotomayor present the game as an introduction to strategic issues in matching. There is a notion of "*straightforward behavior*" in the game. A proposer behaves straightforwardly if their proposals go from the most-preferred receiver to the second-most-preferred receiver, then to the third-most preferred, and so on. A receiver behaves straightforwardly if at each step he or she accepts the most-preferred proposal. Straightforward behavior corresponds naturally to truthful behavior in the centralized mechanism. The strategic issue is whether agents will behave straightforwardly (or truthfully).[9]

We directly adopt the above game within our experimental design (detailed in Section 3). Roth and Sotomayor's use of this game is pedagogical, our reasons are similar. We want subjects to grasp the relation between their actions and the resulting outcomes. Subjects best understand the incentives they face when directly experiencing the steps involved in the matching process. In contrast, with the preference-revelation game, subjects need to map each declared profile into an outcome of the algorithm: This map is complicated, and it is

[8]Our one change to the above is that we recast the men/women in Roth and Sotomayor as proposers/receivers.
[9]We will reserve the word 'truthful' for when we talk about data or simulations with the static preference submission mechanism, and 'straightforward' for when we talk about our dynamic implementation.

difficult to ensure that laboratory subjects have a clear understanding of the DA algorithm in the lab.

A second reason for adopting the above game is related to experimenter demand (see Zizzo, 2010): If we *provide* subjects with a preference ranking, and then proceed to ask them to *submit* a preference ranking, we worry that subjects will infer the experimenters' goals.[10] They may, as a result, act with a different motivation from that we sought to induce. By asking them to present a preference, we present a cue that the experiment is assessing whether they will behave truthfully or not. This cue may trigger behavior related to the consequences of lying, and/or complying with the experimenters' expectations. The resulting experimenter-demand effect is unclear—towards more truth-telling or away from it?—and is inseparable from the behavior we wish to assess.

Moreover, though the main NRMP match is implemented with a direct revelation approach, extant clearinghouses have begun to use dynamic implementations. A notable example is the second-stage of the NRMP match (the Secondary Supplemental Offer and Acceptance Program, the "scramble") introduced in 2012, which uses a multi-day dynamic approach.[11] In addition, certain markets that are thought of as decentralized resemble the type of clearinghouses we implement, when norms of behavior put enough structure on each side's offers and responses. For instance, consider the academic job market in economics. Offers made by departments for tenure-track positions are rarely reneged on, so proposals made can be considered as commitments to match. Candidates can hold on to offers (for a short period at least) while they wait to receive other proposals. Combined with the idea that departments do not make repeat offers to the same candidate, this decentralized market has a similar structure to our experiments.

Theoretically, under some plausible restrictions on behavior, the dynamic game and the direct revelation game induced by the DA algorithm are effectively equivalent. In the Appendix, we describe some of the theoretical background for our investigation as well as the formal requirements for this equivalence.[12]

[10]Private correspondence with the authors of Calsamiglia, Haeringer, and Klijn (2010) indicates this fear is matched to subject behavior in the direct mechanism. In interviews after their experiment, subjects explicitly mention the idea that they thought they should lie.

[11]See `www.nrmp.org/residency/soap/` for details.

[12]In effect the required assumptions are tantamount to selecting Markov-like behavior in the dynamic game alongside a variation of independence of irrelevant alternatives.

Table 1. Example of Market Payoffs

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
w_1	(360,125)	(210,175)	(60,375)	(110,425)	(160,475)	(10,425)	(310,475)	(260,325)
w_2	(160,475)	(360,125)	(260,275)	(210,475)	(60,225)	(110,175)	(10,225)	(310,475)
w_3	(260,375)	(110,325)	(360,125)	(310,325)	(210,425)	(60,475)	(10,375)	(160,375)
w_4	(310,325)	(160,425)	(110,225)	(360,125)	(260,275)	(10,275)	(60,425)	(210,175)
w_5	(260,275)	(310,275)	(160,425)	(60,175)	(360,125)	(10,375)	(210,275)	(110,225)
w_6	(10,425)	(210,375)	(60,325)	(160,375)	(310,375)	(360,125)	(110,175)	(260,425)
w_7	(110,225)	(260,225)	(160,175)	(60,275)	(210,325)	(310,325)	(360,125)	(10,275)
w_8	(260,175)	(210,475)	(310,475)	(10,225)	(160,175)	(110,225)	(60,325)	(360,125)

3. Experimental Design

Our experimental sessions implemented a sequence of markets involving two sides, which we neutrally labeled as *colors* and *foods* in the experiment, but herein we will refer to two-sides of the market as *Workers* and *Firms*.[13] There were 8 roles in each group, totaling 16 subjects in a market. Subjects could match with at most one subject from the opposite group, deriving different monetary payoffs from each match.

Subjects were fully informed on all the potential payoffs for every possible match in the market through a table on their computer screens, as depicted in Table 1, where the first number in each cell is the corresponding worker's payoff in cents, the second number the corresponding firm's payoff.[14] Remaining unmatched resulted in a payoff of 0.

In each experimental market, subjects interacted within a protocol that mimics the DA mechanism with one of the market sides (either the Workers or the Firms) proposing—the Roth-Sotomayor game discussed in Section 2. Subjects on differing sides of the market took turns, each composed of two periods. In the first period of the first turn, each proposer could make (at most) one proposal to any one receiver. In the second period of the first turn, each receiver with proposal(s) could hold on to at most one, rejecting all others. In subsequent turns, proposers who did not have a held proposal from a previous turn could again make offers in the first period. In the second period of subsequent turns, receivers with new proposals chose at most one offer to hold among the new proposals and held proposal, rejecting all others.

In each proposing period, the proposing subjects had 30 seconds to decide whether to propose, and if so to whom.[15] Receivers had 25 seconds to respond to their offers (with a

[13]In the experiment, specific roles on one side were labeled as Red, Blue, etc.; on the other side Apple, Banana, etc.

[14]Full instructions and a list of all markets used are available at: `http://sites.google.com/site/galeshapley/`

[15]Failure to propose in a turn did not alter the proposer's ability to propose in future rounds, unless their non-proposal caused one of the market's end conditions to be met (see footnote 16 below).

failure to respond to any proposal within the time limit interpreted as a rejection of all new proposals).

In order to induce the Roth-Sotomayor game, we imposed a restriction that proposers may not repeat proposals. So, after proposing to and getting rejected by a particular receiver, the proposer could not make subsequent proposals to that same receiver. Each experimental market ended whenever there were no new proposals within a proposing stage.[16]

As markets progressed, turn-by-turn, the subjects observed only their own interactions; they did not observe any proposals/rejections in which they were not directly involved. Subjects in the proposer role knew the precise turn and order in which they had made proposals to the chosen receivers, and similarly the precise turns they were rejected in. They did not, however, observe who else proposed to a particular receiver at any time, which other proposers the receiver had rejected, *etc.* Similarly, receivers observed only the proposals made to them, and their own hold/reject behavior. When the market ended, each held proposal became a match, and the receivers and proposers received their corresponding payoffs (according to the match-payoff table).

Each experimental session was composed of 2 unpaid practice markets followed by a sequence of 15 paid markets. Each market used match payoffs corresponding to one of 6 preference profiles for the participants.[17] A detailed summary of the markets used in the sessions, as well as these markets' characteristics, appears in Table 2. The number of times each market was run appears under the N-column.

The markets were designed to vary over the following dimensions:

Market "Complexity." All but one of our markets have either a unique stable matching or two disjoint, stable matchings. We designed the markets to vary in the number of turns (each 2 periods, proposal/response) required for the DA algorithm to converge under truth-telling, as well as the sensitivity of outcomes to truncation by receivers (the receiving side of the market). The latter is captured in two ways: First, in the column *R-best* we calculate the minimal number of proposers that receivers must truncate in order to achieve the (R)eceiver-preferred stable matching, assuming that proposers behave straightforwardly.[18] Second, in the column *Unstable* we calculate the minimal number of proposers that receivers must

[16] This end condition can have three potential causes: i) All the proposers have held proposals and therefore none is available to make an offer; ii) All proposers without held proposals have no receivers to which they have not made a proposal, so no unheld proposer can make an offer; iii) Some proposers without a held proposal choose not to make a proposal in this turn, and the remaining proposers have no new proposals to make.

[17] Rows and columns were randomly permuted so as to disguise obvious patterns such as the one appearing in the main diagonal of Table 1.

[18] We compute the minimal number $t \in \{1, ..., 8\}$ such that if one receiver truncates the bottom t proposers, then the receiver-optimal stable matching is implemented assuming proposers behave straightforwardly. Smaller truncation values t correspond to smaller necessary deviations from straightforward revelation to implement the receiver-optimal stable matching.

Table 2. Markets Used

Market	Arrangement	Stable Matchings	Truncation		Core Span		Avg. Payoff		DA Turns	N
			R-Optimal	Unstable	P	R	P	R		
(I)	W-F	1	-	1	-	-	\$2.50	\$2.50	8	4
(II)	W-F	1	-	7	-	-	\$2.50	\$3.48	8	8
	F-W	1	-	1	-	-	\$3.48	\$2.50	2	4
(III)	W-F	2	5	8	1.00	1.75	\$2.85	\$2.73	4	4
	W-F Dev 1	1	-	5	-	-	\$2.85	\$2.79	4	8
	W-F Dev 2	1	-	8	-	-	\$2.60	\$3.60	8	8
	F-W	2	4	5	1.75	1.00	\$3.60	\$2.35	1	4
(IV)	W-F	2	1	4	1.00	5.13	\$3.60	\$1.25	1	8
	F-W	2	7	8	5.13	1.00	\$3.81	\$3.10	11	8
(V)[a]	W-F	2	1	3	1.75	2.00	\$3.10	\$2.00	5	28
	W-F Dev 1	1	-	3	-	-	\$2.53	\$2.85	15	8
	F-W	2	4	5	2.00	1.75	\$3.00	\$2.22	6	16
	F-W Dev 1	1	-	5	-	-	\$2.85	\$2.53	6	8
(VI)	W-F	4	7	7	1.00	0.75	\$3.35	\$3.10	3	4
All		1.67	1.83	4.77	1.21	1.23	\$3.04	\$2.64	6.1	**120**

[a]This market was run with marginal payoffs of 20¢ and 50¢ for both the W-F and F-W arrangements.

truncate (jointly and uniformly) to generate an unmatched partner. This measure captures the sensitivity of stable matchings to truncation.

Cardinal Representation. Match payoffs in cents are constructed from each market's ordinal preference profile. The marginal decrease between an agent's n-th and $(n+1)$-th favorite partners is fixed at 50¢ in the majority of markets. In order to gauge the effects of cardinal representations within our markets, we use marginal decreases of just 20¢ in our baseline market, Market (V).[19] The average payment across agents (and across stable matchings when there were two) is between \$2.50 and \$3.20.[20] The average payoffs for proposers and receivers under the proposer-optimal stable matching, which would have been generated under straightforward play by all participants, are given in the *Average Payoff* column for proposers and receivers. Given straightforward behavior, proposers should earn an extra 40¢ per market, varying between \$1.00 less than receivers through to \$2.35 more, depending on the specific market.

Incentives to Manipulate and Core Size. Three markets with multiple stable matchings (Markets (III), (IV), and (V)) are run under both the Worker- and Firm-proposing arrangements.[21] This provides information on the effects from being the proposing side under DA, keeping constant each sides' preferences. The reversed markets are indicated in the *Arrangement* column of Table 2, where *W-F* is the Worker-Proposing arrangement and *F-W* the Firm-Proposing arrangement. In addition, we alter two of our markets, (III) and (V), by switching the position in the ranking of two potential matches for *just one* participant, keeping constant all other preferences. Through this small change we induce a *similar* market with a *unique* stable outcome. For Market (III), two different modifications are introduced to make the worker-optimal and firm-optimal stable matchings from the original market the unique stable outcome (with resulting markets denoted by *W-F Dev 1* and *W-F Dev 2*, respectively, each run with workers proposing). For Market (V), we introduce a modification to make the original worker-optimal stable matching the unique stable outcome. We run this deviation in both the worker-proposing and firm-proposing orientations, which we refer to as *W-F Dev 1* and *F-W Dev 1*, respectively.

Markets also differ in the size of the core. For each proposer we calculate the distance in rank position between their best and worst stable partners, and average these values across all eight proposers. We call the resulting number the proposers' "core span." The analogous calculation is also given for receivers. Core spans vary between 0 (when the stable matching

[19]In theory, payoff representations of preferences do not affect incentives in the complete information DA mechanism, nor do they matter for the set of stable matchings.

[20]For each profile of preferences, we chose payoffs to minimize this average under two constraints: i) the average is above \$2.50; and ii) each subject's payoffs from any match exceeds 5¢.

[21]We also do this for Market (II), which has only a single stable match.

is unique) and 5.13.[22] A larger core span for one side corresponds to greater incentives for achieving that side's optimal stable matching.

Our sessions were run at the California Social Science Experimental Laboratory (CASSEL), and implemented using a variation of the Multi-Stage software. In total, 128 subjects were recruited; all were UCLA undergraduates, and each subject participated in just one session. The average payment per subject was $41 (with a standard deviation of $5), combined with a $5 show-up payment.

4. Aggregate Outcomes

In this section we outline the results from our sessions at the aggregate market level; Table 3 reports a number of aggregate statistics. First, we discuss one of the main motivations for using the DA algorithm in centralized markets—stability of the resulting outcome. Our results demonstrate that stable matchings are *not* the typical outcome. Moreover, we will demonstrate that the specific unstable matches our experimental markets arrive at suggest that proposers, rather than receivers, are the side behaving non-straightforwardly. Second, we examine those markets with multiple stable matchings and investigate the selected matchings. In line with subjects not behaving straightforwardly, we see a large majority of markets end up close to the receiver-optimal stable matching. Finally, we study some tangible outcomes experienced by subjects in our markets, namely time spent and payoffs earned.

4.1. **Proximity to Stable Matchings.** Our experimental markets do not consistently produce a stable outcome. In fact, *just half of the markets result in a stable matching*—48 percent for the markets with a unique stable outcome and 49 percent for those with multiple stable outcomes. The *Stable* column in Table 3 provides the fraction of markets that terminated at a stable matching, broken down by market-arrangement. The table illustrates that the prevalence of unstable outcomes holds across our experimental markets and is not driven by any particular market.

Markets that culminate in an unstable outcome have, by definition, at least one blocking pair for the observed matching. The average unstable matching in our data has 2.9 blocking pairs, and the largest number of blocking pairs in any particular market is 11.[23]

Blocking pairs can be classified into two types. First, there are markets with *available* blocking pairs: blocking pairs that could still form at the final stage of the market, but do

[22]When there are two stable matchings, they were designed to be disjoint—that is, every proposer and receiver's best and worst stable partner are different—so the core span is at least 1 in these markets.

[23]An alternative way of quantifying this distance from stability is to count the number of participants who are part of *some* blocking pair, rather than the overall number of possible blocking pairs (that could entail overlaps in participants). If we were to do that, for markets culminating in an unstable outcome we have an average of 4.0 participants who are part of some blocking pair (the mode is two participants in a single blocking pair, in 22 of the 62 unstable markets).

Table 3. Aggregate Outcomes

Market	Arrangement	Stable	P-Optimal (Closer)	Distance	Unmatched	Δ Payoff		Turns	N
						P	R		
(I)	W-F	25.0%	-	0.71	6.3%	-3.3¢	3.3¢	9.3	4
(II)	W-F	50.0%	-	0.92	1.6%	0.0¢	-17.5¢	8.9	8
	F-W	25.0%	-	1.41	9.4%	-22.4¢	8.6¢	9.0	4
(III)	W-F	50.0%	50.0% (50%)	0.78	3.1%	-22.6¢	-53.2¢	7.3	4
	W-F Dev 1	37.5%	-	1.03	1.6%	-8.7¢	15.1¢	6.0	8
	W-F Dev 2	87.5%	-	0.69	0.0%	0.0¢	-5.5¢	8.4	8
	F-W	50.0%	50.0% (25%)	0.84	6.3%	-58.3¢	-11.7¢	8.0	4
(IV)	W-F	62.5%	0.0% (0%)	1.79	6.3%	-62.5¢	-4.1¢	4.0	8
	F-W	62.5%	100.0% (100%)	1.20	0.0%	-22.7¢	-58.6¢	8.0	8
(V)	W-F	53.6%	0.0% (7.1%)	1.01	3.1%	-64.3¢	-5.7¢	10.7	28
	W-F Dev 1	62.5%	-	1.13	4.7%	-2.5¢	0.8¢	8.3	8
	F-W	18.8%	33.3% (37.5%)	0.86	3.1%	-39.5¢	-25.2¢	11.4	16
	F-W Dev 1	25.0%	-	1.52	6.3%	-44.1¢	34.2¢	10.1	8
(VI)	W-F	75.0%	66.7% (75%)	0.20	0.0%	-15.6¢	-29.7¢	3.5	4
All		48.3%	28.6% (18.3%)	1.05	3.3%	-26.2¢	-10.6¢	8.8	**120**

not. This type of blocking pair necessarily involves unmatched subjects.[24] Alternatively, there are *unavailable* blocking pairs: blocking pairs that cannot form because the proposer in the pair was either previously rejected by the receiver in the pair, or is held by another receiver, and subsequently has no agency to make a proposal to form the blocking pair.

For the 62 unstable markets, 29 have unmatched subjects (see column *Unmatched* in Table 3), while the remaining 33 markets have all the participants matched (with an average of 2 blocking pairs per unstable market). Of the 29 markets in which some subjects end the process unmatched, just 8 markets had an available blocking pair; in the remaining 21 markets, the unmatched proposers were rejected by every blocking receiver.

The observation that most blocking pairs are unavailable suggests that instability is not due to an early termination of the process (say, due to subjects failing to respond in time, or preferring an early close of the market).[25] The high rates of unstable outcomes are by and large due to deviations from straightforward play by some participants in the market. Consider a proposer-receiver blocking pair (w, f) for some matching μ. The blocking pair must be formed as the result of one of two possible deviations from straightforward play: i) the receiver f previously rejected w (equivalent to f submitting a preference report ranking their ultimate match $\mu(f)$ as preferable to w); or ii) proposer w never proposed to receiver f (equivalent to w stating the current match $\mu(w)$ as preferred to f). Of the 181 blocking pairs 57.5 percent have blocking pairs corresponding to category (ii), where proposers have necessarily misstated their preferences. This is suggestive of the substantive misreporting by proposers in our markets. We further examine the behavior that produces these results in Section 5.

Given the prevalence of markets culminating in unstable matchings, it is interesting to see *how far* the resulting matchings are from the set of stable matchings. We use subjects' preference rankings to create a distance measure for all markets at an unstable outcome. Specifically, we measure the average distance in ranking for each individual between their final match (defining the unmatched outcome as rank 9), and the closest rank of a stable match partner. The results are in the column titled *Distance* in Table 3. On average, subjects were approximately one position away from a stable-match partner across all unstable matches, corresponding to an approximate loss of 50¢ per person (the exception being those markets with lower marginal differences between partner payoffs, where this loss was 20¢).[26]

[24]This must be a pair comprised of an unmatched proposer and a receiver such that: i) the receiver had not rejected the proposer; and ii) the receiver is either unmatched or prefers the proposer to her current match.

[25]In fact, as we show below, our experimental markets lasted, on average, a longer number of stages than what would be prescribed by the DA algorithm when preferences are reported straightforwardly.

[26]The overall distance measure for each market arrangement (the average distance in ranking for each individual between their final match and the closest rank of a stable match partner, across *all* realized matchings) may be calculated by multiplying our distance number by the percentage of unstable matchings in the market, as all stable matchings are by definition a distance 0 from a stable matching.

4.2. **Selection of Stable Matchings.** The selection of stable matchings is of particular importance to applications. For example, the NRMP started out following the hospital-proposing DA algorithm. However, after much debate in the medical community, in May of 1997 the board of directors of the NRMP voted to replace the existing matching algorithm with a newly designed, resident-proposing algorithm (that was put into action in 1998), see details in Roth and Peranson (1997). Our experimental data is useful in identifying the role played by each side of the market, since underlying preferences are observed.

We examine those markets that have multiple stable matchings and ask which matching the observed outcome is closest to. The *P-Optimal* column in Table 3 gives the fraction of stable outcomes at the proposer-optimal stable outcome. The figure in parentheses is the fraction of markets in which the outcome was *closer* to the proposer-optimal outcome than the receiver-optimal, measured the same way as the *Distance* column.

For the markets with multiple stable matchings that produced a stable outcome, 28.6 percent are at the proposer-optimal stable matching, the outcome that would result from straightforward play in the DA mechanism. Furthermore, in markets in which roles were reversed (Markets III–V), if anything it is the receiving side in the algorithm that is more likely to achieve their preferred stable outcome.

We note, however, that there is large variation across market arrangements. Market (IV) provides particularly stark observations: all stable outcomes correspond to the receiver-optimal stable matching when the workers propose; when firms propose all the stable outcomes are the proposer-optimal stable matching (i.e., conditional on achieving a stable matching, the same matching is selected in both markets, regardless of the side proposing). In order to glean insight into what is driving these differences across markets, consider the truncation column in Table 2. For this particular market, we see that the worker-proposing arrangement (W-F) is particularly sensitive to truncation, reaching the receiver-optimal stable matching under very small truncations by receivers. Conversely, attaining the receiver-optimal outcome in the *F-W* arrangement requires extreme truncation by the receivers. Inspection of other markets suggests this as a general trend: When truncation requirements are low, the stable matching implemented is the receivers' best. With moderate levels of truncation required, both stable matchings emerge experimentally. When the (collective) truncations required by receivers are extreme, the stable matching generated is the proposers' best.

4.3. **Tangible Outcomes: Time and Payoffs.**

4.3.1. *Time to Termination.* On average, each market takes approximately 9 turns to finish (see column *Turns*), with the average turn taking 21.5 seconds.

Comparing the number of turns observed to the number predicted by truth-telling behavior in the DA mechanism, experimental markets take an extra 2.5 turns to finish; only 24 out

of 120 markets end within the truth-telling number of turns. One simple conjecture to explain "skipping" by proposers is that subjects are trying to shorten the time before a final matching is achieved. These results suggest, however, that any behavior intended to shorten time spent in the experiment was unsuccessful.

4.3.2. *Average Payoffs.* Consider the average proposer in our average market. Conditional on the proposer-optimal outcome being chosen, her expected payoff is \$3.02 per market; if the receiver-optimal stable matching is chosen, her corresponding expected payoff is \$2.57. The observed figures are closer to the latter, lower, prediction: the average payoff of a proposer in our markets is \$2.66. Conducting the same exercise for the receivers' side of the market, the average receiver's expected payoff varies between \$2.66 per market if the proposer-optimal outcome is selected, and \$3.09 under the receiver-optimal stable matching. The observed value is \$2.91, in between these two figures.[27] These figures are consistent with our observations regarding the selection of stable matchings. In particular, payoffs do not coincide with those generated by the proposer-optimal stable matching.

Column Δ-*Payoff* provides the average difference in the actual payment from that of the best outcome by market side (that is, the sub-column corresponding to proposers contains the difference between the average realized proposer's market payoff and the payoff under the proposer-optimal stable matching. Similarly for the sub-column corresponding to receivers).[28] This column contains similar information to the *Distance* and *P-Optimal* columns, but provides additional insights that will tie to the individual-behavior analysis below. In some markets the average matched receivers achieve better outcomes than their most-preferred stable match partner. In these markets, there is a unique stable outcome, and the average matched proposer is faring worse. As will be echoed in the individual analysis below, the reason for these results is that proposers in these markets propose to a receiver that is ranked below their stable match partner, one that values them more highly. In payoff terms, conditional on being matched receivers earn, on average, 6¢ more than the stable-outcome payoff in those markets with a unique stable outcome. In markets with multiple stable outcomes, both sides fare poorly, though receivers are closer in dollar and relative terms to their most-preferred stable outcomes. In fact, at the end of the experiment, we asked subjects to reflect on the experiment and express their preference over having the role of proposer or receiver; 79.6 percent expressed a preference for the receiver role.

4.4. **Market Characteristics and Outcomes.** The previous discussion suggests that there are aspects of the market that predict *which* stable matching is produced. In particular, the manipulation difficulty for receivers (as measured by the level of truncation required to establish their preferred matching) is a good predictor of whether the market

[27]Accounting for unmatched subjects raises these observed averages by approximately 8¢.
[28]These averages are conditional on agents being matched.

Table 4. Descriptive Outcome Regressions

	Turns	Distance	Blocking Pairs	Stable Outcome	Closer to P-Optimal
Market No.	-0.058	-0.176	0.161	0.005	-0.004*
	(0.082)	(0.137)	(0.126)	(0.005)	(0.002)
Low marginals for proposers	0.001	0.101***	0.116**	-0.390***	-0.001
	(0.025)	(0.030)	(0.053)	(0.024)	(0.202)
Low marginals for receivers	0.065***	-0.028*	-0.026**	0.107***	0.274***
	(0.015)	(0.013)	(0.010)	(0.023)	(0.053)
Proposer core span	0.145*	-0.046	-0.011	0.037	0.102***
	(0.080)	(0.138)	(0.118)	(0.049)	(0.044)
Receiver core span	-0.096	0.002	0.060	0.024	-0.048
	(0.056)	(0.091)	(0.078)	(0.020)	(0.055)
R-best Truncation	-0.109	-0.039	-0.094	-0.017	0.135***
	(0.066)	(0.125)	(0.100)	(0.037)	(0.032)
N	120	120	120	120	72

Note: Stable Outcome and Closer to P-Optimal give the marginal effects from a Probit regression; all other columns are elasticities obtained from an OLS regression. Standard errors given in parentheses below the estimates, and are clustered by market. Significance levels indicated as follows: ***-99%, **-95%, and *-90%.

ends at the proposer- or the receiver-optimal stable outcome. We now formalize this idea, and inspect other market characteristics that affect outcomes. Table 4 provides results from descriptive regressions seeking to explain different dimensions of the observed market outcomes, using the characteristics outlined in Section 3 as regressors. The first column outlines the effect these design metrics have on a market's duration, the observed number of turns. The next three measures relate to stability: the distance to a stable outcome; the number of blocking pairs; and a dummy variable indicating whether the final outcome was stable or not. Finally, the last column looks at the proximity to the proposer-optimal matching, where the dependent variable is a dummy indicating that the market outcome is *closer* to the proposer-optimal stable matching, where we restrict the regression to those markets with multiple stable outcomes.

We use the following regressors: *Market No.*, takes values from 1 to 15 and represents the position in the sequence of markets within an experimental session–the first paid market takes value 1, the last market takes value 15. The next two regressors are dummies indicating lower 20¢ marginals (as opposed to the standard 50¢) in the market, for each of the two sides. The final three regressors are metrics from Table 2, corresponding to the average distance (core span) between the extremal stable matchings, for proposers and receivers, respectively, and the truncation required by receivers to produce the receiver-optimal stable matching if proposers act straightforwardly (*R-Optimal Truncation* from Table 2).

We first note that *Market No.* does not have much explanatory power in our regressions, indicating limited learning or convergence throughout an experimental session.

In terms of market attributes, the different columns highlight several points. First, the only significant effect on the number of stages taken to conclude a market are the incentives of receivers to truncate—the smaller the marginal incentive, the greater the number of stages.[29]

Second, the regressions on measures of market stability indicate that low-powered incentives seem to have a strong effect: Low marginals for proposers significantly increase instability across all three measures. Low marginals for receivers have the opposite effect, increasing outcome stability. We return to the link between payoffs and outcomes in Section 6.

Finally, consistent with the observation in Section 4, we find that the greater the required truncation levels, and the weaker the receiver's incentives, the more likely it is that the observed outcome is closer to the *proposer-optimal* matching. Greater proposer incentives (namely, a larger distance between the two stable matchings for the proposers) have the same effect.

[29]This, again, suggests that subjects are not following strategies intended to shorten the length of play. Lower marginal payoffs would, if anything, lead truncation (or skipping) to be less costly. Thus, if impatience were driving results, we would expect these variables to be associated with shorter market activity.

The theoretical framework underlying stable predictions, as decentralized core outcomes, or as outcomes of a centralized clearinghouse à la deferred acceptance, is inherently ordinal. In many applications matchings are associated with cardinal outcomes for the participating individuals: wages in labor markets, school performance or commute time from home for school assignments, etc. Our observations suggest that cardinal specifications may have an important role in determining outcomes.

5. Individual Behavior

The previous section depicts aggregate market outcomes, frequently corresponding to instability. But these aggregate measures are the product of 16 individuals' choice sequences within each market. We now analyze response within the experiment at the market participant level.

An important finding of the paper is that proposers do not behave straightforwardly, in the sense defined in Section 2. That is, their proposals do not track their preference rankings. Receivers' behavior, on the other hand, is largely straightforward: receivers (tentatively) accept proposals from the most-preferred proposers in the vast majority of cases. Figure 1 presents the empirical distribution for straightforward play by proposers and receivers, where each data point represents the fraction of interactions in which a specific subject makes choices according to their induced preference.[30]

The results are striking. The theory predicts that proposers will straightforwardly reveal their preferences, and receivers will strategically misrepresent to achieve better outcomes, most notably (and simply) by truncating preference orderings. In our experiment, over half the subjects acting as receivers behave straightforwardly in *all* their experimental interactions within this role, with two-thirds reporting straightforwardly more than 90 percent of the time. The distribution of truth-telling for proposers is more uniform—and stochastically dominated by that for receivers—with approximately one-third of the proposers behaving straightforwardly less than half of the time. In what follows we analyze individuals' behavior in detail.

5.1. Truncation and Skipping.

If any single receiver (or a group of receivers jointly) were to truncate their preference below the receiver-optimal stable match—listing any proposers ranked below this point as unacceptable—then if other market participants play straightforwardly, the resulting outcome is the receiver-optimal stable matching. Given our data, we

[30] An alternative measure for straightforward play is the Kemeny distance between the revealed preference and the induced one. Calculating this measure at the individual level (where each observation is a subject-market), the results are qualitatively similar to those shown in Figure 1, with the same pattern of stochastic dominance. However, the results for the Kemeny measure are less skewed, reflecting the fact that though subjects deviate from straightforward play, the size of this deviation is generally small with respect to the induced preference.

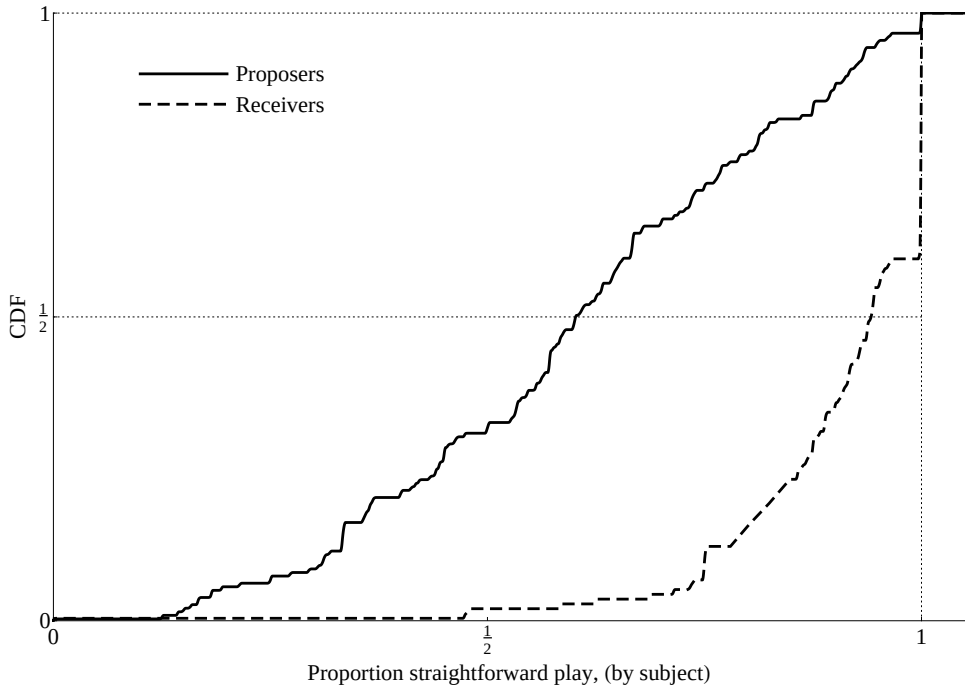


FIGURE 1. Distributions of Straightforward/Truthful Play

can check for the extent of the truncation receiver players are using by direct inference: when an unmatched receiver rejects all those proposing in a turn, this is equivalent to stating that the proposals all came from (purportedly) unacceptable proposers. We do not observe truncations in any other case. For instance, consider the worker-proposing situation where two workers, w and w' have proposed to a particular firm on the same turn, and w' is accepted. In this situation we cannot use revealed preference to infer whether w was acceptable or not, only that w' is preferred to w, and that w' is preferred to no match.

Table 5(A) presents the probability of rejecting *all* those proposing, conditional on the true ranking of the *best* proposer. That is, for any rank k, we track all the events at which a receiver (with no tentative acceptances) receives proposals, the best of which is from their k-th ranked partner. The number of these events across all turns is in the fourth column, and the number in the first turn of the market is in the fifth column. We calculate the fraction of times that *all* these proposals were rejected, both across turns, and in the very first turn. When the proposer is the receiver's first-best (rank 1), this figure is close to zero. In fact, *truncations within the upper-half of the preference ordering are rare.* As the ranking of the best proposal falls (toward 8) the truncation probability increases, reaching a rejection rate of 58.2 percent when the highest-ranked proposer is the worst partner. This truncation behavior does not qualitatively differ between the first and subsequent turns—both exhibit large probabilities only in the final two positions of the preference ordering. The results could, in principle, be influenced by the large number of observations in particular markets

(for instance, the two arrangements *W-F* and *F-W* of Market V). Analyzing each of the markets separately does not drastically change our results.[31]

However, the use of truncation strategies does not provide the complete story. The theory makes clear that proposers have a dominant strategy to straightforwardly reveal their preferences. We now analyze whether proposers follow this dominant strategy, and move in sequence through their preference list. Table 5(B) details the probability with which proposers act non-straightforwardly, that is where they do not propose to the highest-ranked receiver available. The overall probability is 33.8 percent, consistent with our initial observations that a substantial number of proposers do not make offers in order of their true preferences. The table also indicates how non-straightforward play varies with how the proposer is ranked by the straightforward receiver. Specifically, we report the rate at which proposers with an active choice choose to skip their most-preferred available receiver, conditioning on how that receiver ranks the proposer.[32] In order to provide some control over any time effects within a market, we again report separately the probabilities for the first turn within a market (with the fourth and fifth columns denoting the number of observations over all turns and over the first turn, respectively).

The results illustrate a clear pattern in proposal behavior: proposers are not following their dominant strategy. Instead, *proposers are skipping highly ranked receivers who are likely to reject them*. This pattern is qualitatively similar, for behavior in the first turn of a market, matching our stationarity assumptions.[33] Skipping behavior is reduced by 8.6 percent when we compare the first and last five markets in an experimental session (where theses blocks of five have an identical sequence of markets within them), so subjects do learn to skip less as the experiment proceeds. However, quantitatively, the fraction of turns where proposers skip is still large.

In many instances this skipping behavior would be inconsequential for outcomes: for instance, if every proposer were to skip down to their most-preferred stable partner, the game would end in a single turn and yield that stable matching. However, in the first turn 19.5 percent of proposers skip down *below* their optimal stable partner, and 10.2 percent skip down to receivers ranked *below* their worst stable partner. Across all turns, conditioning on the availability of the stable partners, 17.2 percent of proposers skip below their own

[31]For Market (V), the probability of truncating within the top half of the preference ordering is 3.2 percent for the W-F treatment, 3.3 percent in the F-W treatment, and 2.4 percent for all other markets. For the bottom half, the respective percentages are 31.2, 46.2, and 30.6.

[32]For instance, the first row of the table, corresponding to a proposer ranked 1st, details the probability with which the proposer skips down below the best-ranked receiver that has not been ruled out, where that receiver ranks them as their best possible match. When the proposer's rank is 8th, the proposer's straightforward proposal ranks them as the worst outcome among all eight proposers.

[33]Rates of non-straightforward behavior in the first round are not significantly different overall, for either proposers or receivers. Conditioning on how receiver ranks them we do find significant differences for proposer skipping only at the reflected ranks 3rd, 6th and 7th.

TABLE 5. Non-Straightforward Play

(A) Receiver Truncation

Best Proposal ranked as	Prob. of rejecting all (%)		Subsample Size	
	All turns	First turn	All	First
1st	0.2	0.0	551	80
	(0.2)	(-)		
2nd	1.1	0.0	472	118
	(0.5)	(-)		
3rd	3.2	4.5	402	132
	(0.9)	(1.8)		
4th	8.8	12.2	317	115
	(1.6)	(3.0)		
5th	21.0	19.4	119	36
	(3.7)	(6.6)		
6th	21.8	20.0	87	15
	(4.4)	(10.3)		
7th	45.6	63.6	57	11
	(6.6)	(14.5)		
8th	58.2	50.0	55	22
	(6.7)	(10.7)		
All	7.2	9.1	2060	529

(B) Proposer Skipping

Best Receiver ranks proposer	Prob. of skip (%)		Subsample Size	
	All turns	First turn	All	First
1st	6.5	3.1	92	32
	(2.6)	(3.1)		
2nd	21.6	13.9	208	79
	(2.9)	(3.9)		
3rd	33.1	18.4	317	141
	(2.6)	(3.3)		
4th	23.4	24.7	487	186
	(1.9)	(3.2)		
5th	38.4	40.9	383	154
	(2.5)	(4.0)		
6th	27.6	39.4	351	94
	(2.4)	(5.1)		
7th	37.9	61.9	314	42
	(2.7)	(7.5)		
8th	53.4	60.8	483	176
	(2.3)	(3.7)		
All	33.8	35.1	2635	904

proposer-optimal partner, and 8.5 percent below their receiver-optimal partner. We see no qualitative difference between the first and subsequent turns.[34]

6. Noisy Equilibrium

One possible explanation for observed behavior is that subjects are trying to optimize, but are making mistakes. When mistakes are not very costly, they are less likely to be corrected through play. To our knowledge there has been little research on the question of robustness for centralized mechanisms: when are small mistakes on the part of participants likely to lead to unstable/undesirable outcomes? In this section we will define a quantal response equilibrium (QRE) for our matching game—a solution concept incorporating both best response and noise, which has been used extensively to explain non-equilibrium behavior.[35] After calibrating the model's noise-parameter to our outcome data, we will show that the QRE's predictions mirror the patterns of behavior within our data. Outside of helping to understand our experimental results, this behavioral model provides a potentially useful tool for market designers, allowing a way to examine robustness of new clearinghouses to a structured form of mistake by participants.

To illustrate the noisy equilibrium concept we will use Market (V) as a running example. Suppose all proposers and receivers with the exception of w_7 truthfully reveal their preference order. Worker w_7 has the underlying true ranking $f_8 \succ f_7 \succ f_1 \succ f_6 \succ f_2 \succ f_3 \succ f_5 \succ f_4 \succ w_7$. Her most-preferred stable match partner is f_7, while her least-preferred stable partner is f_6 (though she prefers this to being unmatched). Given others' truthful behavior, if w_7 were to skip down her preference order and omit her overall most-preferred firm f_8 (reporting the preference $f_7 \succ f_1 \succ f_6 \succ f_2 \succ f_3 \succ f_5 \succ f_4 \succ w_7$) her final match outcome under deferred acceptance would still be f_7. Skipping down to the most-preferred stable match partner has no effect on her final match—if others report truthfully, f_7 will still be her matched outcome. More generally, *any* group of proposers can *omit* any number of non-stable match partners and the resulting outcome will still be the same proposer-preferred stable matching, so long as receivers truthfully rank.

However, if worker w_7 were to skip down just below her most-preferred stable partner f_7 (for example, providing the ranking $f_6 \succ f_2 \succ f_3 \succ f_5 \succ f_4 \succ w_7$) the final outcome if all others report truthfully switches to the least-preferred stable matching, and w_7's matched partner is f_6. Skipping down the true order even further yields a strictly worse match than f_6, where the resulting matching will always be unstable. In general, given n stable matchings in

[34]Proposer skipping has also been observed in the direct-revelation mechanism, cf. results in Harrison and McCabe (1992) and Featherstone and Mayefsky (2011). This behavior is not the focus of their analyses, but their reported results make its presence clear and at non-negligible frequencies.

[35]For literature on QRE see McKelvey and Palfrey (1995, 1998) and references thereof. These papers also demonstrate the existence of a QRE in our environment by finiteness of our dynamic game (players, choices, and round after which it must end) and our choice of error distribution.

any finite matching market, the matchings possess a well-known lattice structure, and can be jointly ordered by the proposing side from best to worst. Measure the size of each proposers' contiguous skips by the number of stable-match partners omitted from the ranking, k, and look for the max skip over all proposers $\overline{k}$. The final outcome under DA will be the proposers' $\overline{k}+1$-th most-preferred stable matching, so long as $\overline{k} < n$.

This process is illustrated in Figure 2's upper-left panel. For each worker w_i in Market (V) the array illustrates the outcome if they idiosyncratically skip their *k-most-preferred* partner(s) when all other participants truthfully reveal their preference order. Each cell's shading indicates the resulting match outcome, with darker shading for final matches to more-preferred partners, and lighter shading for less-preferred partners.[36]

In addition, in the first panel, a black circle in a cell indicates where the worker's most-preferred stable partner is. For worker w_7, the black dot appears in the second cell (corresponding to firm f_7 their second most-preferred partner). Skipping down to the black dot does not change the outcome when others truthfully reveal: the shading remains constant up to this point for all workers, but it will lighten directly after in the third slot, indicating where the most-preferred firm has been skipped.

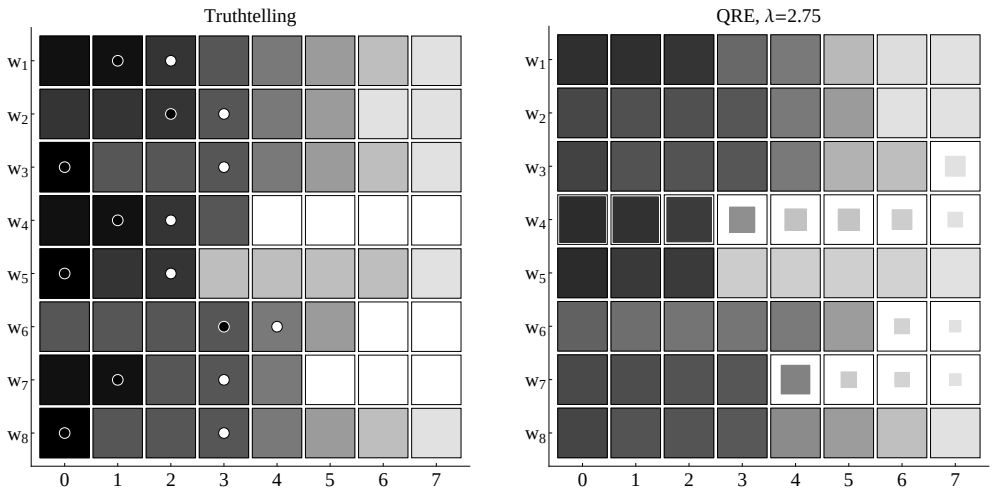
A white dot in a cell represents the critical location for the worker's least-preferred stable partner. For w_7 this is the fourth location, corresponding to firm f_6 being the top of the provided ranking. Skipping to this point leads to the least-preferred stable partner, when others straightforwardly reveal; skipping below this point leads to a strictly-worse outcome, as is indicated by the progressively lighter shading for all workers as they skip below the least-preferred stable partner.

Truthful revelation is a weakly dominant strategy for proposers under DA, so the payoff from skipping down the order must be less than or equal to the payoff from behaving straightforwardly. Moreover, payoffs are weakly decreasing in the size of the skip, so the cell shading necessarily gets lighter as we move from left to right, regardless of other participants' play. However, under truth-telling the pattern revealed in the figure indicates the indifference over skips up to the most-preferred parter, and then again a constant payoff in skips between stable-partners.

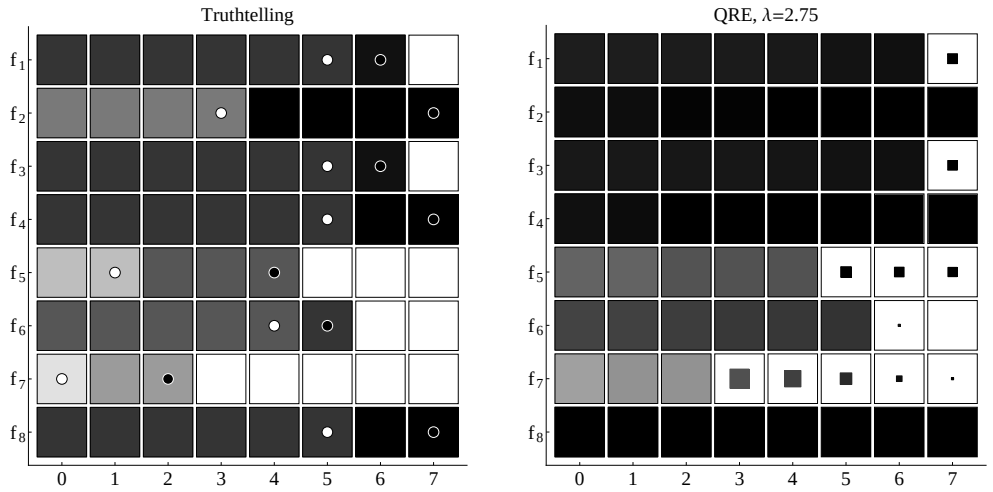
The bottom-left panel provides payoff information for the proposal-receiving side in Market (V). The array illustrates the gain/loss to firm f_i from *truncating* the true preference ordering from below by k places. Unlike the proposers, receivers do not have a dominant strategy to straightforwardly reveal, and truncating can both improve or harm their final outcome. For example, firm f_5 has an underlying preference $w_2 \succ w_8 \succ w_3 \succ w_5 \succ w_1 \succ w_4 \succ w_6 \succ w_7 \succ f_5$, with stable match partners w_5 and w_6. Truncating her preference and dropping the worst-ranked worker w_7 has no effect when others are truthful (the shading

[36]A completely white cell represents the worst outcome in this market, remaining unmatched.

FIGURE 2. Payoffs in Noisy Equilibria



(A) Proposer/Worker Skipping



(B) Receiver/Firm Truncation

Note: Shading represents expected payoff from the corresponding skip/truncation (payoffs are degenerate in the truth-telling arrays and derived from a simulation of size 1,000 in the QRE panels). Darker shades represent higher normalized payoffs. Black circles correspond to a critical points for skipping/truncation beyond which the most-preferred stable match partner will be skipped/truncated; white circles represent the same critical point for the least-preferred stable match partner.

in spots 0 and 1 is identical), and the matched partner is w_6, the truth-telling outcome. The firm obtains a better outcome when she truncates the least-preferred stable partner w_6, shortening her ranking by 2 workers to $w_2 \succ w_8 \succ w_3 \succ w_5 \succ w_1 \succ w_4 \succ f_5$, and obtaining her most-preferred stable partner. However, if f_5 truncates too much and removes the most-preferred stable partner w_6 (truncating 5 spots and ranking $w_2 \succ w_8 \succ w_3 \succ f_5$), then her payoff is reduced, and she will necessarily be unmatched. Similar to the above

panel for proposing workers, cells with black and white circles in them represent the critical locations for truncation when others are truthful—the ranking position of the firm's most- and least-preferred stable-match partners, respectively. The array's shading illustrates the pattern in payoffs, with just three levels per firm: the least-preferred stable partner if they do not truncate enough; an increase to the best stable partner when the firm truncates the least- but not the most-preferred stable parter; and a decrease to being unmatched if the most-preferred stable match partner is truncated.

Truth-telling by all participants is not an equilibrium outcome when there are multiple stable matchings. To examine how best-response changes as others play with noise we use a modified notion of QRE, where we limit the available strategies to block-skips or block-truncations of the true preference, by proposers and receivers of proposals, respectively.[37] Using a logistic-error structure, we assume that if worker w_i expects to get a payoff of π_{ij}^W from skipping at level $j \in \{0, \ldots, 7\}$, then she will play this skip strategy with probability

$$p_{ij}^W(\Pi^W; \lambda) = \frac{\exp\left\{\lambda \cdot \pi_{ij}^W\right\}}{\sum_{k=0}^{7} \exp\left\{\lambda \cdot \pi_{ik}^W\right\}},$$

where λ is a parameter capturing the noisiness of play (a value of zero produces random play, while as $\lambda \to \infty$ behavior tends to best response), and Π^W is the matrix with generic element π_{ij}^W. Similarly the probability of firm f_i using a truncation level j will be assumed to be

$$p_{ij}^F(\Pi^F; \lambda) = \frac{\exp\left\{\lambda \cdot \pi_{ij}^F\right\}}{\sum_{k=0}^{7} \exp\left\{\lambda \cdot \pi_{ik}^F\right\}},$$

where π_{ij}^F represents firm f_i's expected payoff from truncation at level j in cents, and where π_{ij}^F is a generic element of the matrix Π^F.

An equilibrium discipline on the outcome comes from forcing the payoff matrices Π^W and Π^F to be expectations under a consistent set of beliefs over other participants' skips/truncations. Given the mixed-strategies used—the matrices $P^W(\Pi^W)$ and $P^F(\Pi^F)$, calculated via the logistic-error assumption according to the believed payoff matrices—we can calculate the expected payoffs using the deferred-acceptance algorithm. We will denote the map from probabilities over rankings to expected payoffs for each strategy and role as $\begin{bmatrix} \tilde{\Pi}^W & \tilde{\Pi}^F \end{bmatrix} = \phi_{DA}\left(P^W, P^F\right)$.

[37]So if the true ranking is $w : f_1 \succ f_2 \succ \cdots \succ f_n \succ w$ we allow for stated preference $f_i \succ f_{i+1} \succ \cdots \succ f_n \succ w \succ f_{i-1} \sim f_{i-2} \sim \cdots \sim f_1$. Similarly, for the firms we only allow truncations of the true preference $f : w_1 \succ \cdots \succ w_n$ to the truncated preference $w_1 \succ \cdots \succ w_{n-i}$. Each worker/firm therefore has 8 available strategies, in comparison to the 9! possible preference orderings available in the game. This restriction will retain much of the strategic nature of the game, and mirrors observed features in our data, while making the model tractable. We will refer to skipping at level-0 as truthfully listing the preferences, while level-7 will refer to only listing the worst outcome as acceptable. Similarly, truncation at level-0 will be listing the underlying preference, and level-7 will refer to only listing the most-preferred partner as acceptable.

A QRE in skipping/truncation therefore boils down to solving the following fixed point:

$$\begin{bmatrix} \Pi^W & \Pi^F \end{bmatrix} = \phi_{DA}\left(P^W\left(\Pi^W;\lambda\right), P^F\left(\Pi^W;\lambda\right)\right).$$

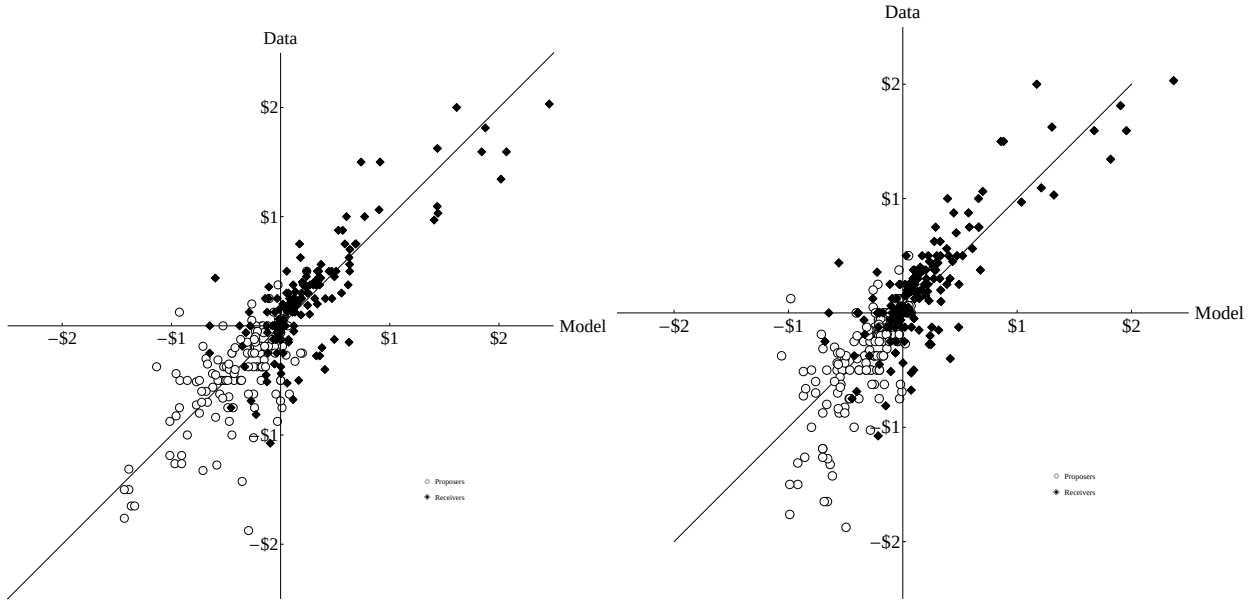
Fixing the noise-parameter λ to be 2.75, the upper- and lower-right panels in Figure 2 illustrate the expected payoffs to the participants for each skip/truncation strategy at one such fixed point of the system.[38] In contrast to the truth-telling panels, the outcomes under QRE are stochastic, and expected payoffs are calculated through a simulation of 1,000 (fixed) draws for the payoff-weighted mistakes. The rightmost panels illustrate two distinct effects, the probability of matching at all, and the preference for those matches, conditional on matching. The probability of matching is indicated by the area of the cell shaded, where a fully shaded cell indicates a certain probability of matching, while the smaller the shaded region the less likely the participant is to be matched at all. The conditional expectation of the match value is indicated by the shaded part of the cell, using the same shading scale as the truth-telling panels.

Comparing truth-telling and QRE for this particular market, we see an increase in receivers' expected outcomes under minimal truncation, and a corresponding reduction in the truth-telling payoff for workers. Particular firms (in this case f_5 and f_7) derive a slight payoff increases from truncating their least-preferred stable partner, but the gains are much smaller when compared to the situation where others are truthful. The remaining firms get no gain from truncation at any level, though they do not suffer large losses from truncating at moderate levels. Truncating the most-preferred stable partner still produces a large expected loss, though firms still occasionally match even when truncating past this extreme (an improvement over the truth-telling panel, where they would certainly be unmatched if they truncated too much).

The change in payoff patterns for firms stems from both an increased chance that the firm's idiosyncratic truncation is unnecessary (the receiver-preferred stable matching would likely occur anyway because of others' deviations) and an increased cost of truncations (an increased likelihood that the relevant stable-match proposer might now skip them).

For the proposing workers there is a corresponding reduction in payoffs when being truthful or skipping a small number of firms in comparison to the truth-telling array. Workers are now mostly indifferent over skips up to or above the receiver-preferred stable match partner, they have no strong incentives not to skip down. However, proposers do have a strong preference not to skip below their least-preferred stable partners, as this corresponds to a large drop in final payoff.

[38]Fixed points are found by iterating to convergence, where the algorithm's initial condition is truth-telling by all participants. In principle, there might be multiple fixed points of this system. Our initial condition therefore serves as a consistent selection device.



(A) QRE Fixed Point (B) One-Step Approximation

FIGURE 3. Simulated and Observed Payoff Difference from Proposer-best matching

The particular value used in the figure, $\lambda = 2.75$, matches the value estimated through our data, obtained through a simulated maximum-likelihood approach. The likelihood of each *market-wide matching* in our experiments is measured via simulation, and this likelihood is maximized over the QRE parameter λ.

The noise parameter λ is fitted with data from 120 *market-wide* matchings. Table 6 provides results for the data and the model across markets. The table's first two columns present the fraction of matchings in each market that are completely stable relative to the true preference, in both the observed and fitted QRE model. We have 20 unique markets (taking into account changes in the marginals that affect the QRE prediction), each with at least four experimental observations. Over these 20 markets there is a correlation of 0.68 between model and data. The next two columns in table 6 examine the markets with multiple stable matchings, and provide the fraction of stable outcomes at the proposer-best matching. Again, the model accurately predicts the precise stable matchings chosen, with a correlation between the observed and simulated frequency of 0.76. Similar levels of fit (a correlation of 0.77) is found when we examine stability at the market participant level, asking how often each particular participant is matched with one of their stable-match partners.

The model has a comparable fit when we examine the payoffs at the participant level. To difference out levels in induced payoffs, we examine the difference in payoff between the realized and the proposer-best stable match payoff. Figure 3a provides a scatter plot, where each point in the plot represents a single market participant in one of the 20 markets, and

Table 6. QRE Model Predictions

Market	Arrangement	Stable Match		P-Best		Core Span (P/R)	
		Observed	QRE	Observed	QRE	True	QRE
(I)	W-F	25%	1%	-	-	0/0	0/0
(II)	W-F	50%	90%	-	-	0/0	0/0
	F-W	25%	12%	-	-	0/0	0/0
(III)	W-F	50%	44%	50%	59%	1.00/1.75	0.77/1.15
	W-F Dev 1	37.5%	31%	-		0/0	0.17/0.21
	W-F Dev 2	87.5%	88%	-	-	0/0	0/0
	F-W	50%	53%	50.0%	4%	1.75/1.00	0.13/0.12
(IV)	W-F	62.5%	91%	0%	0%	1.00/5.13	0.30/0.17
	F-W	62.5%	64%	100%	100%	5.13/1.00	5.07/1.04
(V)	W-F	53.6%	33%	0%	0%	1.75/2	0.18/0.28
	W-F Dev 1	62.5%	53%	-	-	0/0	0.11/0.19
	F-W	18.8%	27%	33.3%	49%	2/1.75	0.83/0.80
	F-W Dev 1	25%	37%	-	-	0/0	0.19/0.19
(VI)	W-F	75.0%	28%	66.7%	89%	1/0.75	0.62/0.62

the location indicates the payoff difference for that participant. The horizontal axis indicates the expected difference from the QRE model, while the vertical axis indicates the average difference for that participant within the experimental data. White circles represent the model-data differences for proposers (8 in each market) while black-diamonds represent the differences for each receiver (8 in each market). The 45-degree line therefore indicates perfect agreement between the model and the data. As the figure indicates, the QRE model's fit is remarkably good (a correlation of 0.88 across all participants). This bears highlighting. A fully stationary behavioral model, fitted to a single free parameter using data only from the final *market-wide* matching selected in our dynamic clearinghouse, has a near 90 percent fit as an explanation for market-participant *deviations* from the truth-telling prediction.

Despite this good fit, the full QRE model might be intractable in many interesting matching situations. The model requires the calculation of a fixed-point, where both the dimension of the fixed point, and the number of steps required to complete the market algorithm, increase substantially with the number of participants. The QRE model might therefore be impractical for assessing the robustness of the deferred-acceptance algorithm in any large market. Furthermore, QRE requires a large degree of sophistication from participants, which may be a shortcoming if considered as a positive behavioral model. We therefore considered an approximation of the QRE model, which is simple to calculate by both a market designer assessing a potential clearinghouse, and by potential market participants. Namely, we inspect a model that is based only on unilateral deviations from truthful play (the first step of our QRE estimation algorithm). We calculate each participants' payoff under a particular skip/truncation, holding others' response as truthful, and use these payoffs to assess the mixed strategies as before. This process provides a good first approximation to our QRE model, and can be calculated with far fewer steps than the fixed-point method, which must be iterated to convergence. As it turns out, this first-pass model of noisy behavior provides a similar fit to our fixed-point QRE results.

Figure 3b provides the same scatter plot for model-data fit, where we use the deviation-from-truth-telling payoff matrix to determine skips and truncations instead of the fixed-point payoffs. Though there are some small differences between Figures 3a and 3b, the main pattern, a strong correlation between model and data, is the same. Similarly, the one-step version of the algorithm produces comparable results for stability, stable-matching selection, and core-size.

7. Connecting Results to the NRMP

The NRMP releases annual reports containing a variety of aggregate statistics on the matching procedure, as well as results from surveys conducted by the NRMP itself.[39] These results seem to bear some strong connections with our experimental data.

In 2013, 49.4 percent of U.S. senior applicants and 42.4 percent of independent applicants filled out the survey administered by the NRMP. When asked to specify different ranking strategies, 34 percent of U.S. senior respondents (12 percent of independent respondents) confirmed the statement "I ranked one or more less competitive program(s) in my first-choice specialty as a 'safety net'." Similarly, 7 percent of U.S. senior respondents (10 percent of independent respondents) confirmed the statement that "I ranked one or more program(s) in an alternative specialty as a 'fall-back' plan." In fact, 6 percent of U.S. respondents (22 percent of independent respondents) confirmed the more global statement regarding the reported preferences that "I ranked the programs based on the likelihood of matching (most likely first, etc.)." Residents confirm that they ranked programs in the order of their preference (98 percent of U.S. seniors and 87 percent of international applicants) but they do not rank all of the programs that they are willing to attend (71 and 47 percent, respectively).

These observations are in line with our experimental observations suggesting that participants in the DA algorithm may not always report their preferences straightforwardly/truthfully. Furthermore, a substantial number of respondents state that they use the *likelihood of matching* as a guide to submit their rankings. This is consistent with our experimental proposers, whose decision over who to propose to each round is closely related to how each receiver ranks them, where they skip past receivers who do not rank them highly.[40]

A possibly more striking aggregate statistic documented by the NRMP pertains to the percentage of matches with a resident's k'th ranked hospital. For low k the fraction of matches is particularly high, and fairly constant across recent years. For instance, between 1997 and 2013, the percentage of U.S. senior applicants being matched with their first-ranked hospital ranged from 48.8 percent (in 2015) to 59.5 percent (in 2000). The percentages of matches with second-ranked hospitals ranged from 14.2 (in 2005) to 15.6 percent (in 2015), the percentage of matches with third-ranked hospitals ranged from 8.1 percent (in 2004) to

[39]See `nrmp.org/match-data/main-residency-match-data/` for historical results from each year, and the 2013 applicant survey.

[40]Naturally, there might be selection issues pertaining to the type of residents who choose to respond to the NRMP survey, which we cannot control for. Furthermore, due to the privacy policies of the NRMP, we cannot gain access to individual backgrounds of respondents, which could allow us to establish various correlations between the general tendency to report in a particular way and residents' attributes.

9.8 percent (in 2014), etc. The figures for independent applicants are similar, though lower (as around a half remained unmatched each year).[41]

To examine how high these numbers are we conducted simulations assuming independent preferences of both the proposing residents and the receiving hospitals. Using participant numbers and number of hospitals and positions derived from the 2013 match, we examined the fraction of first-, second-, and third-ranked matches assuming truth-telling.[42] In our simulations first- through third-ranked positions account for 9.7, 8.7 and 7.7 percent, respectively. Were we to assume positively correlated preferences, these numbers shrink, as there is more competition for each top-ranked slot. The number of first-ranked matches indicated by the NRMP survey would seem to come from a negative correlation in the stated preferences of the residents, a sorting over who proposers are choosing to rank first. This leads either to a conclusion that underlying preferences have a large negative correlation, or that the stated preferences of the proposing side are different from the underlying preferences.

Roth and Peranson (1999) analysis of NRMP data finds small cores, and little scope for manipulation. Their argument that *actual* preferences exhibit small cores hinges on similar core-spans found through simulations and actual ranking data. Their simulations assume independent preferences (with the argument that the core would be even smaller with positively correlated preferences). However, these assumptions would not explain the high degree of matches between residents and their top-ranked hospital. Our experimental data and the QRE model both indicate that rank-order lists specified by participants might exhibit substantial modification from the underlying preference, which naturally manifests itself through small measured cores *and* a higher preponderance of proposers matched to their *stated*-best receiver.[43]

This proposer-based deviation can produce rankings with smaller cores (reflecting smaller gains from truncation by the receiving side) as well as high fractions of proposers matched to their first choices. The last columns in Table 6 provide information on the core-span within our experimental markets, while the penultimate *Induced* column indicates the core span in each market according to the given payoff matrix (both measured in terms of the difference in rank between the best and worst stable outcome, averaged across the eight participants

[41]In 2015, matches of independent applicants with first-, second- and third-ranked hospitals occurred with frequencies of 28.6, 11.5, and 6.7 percent, respectively. Conditional on matching theses numbers are 49.1, 19.7 and 11.5 percent. Qualitatively similar figures occur in the years previous to 2015.

[42]Our simulations had 36,000 residents proposing to 4,000 hospitals/programs, each with 7 slots. Residents had a uniformly determined preference over 15 hospitals, while hospitals had preferences uniformly determined across 60 residents. (These numbers were chosen to approximate aggregate statistics reported by the NRMP.) We simulated the market outcome 10 times.

[43]Across the 20 different markets in our experiment, 42.0 percent of simulated outcomes under our QRE model are proposers being matching to their *stated best* receiver. Measured according to the true preference, just 16.0 percent of the outcomes are between a proposer and their *actual best* receiver. Under truth-telling by all participants this figure would be 15.6 percent.

on that market side). For instance in market (V), the induced preferences yields a core with an average span of 1.75 for the workers, and an average span of 2 for the firms. The last column, provides information for the same markets using the QRE model, where the core-span measure is calculated by simulation. The QRE core spans are much smaller than the induced level in all but one of our markets with multiple stable matches. In Market (V) the workers' QRE rankings would indicate an average core span of just 0.18 for the workers, and 0.28 for the firms when run as worker-proposing, and 0.83 for workers and 0.8 for firms, when run firm-proposing.

8. Conclusion

The paper reports observations from experiments emulating a highly utilized matching clearinghouse, the deferred-acceptance (DA) mechanism. We studied a large set of markets, varying in their complexity, incentives to straightforwardly reveal preferences, and cardinal representations. Several important insights emerge from our experiments. First, less than half of the markets generated a stable matching. Of those markets with multiple stable matchings that did end at a stable outcome, over 70 percent are at the receiver-best stable matching. Since straightforward revelation of preferences generates the proposer-best outcome, these results are suggestive of manipulation. Our second set of insights regard the source of deviations from straightforward behavior. Proposers frequently skipped down their preference ordering, preferring to propose early to those more likely to accept them. Receivers, however, appeared to by and large behave in an effectively straightforward manner, accepting the best offer at each point in time. This is in contrast to the underlying theoretical predictions that proposers behave straightforwardly and receivers do not. Last, we show that market attributes have a significant impact on outcomes. For instance, both the cardinal representation and core size influence whether outcomes are ultimately stable. They also impact the overall distance of observed outcomes from the core, and the number of turns it takes markets to converge to a final outcome.

The study has potentially important practical implications given the wide use of the DA mechanism. In particular, for the approximately 60,000 participants involved in medical-residency matching each year in the United States. The behavior we observe in the lab might mirror medical residents from top programs applying to top-tier residencies, while those from less well-regarded schools aiming at middle-ranked hospitals and below. Naturally, outcomes are then very fragile to mistakes (by residents) regarding how low to aim with their applications, even if hospitals submit their preferences truthfully. While the centralized system is designed to generate stable matchings, such behavior may cause clearinghouses to converge to outcomes that are, in fact, unstable, and for the data derived from them to look less amenable to manipulation.

In order to test the DA mechanism in the laboratory, we implemented a dynamic version of the mechanism. Nonetheless, there are several aspects of the data that suggest the results may be useful for predicting behavior in the field, where a static version of the mechanism is often used. First, we provide a simple model of behavior in the static mechanism that fits many facets of our data and allows us to make out-of-sample predictions. Second, we compare moments generated by the behavioral model to those available in NRMP field data. That a model of behavior in our experiment is consistent with several stylized facts from the field, suggests our findings might be more widespread.

We also note that our results could provide insights on outcomes of particular decentralized matching processes as well. This is the case for markets in which two conditions hold. First, offers can flow only from one side of the market to the other (say, firms can make offers to workers but not vice versa); Second, repeat offers are impossible or prohibitively costly. In such markets, our results suggest that outcomes may not be stable, and their features depend crucially on particular market characteristics.

The paper opens the door for several directions for future research. First, in light of the behavior we observe, it would be important to formally understand how fragile outcomes are to particular skipping heuristics by proposers. Second, while the limited-friction case studied in this paper is a natural first step for inquiry, and matches much of the extant theoretical literature, it would be important to determine how certain frictions, particularly ones pertaining to incomplete information (regarding others' preferences, as well as one's own), may impact behavior and outcomes in centralized clearinghouses. This may be particularly interesting for larger markets, where complete sharing of private information would require massive amounts of communication. In fact, in large markets with incomplete information, other details of the clearinghouse may play an important role—such as pre-application interviews, that are common in, for example, the NRMP.

References

Calsamiglia, C., G. Haeringer, and F. Klijn (2010): "Constrained School Choice: An Experimental Study," *American Economic Review*, 100(4), 1860–74.

Chen, Y., and T. Sönmez (2006): "School choice: an experimental study," *Journal of Economic Theory*, 127(1), 202–231.

Echenique, F., and L. Yariv (2013): "An Experimental Study of Decentralized Matching," Working Paper, Caltech.

Eriksson, K., and P. Strimling (2009): "Partner search heuristics in the lab: Stability of matchings under various preference structures," *Adaptive Behavior*, 17(6), 524–536.

Featherstone, C., and E. Mayefsky (2011): "Why Do Some Clearinghouses Yield Stable Outcomes? Experimental Evidence on Out-of-Equilibrium Truth-telling," Working

paper, Stanford University.

Featherstone, C., and M. Niederle (2011): "School Choice Mechanisms under Incomplete Information: An Experimental Investigation," Working Paper, Stanford University.

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

Haeringer, G., and M. Wooders (2011): "Decentralized job matching," *International Journal of Game Theory*, 40(1), 1–28.

Harrison, G. W., and K. A. McCabe (1992): "Testing Noncooperative Bargaining Theory in Experiments," *Research in Experimental Economics*, 5, 137–69.

Haruvy, E., and M. U. Ünver (2007): "Equilibrium Selection and the Role of Information in Repeated Matching Markets," *Economics Letters*, 94(2), 284–289.

Hoffman, M., D. Moeller, and R. Paturi (2013): "Jealousy Graphs: Structure and Complexity of Decentralized Stable Matching," in *Web and Internet Economics*, ed. by Y. Chen, and N. Immorlica, vol. 8289 of *Lecture Notes in Computer Science*, pp. 263–276. Springer.

Kagel, J. H., and A. E. Roth (2000): "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by A Natural Experiments*," *Quarterly Journal of Economics*, 115(1), 201–235.

Krishna, A., and M. U. Ünver (2008): "Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies," *Marketing Science*, 27(2), 262–282.

Liu, Q., G. J. Mailath, A. Postlewaite, and L. Samuelson (2014): "Stable Matching with Incomplete Information," *Econometrica*, 82(2), 541–587.

McKelvey, R. D., and T. R. Palfrey (1995): "Quantal response equilibria for normal form games," *Games and Economic Behavior*, 10(1), 6–38.

——— (1998): "Quantal response equilibria for extensive form games," *Experimental Economics*, 1(1), 9–41.

Nalbantian, H. R., and A. Schotter (1995): "Matching and Efficiency in the Baseball Free-Agent System: An Experimental Examination," *Journal of Labor Economics*, 13(1), 1–31.

Niederle, M., and L. Yariv (2011): "Matching Through Decentralized Markets," Working Paper, Caltech.

Pais, J., and Á. Pintér (2008): "School choice and information: an experimental study on matching mechanisms," *Games and Economic Behavior*, 64(1), 303–328.

Pais, J., Á. Pintér, and R. F. Veszteg (2011a): "College Admissions and the Role of Information: An Experimental Study," *International Economic Review*, 52(3), 713–737.

——— (2011b): "Decentralized matching markets: a laboratory experiment," Working Paper, Technical University of Lisbon.

Roth, A. E., and E. Peranson (1997): "The Effects of the Change in the NRMP Matching Algorithm," *JAMA: The Journal of the American Medical Association*, 278(9), 729–732.

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(4), 748–780.

Roth, A. E., and M. Sotomayor (1990): *Two-sided Matching: A Study in Game-Theoretic Modeling and Analysis*, vol. 18 of *Econometric Society Monographs*. Cambridge University Press, Cambridge England.

Wang, S. W., and X. Zhong (2012): "The Five Ws of Preference Manipulations in Centralized Matching Mechanisms: An Experimental Investigation," In preparation, University of Pittsburgh.

Zizzo, D. (2010): "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 13(1), 75–98.