Two-sided matching with interdependent values*

Archishman Chakraborty[†] Baruch College, CUNY

Alessandro Citanna[‡] HEC – Paris

Michael Ostrovsky[§] Stanford University

December 13, 2007

Abstract

We introduce and study two-sided matching with incomplete information and interdependent valuations on one side of the market. An example of such a setting is a matching market between colleges and students in which colleges receive partially informative signals about students. Stability in such markets depends on the amount of information about matchings available to colleges. We show that when colleges observe the entire matching, a stable matching mechanism does not generally exist. When colleges observe only their own matches but not those of other colleges, a stable mechanism exists if students have identical preferences over colleges, but may also fail to exist if students have different preferences.

Keywords: interdependent values, stability, matching.

JEL classification numbers: C78, D82.

*We would like to thank Atila Abdulkadiroglu, Dirk Bergemann, Francoise Forges, Drew Fudenberg, Christine Jolls, John Lazarev, Dan Levin, Paul Milgrom, Muriel Niederle, Ariel Pakes, Parag Pathak, Andrew Postlewaite, Ronnie Razin, Al Roth, Michael Schwarz, Ilya Segal, Tayfun Sonmez, Adam Szeidl, Utku Unver, and Richard Zeckhauser for useful discussions. This paper stems from and adds to results that previously circulated in two separate papers. The paper was completed while the second author was visiting the Graduate School of Business of Columbia University, New York. Archishman Chakraborty and Alessandro Citanna acknowledge the financial support of the NSF, grant #0617850.

[†]archishman_chakraborty@baruch.cuny.edu

[‡]citanna@hec.fr

[§]ostrovsky@gsb.stanford.edu

1 Introduction

The literature on two-sided matching, starting with Gale and Shapley (1962), has assumed that agents on each side of a matching market have enough information to rank agents on the other side. Some of the papers, such as Roth (1989), allow for uncertainty over other agents' preferences, but do not allow for uncertainty over one's own preferences.

Such uncertainty is however quite pervasive. An interview may still leave an employer uncertain about the candidate's aptitude and skills. Letters of recommendation give a college only limited information about a prospective student. Courtship does not necessarily reveal everything about a potential spouse. In all these situations, agents on one or both sides of the market are uncertain about the intrinsic qualities of potential matches, and so about their preferences over these matches.

As a result, agents may be able to infer additional information from the actions of others, leading them to revise their evaluations and, possibly, their own actions. Indeed, top law schools tend not to hire a new professor until he or she has received several offers from other law schools. A college senior who can credibly convey existing job offers to a recruiter finds it easier to get a new offer. An acceptance of a paper for publication by one law journal may make it easier for the author to publish it in a better one.[1] Of course, the reverse is also true: a paper rejected by several journals can be much harder to publish if, for example, the editors can get an estimate of how many journals have previously rejected the paper by how old it is.

This paper analyzes some of the new effects arising in two-sided matching when valuations are interdependent, in the context of student-school market in which school A's estimate of the value of student s depends on the information of school B. In such environments, the matching that obtains depends on the different agents' information about the state of the world. We ask if there exist centralized matching mechanisms that are stable, that is, elicit information truthfully and are also immune to objections and rematching by some participating agents, based on posterior beliefs updated after observing all or part of the realized outcome.[2]

[1] Simultaneous submission of papers to multiple journals is the prevalent practice in law. E.g., the rules for "expedited" review at the University of Chicago Law Review are as follows: "If another journal has offered to publish your article, to request an expedited review, email[...] or call the articles office... Please include the following information: the author's name, the article's title, a contact phone number, the journal making the offer, and the deadline for the expedited review to be completed." (http://lawreview.uchicago.edu/submissions/index.html)

[2] The current paper largely follows the "mechanism design" strand of the two-sided matching literature, which typically tries to find a mechanism implementing a stable outcome or conditions under which a stable matching exists. This is in contrast to several recent papers that take a different approach and study the effect of uncertainty and interdependencies in valuations in specific dynamic matching games. Lee (2004) argues that interdependencies in valuations give rise to an adverse selection problem in college admissions, which in turn makes it optimal for colleges to use different thresholds for regular and early-admission applicants. Nagypál (2004) and Chade, Lewis, and Smith (2007) study the behavior of students in the application process when they are uncertain about their own qualities and applications are costly. Chade (2006) studies stationary equilibria in a dynamic random matching game in which agents get imperfect signals about the qualities of their potential matches and describes a phenomenon he calls "acceptance curse": an acceptance decision by an agent conveys negative information to his or her potential partner. Finally, Hoppe, Moldovanu, and Sela (2005) study equilibria of a matching game in which men and women have private information about their own qualities, send costly signals based on this information, and are matched assortatively based on these signals. In the mechanism design literature, notions such as durable decision rules (Holmstrom and Myerson, 1983) and posterior efficient collective choice (Forges, 1994) bear a conceptual resemblance to our analysis of stable matching mechanisms that are immune to objections based on information revealed by the mechanism itself.

We show that even when student preferences are known, and the incomplete information pertains only to the evaluation of students by schools, stable mechanisms may not exist. In general, existence hinges on two issues: (i) the diversity of student preferences and (ii) the transparency of the mechanism, i.e., what is observed about the outcome of the mechanism by participating agents at the time they may raise objections.[3] Indeed, when student preferences over schools are commonly known but diverse, there does not exist a stable mechanism in general, even when each participating agent observes only the smallest feasible part of the overall matching outcome, i.e., its own match.

In contrast, when preferences on the student side of the market are identical, existence depends on what is observed. When each agent only observes its own match, leading to a notion we call "weak" stability, stable matching mechanisms exist. When the entire realized matching outcome is observed by all agents, a notion we call "strong" stability, stable mechanisms do not generally exist. Strong stability may be thought of as the case where the mechanism designer publicly announces the profile of matches after eliciting reports from the schools, while weak stability corresponds to the case where the designer communicates privately to each agent their own match.

We demonstrate non-existence via non-degenerate examples. We demonstrate existence constructively via a serial dictatorship algorithm under which the match for any school is determined based on the information held only by that school and the schools ranked higher according to the common student preferences. This shows that observability matters for existence in environments with homogeneous student preferences precisely because the mechanism may reveal too much information originally held by lower-ranked schools to the higher-ranked ones. When each school observes the entire profile of matches, and information is to be aggregated truthfully, a higher-ranked school will sometimes be able to infer enough information about the signals of lower ranked schools to object, precluding, in general, the existence of strongly stable mechanisms. But if each school observes only its own match, it is possible to construct mechanisms (namely, serial dictatorships) that manage to elicit truth-telling while at the same time ensuring that the match of any school is independent of information held by lower-ranked schools, regardless of the preferences of the schools and the nature of interdependencies between signals and values. More generally, an upper bound on the amount of observed information about the realized profile of matches consistent with existence is one where each school only observes the realized matches of all schools higher, but not lower, than itself in the student rankings.

Finally, we show that serial dictatorship can be implemented via an extensive form game that is independent of the details of the environment such as preferences and priors on signals and states.[4] In this game, the market clears from the top down, with schools that are considered better by students making offers earlier. This appears to be in broad conformity with the observed behavior

[3] The restriction to one-sided incomplete information environments is necessary for positive existence results given what is known about impossibility of stable mechanisms even when values are independent and private (see Roth, 1989). While matches are observable and potentially informative about the state of the world, we assume that private signals are not publicly observable or verifiable, making necessary the truthful revelation of information.

[4] Wilson (1987) discusses the desirability of creating mechanisms that are independent of the details of the environment as much as possible.

in some labor markets, such as the AEA market for assistant professors in economics and related disciplines.

The remainder of the paper is organized as follows. Section 2 describes the setup. Section 3 introduces definitions of stability in the environment with interdependent valuations. Sections 4, 5, and 6 present our results on existence and non-existence of stable mechanisms in various settings. Section 7 explains why restricting attention to direct revelation mechanisms is without loss of generality in our setting. Section 8 concludes.

2 Model

Consider a one-to-one matching market between students and schools. The set of students is denoted by $\mathbf{S} = \{1, ..., S\}$ with typical element s and the set of schools is denoted by $\mathbf{I} = \{1, .., I\}$ with typical element i.

Each students s has an unobserved quality $q_s \in \mathbf{Q}$, where $\mathbf{Q}$ is a finite set. Quality q_s can encompass a variety of student characteristics, and as we explain below, different schools may have different preferences over student qualities. Neither schools nor students know the quality q_s of any student s, but each school i receives a private signal $x_{i,s} \in \mathbf{X}$, where $\mathbf{X}$ is also a finite set. Let $x_s = (x_{i,s})$ be the vector of signals associated with each student s, let $x_i = (x_{i,s})$ be the vector of signals received by each school i, and let x be the matrix of signals received by all schools about all students. Let Pr be the joint probability distribution over signals and qualities that is strictly positive for all x.[5]

A two-sided one-to-one *matching* m is a function from $\mathbf{S} \cup \mathbf{I}$ to itself such that: (i) for any agent a, $m(m(a)) = a$; (ii) if $m(i) \neq i$ for $i \in \mathbf{I}$, then $m(i) \in \mathbf{S}$; and (iii) if $m(s) \neq s$ for $s \in \mathbf{S}$, then $m(s) \in \mathbf{I}$. In other words, the set of schools and students is broken into school-student pairs (i, s) for whom $m(s) = i$ and $m(i) = s$ and unmatched agents a for whom $m(a) = a$. Let $\mathbf{M}$ be the set of all matchings.

Students have preferences over their matches. These preferences are publicly known: student s values a match to school i at $v_{s,i}$. We normalize the value to a student from staying unmatched to zero and assume that students' preferences are strict: for any student s and pair of schools i and i', $v_{s,i} \neq v_{s,i'}$ and $v_{s,i} \neq 0$. If $v_{s,i} < 0$, school i is unacceptable for student s; otherwise it is acceptable.

The payoffs of schools depend on students' qualities and signals: school i derives utility $w_{i,s}(x_s, q_s)$ from matching with student s if schools receive signals x_s about the student and the student's quality is q_s. We normalize the value to a school from staying unmatched to zero.

Schools and students have von Neumann-Morgenstern preferences. In particular, school i's expected utility from matching with student s conditional on signal x_s and additional information y is equal to $u_{i,s}(x_s, y) = \sum_{q_s} w_{i,s}(x_s, q_s) Pr(q_s|x_s, y)$.

[5] Notice that we allow informational spillovers by letting signals and qualities to be correlated not only across schools but also across students. The full support assumption is made only to simplify proofs.

3 Stability

The classic concept of stability under private values is defined on matchings themselves: an agent can "verify" stability by checking that his current match is acceptable and by making offers to the agents on the other side of the market preferable to his current match. Thus, a mechanism used to arrive at a matching does not influence whether the matching is stable or not. In contrast, with interdependent values, schools will update their beliefs about student qualities differently under different mechanisms (and under different amounts of information available to the schools), and so a particular matching may be stable after one mechanism and unstable after another. Therefore, we need to define the concept of stability on matching mechanisms, not on matchings themselves, and take into account the amount of information available to the schools.

3.1 Definitions

We begin by formalizing the concept of a "matching mechanism." We define it as a centralized direct revelation mechanism, in which schools report their signals to the mechanism designer and the designer then proposes who should be matched with whom. More formally, a *direct revelation matching mechanism* μ is a function from the set $\mathbf{X}^{SI}$ of reported signal profiles to the set $\Delta(\mathbf{M})$ of probability distributions over matchings. That is, mechanism μ takes as arguments signals reported by schools and returns a matching, possibly stochastically. We denote by $\mu(m; x)$ the probability that the mechanism returns matching m given the profile of signals x reported by schools.

We now turn to our solution concepts: weak and strong stability. Under the classic definition of pairwise stability in the setting with private values, matching m is called stable if no agent wants to drop his match and no school-student pair wants to match with each other (instead of the partners assigned to them under matching m) even if they have an opportunity to do so. In our setting, we define stability in a similar way, on matching mechanisms, taking into account how much information is available to the schools.

Specifically, let μ be a matching mechanism. Take any student s and school i. Define the extensive-form game $\Gamma_w(\mu, s, i)$ as follows. The players are all students in $\mathbf{S}$ and all schools in $\mathbf{I}$. The game consists of the following stages:

1. Nature selects the qualities q of students and the matrix x of signals about the students observed by the schools, according to the commonly known distribution Pr.

2. Schools report their signals, $\hat{x}$, to the mechanism.

3. Based on the reports, the mechanism, μ, generates a matching, m.

4. Each player a observes his assigned match, $m(a)$.

5. At this stage, school i chooses one of several actions. If $m(i) \neq i$, school i can choose to unilaterally drop its assigned match. Alternatively, if school i and student s are unmatched,

i.e., $m(i) \neq s$, school i can choose to make an offer to student s. Finally, the school can choose not to take either of these two actions, keeping its assigned match.

6. If in the previous stage the school made an offer to student s, that student can either accept or reject the offer.

The payoffs in game $\Gamma_w(\mu, s, i)$ are as follows. If the school did not drop its assigned match or did not make an offer, or it did make an offer but the offer was rejected by the student, each agent's payoff is equal to its utility from matching with its assigned partner. If the offer was made and accepted, then school i gets utility $w_{i,s}(x_s, q_s)$ and student s gets utility $v_{s,i}$ from matching with each other; their assigned partners, $m(s)$ and $m(i)$, get utility 0 from staying unmatched, and all other agents' payoffs are equal to their utilities from matching with their assigned partners. If school i dropped its assigned match, then both i and its assigned match, $m(i)$, get payoff 0 from staying unmatched and all other agents' payoffs are equal to their utilities from matching with their assigned partners.

Game $\Gamma_w(\mu, s, i)$ will be used to define the concept of weak stability: schools observe only their own matches. For strong stability, we define game $\Gamma_s(\mu, s, i)$. It is the same as $\Gamma_w(\mu, s, i)$, except for stage 4:

4^s. Each player a observes the entire matching m generated by the mechanism.

All other stages are identical to those in game $\Gamma_w(\mu, s, i)$.

We are now ready to define our concepts of stability.

Definition 1 *Matching mechanism μ is weakly stable if it never matches a student to an unacceptable school and for any student s and school i, there exists a Perfect Bayesian Equilibrium of game $\Gamma_w(\mu, s, i)$ in which all schools report their signals truthfully and school i does not make an offer to student s and does not drop its assigned match on the equilibrium path.*

Definition 2 *Matching mechanism μ is strongly stable if it never matches a student to an unacceptable school and for any student s and school i, there exists a Perfect Bayesian Equilibrium of game $\Gamma_s(\mu, s, i)$ in which all schools report their signals truthfully and school i does not make an offer to student s and does not drop its assigned match on the equilibrium path.*

In essence, our notions of stability embed three requirements. The first is incentive compatibility: schools do not have an incentive to lie about their signals. The second is no rematching after truthful information revelation: no school can report its signals truthfully and then profitably rematch after the mechanism proposes a particular matching (assuming other schools report signals truthfully as well). The third requirement is a combination of the first two: no school can benefit by first misrepresenting its signals and then rematching.

The third requirement is a necessary part of an internally consistent definition of stability: if players can misrepresent their signals and are also allowed to rematch, they can also anticipate the possibility of rematching when misreporting their type. Such "anticipated rematching" may in turn affect their incentives to reveal their signals truthfully. In other words, once we move away from the framework of classical matching and introduce interdependent values and updating of beliefs based on the mechanism output, all three requirements must be included in a definition of stability.

The possibility of anticipated rematching, in particular, makes the details of the extensive form game representing the matching process important, as it determines the amount of information available to agents at the rematching stage and does not give the mechanism designer the final say about the realized allocation.[6] As a result, one cannot immediately invoke the revelation principle and the results of Myerson (1979, 1986) on the equivalence of direct- and indirect-revelation mechanisms. Nevertheless, as we discuss in Section 7, a slight modification of these arguments does in fact allow us to conclude that the restriction to direct revelation mechanisms in our setting is without loss of generality.

The rematching stages of Γ_w and Γ_s are also quite specific. As alternatives, we could have included uncertainty over who can rematch and with whom, allowed multiple schools to rematch in an exogenous or endogenous order, or considered multiple rounds of rematching. These different definitions would, in principle, produce different sets of stable mechanisms. We feel, however, that our definition provides the best starting point for the analysis of stability under incomplete information, because it focuses on the simplest, most basic deviations: the ones involving at most one pair of agents and at most one chance at rematching.[7]

3.2 Information availability and stability

We now state and prove our first result, which shows that transparency is a key determinant of stability in environments with interdependent values. All proofs are in Appendix A.

Proposition 1 *Any strongly stable matching mechanism μ is also weakly stable.*

While the proof contains some important technical details, the key intuition behind the result is straightforward. In fact, a more general result is true: if a matching mechanism is stable under some amount of information, it will also be stable under any coarser amount. Indeed, consider an event E corresponding to the coarser information regime, consider its partition into events E_1, E_2, etc., in the finer information regime, and consider the same mechanism in the two information regimes. If a school has an incentive to rematch or drop its assigned match after observing E in the coarser information regime, then such an action must also be profitable after observing at least one of the finer events E_k, making the mechanism unstable in the finer information regime as well.

[6] For example, a uniform random matching mechanism that ignores all reports and assigns every student to every school with equal probabilities is not necessarily stable.

[7] This approach follows in spirit that of the original definition of stability in two-sided matching markets: In a matching, a man-woman pair m_1 and w_1 decide whether to marry each other and drop their assigned spouses as if they were the only ones allowed to rematch and no further rematchings were allowed. They do not, for instance, consider the possibility that if they marry each other, their spouses will then go and marry someone else, those people's spouses will then go on and steal someone else's spouses, and so on, and eventually the chain of rematchings will come back to hurt m_1 or w_1. Likewise, a woman does not consider unilaterally dropping an acceptable husband hoping that the subsequent chain of rematchings will allow her to get a better one.

4 Strong stability

We now turn to exploring conditions under which stable mechanisms exist. We begin our analysis by studying the more restrictive solution concept—strong stability. We show that in general, strongly stable mechanisms may fail to exist, even if we assume that the preferences of students over schools are identical. We show this by presenting two examples of settings for which no such mechanism exists. In the first example, we show that there is a whole class of matching markets for which no such mechanism exists: markets with two schools, three students, and two quality levels and signals. In the second example, we show that a stable mechanism would not generally exist even under a somewhat relaxed solution concept that only allows a school to either misrepresent its signal or rematch, but not both.

Example 1

Consider a market with two schools, 1 and 2, and three students, s_1, s_2, and s_3. Each student is either low type (quality $q_l > 0$), with probability θ, or high type ($q_h > q_l$), with probability $1 - \theta$. If the student is high type, each school gets signal H about him with probability p_h, and signal L with probability $1 - p_h$. If the student is low type, each school gets signal H with probability p_l, and L with probability $1 - p_l$, where $0 < \theta$, p_l, $p_h < 1$ and $p_l < p_h$. Students like school 1 more than school 2: for any s, $v_{s,1} > v_{s,2}$. Schools' preferences are identical, with $w_{i,s}(x_s, q_s) = q_s$. Note that the information structure is such that a student with two H signals is in expectation strictly more valuable to a school than a student with one H and one L signals, who in turn is strictly more valuable than a student with two L signals. Note also that every student is acceptable.

Claim 1 *The above matching market does not have a strongly stable mechanism.*

The key idea behind the proof is the tension that arises when there is exactly one student with two H signals: assigning this student to the more desirable school, 1, gives the less desirable school, 2, incentives to conceal its information in the reporting stage, in the hopes of obtaining the student later. Assigning this student to school 2 gives to school 1 incentives to steal school 2's assigned match. As the details of the proof show, the possibility of assigning the student stochastically does not help achieve stability.

The proof of Claim 1 makes use of the fact that a school may first misrepresent its signals and then make a rematching offer to a student. A natural question is whether this is necessary: would a stable matching mechanism exist if a school was only allowed to take one of these two deviations. While we do not know the answer to this question for the class of matching markets described in Example 1, we can show that at least for some markets, such a mechanism still does not exist, and hence there would be no general stable mechanism even under the weaker definition. The following example presents such a market.

Example 2

We take an environment with one student s and two schools 1 and 2, with two signals for each school $\mathbf{X}_1 = \{T, B\}$, $\mathbf{X}_2 = \{L, R\}$. The following table provides the payoffs for the two schools from matching with the student for each signal profile, with school 1 (row school) listed first:

	L	R
T	$-1/2, 1$	$1, 1$
B	$-9/8, -1/2$	$1, 1$

The student prefers school 1 to school 2 and the joint distribution of signals is i.i.d. uniform.[8]

Claim 2 *The above matching market does not have a strongly stable mechanism.*

In Example 2, strong stability demands that the matching mechanisms be nonmonotonic in the lower school's signals, creating an unresolvable tension with incentive compatibility. Indeed, since the arguments underlying Example 2 rely only on the inconsistency of truth-telling with no rematching after truth-telling, this non-existence result carries over to milder definitions of stability that do not include the "anticipated rematching" feature.

Finally, even though Examples 1 and 2 are robust and can be embedded into larger matching markets, they do not exclude the possibility that strongly stable mechanisms exist in some non-degenerate markets. Indeed, Example 4 in Appendix B presents a matching market with a generic information structure for which a stable matching mechanism does exist and a non-trivial amount of information is exchanged by the schools.

5 Weak stability with heterogeneous student preferences

As the results of the previous section show, there does not exist a strongly stable matching mechanism, even under restrictive assumptions on the preferences of students. We turn now to studying a less demanding solution concept: weak stability. In this section, we show that when students' preferences over schools differ, even a weakly stable matching mechanism may not exist. As we show in the next section, when students have identical preferences, such a mechanism does exist.

The intuition behind the next example is simple. A school that is low in the ranking of its most preferred student (say, student 1) may lie in order to mislead a school higher in that student's rankings that student 1 is not worth getting, but another one is. Since any stable mechanism must assign a student to the top-ranked available school if the school thinks that student is best, and since student rankings are not identical, there is room to change the available set of matches through lying.

[8]The example is numerical for expositional convenience. Its underlying logic extends to generic payoffs and distributions.

Example 3

Consider an environment with two schools, 1 and 2, and two students, s_1 and s_2. Students' preferences $v_{s,i}$ are such that $v_{s_1,1} > v_{s_1,2} > 0$ and $v_{s_2,2} > v_{s_2,1} > 0$.

There is no uncertainty about student s_1's quality. Student s_2's quality is equal to $q_2 \in \{-2, 2\}$, with probability $\frac{1}{2}$ each. School 1's signal about q_2 is uninformative. School 2's signal about q_2 is perfectly informative; it is equal to q_2 with probability 1. Schools' preferences are summarized as follows:

$$\begin{aligned} w_{1,s_1}(x_{s_1}, q_{s_1}) &= 1, \\ w_{1,s_2}(x_{s_2}, q_{s_2}) &= -q_2, \\ w_{2,s_1}(x_{s_1}, q_{s_1}) &= 3, \\ w_{2,s_2}(x_{s_2}, q_{s_2}) &= q_2. \end{aligned}$$

Claim 3 *The above matching market does not have a weakly stable mechanism.*

In fact, just like in Claim 2, the proof of Claim 3 only uses rematching or misreporting and does not rely on deviations involving both. Hence, the result would continue to hold even under a weaker definition of stability.

6 Weak stability with homogeneous student preferences

Section 5 shows that with heterogeneous student preferences, stable mechanisms do not generally exist. We now turn to matching markets in which the preferences of all students are identical: they all agree which school is the most desirable, the second most desirable, and so on. Formally, for any students s and s' and school i, $v_{s,i} = v_{s',i} \equiv v_i$. School preferences are as before. We focus on weak stability since, as shown before, if agents observe the entire matching (rather than their own match), there is no mechanism that always produces a stable matching even with identical preferences of students over schools.

It turns out that in this setting, weakly stable mechanisms do exist. Indeed, consider the following mechanism, based on a simple *serial dictatorship* (SD) algorithm. Let i_1 be the top-ranked school. Compute each student's value to school i_1 based on the school's own signal, and assign to school i_1 the student with the highest value (s_1). For the second-ranked school (i_2), compute the value of each remaining student based on the school's own signal as well as on the observed match of i_1. Assign to school i_2 the student with the highest value (s_2). For the third-ranked school, i_3, compute the value of each remaining student based school i_3's own signal and on the fact that student s_1 went to school i_1 and s_2 to school i_2. Proceeding in this fashion, we assign a match to each school. We denote the resulting mechanism as μ^{SD}.[9]

[9]Note that there are other ways to implement analogues of serial dictatorship in this environment, requiring

Theorem 4 *Under homogeneous student preferences, mechanism μ^{SD} is weakly stable.*

The intuition behind this result is that the only way a school may profit from lying and/or rematching is to change its set of available students or to get information from schools below them in a student's rankings via the observed outcome. Under serial dictatorship, however, the (observed) matching outcome for any school does not depend on the reports sent by schools ranked lower, implying also that a school cannot expand its available set of students by misreporting.

The arguments underlying Theorem 4 also make precise the degree of transparency under which stable mechanisms generally exist. In particular, if schools observe the profile of matches for lower ranked schools, a higher ranked school will typically be able to infer information held by lower ranked schools from observing the matching outcome of lower ranked schools. It follows that existence is guaranteed in general, if each school observes at most the matches of all higher (but not lower) ranked schools.

Note also that in many plausible situations, the information held by lower ranked schools may not be relevant for the evaluation of students by higher ranked schools. For instance, higher-ranked graduate schools may be better at evaluating the suitability of a student for graduate school. In such cases, even if each school observes the entire profile of matches (including those of lower ranked schools), stability will not be an issue. In this sense, the proof of Theorem 4 also identifies economically interesting environments for which strongly stable mechanisms will exist.

6.1 Detail-free implementation

In Theorem 4 we implemented a weakly stable matching via a centralized direct mechanism. A mediator implementing this mechanism needs to be able to understand the various signals that schools can receive about students, compute expected payoffs to schools from students based on these signals, and so on. We will now present a simple way to implement this mechanism in a detail-free way. The only information that will be required to run the mechanism is the ranking of schools by the students; if this ranking is not known to the mediator (but he knows that such a ranking exists), before running the mechanism he can simply solicit this ranking from the students, as long as there are at least three students in the market.[10]

The game proceeds in rounds, $t = 1, 2, ..., I$. In each round t, school i_t is allowed to make an offer to at most one student; it is also allowed to make no offers (recall that i_1 is the most desirable school, i_2 is the second most desirable, and so on). The student to whom the offer was made can

less information. For instance, in assigning matches to low-ranked schools, the mechanism can condition only on which set of students was matched to higher-ranked schools without specifying which student went to which school. Theorem 4 would continue to hold under this formulation, although the resulting matching will not in general be the same. We use this particular version of serial dictatorship because it can be implemented using the game described in Section 6.1.

[10]The mediator can use the ranking reported by the majority of students as the true one, if such a ranking exists, and an arbitrary ranking otherwise. Then with three or more students present, truthful reporting by all students will be an equilibrium, since one student cannot change the behavior of the mediator by misreporting.

accept it or reject. After the round is over, schools $i_{t+1}, \dots, i_I$ are told which match, if any, was formed in round t, and the mechanism then proceeds to stage $t+1$.

The following profile of beliefs and strategies constitutes a Perfect Bayesian Equilibrium of this game and implements the serial dictatorship matching outcome. Each school, having observed what happened in the previous stages, updates its beliefs on the signals of other schools using Bayes' rule whenever possible, finds the student with the highest expected payoff according to these updated beliefs, and makes an offer to that student (unless the payoff from matching with that student is negative, in which case the school makes no offer). If what a school observes in the previous stages has zero probability according to its priors (i.e., someone has deviated from the prescribed strategies), it takes the largest t for which the matches of schools i_1 through i_t have a positive probability according to its priors and updates the beliefs based only on the matches of schools i_1 through i_t, ignoring the subsequent matches.[11] Having computed these beliefs, the school makes the offer to the student with the highest expected payoff, provided that payoff is positive. Students always accept offers from acceptable schools and reject offers from unacceptable ones.

It is easy to check that no school or student has an incentive to deviate from the above strategies, and that if they follow them, the outcome will be the same as in direct revelation mechanism μ^{SD}. Moreover, even if the game was augmented by one rematching stage, in the same way as the direct revelation mechanism was in the previous sections, the profile of strategies above and no rematching on the equilibrium path would still constitute an equilibrium.

The construction above presents just one detail-free game that implements the serial dictatorship outcome in a perfect Bayesian equilibrium. In this game, better schools make earlier offers. We find this to be in broad conformity at least with anecdotal evidence from some observed matching markets. There may of course exist other stable mechanisms implementable via different detail-free extensive games. We leave a more detailed characterization of such rules and games for future research.

7 Indirect Mechanisms

Most of the results in this paper use the definitions of stability for centralized direct revelation mechanisms. The only exception is Section 6.1, which shows how a particular direct mechanism can be implemented in a detail-free way. This raises a natural question: would the set of implementable outcomes become larger if in addition to direct-revelation mechanisms we considered a wider class of games and mechanisms, involving multiple stages, communication between players, and so on? In this section we argue that restricting attention to direct revelation mechanisms is in fact without loss of generality. The reasoning is along the lines of the standard revelation principle arguments (Myerson, 1979, 1986), and so we omit formal details.

For concreteness, we will discuss the concept of strong stability. The discussion of weak stability would be completely analogous. Consider a matching market for which there does not exist a

[11] Any other internally consistent way of updating beliefs after zero-probability events would also work.

strongly stable direct revelation mechanism. Can we achieve stability with an indirect mechanism?

First, we need to specify what we mean by an indirect mechanism and by its stability. We say that an indirect mechanism is any finite extensive form game F in which the actions available to players are independent of their types and at the end of which, at each node, every agent is matched to someone (possibly himself) and observes the matches of everyone. Of course, during the course of play an agent may have observed some additional information as well, including his own actions. Crucially, however, the information partition at the end of this game is always at least as fine as what the final matching is.

An indirect mechanism is stable if there exists a profile of strategies for all players, σ, such that for any game $\Gamma(\cdot)$ extending F with a single rematching stage (just like in the direct revelation mechanism case), there exists a Perfect Bayesian Equilibrium in which first all players play according to σ and then never rematch on the equilibrium path. Thus, we can think of this indirect mechanism as implementing a matching according to game F and profile of strategies σ in a stable way.

Now suppose there is a stable indirect mechanism (game F with a corresponding profile of strategies σ) for a particular matching market. We will now argue that there also exists a stable direct revelation mechanism for this market as well.

Consider the following direct revelation mechanism *with messaging*, ν: (i) schools observe their signals and report them to the central mediator; (ii) the mediator, in the background, runs game F in accordance with these reports and strategies σ; (iii) the mediator outputs to the schools and students their final matches *and also all the extra information that they would have observed along the way if they played the game themselves.*

Note that this mechanism ν is stable, since any feasible lying and/or rematching deviation in this mechanism would have been feasible in the original indirect game as well, where an agent can simply "pretend" that he received different signals and "imitate" someone of a different type.

Now consider a direct revelation mechanism, μ, which maps reported signals to matchings with the same probabilities as mechanism ν, but at the end outputs to the schools and students only the final matching, *without* all the additional information that they would have observed along the path of play in game F. In this mechanism, the information sets of players are coarser than those under mechanism ν, and mechanism ν is stable, so by the arguments underlying Proposition 1, mechanism μ is stable as well.

8 Conclusion

In this paper, we introduced and studied a notion of stability for matching markets in which agents on one side of the market have imperfect information and interdependent values over the agents on the other side of the market. Our results suggest that when the entire matching outcome is observable, stability is hard to obtain in general, unless of course full enforcement of the outcome is available or higher ranked schools are ranked higher precisely because they are better able to identify suitable students. In such cases, the information held by lower ranked schools is unlikely to matter for evaluations by higher ranked schools, and stable rules will exist even if most or all of the matching profile is observable at the rematching stage. Similarly, if a journal is more prestigious precisely because it is better able (in the sense of a sufficient statistic) to evaluate the quality of a paper relative to less prestigious journals, instability is unlikely to be an issue in the journal submission game. On the other hand, when lower-ranked schools have information that may be valuable to higher-ranked ones, the question of stability hinges crucially on features of the mechanisms by which agents are matched, in particular, their transparency.

We considered the effects of interdependent values when matching is one-to-one, though the conclusions we obtain can be generalized to many-to-many matching problems. Also, we focused on the notion of pairwise stability, allowing schools and students to object to a proposed assignment in pairs or alone. While pairwise stability and various notions of group stability may not necessarily coincide here, the restriction to pairwise blocks is natural in the matching context as argued, among others, by Roth and Sotomayor (1990) and Crawford (1991).[12] The study of stability in matching markets under various notions of group stability goes beyond the scope of the present paper, but is one possible avenue for future research.

[12] For example, Roth and Sotomayor (1990, p. 156): "..., identifying and organizing large coalitions may be more difficult than making private arrangements between two parties, and the experience of those regional markets in the United Kingdom that are built around stable mechanisms suggests that pairwise stability is still of primary importance in these markets." See also Crawford (1991, p. 394).

Appendix A: Proofs

Proof of Proposition 1

Suppose mechanism μ is strongly stable, and consider some i', and s'. By strong stability, there exists a Perfect Bayesian Equilibrium of game $\Gamma_s(\mu, s', i')$ with truthtelling and no rematching on the equilibrium path of the game. We will now show that the same is then true about game $\Gamma_w(\mu, s', i')$.

To see this, note that any second-stage information set for i' in Γ_s is fully specified by his signal $x_{i'}$, his report $\widehat{x}_{i'}$, and the observed profile of matches $\{m(i)\}_{i\in\mathbf{I}}$, while in Γ_w this information set is fully determined by $x_{i'}$, $\widehat{x}_{i'}$ and the school's own match $m(i')$. Crucially, given μ, $x_{i'}$, and $\widehat{x}_{i'}$, the second-stage information of i' in Γ_s is always finer than in Γ_w.

Also, by the full support assumption on signals and qualities, the second-stage beliefs of i' in both games are uniquely pinned down by Bayes' Rule, and for any set of reports $\widehat{x}$ and for any output of the matching mechanism given these reports, the outcome observed by i' is consistent with the assumption of truthtelling by other players. In other words, i' will never observe zero-probability events if he believes that other schools are reporting their signals truthfully.

Consider now the following set of strategies in game $\Gamma_w(\mu, s', i')$. Student s' accepts an offer from school i' if and only if he prefers it to his assigned match. All schools always report signals truthfully. School i' never offers student s' to rematch and never drops its assigned student if it reported its signal truthfully in the first stage; if it misreported in the first stage, then in the rematching stage it computes the expected values of keeping its student and making rematching offers to all students (taking into account all available information and the fact that if those students accept, that reveals some information about their assigned matches and thus potentially about their values) and decides whether to keep the assigned student, drop him, or try to rematch with someone else based on that calculation. Let us show that this profile of strategies is a PBE of $\Gamma_w(\mu, s', i')$.

First, note that if i' has sent a truthful report $\widehat{x}_{i'} = x_{i'}$ in the first stage of Γ_w, he cannot find it strictly profitable to object at the second stage at any information set corresponding to his own observed match $m(i')$. For if he could, since his beliefs are pinned down by Bayes' Rule, by the law of iterated expectations and the inclusion property of information sets in Γ_w relative to Γ_s established above, there must also exist a profitable second stage deviation for i' in Γ_s, given the same truthful report $\widehat{x}_{i'} = x_{i'}$, after observing some profile of matches $\{m'(i)\}_{i\in\mathbf{I}}$ satisfying $m'(i') = m(i')$. This shows that there will be no blocking by i' after a truthful report, given that he expects others to report truthfully.

Now consider the first stage reporting strategy for i'. If for some realized signal $x_{i'}$, i' finds it strictly profitable in Γ_w to send a false report in the first stage and then possibly block in the second stage in a sequentially rational manner as a function of his beliefs given the observed $m(i)$, then the same deviation is also feasible for i' in Γ_s, using again the inclusion property on information sets in Γ_w relative to Γ_s. Furthermore, such a deviation is also strictly profitable in interim expected terms at the first stage of Γ_s, a contradiction with the hypothesis that truthtelling by all players and no rematching on the equilibrium path was a part of a PBE of Γ_s.

Finally, consider the reporting strategy of $i \neq i'$. If for some realized signal x_i, school i has a strictly profitable deviation $\widehat{x}_i \neq x_i$ in Γ_w, then the same deviation is profitable in Γ_s, a contradiction. Note that we have used the fact that i' never objects after any report by i in Γ_w or Γ_s, which follows from the assumption that he does not object after the truthful report and the full support assumption on signals.

Proof of Claim 1

We divide the argument into several steps. Suppose μ is a strongly stable mechanism for the market described in Example 1.

Step 1. *Suppose school 1 observes one H (say, s_1) and two Ls (s_2 and s_3). Then μ can match school 1 with s_2 or s_3 with positive probability only if school 2's signal about s_1 is L and its signal about school 1's match is H.*

To see that, notice that school 1 cannot become worse off by rematching with s_1 (since s_1 has *at least* one H, while each of the remaining students has *at most* one H), and so for it to be willing not to rematch, the probability of benefiting from rematching has to be zero. Therefore, μ cannot match school 1 with, say, s_2 if student s_1 is better, i.e., school 2's signal about s_1 is H, or school 2's signal about s_2 is L.

Step 2. *If school 2 is matched to a student about whom its signal is L (say, s_1), while a student about whom its signal is H (say, s_2) remains unmatched, it has to be the case that school 1's signal about s_1 is H and its signal about s_2 is L.*

The logic is similar to that of Step 1: school 2 cannot lose by rematching with s_2, and so to have no incentive to rematch it has to be sure that the probability of benefiting by rematching is zero. The presence of an unmatched student is necessary, so we need $n \geq 3$.

Step 3. *Suppose school 1 observes two Hs (say, s_1 and s_2) and one L (s_3). Then μ must assign a student with two H signals to school 1, whenever there is such a student.*

To see that, notice that school 1 can guarantee itself a student with two H signals, if there is one. Indeed, school 1 can lie by "inverting" all its signals and reporting $\hat{x}_1 = (L, L, H)$. If it subsequently gets matched with s_1 or s_2, by Step 1 it knows that school 2 has observed H about that student, who is therefore an HH one. If school 1 gets matched to s_3, it should rematch with the student matched to school 2: by Step 2, school 2's signal about that student is H, unless school 2's signals about both s_1 and s_2 are L, in which case there are no HH students anyway. Note that the argument in this step hinges upon the fact that there are three students in the market.

Step 4. Now suppose school 2 observes two Hs (say, s_1 and s_2) and one L (s_3): $x_2 = (H, H, L)$. By Steps 2 and 3, under truthful reporting with no rematching, school 2 gets matched to an HL student when there are less than two HH students, and to an HH student when there are two. But then school 2 can strictly increase its payoff by the following deviation: it should first "invert" the signals, reporting $\hat{x}_2 = (L, L, H)$; then (i) if it is matched to s_3, it should rematch with the unassigned student, (ii) if school 1 is matched to s_3, school 2 should rematch with the unassigned student with probability 1/2 and keep its assigned match also with probability 1/2, and (iii) if s_3 remains unassigned, school 2 should not rematch.

The following table shows the different scenarios arising for each of school 1's possible signals, depending on which students mechanism μ assigns to the schools.

Signals x_1	Match of school 2
HHH	HH
HHL	HH
HLH	HL with prob. 1/2, HH with prob. 1/2
LHH	HL with prob. 1/2, HH with prob. 1/2
HLL	HL or HH
LHL	HL or HH
LLH	HL
LLL	HL

As the table shows, under this deviation, school 2 always gets matched to an HL or HH student, and it gets matched to an HH student whenever there are two of them. Therefore, this strategy is at least as good as truthful reporting without rematching. But in addition, under this deviation, school 2 sometimes gets matched to an HH student with strictly positive probability (1/2) when there is exactly one such student (e.g., when $x_1 = HLH$ or $x_1 = LHH$), which is a strict improvement over truthful reporting and no rematching.

Proof of Claim 2

For convenience we reproduce the payoff table and recall that the signal distribution is i.i.d. uniform:

	L	R
T	$-1/2, 1$	$1, 1$
B	$-9/8, -1/2$	$1, 1$

We summarize a mechanism by four probability distributions over the student's matches, one for each signal profile, as depicted by the table below:

	L	R
T	α	β
B	γ	δ

Thus $\alpha = (\alpha_s, \alpha_1, \alpha_2)$, with α_i equal to the probability that $m(s) = i \in \{s, 1, 2\}$ for (reported) signals T for school 1 and L for school 2, etc. In what follows, by objections we mean only those that may occur on the path of play (i.e., after truth-telling), and derive necessary properties of any stable mechanism.

Step 1. *School 2 cannot match with the student with positive probability* (i.e., $\alpha_2 = \beta_2 = \gamma_2 =$

$\delta_2 = 0$).

Notice first that if either α_2 or β_2 is strictly positive, we must have $-\frac{1}{2}\alpha_2 + \beta_2 \leq 0$, since otherwise school 1 will object after updating its beliefs when it observes $m(s) = 2$ and $x_1 = T$. Similarly, if either γ_2 or δ_2 is strictly positive, we must have $-\frac{9}{8}\gamma_2 + \delta_2 \leq 0$, otherwise school 1 will object upon observing $m(s) = 2$ when $x_1 = B$. If, contrary to the claim, school 2 obtains the student with positive probability for some signal profile then either α_2 or β_2 is strictly positive or γ_2 or δ_2 is strictly positive, allowing us to conclude that

$$\beta_2 + \delta_2 \leq \frac{1}{2}\alpha_2 + \frac{9}{8}\gamma_2$$

On the other hand, for school 2 to reveal its signal truthfully when $x_2 = R$ we need

$$\beta_2 + \delta_2 \geq \alpha_2 + \gamma_2$$

Combining the last two inequalities we have $\frac{1}{8}\gamma_2 \geq \frac{1}{2}\alpha_2$. But, if either α_2 or γ_2 is strictly positive, to prevent school 2 from objecting when $x_2 = L$ and it has been matched with the student, we need $\alpha_2 - \frac{1}{2}\gamma_2 \geq 0$. It follows that $\alpha_2 = \gamma_2 = 0$, implying in turn that $\beta_2 = \delta_2 = 0$.

Step 2. *The student cannot remain unmatched with positive probability* (i.e., $\alpha_s = \beta_s = \gamma_s = \delta_s = 0$).

When $x_2 = R$, to prevent school 2 from objecting after observing that the student is unmatched, we need $\beta_s = \delta_s = 0$. Since $\beta_2 = \delta_2 = 0$ from step 1, we have $\beta_1 = \delta_1 = 1$. Using this in school 1's truth-telling constraints when $x_1 = T$ and $x_1 = B$ respectively,

$$-\frac{1}{2}\alpha_1 + \beta_1 \geq -\frac{1}{2}\gamma_1 + \delta_1$$

$$-\frac{9}{8}\gamma_1 + \delta_1 \geq -\frac{9}{8}\alpha_1 + \beta_1$$

we see that $\alpha_1 = \gamma_1$ and so $\alpha_s = \gamma_s$, since $\alpha_2 = \gamma_2 = 0$ from step 1. However, to prevent school 2 from objecting $x_2 = L$ after observing that the student is unmatched, we must also have $\alpha_s - \frac{1}{2}\gamma_s \leq 0$, whenever either α_s or γ_s is strictly positive. It follows that $\alpha_s = \gamma_s = 0$.

Step 3. *A stable mechanism does not exist.*

From the previous two steps we see that $\alpha_1 = \beta_1 = \gamma_1 = \delta_1 = 1$. But school 1 will object to being matched with s for sure when $x_1 = B$.

Proof of Claim 3

We identify $x = x_{2,s_2}$, without loss of generality, since x_1 and x_{2,s_1} are uninformative. There are two possible signals reported to the mechanism, $x = 2$ and $x = -2$, and seven possible matchings $(m(1), m(2))$ the mechanism can output: (s_1, s_2), $(s_1, 2)$; (s_2, s_1), $(s_2, 2)$; $(1, s_1)$, $(1, s_2)$, and $(1, 2)$. Thus, any mechanism can be represented by two 3x3 matrices of numbers between 0 and 1, each representing the probability of a particular matching given a particular signal; all but the last of the diagonal elements are zero —the outcomes (s_k, s_k) for $k = 1, 2$ are impossible. The proof proceeds by repeatedly using stability arguments to show that all of these numbers have to be equal to zero (i.e., "crossing out" the corresponding matching-signal pairs), a contradiction.

Since school 1's payoff from matching with student s_1, $w_{1,s_1}(x_{1,s_1}, q_{s_1})$, is always equal to 1 and its payoff from staying unmatched is equal to zero, and since student s_1 prefers school 1 to school 2, a stable mechanism cannot leave school 1 unmatched, i.e., any stable mechanism must have $\mu(m; x) = 0$ if $m(1) = 1$, for any x. Likewise, since school 2's payoff from matching with student s_1, $w_{2,s_1}(x_{2,s_1}, q_{s_1})$, is always equal to 3 and its payoff from staying unmatched is equal to zero, and since student s_1 prefers school 2 to staying unmatched, a stable mechanism cannot leave both school 2 and student s_1 unmatched, i.e., any stable mechanism must have $\mu(m; x) = 0$ if $m(2) = 2$ and $m(s_1) = s_1$. Also, since $w_{2,s_2}(x_{2,s_2}, q_{s_2}) = q_{s_2} = x_{2,s_2}$ and $w_{2,s_1}(x_{2,s_1}, q_{s_1}) = 3$, any stable mechanism must have $\mu(m; x) = 0$ if $m(2) = s_2$ and $x = -2$ or if $m(2) = 2$ and $x = 2$.

That, in turn, implies that $\mu(m; -2) = 0$ whenever $m(1) = s_1$: otherwise, in game $\Gamma_w(\mu, s_2, 1)$, after observing a match to s_1, could strictly improve its payoff by making a rematching offer to s_2, because it knows that the offer will only be accepted when s_2 is unmatched, i.e., $x = -2$. This is a key step in the proof, which allows a school to infer some additional information about the state of the world from a student's acceptance decision.

Note that the above arguments imply that $\mu(m; -2) = 1$ for matching $m_1 = [m(1) = s_2, m(2) = s_1]$, i.e., the mechanism is deterministic when $x = -2$. Also, when $x = 2$, neither school can remain unmatched, and so the only matchings that the mechanism can output are m_1 above or $m_2 = [m(1) = s_1, m(2) = s_2]$. But notice that school 2 strictly prefers the former matching, m_1, to m_2, and so the mechanism cannot output matching m_2 with positive probability: that would give school 2 an incentive to report $\hat{x} = -2$ whenever its true signal is $x = 2$. Hence, the only remaining candidate for a stable mechanism is deterministic mechanism μ that always outputs matching m_1 for any report of school 2.

This mechanism, however, is also unstable: in game $\Gamma_w(\mu, s_1, 1)$, given truthtelling, school 1's expected payoff is zero, while its payoff from making an offer to s_1 is one, since s_1 will accept the offer $(v_{s_1,1} > v_{s_1,2} > 0)$. Hence, school 1 and student s_1 would block.

Proof of Theorem 4

For the proof, we can focus on the case where (1) v_i is strictly increasing in i, and (2) $v_i > 0$ for all i. The first assumption is an innocuous relabeling, while the second can be assumed since schools with $v_i \leq 0$ can be ignored from the analysis without altering anything. For each x and $i < I$, let $y_i(x) = \{x_j\}_{j>i}$, and let $y_I(x) = \varnothing$; for all i, let $y_{i+}(x) = \{x_j\}_{j\geq i}$. We proceed in steps.

Step 0. The SD algorithm.

Fix the realized vector of signals x and let $\mathbf{S}^*_I(x) \equiv \mathbf{S}^* = \mathbf{S} \cup \{I\}$ be the set of possible matches for school I. The SD algorithm assigns students to schools iteratively —possibly leaving schools unmatched in the process.

Match I to $m^*(I; x) = \arg\max_{s\in\mathbf{S}^*_I(x)} u_{I,s}(x_I)$. Note that $m^*(I; x)$ depends only on y_{I+}.

Iteratively, for $i < I$ define $\mathbf{S}_i^*(x) = \mathbf{S}_{i+1}^* \cup \{i\} \backslash (\{m^*(i+1;x)\} \cup_{j \geq i} \{j\})$. This is the set of matches available to school i at x, and as a function of x it is measurable with respect to y_i. Match school i to

$$m^*(i;x) \in \arg\max_{s \in \mathbf{S}_i^*(x)} u_{i,s}(x_i, \{m^*(j)\}_{j>i}),$$

where conditioning on $\{m^*(j)\}_{j>i}$ is a short-hand for conditioning on the event $\{x : m^*(j) = m^*(j;x), j > i\}$, i.e., "the set of x's consistent with the same matches $\{m^*(j)\}_{j>i}$ for schools $j > i$," and so the same set of matches $\mathbf{S}_i^* = \mathbf{S}_i^*(x)$ is available for i. Again, $m^*(i;x)$ depends only on y_{i+}. When there are ties in a school preferences over students, the set of maximizers may sometimes not be a singleton; in this case, $m^*(i;x)$ is selected via uniform randomization over the set of maximizers. Once school $i = 1$ is assigned its match, any remaining student is left unmatched. This defines a matching for each x. Call the associated (degenerate) mechanism μ^{SD}.

We use the crucial informational assumptions and measurability restrictions on $m^*(i)$ and $\mathbf{S}_i^*$ repeatedly in what follows. In particular, for each i, $m^*(i)$ depends only on x_i given $\mathbf{S}_i^*$.

Fix the game $\Gamma_w(\mu^{SD}, s', i')$ for arbitrary $s' \in \mathbf{S}$, $i' \in \mathbf{I}$. Consider the following strategy profiles:

– at the first stage, $\hat{x}_i(x_i) = x_i$, all i; i.e., there is truthtelling;

– at the second stage, no blocking occurs at all information sets where agents believe $\hat{x}_i(x_i) = x_i$, all i.

We do not need to specify the strategies otherwise. We are going to show that the suggested strategy profile is a PBE of the $\Gamma_w(\mu^{SD}, s', i')$ game by using the one-shot deviation principle. That is, the following steps show that:

Step 1: i' does not benefit from objecting to its match $m(i') = m^*(i')$, given that it has reported its signal truthfully and anticipating truth-telling by others;[13]

Step 2: no school i (including i') with a realized signal x_i profits from making a false report $\hat{x}_i(x_i) \neq x_i$, assuming other schools report truthfully and that no one blocks the resulting matching outcome;

Step 3: no school i (including i') with a realized signal x_i profits from making a false report $\hat{x}_i(x_i) \neq x_i$, and then school i' objects.

Step 1. No objections after truthtelling.

Notice that $s' \neq m(i')$ will accept i''s offer only in states where $s' \in \mathbf{S}_{i'}^*$. Since $\{m^*(j)\}_{j>i'}$, hence $\mathbf{S}_{i'}^*$, do not depend on $x_{i'}$, by the definition of $m^*(i')$,

$$u_{i',s'}(x_{i'}, m^*(i') = m(i'), s' \in \mathbf{S}_{i'}^*, \{m^*(j)\}_{j>i'}) \leq u_{i',m(i')}(x_{i'}, m^*(i') = m(i'), s' \in \mathbf{S}_{i'}^*, \{m^*(j)\}_{j>i'})$$

for each possible $\{m^*(j)\}_{j>i'}$, then each $\mathbf{S}_{i'}^*$, implying that the same inequality holds if we take expectations over the $\{m^*(j)\}_{j>i'}$, conditional on the event $\{x_{i'}, m^*(i') = m(i'), s' \in \mathbf{S}_{i'}^*\}$, where we have used the fact that the extra events $m^*(i') = m(i')$, $s' \in \mathbf{S}_{i'}^*$ contain no additional information

[13] No student will object to its proposed school since $v_i > 0$ for all i. More generally, allowing no moves for s' is harmless a priori.

for i' given $x_{i'}$, $\{m^*(j)\}_{j>i'}$. But then objecting after truth-telling is not profitable for i'.

Step 2. Truthtelling.

This follows from observing that $\{m^*(j)\}_{j>i}$ and $\mathbf{S}_i^*$ defined in Step 0 do not depend on i's report, implying via the definition of $m^*(i)$, that

$$u_{i,m^*(i)}(x_i, \{m^*(j)\}_{j>i}) \geq u_{i,m'(i)}(x_i, \{m^*(j)\}_{j>i})$$

where, with $x_i' = \hat{x}_i(x_i)$, $m'(i) = \arg\max_{s \in \mathbf{S}_i^*} u_{i,s}(x_i', \{m^*(j)\}_{j>i})$ is the matching outcome for i' after the false report x_i' given $\{m^*(j)\}_{j>i}$. Since this is true for each possible$\{m^*(j)\}_{j>i}$, it is also true on average, i.e., in interim terms conditional on only x_i. Therefore no i can strictly gain in expected terms simply by sending a false report.

Step 3. No profitable objections after lying.

To complete the proof, it suffices to show that school i' does not gain from sending a false report $\hat{x}_{i'}(x_{i'}) \neq x_{i'}$ and then objecting. This follows again from the fact that $\{m^*(j)\}_{j>i'}$, hence $\mathbf{S}_{i'}^*$, do not depend on $x_{i'}$. In particular, by the construction of μ^{SD} we have

$$u_{i',m^*(i')}(x_{i'}, m'(i'), s' \in \mathbf{S}_{i'}^*, \{m^*(j)\}_{j>i'}) \geq u_{i',s'}(x_{i'}, m'(i'), s' \in \mathbf{S}_{i'}^*, \{m^*(j)\}_{j>i'})$$

and

$$u_{i',m^*(i')}(x_{i'}, m^{'}(i'), s' \in \mathbf{S}_{i'}^*, \{m^*(j)\}_{j>i'}) \geq u_{i',m'(i')}(x_{i'}, m^{'}(i'), s' \in \mathbf{S}_{i'}^*, \{m^*(j)\}_{j>i'})$$

where $m'(i') = \arg\max_{s \in \mathbf{S}_{i'}^*} u_{i,s}(x_{i'}', \{m^*(j)\}_{j>i'})$, and we use the fact that the extra events $m^{'}(i')$, $s' \in \mathbf{S}_{i'}^*$ convey no additional information to i' given $\{x_{i'}, \{m^*(j)\}_{j>i'}\}$. Then, taking expectations at the interim stage (i.e., over $\{m^*(j)\}_{j>i'}$, conditional on $x_{i'}$), i' cannot do strictly better to send a false report $x_{i'}'$ and then following the equilibrium strategy, compared to telling the truth and not blocking. Since i' and s' are arbitrary, this concludes the proof.

Appendix B: Market with a Strongly Stable Mechanism

In this Appendix, we present an example of a non-degenerate market for which a strongly stable mechanism exists. This mechanism incorporates information from both schools in determining their matches.

Example 4

Consider a market with two schools, 1 and 2, and ten students. School 2 is the better school. Students have three possible qualities: q_l, q_m, q_h. The probability of quality q_h is equal to $\Pr(q_h) = \frac{1}{2} - \frac{1}{10} + \epsilon_h$, $\Pr(q_m) = \frac{2}{10} + \epsilon_m$, and $Pr(q_l) = \frac{1}{2} - \frac{1}{10} + \epsilon_l$, where $\epsilon_l + \epsilon_m + \epsilon_h = 0$. Quality $q_h = \delta_h$, $q_m = -1$, and $q_l = \delta_l$ (note that when δ_h and δ_l are close to zero, q_m is the worst quality). Each firm observes one of two signals about each worker: H or L. Conditional on true qualities, signals are independent, with distributions given as follows:

$$\Pr(H|q) = \begin{cases} 1-\lambda & \text{if } q = q_h \\ \frac{1}{2}+\phi & \text{if } q = q_m \\ \nu & \text{if } q = q_l, \end{cases}$$

where $\lambda, \nu > 0$. Schools' ex post payoffs are $w_{is}(x_s, q_s) = q_s$.

Consider the following mechanism μ:

- if for each student the two schools' signals are different, match schools with students randomly;
- if for two or more students the two signals are identical, match schools with those students randomly;
- if for exactly one student the two signals are identical, match this student with school 2, and match a random student from the remaining set with school 1.

Claim 5 *Mechanism μ is strongly stable for all ϵ_h, ϵ_m, ϵ_l, λ, ϕ, ν, δ_h, δ_l sufficiently close to zero.*

Proof. It is sufficient to show that when all parameters denoted by Greek letters are equal to zero, compliance by all players is a strict equilibrium, i.e., each school is strictly worse off if it deviates. Then, by the continuity of payoffs, it remains a strict equilibrium for small parameter changes.

When parameter values are zero, schools are indifferent among all students who have two identical signals, and strictly prefer them to all students who have two different signals. For a given school and a student, the probability that the other school's signal about the student is the same is equal to $\frac{4}{5} + \frac{1}{5} \cdot \frac{1}{2} = \frac{9}{10}$.

School 2 can never benefit by deviating, since it gets a student with identical signals whenever there is one. However, for any deviation there is a positive probability of decreasing the payoff, and so deviating makes school 2 strictly worse off.

To see that school 1 is also strictly worse off if it deviates, notice that if it doesn't, it gets a student with two identical signals with probability $1 - \left(\frac{1}{10}\right)^{10} - 10 \cdot \frac{9}{10} \cdot \left(\frac{1}{10}\right)^9 = 1 - \frac{91}{10^{10}}$. School 1 is strictly worse off if it rematches after truthfully reporting signals; likewise, if it misreports all signals, it then has to essentially rematch with a random student and is also strictly worse off. Thus, any possible profitable deviation has to involve misreporting some, but not all of the signals. Without loss of generality, assume that it misreports signals about students from s_1 to s_n. With probability $\frac{9}{10}$, school 2's signal about student s_{10} is the same as school 1's. In that case, this student will be assigned by the mechanism to school 2 with probability at least $\frac{1}{10}$. Then, with probability $\frac{1}{10}$, school 2's *true* signal about student s_1 is different from school 1's, and therefore the *reported* signals are identical. Thus, student 1 gets assigned to school 1 with probability at least $\frac{1}{9}$ (because student s_{10} is already assigned to school 2). Summarizing, the probability that the mechanism assigns student s_{10} to school 2 and student s_1 to school 1, and that student s_{10} has two identical true signals and student s_1 has two different true signals is at least $\frac{9}{10} \cdot \frac{1}{10} \cdot \frac{1}{10} \cdot \frac{1}{9} = \frac{1}{10^3}$. If school 1's strategy is not to rematch in this situation, it is clearly worse off than under compliance. If school 1's strategy is to rematch, it gets a student with two different signals with probability $\frac{1}{10}$ (since one can think of school 1's strategy as "misreport as above; if school 2 gets matched to student s_{10} and I get matched to student s_1, rematch with student s_j," where s_j is chosen in advance and hence has a $\frac{1}{10}$ chance of having two different signals), and so its total expected probability of not getting a student with two identical signals is at least $\frac{1}{10^4} \gg \frac{91}{10^{10}}$.

Note also that since the equilibrium is strict, adding sufficiently small random disturbances to the mechanism in Example 4 does not change schools' incentives, and therefore there is a continuum of different stable mechanisms.

References

[1] H. Chade (2006) Matching with noise and the acceptance curse. J. Econ. Theory 129, 81-113.

[2] H. Chade, G. Lewis, L. Smith (2007) The college admissions problem with uncertainty. Mimeo.

[3] V. Crawford (1991) Comparative statics in matching markets. J. Econ. Theory 54, 389-400.

[4] F. Forges (1994) Posterior efficiency. Games Econ. Behav. 6, 238-261.

[5] D. Gale, L. Shapley (1962) College admissions and the stability of marriage. Amer. Math. Monthly 69, 9-15.

[6] B. Holmstrom, R. Myerson (1983) Efficient and durable decision rules with incomplete information. Econometrica 51, 1799-1820.

[7] H. Hoppe, B. Moldovanu, A. Sela (2005) The theory of assortative matching based on costly signals. Mimeo.

[8] S-H. Lee (2004) Early admission program: Does it hurt efficiency? Chapter 1 of Three essays on applied microeconomics, Ph.D. Thesis, University of Pennsylvania.

[9] R. Myerson (1979) Incentive compatibility and the bargaining problem. Econometrica 47, 61-73.

[10] R. Myerson (1986) Multistage games with communication. Econometrica 54, 323-358.

[11] E. Nagypal (2004) Optimal application behavior with incomplete information. Mimeo.

[12] A. Roth (1989) Two-sided matching with incomplete information about others' preferences, Games Econ. Behav. 1, 191-209.

[13] A. Roth, M. Sotomayor (1990) Two-sided matching. Cambridge, MA: Cambridge University Press.

[14] R. Wilson (1987) Game-theoretic analyses of trading processes. In Advances in economic theory: Fifth world congress, ed. Truman Bewley. Cambridge: Cambridge University Press.